

# Towards embodied and situated virtual humans

Stéphane Donikian<sup>1</sup> and Sébastien Paris<sup>2</sup>

<sup>1</sup> IRISA, Campus de Beaulieu, 35042 Rennes cedex - France - donikian@irisa.fr

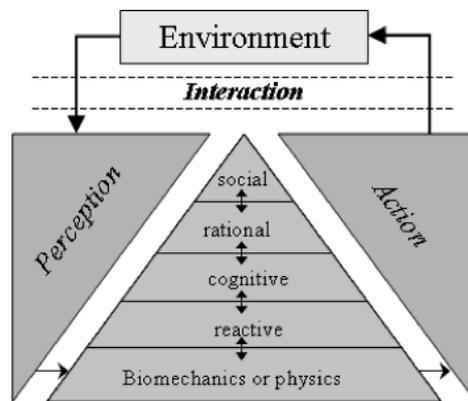
<sup>2</sup> Laval University, Department of Computer Science and Software Engineering,  
Quebec, Canada - paris Sebastien@free.fr

**Abstract.** The main objective of the Bunraku project-team is to develop cross fertilization of researches in the fields of virtual reality and virtual human. Our challenge is to allow real and virtual humans to naturally interact in a shared virtual environment. This objective is very ambitious as it requires developing and federating several research fields. However, it is fundamental for several areas such as taking into account the human activity in manufacturing, the individual or collective training process, or the human study in cognitive sciences. In this paper, we will focus on the research done so far in the team that are concerning autonomous virtual humans.

## 1 Introduction

Virtual Human has at least two usages in a virtual reality environment, the first one is to represent a real human in the virtual world and it is then called an avatar, and the second one is to live autonomously inside the virtual environment, that is to say, perceive, decide and act on its own. Both uses require a geometrical representation of the human, but in the second case it is also necessary to model the behavioural activity instead of reproducing the activity of a real character. Modelling the human behaviour requests to take into account a certain number of topics such as understanding mechanisms underlying functions such as natural language production and perception, memory activity, multi-sensory perception, muscular control and last but not least the production of emotions. In short, it is necessary to be interested in the operation of various faculties that constitute together the human spirit, without forgetting their relation with the body. In complement of the study of these general mechanisms underlying any human behaviour, the work should also concern the study of human faculties in dedicated activities such as navigating in a city, using a work instrument or conducting a structured interview. The comprehension of the human behaviours requires competence in fields as varied as neurosciences, psychology or behavioural biology. Two types of approaches can be distinguished. The first one, known as the symbolic approach, consists in modelling the human behaviour in an abstract way in the form of modules describing each one a mechanism and relations of sequentialism or existing parallelism between them. It seeks to describe the mental processes by using symbols, judgements and mechanisms of logical inference. The second approach, known as systemic, consists to look inside the cerebral activity of patients subjected to various stimuli,

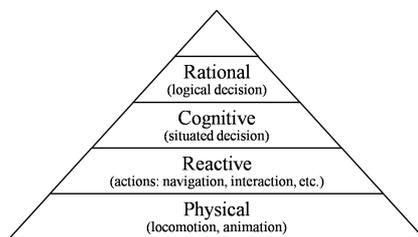
according to well defined operational protocols. It is focusing more on concepts of signal transmission in networks, control and state feedback. The two approaches have different advantages: the first makes it possible to be abstracted from the biophysics processes present within the brain and to propose a modelling of the behaviour based on competence, while the second approaches, nearer to the neuro-physiological data, will be adapted to the modelling of the neuronal and sensori-motor activities. None of the models proposed in the two approaches is completely satisfactory to model the human behaviour in its whole. Indeed, our problem is not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviours of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists or correspond to an imaginary universe described by a scenario writer. However, the proposed architecture should mimic all the human intellectual and physical functions.



