Directing Autonomous Digital Actors

13 avril 2015

Warning : the comments are turned on. All directives and guidelines appear.
To turn them off, comment the line `\newenvironment{xcomment}{\em}{}`
and uncomment the line `\usepackage{xcomment}`

Table des matières

1 Résume de la proposition de projet / Executive summary

Recopier le résumé utilisé dans le document administratif et financier.

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

Décrire le contexte économique, social, réglementaire' dans lequel se situe le projet en présentant une analyse des enjeux sociaux, économiques, environnementaux, industriels' Donner si possible des arguments chiffrés, par exemple, pertinence et portée du projet par rapport Ãă la demande économique (analyse du marché, analyse des tendances), analyse de la concurrence, indicateurs de réduction de coÃżts, perspectives de marchés (champs d'application, '). Indicateurs des gains environnementaux, cycle de vie'

Animating virtual agents with expressivity is a big 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 charac-ters 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) is 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.

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.

1.2 Positionnement du projet / Position of the project

Préciser : positionnement du projet par rapport au contexte développé précédemment : vis-Ãă-vis des projets et recherches concurrents, complémentaires ou antérieurs, des brevets et standards' indiquer si le projet s'inscrit dans la continuité de projet(s) antérieurs déjÃă financés par l'ANR. Dans ce cas, présenter brièvement les résultats acquis, positionnement du projet par rapport aux axes thématiques de l'appel Ãă projets, positionnement du projet aux niveaux européen et international.

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

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

Character animation has been tackled through various approaches in the past. To name a few, chosen among those that are directly related to DADA, we can cite : embodied conversational agents (ECA), ie autonomous virtual characters (Cassell 2000) ; statistical models learned from motion capture examples (Lee 2002) ; physically-based animation (Liu 2006) ; and speech-driven animation (Ding 2013). 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 con-trast,

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 proce-durally using autonomous actors is a valuable goal because it would make it possible for each performance to be unique, which is widely regarded as an important quality to ensure liveliness and immersion, while maintaining a high level of directorial control. Merging both approaches would allow creating autonomous actors able to follow a script (specified in high-level command-like language) that give the main directions the actors ought to follow while adapting their behaviors autonomously to the virtual environment they are placed in that includes objects and other actors.

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

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

1. Animating autonomous actors procedurally. A key idea in DADA is to separate the animation model into a proxe-mic component regulating how actors interact with each other and the audience, and a kinesic component regulating how actors use their body language to communicate moods and expressions (Tannenbaum 2014). The proxemic component of animation will drive the positions and orientations of actors on the stage as well as their gaze direc-tions. This component will be driven by a model encompassing the social relations between and the emotional states of the autonomous actors. The kinesic component of animation will drive all other degrees of freedom of the virtual actors. This component will be driven by parametric statistical models trained from an existing motion cap-ture data-set. The separation between the two components is expected to yield important benefits in terms of ex-pressivity and composability.
2. Synchronizing virtual actors to a single story-line using a story-driven architecture of actors following a scripted sequence of instructions (Pinhanez 2000). In contrast to previous works, which used programming languages (Ma-teas 2002), 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, under the supervision of Georges Gagneré. Starting from a selection of play scripts in various genres and with increasing complexity, theatre experts will use the DADA tools to create virtual theatre performances in the Unity game engine, including stage movements and actions (entering, exiting, sitting down, standing up, taking and putting objects on the stage) ; body language expression of the personalities, moods and emotions of the characters ; and believable gaze, proxemics and action/reaction behaviors between actors.

The expected results of DADA will be (1) a virtual theatre company of autonomous actors with a large vocabulary of expressive animation skills ; and (2) a prototype system for directing

arbitrary dramatic plays, amenable to a variety of digital storytelling applications. Results will be integrated into Unity3D which is already used by the GRETA plat-form at Telecom ParisTech and the virtual cinematography framework developed by the IMAGINE team at Inria. Re-sults 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.

