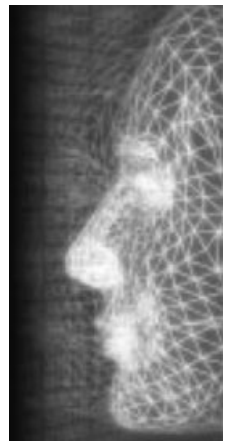
COMPUTER ANIMATION AND VIRTUAL WORLDS
Comp. Anim. Virtual Worlds 2004; **15:** 311–318 (DOI: 10.1002/cav.34)

An Artificial Life Environment for Autonomous Virtual Agents with multi-sensorial and multi-perceptive features



*By Toni Conde and Daniel Thalmann**

Our approach is based on the multi-sensory integration of the standard theory of neuroscience, where signals of a single object coming from distinct sensory systems are combined. The acquisition steps of signals, filtering, selection and simplification intervening before proprioception, active and predictive perception are integrated into virtual sensors and a virtual environment. We will focus on two aspects: 1) the assignment problem: determining which sensory stimuli belong to the same virtual object and (2) the sensory recoding problem: recoding signals in a common format before combining them. We have developed three novel methodologies to map the information coming from the virtual sensors of vision, audition and touch as well as that of the virtual environment in the form of a 'cognitive map'. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: artificial life; perception; virtual sensors; virtual environment; virtual autonomous agents

Introduction

An Autonomous Virtual Agent (AVA) situated in a Virtual Environment (VE) is equipped with sensors for vision, audition and touch that inform it of the external VE and its internal state. An AVA possesses effectors to let it exert an influence on the VE and control architecture to coordinate its perceptions and actions. The behavior of an AVA is adaptive as long as the control architecture allows it to maintain its variables in their validity zone.

Our *Artificial Life Environment (ALifeE)* based on an original approach inspired by neuroscience equips an AVA with the main virtual sensors in the form of a small *nervous* system.[1] The control architecture is kept simple to optimize the management of the AVA's virtual sensors and perception. The processes of filtering, selection and simplification are carried out after obtaining the sensorial information. This approach allows us to achieve some persistence in the form of a 'cognitive map'.

*Correspondence to: Daniel Thalmann, Virtual Reality Lab, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland.
E-mail: daniel.thalmann@epfl.ch

The 'mental processes' of an AVA can be simulated. *Behavioral animation* includes the techniques applied to make an AVA intelligent and autonomous, to react to its VE and to make decisions based on its perceptive system, its short-term memory and long-term reasoning. *Intelligence* is the ability to plan and carry out the tasks based on the model of the current state of the VE.

Our objective is to permit the AVA to explore unknown VEs and to construct *mental structures and models, cognitive maps or plans from this exploration.* Once its representation has been created, the knowledge can be communicated to other AVAs. Each AVA perceives objects and other AVAs with the help of its VE, which provides information concerning their nature and positions. The behavioral model decides which action the AVA should take (such as walking or handling an object) and then uses the knowledge.

Contributions. In this paper, we present three novel methodologies:

- The first technique integrates several virtual sensors in the same multi-sensorial control architecture.
- The second technique integrates different perception types of an AVA coupled with its sensors.

- The third technique allows the fusion of multisensorial information in the form of two 'cognitive maps' and 'internal visual memory' before the AVA's learning and evolving processes.

Document organization: We will begin by examining background research work. Then we will introduce our methodology. Following that, we will present our realization and experimental results. Finally, we will conclude and suggest possible directions for future work.

Virtual Sensors Background

An AVA should be equipped with virtual sensors for vision, audition and touch. These sensors constitute a starting point to implement behavior such as direct vision during a move, handling of objects and responding to sounds or words.

Our *ALifeE* integrates in an important way the main virtual sensors of an AVA as in the following research: *Virtual Vision* proposed by,[2–4] *Virtual Audition* proposed by[5] and *Virtual Touch* proposed by.[6] After acquiring the information, the basic perceptive part of the AVA is carried out by the *Flexible Perception Pipeline* approach proposed by.[7]

Synthetic Vision

Synthetic vision is the main information channel between the VE and the AVA.[8,9] The AVA perceives its virtual scene from a window and the resulting information is sufficient for local navigation. The mixture of image recognition and representation of its VE allows the AVA to react in real time. Noser and al.[3] have used an *octree* as an internal representation of the AVA's view of its VE. These approaches propose a visual memory of the AVA in a 3D environment composed of static and dynamic objects.