**Fig. 1.** The behavioural pyramid adapted from A. Newell[15].

The important bibliographical study made during last years in the field of cognitive sciences pointed out the diversity, even the antagonism, of the approaches and the absence of federator models to propose an architecture allowing to connect together the various functions used in the human behaviour even for the simplest. The various approaches generally focus on one or the other of the functions or on a particular method of their relation. No theory exists for determining either the necessary or sufficient structures needed to support particular capabilities and certainly not to support general intelligence. There is however a general agreement on the decomposition into several layers (cf figure 1) or bands going from very low level control loops (reactive level), providing very fast responses to stimuli (sensory-motor control), to higher levels such as the cognitive one manipulating and reasoning on symbols, and the social one including personality, emotions and relation between humans. The reactive layer does not need to explicitly manipulate an abstract representation of the world,

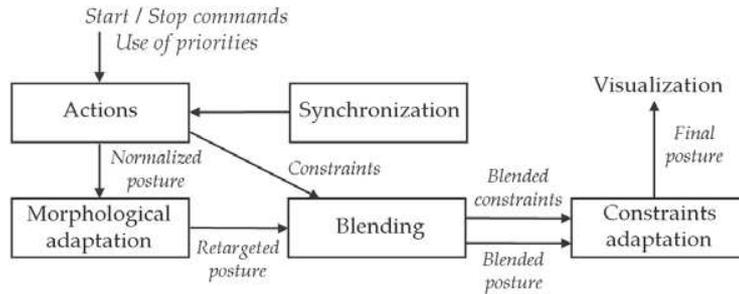
while the cognitive layer manages the abstract knowledge representation system. One of the main difficulties in the existing models is that usually symbols manipulated at the upper levels are not grounded into the world in which the lower level is making the virtual human react to its environment. The symbol grounding problem and the Chinese room problem are well known in artificial intelligence. We are proposing the first cognitive architecture proposing a full bidirectional link between four of the five layers: biomechanical, reactive, cognitive and rational (cf figure 2). Each layer exchanges specific information with the directly upper layer to inform it of some imposed constraints, and also with the directly lower layer to control it. Thereby, each layer is independent from the others, only requiring a set of identified data to work. Let us present now those layers, starting with the lower one.



**Fig. 2.** Our multilayered model of decision.

## 2 Biomechanical Layer

The objective of this layer is to compute believable movements for human-like characters in real-time and interactive environments. Controlling the motion of virtual humans requires addressing four key points. The first one focuses on post-processing of raw data provided by motion capture systems [9] captured on a unique skeleton whereas a target skeleton will generally differ from this original one. The second point [8] consists in retargeting motion by using morphological matching and adaptation to the environment. It involves the solving of constraints in space and time but it does not meet the interactivity requirements. A third solution [2] is based on a frame by frame inverse kinematics approach, that does not guarantee a real-time animation for a high number of degrees of freedom or a high number of tasks to be respected. However, this solution allows the use of biomechanical knowledge when adapting the motion. The last point consists in using a mechanical model based on dynamical laws of motion. Once basic actions are available, it is possible to combine them in order to solve more complex tasks. For example, motion planning requires interpolating various kinds of gaits [20] or generating footsteps in order to drive a character from an initial to a goal configuration.



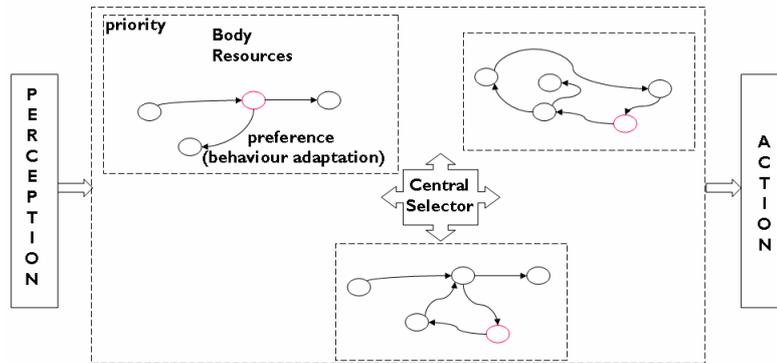
**Fig. 3.** General Architecture of our Animation Engine.