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

1.3 Etat de l'art / State of the art

Présenter un état de l'art national et international, en dressant l'état des connaissances sur le sujet. Faire apparaÃőtre d'éventuelles contributions des membres de la proposition de projet Ãă cet état de l'art. Faire apparaÃőtre d'éventuels résultats préliminaires. Inclure les références bibliographiques nécessaires au Âğ 7.

1.3.1 WP1. Related works

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 [40].

Full body kinematic animation (or control) consists in animating the full body of an avatar while he is performing actions such as walking, dancing, sitting etc. Although there has been lots of work on this subject it is still a challenging problem due to the high dimensionality of the character's configuration. Data-driven approaches are very popular here and make of use motion-capture data to learn animation models which, once learned, may be used to animate a virtual character to perform a given task. Many systems have been proposed for producing animation models and controlers, they usually are based on statistical models such as Hidden Markov Models (HMMs) [27] and Conditional Random Fields (CRFs) [25, 6]. 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 [**?**]. Locomotion controllers have been proposed that concatenate motion clips from a motion capture dataset to produce an animation that is smooth [Treuille et al. 2007 ; Mc-Cann and Pollard 2007]. 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 [**?**]. 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 [27, 25, 6, **?**]. At the end, recent work has demonstrated such mo-

dels for the case of locomotion believable controllers, gesture controllers (Levine 2010) and face controllers (Ding et al., 2013). 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 [28], 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 [28, 6]. A relevant technology for this are Gaussian Process which have been extended for dealing wiuth dynamic data in [44].

These latter models are not far from recurent neural networks, and to Long Short Term Memory neural networks in particular [20, 18], that have been shown recently to work well for complex signals such as speech and handwriting, for recognition tasks [17] as well as for synthesis tasks [16]. 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 [19].

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

Décrire les objectifs du projet et détailler les verrous scientifiques et techniques Ãă lever par la réalisation du projet. Insister sur le caractère ambitieux et/ou novateur de la proposition. Décrire éventuellement le ou les produits finaux développés, présenter les résultats escomptés en proposant si possible des critères de réussite et d'évaluation adaptés au type de projet, permettant d'évaluer les résultats en fin de projet.

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

A titre indicatif : de 5 Ãă 10 pages pour ce chapitre, en fonction du nombre de tÃćches

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

Présentez le programme scientifique et justifiez la décomposition en tÃćches du programme de travail en cohérence avec les objectifs poursuivis. Utilisez un diagramme pour présenter les liens entre les différentes tÃćches (organigramme technique) Les tÃćches représentent les grandes phases du projet. Elles sont en nombre limité. Le cas échéant (programmes exigeant la pluridisciplinarité), démontrer l'articulation entre les disciplines scientifiques. N'oubliez pas les tÃćches correspondant Ãă la dissémination et Ãă la valorisation, Ãă décrire en détails au Âğ4.

Work will be divided into four main work packages : (1) procedural animation of isolated actors ; (2) procedural anima-tion of interaction between actors ; (3) authoring and real-time control ; (4) user evaluations. Through the authoring tool (WP3), a script is elaborated by a

theater director (WP4) ; it gives direction to group of actors which act out autono-mously 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).

thierry : ce passage lÃă doit probablement etre remontÃľ au dessus pour l'explication gÃľnÃľrale du flux entre WPs. This workpackage focuses on animation models for isolated actors. The inputs that are used by the methods to be developped in this WP are procedural animation scenario as ouput 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 almong 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*. .

2.1.1 WP1. Kinesic component

This WP aims at creating multi-modal statistical models of individual body movements from annotated, mainly from mocap data, to generate novel expressive animation suitable for dramatic performances. To do so we will tackle few difficult and open problems : Learning full body animation models for many settings including emotional state and actor's profile (morphology, expressivity level etc). Moreover while the animation model will be learned from a limited number of actors' data we want it to be able to be remapped to other actors. Next we will investigate learning animation models for new gestures and activities from only few training samples which will allow enriching the system easily by avoiding the costly and tedious task of gathering a large corpus of training data as usually required in statistical machine learning.