The perception of the AVA in its VE is communicated by vision and sound, sometimes by sensorial tactile information. Its behavior, as in humans, is strongly influenced by data supplied by its sensors and its own intelligence for certain ends such as: extraction, simplification and filtering, which depend on perception criteria associated with each sensorial modality. All of this is integrated in the perceptive part of our *ALifeE*.

The AVA explores an unknown environment constructed on mental models as well as a 'cognitive map' based on this exploration. Navigation is carried out in two ways: globally (with the pre-learned model of the VE, a few changes and the search for performance with a *path planning* algorithm) and locally (with direct acquisition of the VE). A 3D geometrical model in the form of a grid is implemented with the help of an *octree* combined with the approach proposed by.[3,4]

Synthetic Audition

In real life, the behavior of people or animals can be strongly influenced by sounds. Wenzel[10] calls this 'the function of the ears is to point the eyes'. Audition is a temporal sensor, which is very sensitive to changes in acoustic signals. We can locate objects in space and, even more specifically, when they move. Moreover, acoustic signals carry a lot of semantic and emotional information; they inform us about sound sources relative to us as well as the propagation of sound paths in an acoustic environment. In our *ALifeE*, the restitution of sound must be very effective to react to sound events perceived by the AVAs in each frame.

The most important properties of a sound source in terms of computer knowledge are: 3D position of the source in the world, orientation of the sound source, cone of propagation, distance between the listener and the source, Doppler effect, volume and frequency, occlusion, obstruction and exclusion.

All these parameters can be set to filter the sound source depending on the simulation conditions. Regarding 3D sound, one may believe that it is sufficient to place the sound source in a 3D world without taking care of its direction.[11] However this is too big a limitation, especially for reverberations and reflections. In our *ALifeE*, we represent the sound propagation with a cone. This solution gives us the flexibility to set specifically the different filters for each sound source.

Synthetic Touch

Sensorial tactile information can be used to push buttons or to touch and handle objects. The simulation of this kind of sensor resembles the collision detection proposed by.[12] We have rather opted for the process described by[6] with V-Collide collision detection approach.

The V-Collide approach performs efficient and exact collision between triangulated polygonal models. It uses a 2-level hierarchical approach:

- The top level eliminates from consideration pairs of objects that are not close to each other,
- The bottom level performs exact collision detection down to the level of the triangles themselves.

Synthetic World

In synthetic vision, the vision models for an AVA are different from those used in behavioral robotics. A robot can only acquire information from its environment through these sensors, which limits its behavior in navigating and avoiding obstacles. In a VE, supplementary information can be extracted and dealt with according to the perception model chosen for the AVA. This makes it faster and 'intelligent' during actions. To optimize the model we chose a certain type of representation of the virtual world where the AVA maintains for example its vision at a low level system.

Nevertheless, because of scale and plausibility constraints of its autonomous behavior, an AVA restricts its perception locally in relation to the VE as a whole. This approach is used more specifically when there are a lot of AVAs as described by.[8] Most important, the choice of the method must reflect an AVA in a VE close to reality.

Perception Background

As mentioned above, considerable restrictions appear when the actions produced by the AVA require dynamic knowledge of the VE with a perception system. One of these restrictions concerns the reflex actions, which require perception, but not memory concerning what has been perceived. In the classical approaches different behaviors implement their own perception mechanisms.

Several methods like the one used by[7] have been proposed to implement perception. The AVA maintains a perception puzzle, each piece corresponding to a specific *virtual sensor*. A pipeline is composed of filters to extract relevant information from the data supplied by the related *sensor*. Our attempt to model an approach of unified perception is described by.[13] We will define the main ideas for the implementation of our *ALifeE* in the following subsections.

Proprioception

Proprioception is inspired by the human immune system composed of functional layers combining rapid and archaic mechanisms of innate immunity and slower mechanisms of acquired or adaptive immunity. The response time depends on the anterior exposition to pathogen.