We have proposed methods to synchronize, merge and blend several captured motions (see figure 3)[13]. Each motion is seen as an action, which provides a succession of postures and constraint equations. The role of the synchronisation module is to perform dynamic time warping if actions are not compatible together. The system is then able to make the character step in environments, and move around without any a priori knowledge and without any pre-computation of possible transitions. As the target character is generally different from the original one, we have proposed motion retargeting methods based on normalised representations of the skeleton [10]. An efficient inverse kinematics and kinetics solver based on this specific representation of the skeleton solves the geometric constraints, such as the contact constraint of the feet on the ground. It enables animating more than 500 characters in real time (30Hz with a commodity PC). Moreover, the morphology can also modify the gesture, which requires designing a function linking the motion to morphology. This is achieved by interpolating gestures, which belongs to a database indexed by morphological parameters, such as limb lengths and proportion between segment lengths [21]. This method was applied to the simulation of locomotion for several different hominids. Whatever the used kinematical method, ignoring dynamics may lead to some artefacts. Dynamical parameters (mass and inertia), when taken into account, lead to more realistic motions. However, it entails more complex computations and related models are more difficult to control. Although we have introduced those parameters in secondary tasks of inverse kinematics offering more plausible motions, but it cannot generate a motion that always respects external forces. One of the main problems that still remain consists in understanding and modelling the laws that capture the naturalness of human motion in order to simulate realistic gestures.

### 3 Reactive Layer

Autonomous virtual humans are able to perceive their environment, to communicate with others and to perform various activities in accordance with the nature of the environment and with their intentions. Different approaches have

been studied for the decision part of the reactive layer: sensor-effector or neural networks, behaviour rules, finite automaton. However, none of these approaches were sufficient to handle concurrency and time management in behaviours. The various approaches can be shared within a dataprocessing model implementing a hierarchy of "perception-decision-action" loops in which the decisionmaking process is real-time, and made up of concurrent and hierarchical state-machines having different scales of time. We have proposed the HPTS (Hierarchical Parallel Transition Systems) model [6]. It is composed of a language allowing the description of reactive behaviours through finite state machines and a runtime environment handling parallel state machine execution and offering synchronization facilities. States and transition can be redefined in the inheritance hierarchy and the state machines can be augmented with new states and transitions. The compilation phase translates a state machine in a C++ class that can be compiled separately and linked through static or dynamic libraries. The runtime kernel handles parallel state machine execution and provides synchronization facilities.



**Fig. 4.** Automatic coordination of reactive behaviours.

V. Decugis and J. Ferber [5] combine reactivity and planning capabilities in a real-time application, by extending the Action Selection Mechanism (ASM) proposed by P. Maes [12] into hierarchical ASMs. In the contrary of the ASM approach, W.J. Clancey [3] is defending that an activity does not have to be interrupted when another one requests to be performed. Instead, he defends the notion of competitive activation mechanism. In the above mentioned models, it is relatively complex to reproduce this type of interlaced behaviours, due to their mutual inhibition. We have proposed a new approach whose objective was to launch the required behaviours and let them arrange automatically according to the circumstances, the desires and the physical constraints of the agent [11]. Three new concepts were introduced within our reactive model: resources used at each step of behaviour, behaviour priority upon time, and degrees of preference

allowing the adaptation of each behaviour to its context. A scheduler or central selector (cf figure 4) is trying, as well as possible, to respect the allowance of the resources according to the current priority of each behaviour. The description of a new behaviour, thanks to the mechanism of deadlock avoidance, does not require the knowledge of other already described behaviours. In this model, known as HPTS++, the behaviour coordination becomes thus generic and automatic.