A first scientific lock lies in the design of generic body controllers able to synthesize the animation of the full body of a character for many settings (combination of action, emotion and actor's profile). It is an elegant way for producing smooth animation of complex motions which usually requires artifical smoothing and postprocessing. It is also a relevant modeling framework for learning from limited datasets. Indeed gathering a dataset including enough training samples for every combination of (action, emotion, actor profile) is unlikely. Defining generic models should allow to maximize the exploitation of the training data for learning, which is a key issue here. The idea is to build models that take as input contextual variables that encode the setting in such a way that the animation model for a particular action may be learned from all samples of this action whatever the mood, and from all samples of any action performed with this particular mood. One main idea for doing so consists in extending to full body animation the idea of contextual models [38, 8, 9] which are a variant of Hidden Markov Models (HMMs) that have shown strong potential for designing face controllers. Contextual HMMs are HMMs whose parameters (means of Gaussian distribution transition probabilities etc), are defined as a (learned) function of contextual variables. One Contextual HMM may be viewed as a continuum of HMMs, one model for every possible value of contextual variables. These models will serve as a baseline. Next, we will investigate the use of continuous state space models and particularly of (deep) recurrent neural networks. Such models have shown strong abilities for dealing with complex signals like speech and handwriting [16]. The main difficulty lies here in imagining ways to integrate the idea of using contextual variales in these models in order to

integrate contextual information that would modify the behaviour of the models. Finally we plan to explore alternative strategies such as using neuro muscular based models following ideas like the one of deltalognormal models from [11, **?**] which allow recovering the sequence of neuromuscular commands that generated a handwritten gesture.

A second lock will concern the animation of the face in dialogue situations. Given that we have worked previously with three complementary methods, we will focus here on how to mix our face animation models : mocap based animation, video-based animation, and procedural anmation. The latter animation model is already working and part of the GRETA system. The mocap based animation model will be easily built on previous works by the team [**?**]. Finally starting from previous work on visual prosody we will design the third model ... Ideally, this should be done without MOCAP data, using only audio and video processing, possibly enhanced with depth (kinect). To be continued (RÃľmi)...We will explore strategies for optimally combining these three animation models...

Finally we want our animation models easily extendabke to new activities 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 models, and particularly statistical models from few samples is a key and open issue [23]. We will mainly explore two ways that aim at favoring transfer from learning one gesture model to learning another gesture model. First, few preliminary works have shown that contextual markovian models such as the ones proposed in [38] for gesture recognition could be defined in such a way that the data from all gesture classes could be exploited to learn models for all gesture classes. Second using continuous state space models with a low dimensional state space such as Reccurrent neural nets (corresponding to the degree of freedom of body poses) should permit characterizing a particular gesture as its dynamic in this latent space whose limited dimension would enable learning from few samples.

All along the project we will rely as much as possible on existing datasets. For instance Mocap data of considered actions have already been recorded by C. Pelachaud within the project Feder Anipev (http ://www.anipev.com/). The corpus EMILYA (EMotional body expressIon in daiLY Actions databaseBodily Emotional Actions Behavior) (Fourati, 2014) is constituted of 7 actions performed by 11 actors with 8 emotions. The actions encompass everyday actions such as walking, carrying an object, and sitting. The emotions cover the positive and negative spectrum.

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

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

Physical distance between actors, their body orientation toward each other, gaze direction,

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

2.1.3 WP3. Performance authoring and real-time execution.

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

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

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

2.1.4 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).

2.2 Management du projet / Project management

Préciser les aspects organisationnels du projet et les modalités de coordination (si possible individualisation d'une tÃćche de coordination).

2.3 Description des travaux par tÃćche / Description by task

Pour chaque tÃćche, décrire : les objectifs et éventuels indicateurs de succès, les personnes impliquées, le programme détaillé des travaux, les livrables, les contributions des personnes (le qui fait quoi), la description des méthodes et des choix techniques et de la manière dont les solutions seront apportées, les risques et les solutions de repli envisagées.

2.3.1 WP1 Kinesics

WP1	Kinesic component
Responsable	ECM
Participants	Inria, Telecom ParisTech
Duration	
Objectives	
Content	
Task 11	Full body animation
Task12	Interaction animation
Task13	Learning from few samples

Deliverables	Name and content	Date
L1.1	Report on the state of the art for statistical models for animation synthesis	
L1.2		
L1.3		

The aim 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 it will explore the ability to extend these models in order to deal with only few training data by relying on transfer learning strategies.