Our proprioception is based mainly on the integration in a same model of endogenous variables, homeostasis and reinforcement learning as proposed by[14] and adapted for the *ALifeE*:

- The notion of 'endogenous' variables captures information related to the internal state of the AVA that is influenced by both the AVA's perceptions and actions but not restricted to them. Additional influences permit differentiation between the effects of similar perception inputs, on the resulting AVA's action. The existence of these variables constitutes an important type of cognitive intermediary between the sensory and virtual human controller poles of the behavior loops.
- The notion of 'homeostasis' describes variables whose temporal dynamics must guarantee their keeping within pre-determined boundaries. Since exceeding these boundaries in human beings would result either in significant discomfort or in the death of the AVA, actions are taken to prevent these variables from departing from the set-value.

In order to be truly autonomous, the AVA must not just be capable of intelligent action, but also be self-sustaining. The third aspect of the model is the use of a reinforcement learning mechanism. This enables the discovery of a sequence of actions, which allows the AVA to remain viable despite the strong constraints exerted by the VE and the AVA's own endogenous variability.

Active Perception

In a VE, an AVA requires a combination of perception and action to behave in an autonomous way. The perception system provides a uniform interface to various techniques in the field of virtual perception, including synthetic vision, audition and touch. In usual approaches, different behaviors implement their own perception mechanisms, which leads to computation duplication when multiple behaviors are involved. Basically, an AVA maintains a set of perception pipelines, each corresponding to a particular type of virtual sensor.[7] A pipeline is composed of filters that coordinate themselves in order to extract relevant information from the data sent by the associated virtual sensor.

Predictive Perception

The faculty of predicting, one of the main activities of the human brain is an essential notion in the perception of an AVA. It plays a basic role in *active perception* by

giving the AVA the possibility to direct it's look and attention elsewhere.

This prediction can be found in humans when anticipating the path a ball they should catch will take, avoiding mobile obstacles, preparing the body to wake up during the final hours of sleep, or even in the absurd effectiveness of a placebo.[15]

To obtain an active perception like directing the look and attention elsewhere, prediction functions must be organized. In the visual system alone, certain zones of the cortex deal with outlines, others with forms, movement, distance or color. These processes are unconscious.

Our model of predictive perception is based on the mathematical theory of the observer. Algorithms are used to predict from partial measures, often external and with sound effects, the internal state of a non-linear system. An observer is typically composed of a system simulation that uses an internal model that may be approximate. It is guided and corrected by the measures taken on by the system. In problems of *active perception* and in certain circumstances, the observer also allows the selection of the measure or combination of measures to be carried out. This is particularly useful in improving the estimation of the system state at a given moment; this is inspired by the human nervous system.

Methodology

An AVA is not only situated in an environment with the help of virtual sensors but it must be equipped with behavior, perception and memorization faculties to make it autonomous and 'intelligent'. Our objective is to give an AVA the ability to explore an unknown environment and thus construct mental models and 'cognitive maps'. During or after the construction of these models, the AVA can carry out many functions successfully, for example 'path-planning', navigating and 'place-finding'.

Our *ALifeE* model is based on the multi-sensory integration of the standard theory of neuroscience (Figure 1). Signals related to the same virtual object but coming from distinct sensory systems are combined. We will focus on two aspects: 1) the assignment problem: determining which sensory stimuli belong to the same virtual object and 2) the sensory recoding problem: recoding signals in a common format before combining them.[1]

We have developed a multi-sensorial approach based on a 3D geometric model with a grid implementing an *octree* since humans do not carry out spatial reasoning based on a continuous map, but on a discrete one.[16] Our computation of the *octree* was achieved using the fast *voxelisation* method developed by.[17]

Our goal was to introduce the equivalent of a small nervous system into the control architecture, thus linking its sensors and its effectors. Learning can modify the organization of the control architecture and that of the evolving process at the same time. The latter will be the object of our future research, as these processes are the main adaptive ones that nature has invented to ensure the survival of living beings.