## 4 Cognitive Layer

As humans are deliberative agents, purely reactive systems are not sufficient to describe their behaviour, and it is necessary to integrate the cognitive aspect of behaviour (beliefs and intentions) [7]. In order to do so, Badler et al. [1] propose to combine Sense-Control-Action loops with planners and PaT-Nets. In this system, like in others [16], the action is directly associated with each node, which doesn't allow the management of concurrency. Other planning systems elaborate plans by reasoning on actions and their associated resources [14]. However, such kind of approach is based on an omniscient point of view, where the agent can obtain any information from the informed environment to plan its path, which limits the realism of the simulation. Another approach, based on artificial vision, is using information retrieval from an image captured by a camera located on the head of the virtual human. This process, using the well known Zbuffer computes the perception of a virtual human. N. Courty [4] has extended this kind of approach by blurring the peripheral vision area in the perceived image and by introducing a saliency map calculation. Salient points are extracted from the perceived image and used in the visual attention guidance of a virtual human navigating without any goal in an urban environment. C. Peters et al. [19] use also an artificial vision model, but in their case it is coupled to a human memory model allowing to manage scanning and object retrieval techniques inside an apartment. A human being navigating in an environment is confronted to the problem of his own spatial localization. An approach is to consider that he or she is using a geocentric spatial cognitive map to navigate. Indeed, most of the cognitive works on that subject state that the knowledge a pedestrian can have of his environment differs a lot from what is really the environment: his perception and memory are most of the time incomplete and distorted [23]. Then it seemed interesting to us to study what should be added to the classical navigation architecture in behavioural animation, to simulate more realistic navigation behaviour. After a careful study of various works in the cognitive psychology field, the introduction of a spatial cognitive map and a human-like memory appeared necessary. The map is necessary to restrict the omniscience of the agent, and it allows the model to exhibit a more realistic navigation behaviour, due to the fact that planning is computed with incomplete knowledge of the environment. As the memory is dedicated to take into account a temporal factor in the simulation, the major consequence is that, paths taken to go from one location to another will not be the same at different moment during the simulation as the state of the agents memory is continuously evolving [22].

## 5 Rational Layer

This layer is in charge of the logical decisions of the autonomous agent. Decisions are based on un-embodied notions describing the goals and sub-goals of the agent at a conceptual level. This layer also organises these goals to find the feasible actions independently of their location. The logical decision of the autonomous agent is based on concepts describing its goals in the environment. These concepts are unified by an architecture which defines the way they operate together: BIIO for Behavioural Interactive and Introspective Objects. BIIO manages the concepts as objects in the sense of object oriented languages, i.e., by allowing the creation of a hierarchy of concepts specialising or unifying already existing ones. Another characteristic of BIIO is that any concept is endowed with introspective abilities, and thus able to communicate its components to others. This architecture also proposes a predefined set of concepts related to interaction behaviours. We define an interaction as any action between an autonomous agent and another embodied element situated in the environment, such as another autonomous agent or an equipment.

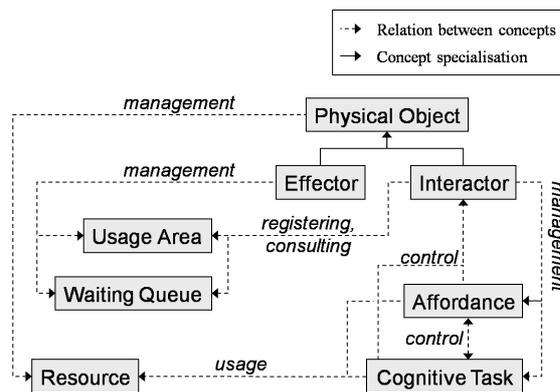


Fig. 5. Predefined set of concepts introduced within BIIO

The set of concepts we introduce are defined in figure 5. *Physical Object* describes a situated object whose location is identified inside the environment. This concept is specialised in two other concepts: *Effector* and *Interactor*. An effector represents something which sustain an interaction, i.e. which participate to an interaction without having initiated it. represents an autonomous agent which can realise an interaction. An interactor is able to manage two concepts representing its behaviours: the *affordance* and the *cognitive task*. An affordance describes an interaction between an interactor and an effector, and so a behaviour of the interactor which is directly observable. These interactions are defined globally in the virtual environment, allowing to easily add new affordances without having to modify the effectors nor the interactors. Thereby, effectors and interactors

are able to collect the affordances they are compatible with thanks to the introspective abilities provided by BIIO. Then, they have some generic processes to manage the collected affordances.