The workpackage is divided into three subtasks which are dedicated to the animation of the full body of a character, to the animation of specific parts of its body which are engaged in an interaction with another character (mainly the face) by combining few animation models, and to the specific strategies that will be explored for learning such models from few training data.

Task 1.1. Generic full body animation model We will first focus on the design of generic body controllers able to synthesize the animation of the full body of a character (through the sequence of mocap representation) for a given procedural animation scenario as output by WP2, i.e. a sequence of actions realized with a particular mood context and for a specific character profile (morphology, expressivity level).

To start we will consider that there is one model per action and that the animation produced by such a model should take into account, as inputs, the mood context as well as the character profile context, later on we will investigate one model for all settings (action, mood, charecter profile). We will investigate modeling frameworks that allow taking into account the contextual variables (e.g. mood components and profile components) as few inputs which are mixed to produce an animation. This will enable learning from a limited combination of (mood, profile) settings while allowing extrapolating to any other combination (mood, profile). Whatever the models under investigation we will pay attention to focus on strategies that enable synthesizing smooth transition between successive actions, moods, or gestures. We plan to investigate the following lines of research :

Firstly we will investigate *contextual markovian models* where gaussian probability density functions associated to states are parameterized by (i.e. defined as a function of) contextual observation (mood and profile information). Recent work has demonstrated such models for the case of locomotion believable controllers, gesture controllers (Levine 2010) and face controllers [38, 8, 9]. We will aim to generalize these works to more general action controllers, including such actions as : sitting, standing, walking, grasping, taking and putting objects, in a variety of expressions and moods. These models will serve as a baseline for evaluating new modeling approaches.

Second, we will investigate the use of *(deep) neural networks* and of dynamic versions of these (i.e. recurrent neural nets) which have demonstrated strong abilities to model, to classify and to synthesize complex signals such as speech or handwriting [**?**, **?**, 17, 16]. These models are related to what is called *representation learning* which emerged in the last few years as a key topic in the machine learning community [1] (Contardo 2014). One main difficulty will be to integrate the use of contextual information as input in order to modify the behaviour of the models. We plan to extend the principle of contextual markovian models to neural nets by investigating ideas like designing bilinear layers in the neural net where weights could be defined as a function of the contextual input, inspired by works like [48, 21]. At last we will investigate *low dimensional state space models* such as neuro muscular based models following ideas like [**?**] which aims at recovering from a handwritten signal the sequence of neuromuscular commands that generated the handwriting signal. The underlying idea here is to exploit such models in order to work in a new representation space, the space of neuromuscular commands that generate motion, rather than on the observed motion itself. Although such models have not been used to model complex gestures up to now it is expected that they could be robust enough to provide good estimation of the command sequence. The main advantage of such a change of representation space is an expected reduction of the dimension of this space (as in [44]), enabling easier learning from few samples and transfer learning (as will be investigated in task 13).

Task 1.2. Combining models for face animation The second task focuses on learning models of gesture and facial expressions in dialogue situations. It is dedicated to the combination

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

of animation models, whch is a diifuckt and open question, with a focus on the animation of the face. We will start from available face animation models in the consortium : a mocap based animation [8], a video-based animation [**?**], and a procedural animation [32] (integrated in the GRETA system).

All of these models types have pros and cons. While statically-driven models are more prone to produce natural looking animation, cognitive models capture more precisely the semantic emotional behaviors to communicate. These latter ones are often event-driven ; that is they compute a behavior only when a given communicative function is specified. Statically driven models produce animation continuously that captures the communicative colour of the message to convey but they have difficulty to compute behaviors which have specific meaning. As a result, virtual agents driven by cognitive-like system are able to convey more precise displays while those driven by statistical models look more natural and lively [24].