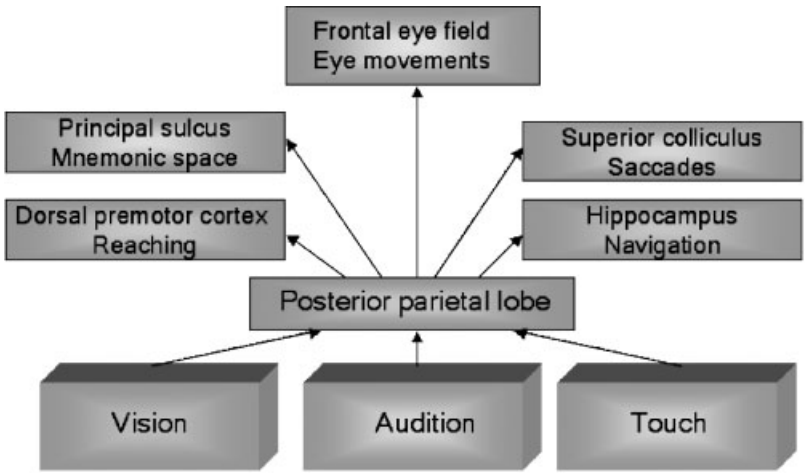


Figure 1. A schematic representation of the standard theory for multi-sensory spatial integration and sensory-motor transformations. Sensory modalities encode the location of objects in reference frames that are specific to each modality. Multi-sensory integration occurs in multiple modules with the parietal cortex.

An AVA is able to explore its VE and work out a mental representation of its spatial organization in the form of a 'cognitive map'. It can then use it to locate itself and reach a given goal. This ability is based on the use of a performing visual system as described before. An AVA learns a more general *model of the virtual world* during its interactions with the VE. This model helps it anticipate how the VE changes depending on actions that are performed.

An AVA is *situated* in a simulated VE with sensors for vision, audition and touch, which inform it of its external VE (active and predictive perception) or its internal state (proprioception). An AVA has effectors permitting it to act on the VE and a control architecture that *coordinates its perceptions and actions*. The behavior of an AVA is adaptive as long as the control architecture allows it to maintain its essential variables in their viability zone (e.g. a corrective action was accomplished at point B, to avoid leaving the viability zone at point A). The control architecture plays the role of a motivational system when it is used to choose successive goals that the AVA is trying to reach or to arbitrate between conflicting goals.

The auditive position of an object is predicted from its visual position. This requires the transformation of a reference system whose origin is a vision coordinate (eye position) to a reference system whose origin is an audition coordinate (head position). The comparison of the results can be used to determine whether the signals from the two types of virtual sensors belong to the same object.

Our multi-sensorial approach (Figure 2) integrates the behavior model of an AVA. The control architecture is standardized with sub-modules covering the different techniques necessary for the artificial simulation of the AVA's behavior.

A major problem in behavioral learning is the introduction of automatic learning techniques in multi-agent systems. It is a challenge to teach multi-agent systems how to behave, interact or get organized in order to improve their collective performances in carrying out a task.

In this context, two major obstacles are encountered:

- The choice of a learning technique relevant to the task and the level. It should allow comparison between similar problems.
- The choice of a learning protocol.

Our *ALifeE*'s objective is to combine these two generic sub-parts, namely a specific form of learning with an AVA's simulator.