**Fig. 6.** Some virtual humans waiting for the use of a vending machine.

A cognitive task describes an internal behaviour of the interactor, which is not directly observable. A resource symbolises a property of a physical object, such as electricity for an equipment, or the ability to move for an autonomous agent. A resource can be transitory or permanent, and can have multiple internal properties which might evolve. Both behavioural concepts (affordance and cognitive task) use resources to define their run, as for example to manage their competition. A usage area describes the positioning of an interactor in order to use an effector. A waiting queue describes the way the interactors must socially organise in order to wait for a specific interaction with an effector (cf figure 6). Waiting queues are managed on the same way than usage areas. In this approach, the situated decision is managed by a specialisation of the concept of cognitive task, *interact*, which is related to interactions. The goal of this specialised behaviour is to create a connection between the abstract decision of the interactor and its embodied abilities: perception, path planning[17], and navigation[18]. In fact, this cognitive task handles the competition between the interactors active affordances for the control of the agents physical abilities. *Interact* takes place as a hierarchical state machine, where each automaton has a specific role (cf figure 7).

The situated decision ability of our agents allows us to simulate complex behaviours with a really simple description phase: we just have to create the equipments (effectors) with their corresponding affordances, and then to assign an agent its final goal. For example, the behaviour of an outgoing passenger is shown in figure 8: the final goal assigned to the interactor is to reach a desti-

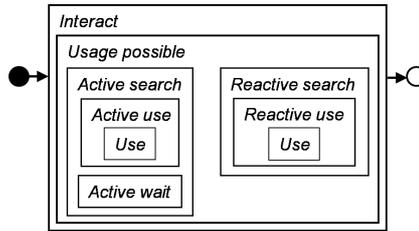


Fig. 7. The hierarchical state machine of *interact*

nation by train, producing a set of sub-goals in order to buy a ticket (which creates a ticket resource added to the interactor), check the departure board, and punch the ticket (which modifies a property of the ticket resource); all of these interactions are performed with equipments which are chosen thanks to the interactors situated cognition ability.

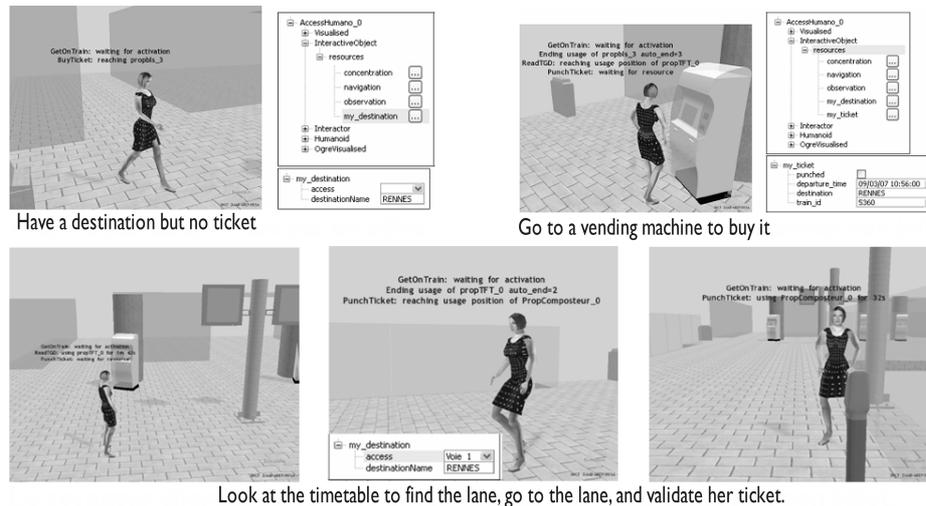
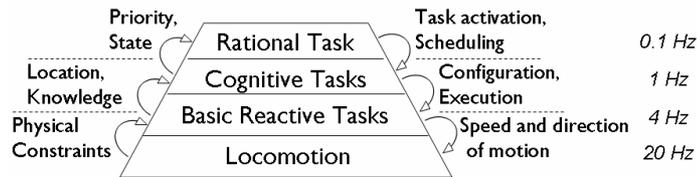


Fig. 8. Example of the goal oriented behaviour of an outgoing passenger.