We will explore ways to combine few such animation models which remains an open question today, be it for animating the face ot the full body [**?**]. We will explore strategies and implement these within the Greta framework where communicative intentions and emotions are represented with the FLM language while multimodal behaviors with BML [43]. The merge of multiple animation models may be performed as a weighted blend of the animations produced where the weights might be context dependent and tuned either manually or automatically, alike in [40]. 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 animatyion model and use when neccessary animatyions produced by the other models.

Task 1.3. Learning from few samples We will mainly investigate two approches for extending approaches developped in task T11 to enable learning from few samples. The first strategy consists in extending the idea of context variables that models of task T11 rely on in order to design a global model for all actions. In the case of markovian model for instance this means that instead of definig one model per action one could define a unique global markovian model where every state would stand for a particular position of the body and performing an action would correspond to following a path (i.e. a state sequence) in this big model. Making transition probablities dependent on the action to perform such a big model would be instaiated as an action model by considering a bundle of paths only in this model. Doing so one could expect that all the training data (whatever the action it corresponds to) could be exploited to learn all the states of this big markovian model, hence implementing some kind of transfer learning between actions. A new action would correspond then in a bundle of paths in this model and could be learnt from few samples only. Preliminary works that we did let us expect that such a strategy would work with statistical markovian models (Ding et al., 2013). In this case the above idea is implementing by introducing new contextual variables, which might be at the simplest on-hot indicators of the action to perform (a vector with zeros everywhere but at the position of the action number), that modify the gaussian densities. We will first investigate this strategy deeper for contextual markovien models then we will extend this approach to neural networks...

Second we will explore the use of using continuous state space models with a low dimensional state space (e.g. corresponding to the degree of freedom of body poses or to the neuro muscular commands) which should permit characterizing a particular motion or gesture as its dynamic in this latent space whose limited dimension would enable learning from few samples...

Task14 Thierry : this should probably be moved to WP3

- Combining full body animation and interaction animation
- Mettre de la diversitÃľ dans l'animation pour ne jamais reproduire la mÃłme animation

We will pay particular attention to design models capable of generating real animations. Indeed synthezing from statistical models usually resumes to finding the most likely animation sequence in a given situation, which may yield to too similar and unrealistic animations. Actually one would be pretty much interested in synthezing animations that are both likely given the learnt statistical models but also exhibiting the variability one can observe in human motion and gestures. Introducing such a stochastic component in the synthesis while maintaining a high quality animation level is not straightforward and is an open question that we will have to solve.

Partners' roles bla bla

2.3.2 WP2 Proxemics

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 incorporate two models : statistical model as developed in WP1 and procedural model developed within the Greta plateform.

Task 2.1 : Group behaviors during multi-way conversation In this task we will model multi-party conversation behaviors. We will focus on turn-taking management. While indication of what the agents would say to whom and when will be provided by a script (Task 3.1 and Task 4.X), the turn-taking model will instantiate which behaviors the agents will display. Gaze, body orientation, position in space are important cues for indicating who has the turn, who wants to keep it, to give it to someone, who listens ? We will extend an existing turn-taking model (Ravenet et al., 2014) that is based on Sack ?s model (Sack et al, 1974), that embeds F-Formation (Kendon, 1990) 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 (Schegloff, 2000) 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).

A version of this model will be instantiated to model tri-partite interaction between two virtual actors and the audience (viewed as a virtual actor taking part of the interaction).

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

etc.

This task will model agents ? behavior when moving around in the environment. The animation of the virtual agent doing some tasks will be given by WP1. It will not focus on path planning as this information will be provided by a script (Task 3.1 and Task 4.X). Rather it will model how agents perform displacement in social settings. Gaze direction, body orientation and spatial distance to other agents will be computing for different ?steering behaviors ?. These features will be modeled through different synchronization mechanisms : moving in synch, moving ahead, following, etc. They evolve dynamically in function of each agent ?s position and orientation in space. The basic animation of the agent, ie without any influence from surrounding agent, is given by WP1. To simulate the dynamic evolution of agents' behaviors we will make use of Neural Network simulation (Prepin 2013) where we can render how behaviors of one actor can act on behaviors of other actors (eg walking powerfully toward an actor with an angry expression will result in moving backward of another actor with a less dominant attitude. Mutual coupling of behaviors will be modeled as emerging from such action-reactive behavior simulation (Prepin 2013) ensuring not only the synchronization between actors ? behaviors but also their mutual influence.