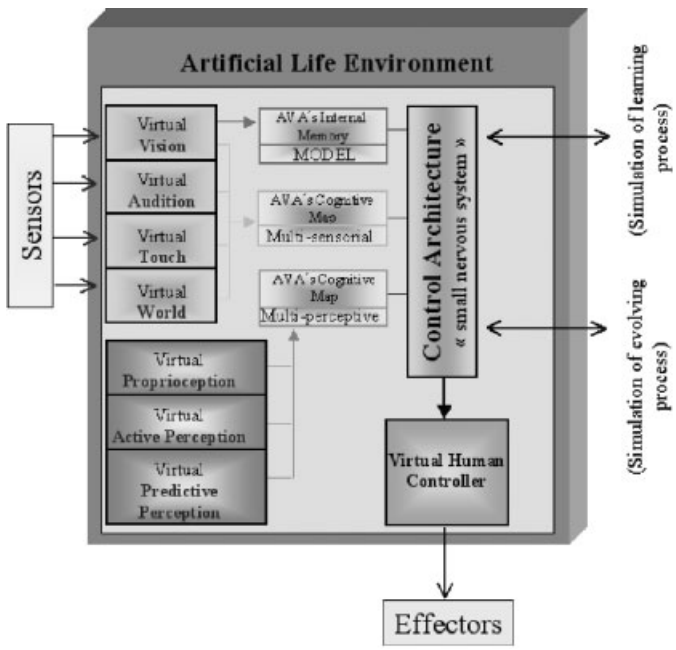


Figure 2. A schematic representation of our ALifeE. Virtual Vision discovers the VE, constructs the different types of Perception and updates the AVA's Cognitive Map to obtain a multi-perceptive mapping. Then the Control Architecture uses both the 'cognitive maps' and the 'memory model' to interact with the learning, development, and control processes of the AVA (Virtual Human Controller).

Realization

Integration of Virtual Sensors

The modeling of an AVA gaining its independence with regard to its virtual representation remains an important theme in research and is very close to autonomous robotics. It helps also to understand and model human behavior. The AVA collects information only through the virtual sensors described earlier.

We assume that vision is the main canal of information between the AVA and its environment as indicated by the standard theory in neuroscience for multi-sensorial integration.[18]

The sensorial modalities update the AVA's cognitive map to obtain a multi-sensorial mapping. For example, visual memory in the AVA's internal memory is used for a global move from point A to point B. Should obstacles be present, it would have to be replaced for a local move by direct vision of the environment.

In our *ALifeE* approach, we tried to integrate all the multi-sensorial information from the AVA's virtual sensors. In fact, an AVA in a VE may have different degrees of autonomy and different sensorial canals depending on the environment. For instance, an AVA moving in a VE represented by a well-lit room will use primarily the sensorial information of vision. However if the light is turned off, the AVA will appeal to the acoustic or tactile sensorial information in the same way a human would move around in a dark room.[19]

From this observation we derive the hypotheses underlying our *ALifeE framework* approach. They are backed up by the latest research in neuroscience,[1] which describes a partial re-mapping at the behavioral level of the human including:

- *Assignment*: the prediction of the acoustic position of an object from its visual positions requires a transformation from its *eye-centered* (vision sensor) coordinates to its *head-centered* ones (auditory sensor). The comparison of these two types of results can be used to determine whether the acoustic and visual signals are directly connected to the same object.
- *Recoding*: the choice of the reference frame to integrate the sensorial signals.

Integration of Perception

We have also used an *octree* to represent different types of perception (proprioception, active and predictive perception) in the form of *percepts*.[13]

Experimental Results

Using different scenarios we were able to confirm that our *ALifeE* integrates sensorial and perceptual modalities in a coherent way. Figure 3 provides the overall picture used for our experimental results.

We met our objectives using the three novel methodologies defined in this paper. We have applied this approach to a non-parametric Bayesian learning methodology, more particularly the continuous k-nearest neighbor (*k-NN*) method by using a high-speed search algorithm.[20]

Semantic Information for Virtual Sensors and Perception

All the sensorial and perceptive information used in the *ALifeE framework* for our application with the *k-NN* algorithm are synthesized in Figure 3.

Examples of Multi-Sensory Integration

We have used the *ALifeE framework* for the exploration of an unknown VE by an AVA. The AVA's Internal Visual Memory allows it to move around and navigate (Figures 4 and 5).

Future Work and Conclusions

By combining virtual sensors and perception in the same framework, we believe that we can significantly reduce the dimensionality and complexity of the data necessary in the learning of an AVA. This constitutes the objective of our present and future research. The proposed approach is a contribution to the 'curse' of dimensionality, with certain learning methods.

Apart from very simple VEs, certain learning methods imply memories of unrealistic size. As the states and actions are often infinite, it is possible to represent different functions (for example, the utility function in reinforced learning) by value tables. However since in many applications the space of states and sometimes also that of actions is continuous, the direct usage of tables is impossible.