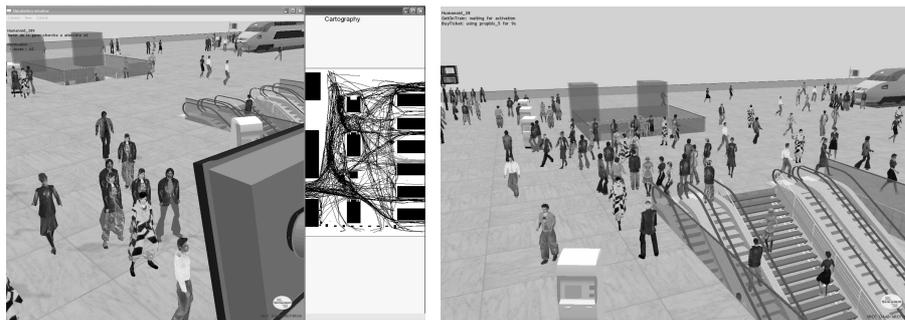
Both models of decision presented in this paper, logical and situated, have been unified in a multilayer model of virtual agent (cf figure 9). The *interact* cognitive task manages embodied and situated behaviours, like navigation, with a purely logical process based on the abstract concept of affordance. This task can handle generically any affordance with a defined hierarchical state machine, and is able to take into account new affordance without being modified. A final problem needs addressing, which consists in choosing an effector inside the environment. This second part of the situated decision process is used by the active

search automaton of the interact task. Its goal is to find the best couple of affordance and effector considering the relative priorities of the active affordances of the interactor, and the distance to travel to reach the effector. To do so, a path planning is performed inside the environment, whose ending condition is not a position, but an effector which is compatible with an active affordance.



**Fig. 9.** A multilayered architecture of an embodied and situated virtual human.

The performances of the decision process are very good, with lesser than 0.5ms to take an initial decision, and with a negligible cost while only refreshing that decision (while walking towards a selected effector). Such high performances allow us to simulate one thousand of people in real time on a standard computer with full behavioural abilities, including the reactive and physical layers. This result is obtained with graphical output, which is responsible for approximately 50% of the total computation time. Left image of figure 10 is illustrating some characters waiting for the lane of their train in front of an information board. A window allows also to show trajectories followed by all pedestrians in the train station with the ability to discriminate by using a different colour those who will take a specific train. The right image illustrates the different activities of people inside the train station.



**Fig. 10.** User activities in a train station.

## 6 Conclusion

The principal key words of the whole work done so far on virtual humans are: how to model it, animate it, control it, and define its interactions with its environment and its congeneric. We wish in an immediate future to gather the whole of the work completed around the virtual human ranging from reactive to social behaviours and offering facilities for perception and action modelling. Those models have to take into account human behaviour characteristics in order to produce complex and believable movements and behaviours. Moreover, it has to respect the virtual reality constraints: real-time control and interactivity. To combine autonomy and believability, we are working on a unified architecture to model individual and collective human behaviours. This architecture should include reactive, cognitive, rational and social skills and manage individual and collective behaviours. This is one of our current challenges. We have proposed with BIIO a first answer to the symbol grounding problem by combining in a generic way cognition with a reactive embodied and situated human character. To tackle the embodiment of cognitive symbols we have developed a complex hierarchy of perception-decision-action loops, by managing the bidirectional exchange of information between the different levels. We will continue to develop this model by adding the social layer to the pyramid. We plan also to develop an audio-visual attentive coordination, integrating the management of human memory, the filtering of attention and the cognitive load. At the moment, motion control and behaviour model are usually seen as two independent software components, which does not allow a good solution for prediction and adaptation to the context. We will have to provide a better integrated solution, allowing the management of adaptable and expressive models of motion.