As for Task 2.2, a version of this model will be instantiated to consider the audience as one virtual actor.

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 (Huang, 2012). 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.

2.3.3 WP3 Authoring

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

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

Part of this language will be devoted to dialogue

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

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

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

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

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

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

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

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

2.3.4 WP4 User evaluation and validation

Task41 Scenarios.

Writing scenes with didascalia

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

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

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

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

Task42 Validation of the interaction.

Is the dramatic language adequate ? useful ? efficient ?

Is the dramatic score interface adequate ? useful ? efficient ?

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

Task43 Validation of the animation

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

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

Combination of dialogue and action for scenes with 2 actors

Additional notes

We will dedicate joint research between Inria and LIF to make it easy to extend our database of actions and attitudes using video, rather than motion capture. This will necessitate

fundamental research in transfer learning (so that the sparse data obtained from video can benefit from the dense data obtained with motion capture) and video processing. Following the methodology of gesture controllers [26], where the gesture are controlled directly by speech prosody features extracted from real actors voices, it appears possible to drive expressive and plausible gestures and body movements from visual signatures of actions and attitudes extracted from example videos.

We will use our previous work in actor and action recognition [46, 45, 14] to detect and recognize actors and their actions in real movies ; and extract visual signatures of the corresponding actions and attitudes. Based on this analysis, we will learn joint statistical models for driving gesture controllers from those video signals.

Combining proxemics and kinesics components can be done along the lines of Mitake et al. [31], where the degrees of freedom of a virtual character are separated into six parameters for rigid body simulations, and four parameters for encoding multi-dimensional keyframe animations. Similarly, we would like to hide the complexity of high-dimensional character animation (with 40-60 degrees of freedom) behind a small number of control parameters. We will extend rigid body simulations to include proxemic interaction forces in WP2. And we will replace keyframe animations with statistical models learned from data in WP1.

One promising avenue for research will be to design strategies for controlling the proxemic components of character animation using the rigid motion of the head, rather than the full body. Sreenivasa et al. [41] 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.).

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

Présenter sous forme graphique un échéancier des différentes tÃćches et leurs dépendances (diagramme de Gantt par exemple). Présenter un tableau synthétique de l'ensemble des livrables du projet (numéro de tÃćche, date, intitulé, responsable). Préciser de faÃğon synthétique les jalons scientifiques et/ou techniques, les principaux points de rendez-vous, les points bloquants ou aléas qui risquent de remettre en cause l'aboutissement du projet ainsi que les réunions de projet prévues.

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

A titre indicatif : 2 pages maximum pour ce chapitre. Présenter les stratégies de valorisation des résultats : la communication scientifique, la communication auprès d'autres communautés scientifiques et du grand public, notamment la promotion faite Ãă la culture scientifique et technique. Si un budget spécifique est prévu Ãă cet effet, le spécifier et l'identifier dans une tÃćche

de la proposition (voir Âğ 3.1). les résultats attendus en matière de valorisation, les retombées scientifiques, techniques, industrielles, économiques, ' la place du projet dans la stratégie industrielle des entreprises partenaires du projet, les autres retombées (normalisation, information des pouvoirs publics, formation dans l'enseignement supérieur, ...), les échéances et la nature des retombées technico-économiques attendues, l'incidence éventuelle sur l'emploi, la création d'activités nouvelles, '

Présenter les grandes lignes des modes de protection et d'exploitation des résultats.

4 Description de l'equipe / Team description

A titre indicatif : 2 pages maximum pour ce chapitre.