Our approach is closely related to the work proposed by[21] which tries to understand the perceptual organization of the sensory field in such a way that it delivers the highest utility to the AVA. Learning how to control the

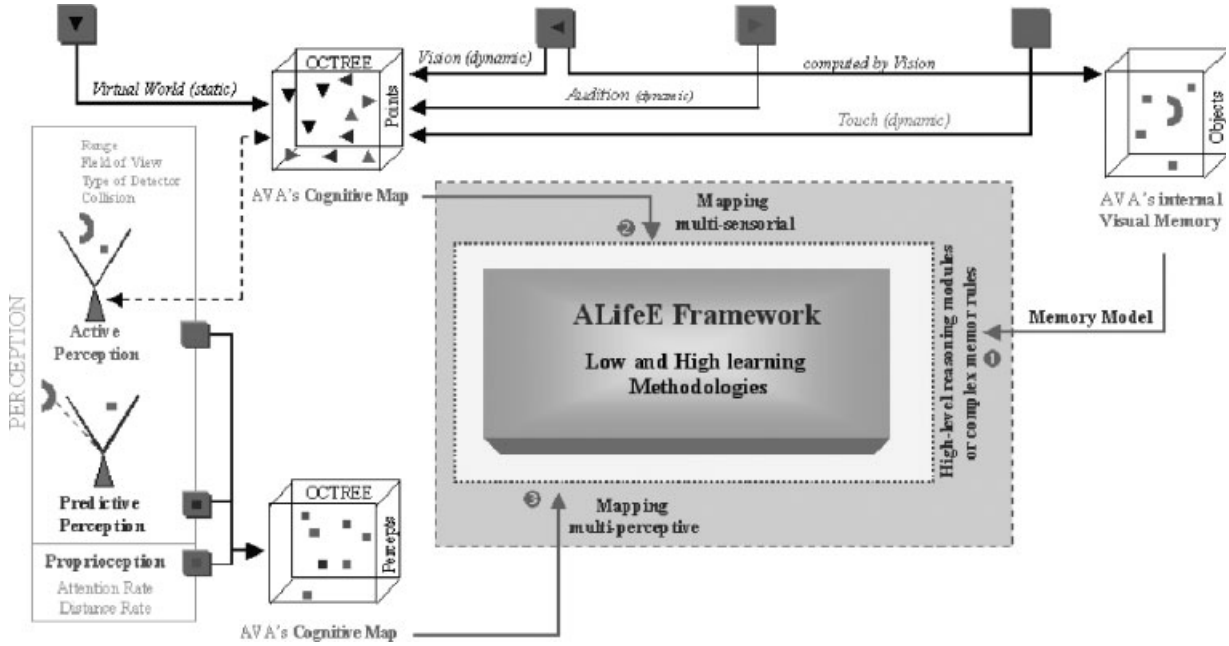


Figure 3. The architecture (ALifeE) used for our experimental results. The semantic information coming from ALifeE multi-sensorial mapping, multi-perceptive mapping and memory models is used by learning processes to establish high-level reasoning modules or complex memory rules.

effectors (Figure 2) for any given set of virtual sensor readings can be a difficult problem and is the primary focus of many machine-learning algorithms (e.g. reinforcement learning). Sensations and perceptions link the AVA's brain to the world and allow it to provide mental representations of reality.

However, there are some weaknesses in our approach. The 'cognitive maps' give only an approximate mapping. And, the 'cognitive map' inputs must be processed with care. The approach presented here is part of a more complex model, which is the object of our research. The goal is to realize a *Virtual Life Environment* for an AVA

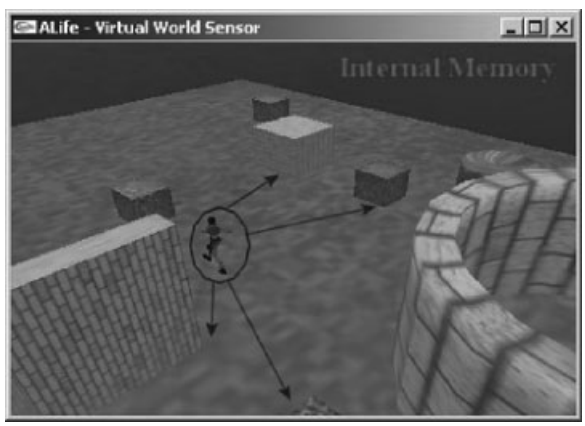


Figure 4. Application of learning an unknown VE. The yellow-colored graphic objects are those discovered by the AVA and are memorized in the AVA's Internal Visual Memory.