## References

1. N. Badler, B. Reich, and B. Webber. Towards personalities for animated agents with reactive and planning behaviors. *Lecture Notes in Artificial Intelligence, Creating Personalities for synthetic actors*, pages 43–57, 1997.
2. B. Le Callenec and R. Boulic. Interactive motion deformation with prioritized constraints. In *ACM/Eurographics Symposium on Computer Animation*, pages 163–171, Grenoble, France, august 2004.
3. William J. Clancey. Simulating activities: Relating motives, deliberation, and attentive coordination. *Cognitive Systems Research*, 3(3):471–499, 2002.
4. N. Courty. *Animation référencée vision : de la tâche au comportement*. PhD thesis, INSA, Rennes, France, Novembre 2002.
5. V. Decugis and J. Ferber. Action selection in an autonomous agent with a hierarchical distributed reactive planning architecture. In Katia P. Sycara and Michael Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pages 354–361, New York, Mai 1998. ACM Press.
6. S. Donikian. Hpts: A behaviour modelling language for autonomous agents. In *Fifth International Conference on Autonomous Agents*, pages 401–408, Montreal, Canada, may 2001. ACM Press.

7. J. Funge, X. Tu, and D. Terzopoulos. Cognitive modeling: Knowledge, reasoning and planning for intelligent characters. In *Siggraph99*, pages 29–38. ACM Press, August 1999.
8. M. Gleicher. Retargetting motion to new characters. In M.F. Cohen, editor, *ACM SIGGRAPH*, pages 33–42, 1998.
9. L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Transactions on Graphics*, 21(3):473–482, 2002.
10. R. Kulpa, F. Multon, and B. Arnaldi. Morphology-independent representation of motions for interactive human-like animation. *Computer Graphics forum*, 24(3), 2005.
11. F. Lamarche and S. Donikian. Automatic orchestration of behaviours through the management of resources and priority levels. In *First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, pages 1309–1316. ACM Press, July 2002.
12. P. Maes. Situated agents can have goals. In Patti Maes, editor, *Designing Autonomous Agents*, pages 49–70. MIT Press, 1990.
13. S. Ménardais, F. Multon, R. Kulpa, and B. Arnaldi. Motion blending for real-time animation while accounting for the environment. In *IEEE CGI*, june 2004.
14. A. Nareyek. A planning model for agents in dynamic and uncertain real-time environments. In *AIPS-98 Workshop on Integrating Planning, Scheduling and Execution in Dynamic and Uncertain Environments*, pages 7–14, 1998.
15. A. Newell. *Unified theories of cognition*. Harvard University Press, Cambridge, Massachusetts, 1990.
16. H. Noser and D. Thalmann. Sensor based synthetic actors in a tennis game simulation. In *Computer Graphics International '97*, pages 189–198. IEEE Computer Society Press, 1997.
17. S. Paris, S. Donikian, and N. Bonvalet. Environmental abstraction and path planning techniques for realistic crowd simulation. *Computer Animation and Virtual Worlds*, 17(3-4):325–335, 2006.
18. S. Paris, J. Pettré, and S. Donikian. Pedestrian reactive navigation for crowd simulation: a predictive approach. *Computer Graphics Forum*, 26(3):665–674, september 2007.
19. C. Peters and C. O'Sullivan. Synthetic vision and memory for autonomous virtual humans. *Computer Graphics forum*, 21(4):743–752, 2002.
20. J. Pettré, J.-P. Laumond, and T. Siméon. A 2-stages locomotion planner for digital actors. In *Proc. of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'03)*, pages 258–264, 2003.
21. N. Pronost, G. Dumont, G. Brillon, and G. Nicolas. Morphological and stance interpolation in database for simulation of bipedalism of virtual humans. *The Visual Computer*, 22(1), 2006.
22. R. Thomas and S. Donikian. A spatial cognitive map and a human?like memory model dedicated to pedestrian navigation in virtual urban environments. In *Spatial Cognition*, volume 4387 of *Lecture Notes in Computer Science*, pages 421–438, Bremen, Germany, september 2006. Springer-Verlag.
23. B. Tversky. Structures of mental spaces. In *3rd International Space Syntax Symposium*, Atlanta, USA, 2001.