LIF Two main researchers from the QARMA team will participate to the project. **Thierry Artières** is a professor at University of Aix-Marseille, and a member of the *QARMA team* (eQuipe AppRentissage et MultimÃľdia) at LIF (Laboratoire dâĂŹInformatique Fondamentale). One of his major research topic concerns machine learning for multimedia applications, more particularly for sequences and signals, either for classification, pattern discovery, sequence labeling and sequence synthesis, with strong experience with various signals such as speech, bioacoustics, handwriting, gestures, eye movements, WII signals, Kinect and motion capture data. He is author or co-author of about sixty papers and articles in top ranked international conferences (NIPS, ICML, AISTAT, ICASSP, EMNLP) and journals (IEEE PAMI, JMLR, Pattern Recognition) in the fields of theoretical as well as applied machine learning (speech and handwriting recognition, user modeling) and artificial intelligence. **Valetin Emiya** is assistant professor in the QARMA team at LIF since 2011. He has conducted research in audio processing and sparse models for 8 years and has strong connexion with the signal processing group at I2M Lab in Marseille. His current works on models and algorithms for audio inpainting (see [2, 1] and project ANR JCJC MAD), i.e. interpolation and extrapolation in audio sequences. This works are currently being extended to the extrapolation of gesture for the control of electronic musical instrument and contemporary music creation, through the Progest project by GdR ISIS (2014-2016) in collaboration with the gmem Centre National de CrÃľation Musicale (http://www.gmem.org/index.php?option=com_content&view=article&id=5580144&Itemid=13660).

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

Fournir les éléments permettant d'apprécier la qualification des personnes impliquées dans la proposition de projet (le pourquoi qui fait quoi). Il peut s'agir de réalisations passées, d'indicateurs (publications, brevets), de l'intérÃłt pour le projet' Montrer la complémentarité et la valeur ajoutée des coopérations entre les différents participants. Le cas échéant, l'interdisciplinarité et l'ouverture Ãă diverses collaborations seront Ãă justifier en accord avec les orientations du projet.

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

Name	First name	Position	Field of research	PM	Contribution to the proposal
Artières	Thierry	Pr	Machine Learning	30	WP1 (task leader), WP2 and WP3,
Emyia	Valentin	Assistant Pr	Machine Learning and Signal Processing	10	WP1

TABLE 1 – Qualification and contribution of each partner

actors, as part of Adéla Barbulescu thesis (Barbulescu 2014). Results of the two theses will be exploited in the project.

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

0,5 page maximum Fournir les éléments permettant de juger la capacité du coordinateur Ãă coordonner le projet.

Rémi Ronfard is a senior researcher at Inria in the IMAGINE team, whose research is devoted to designing novel interfaces between artists and computers (Intuitive Modeling and Animation for Interactive Graphics & Narrative Envi-ronments). He has a 20 year experience in industry and academia in France, Canada and USA, and has directed an R & D team on virtual cinematography at Montreal-based startup Xtranormal Technologies. He will be acting as coordinator of DADA.

4.3 Qualification, rÃť le et implication des participants / Qualification and contribution of each partner

(1 page maximum) Qualifier les personnes, préciser leurs activités principales et leurs compétences propres.

Pour chacune des personnes dont l'implication dans le projet est supérieure Ãă 25% de son temps sur la totalité du projet (c'est-Ãă-dire une moyenne de 3 hommes.mois par année de projet), une biographie d'une page maximum sera placée en annexe du présent document qui comportera : Nom, prénom, Ãćge, cursus, situation actuelle Autres expériences professionnelles Liste des cinq publications (ou brevets) les plus significatives des cinq dernières années, nombre de publications dans les revues internationales ou actes de congrès Ãă comité de lecture. Prix, distinctions Si besoin, pour chacune des personnes, leur implication dans d'autres projets (Contrats publics et privés effectués ou en cours sur les trois dernières années) sera présentée et fournie en annexe du présent document. On précisera l'implication dans des projets européens ou dans d'autres types de projets nationaux ou internationaux. Expliciter l'articulation entre les travaux proposés et les travaux antérieurs ou déjÃă en cours.

Thierry Artières is a professor at University of Aix-Marseille, and a member of the QARMA team (eQuipe Ap-pRentissage 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.