Figure 5. Application of Multi-Sensory VE (Vision—Touch). Snapshot of an AVA (pink-colored triangle) using multi-sensorial information to move around in a VE with k-NN algorithm. Inside the red-colored pentagon: yellow-colored triangle for 1-NN and cyan-colored triangles for 2-NN.

including different interfaces and sensorial modalities coupled with different evolving learning methodologies.

ACKNOWLEDGEMENTS

This research has been partially funded by the Swiss National Science Foundation.

References

1. Pouget A. A computational perspective on the neural basis of multi-sensory spatial representations. *Nature Reviews/ Neuroscience* 2002; **3**: 741–747.
2. Renault O, Magnenat-Thalmann N, Thalmann D. A Vision-based Approach to Behavioural Animation. *Journal of Visualization and Computer Animation* 1990; **1**: 18–21.
3. Noser H, Renault O, Thalmann D, Magnenat Thalmann N. Navigation for Digital Actors based on Synthetic Vision, Memory and Learning. *Computers and Graphics* 1995; **1**: 7–19.
4. Kuffner JJ, Latombre JC. Fast Synthetic Vision, Memory, and Learning Models for Virtual Humans. In *Proceedings of Computer Animation, IEEE*, 1999; 118–127.
5. Noser H, Thalmann D. Synthetic Vision and Audition for Digital Actors. In *Proceedings of Eurographics*, 1995; 325–336.
6. Hudson T, Lin M, Cohen J, Gottschalk S, Manocha D. V-COLLIDE: Accelerated Collision Detection for VRML. In *Proceedings of VRML*, 1997; 119–125.
7. Bordeux C, Boulic R, Thalmann D. An Effcient Perception Pipeline for Autonomous Agents. In *Proceedings of Eurographics*, 1999; 23–30.
8. Reynolds CW. An evolved, vision-based behavioral model coordinated group motion. In *From Animals to Animats, 2nd International Conference on Simulation of Adaptive Behavior*, MIT Press, 1993; 384–392.
9. Tu X, Terzopoulos D. Perceptual Modeling for the Behavioral Animation of Fishes. In Pacific Graphics, 1994. Ed. World Scientific.
10. Wenzel EM. Localization in Virtual Acoustics Displays. In *PRESENCE* 1992; **1**(1): 80–107.
11. Carollo C. Sound Propagation in 3D Environment. *Ion Storm*, 2002.
12. Huang Z, Boulic R, Magnenat Thalmann N, Thalmann D. A Multi-sensor Approach for Grasping and 3D Interaction. In *Proceedings of Computer Graphics International*, Academic Press, 1995; 235–254.
13. Conde T, Thalmann D. An Integrated Perception for Autonomous Virtual Agents: Active and Predictive Perception. *Techincal Report*. School of Computer and Communication Science. Swiss Federal Institute of Technology, Lausanne.
14. Bersini H. Reinforcement Learning for Homeostatic Endogenous Variables. In *Proceedings of 3rd Int. Conference On Simulated & Adaptive Behaviour*, MIT Press, 1994; 325–33.
15. Linas R. *I of the Vortex: From Neurons to self*, MIT Press, 2001.
16. Sowa JF. *Conceptual Structures*, 1964. Ed. Addison Wesley Company.
17. Karabassi EA. A Fast Depth-Buffer-Based Voxelization Algorithm. *Journal of graphic tools* 1999; **4**(4): 5–10.
18. Elfes G. Occupancy Grid: a Stochastic Spatial Representation for Active Robot Perception. In *6th Conference on Uncertainly in AI*, 1990.
19. Strösslin Th, Krebser Ch, Arleo A, Gerstner W. Combining Multimodal Sensory Input for Spatial Learning. In *Proceedings of ICANN*, LNCS 2415. Springer-Verlag, 2002; 87–92.
20. Mitchell T. *Machine Learning*, 1997. Eds. McGraw Hill.
21. Yvanov YA. *State Discovery for Autonomous Learning*. PhD Dissertation, MIT, 2002.