Georges Gagneré is stage director (www.didascalie.net) and associate professor in Paris 8 University's performing arts department, working in the laboratory "Scènes du monde, création, savoirs critiques" (EA 1573), with full professor Jean-FranÃğois Dusigne, ex-actor of Théâtre du Soleil, and international expert in directing actor theory and practice. He works closely with the digital artist and associated professor Cédric Plessiet from Paris 8 's INREV research laboratory (EA4010- digital image and virtual reality), directed by Marie-Hélène Tramus, full professor and scientific director of Labex Arts and Human Mediations (www.labex-arts-h2h.fr) linking together artistic practice with cognitive sciences and human mediations.

Catherine Pélachaud is Director of Research at CNRS in the laboratory LTCI, TELECOM ParisTech. She has pub-lished over 150 papers and chapters in internationally recognized conferences and journals. She has participated in several national and European projects related to multimodal communication, to believable embodied conversational agents, emotions and social behaviors. She has developed an open-source virtual agent system, Greta, which is used by several international teams for research and teaching purposes.

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

On présentera ici la justification scientifique et technique des moyens demandés dans le document de soumission tel que synthétisé et rempli en ligne sur le site de soumission dans la fiche tableaux récapitulatifs du document administratif et financier tel que rempli en ligne sur le site de soumission. Justifier les moyens demandés en distinguant les différents postes de dépenses. (2 pages maximum)

5.1 équipement / Equipment

Préciser la nature des équipements et justifier le choix des équipements (un devis pourra Ãłtre demandé si le projet est retenu pour financement). Dans le cas oÃź les achats doivent Ãłtre complétés par d'autres sources de financement, indiquer le montant et l'origine de ces aides complémentaires, et le pourcentage demandé Ãă l'ANR sur le présent projet.*

5.2 Personnel / Staff

Le personnel non permanent (thèses, post- doctorants, CDD...) financé sur le projet devra Ãłtre justifié. Fournir les profils des postes Ãă pourvoir pour les personnels Ãă recruter.

Pour les thèses, préciser si des demandes de bourse de thèse sont prévues ou en cours, en préciser la nature et la part de financement imputable au projet.

5.3 Prestation de service externe / Subcontracting

Préciser : la nature des prestations, le type de prestataire.

5.4 Missions / Travel

Préciser : les missions liées aux travaux d'acquisition sur le terrain (campagnes de mesures'), les missions relevant de colloques, congrès'

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

Préciser la nature des prestations.

5.6 Autres dépenses de fonctionnement / Other expenses

Toute dépense significative relevant de ce poste devra Ãltre justifiée.

6 Références bibliographiques / References

Inclure la liste des références bibliographiques utilisées dans la partie Etat de l'art et les références bibliographiques des partenaires ayant trait au projet.

Références

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

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

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

[6] 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.

[7] G. Contardo, L. Denoyer, T. Artieres, and P. Gallinari. Learning states representations in pomdp. In *ICLR*, 2014.

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

[9] 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.

[10] M. Dontcheva, G. D. Yngve, and Z. Popovic. Layered acting for character animation. *ACM Trans. Graph.*, 22(3) :409–416, 2003.

[11] 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.

[12] N. Fourati and C. Pelachaud. Head, shoulders and hips behaviors during turning. In *Worshop on Human Behavior Understanding (HBU), Springer LNCS 8212*, pages 223–234, 2013.

[13] G. Gagneré, R. Ronfard, and M. Desainte-Catherine. La simulation du travail théâtral et sa notation informatique. In *La notation du travail théâtral : du manuscrit au numérique*, 2012.

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

[15] 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.

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

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

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

[19] 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.

[20] 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.

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

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

[23] 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.

[24] 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.

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

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

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

[28] 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.

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

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

[31] 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.

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

[33] C. O'Reilly and R. Plamondon. A globally optimal estimator for the delta-lognormal modeling of fast reaching movements. 42(5) :1428–1442, 2012.

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

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

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

[37] 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.

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

[39] R. Ronfard. Notation et reconnaissance des actions scéniques par ordinateur. In *La notation du travail théâtral : du manuscrit au numérique*, 2012.

[40] 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.

[41] 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.

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

[43] 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.

[44] 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.

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

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

[47] J. Wiggins. A psychological taxonomy of trait-descriptive terms : The interpersonal domain. 33, 1979.

[48] 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.