

INTERAGIR AVEC DES ÊTRES VIRTUELS

Daniel THALMANN
Computer Graphics Lab, EPFL
CH 1015 Lausanne Switzerland
thalmann@lig.di.epfl.ch
<http://ligwww.epfl.ch>

La Téléprésence est l'avenir des systèmes multimedia et elle permettra à des participants de partager des expériences privées et professionnelles telles que des réunions, des jeux, des fêtes. Les humains virtuels ou acteurs de synthèse ont un rôle clé à jouer dans ces Environnements Virtuels partagés et une interaction réelle avec eux est un grand défi.

1. Introduction

Les nouvelles technologies nous ont permis depuis quelques années de créer des êtres virtuels et de les animer. Plus récemment, la Réalité Virtuelle nous a amené la possibilité de s'immerger dans les mondes virtuels et d'y rencontrer des êtres virtuels. Enfin grâce aux recherches en intelligence artificielle et en vie artificielle, les êtres virtuels sont capables d'une certaine autonomie. Dans cet article, nous essayons d'expliquer où en sont les recherches actuelles dans le domaine des humains virtuels.

Pour commencer, étudions comment créer un être virtuel du point de vue géométrique. Il faut considérer essentiellement que nos êtres sont formés de deux parties la tête et le corps. Cette séparation arbitraire repose en fait sur les considérations suivantes:

1. La tête et plus particulièrement le visage est un élément essentiel pour reconnaître une personne. La forme du visage et les traits doivent être très fidèles pour qu'on reconnaisse la personne. Les expressions et l'animation du visage sont très subtiles. Un simple sourire caractérise une personne. Par contre, le corps n'est pas aussi caractéristique et il suffit souvent de données anthropométriques pour caractériser les gens: taille, poids, tour de hanches, de taille, de poitrine ...
2. Pour animer un visage, la base est le muscle. Ainsi un sourire n'est que la conséquence de déformations musculaires. L'animation du corps repose essentiellement sur la variation des angles d'articulation: genou, coude, épaule etc... On créera des expressions à partir des déformations musculaires.

On va donc créer des outils différents pour fabriquer et animer une tête ou un corps. Pour construire une tête, nous nous servirons d'un programme de sculpture. Nous partons soit d'une tête existante, soit d'une forme de base, sphère par exemple, et nous modifions directement la forme en tirant sur la surface comme on tire sur une membrane. On pourra ainsi allonger le nez ou gonfler les joues. De même, on aura ainsi des expressions correspondant à différentes émotions et état d'âme: sourire, angoisse, peur, joie, tristesse, colère (Figure 1). Pour le corps, on créera une vraie musculature et une peau sera comme tissée sur cette musculature. Pour construire et animer des corps humains réalistes, l'approche la plus efficace est un modèle multicouches. Il contient une couche de squelette, des couches intermédiaires qui simulent le comportement

physique des muscles, des os, des tissus mous, etc., et une couche de peau. Jianhua et Thalmann [1] ont ainsi décrit une approche multi-couches. Des surfaces implicites sont employées pour simuler le comportement grossier des os, des muscles et des tissus mous. Ils sont attachés aux articulations du squelette et arrangés dans une approximation basée sur l'anatomie. Les surfaces de peau sont automatiquement obtenues par sectionnement en tranches des surfaces implicites. La Figure 2 montre un corps bâti selon cette méthode.

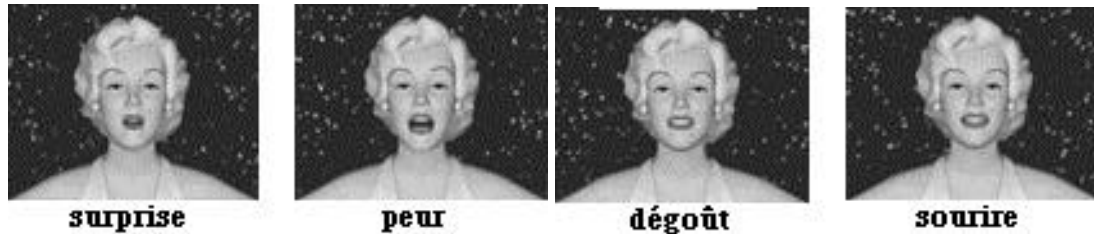


Figure 1. Expressions faciales



Figure 2. Corps construit avec des surfaces implicites

2. Classification des humains virtuels

Nous avons proposé une nouvelle classification [2] des humains virtuels avec quatre types:

- les avatars
- les acteurs guidés
- les acteurs autonomes
- les acteurs perceptifs interactifs

2.1 Les Avatars

Dans notre terminologie, les avatars correspondent à des représentations fidèles de l'utilisateur (Figure 3). La technique consiste à enregistrer directement les mouvements de

l'usager et à les transposer sur un humain virtuel qui devrait avoir aussi son apparence. Pour implanter de tels avatars, il est nécessaire d'avoir des séries de capteurs comme les Flock of Birds.



Figure 3. Avatar et l'utilisateur

2.2 Les acteurs guidés

Les acteurs guidés sont des acteurs qui sont conduits par l'utilisateur mais sans que leurs mouvements ne correspondent directement à ceux de l'utilisateur. Ils sont basés sur le concept de la métaphore directe temps réel [3]. Pour comprendre le concept, prenons un exemple de métaphore traditionnelle: le contrôle d'une marionnette. Une marionnette peut être définie comme une poupée avec des membres déplacés par des fils ou des ficelles. Les doigts humains sont employés pour diriger le mouvement de la marionnette. En Réalité Virtuelle, le meilleur exemple de guidage d'acteurs est la navigation réalisée dans les systèmes d'Environnements Distribués (voir Section 3.9).

2.3 Les acteurs autonomes

2.3.1 Perception-action

Grâce à des lois du comportement, il est possible de définir des acteurs capables d'agir par eux-mêmes, d'être autonomes. Ce domaine est celui où la recherche de pointe se fait. A titre d'exemple, nous allons prendre le cas de l'évitement d'obstacles, un problème que chacun d'entre nous résolvons chaque jour un très grand nombre de fois. Les roboticiens ont introduit des techniques mathématiques qui permettent d'évaluer à partir d'une configuration donnée le ou les chemins possibles pour aller d'un point à un autre en évitant les obstacles. Cette démarche très valable pour un bras de robot se déplaçant mécaniquement dans un environnement connu et statique, est peu applicable au comportement d'un humain dans un environnement souvent mal connu et parfois dynamique. Lorsqu'on se déplace sur un trottoir, il ne faut pas seulement éviter de se cogner aux parcomètres, mais surtout éviter le piéton qui vient en face; sans oublier que ce piéton va aussi tenter de vous éviter. Ce type de comportement, de nature dynamique, ne peut être simulé en animation qu'avec une démarche de type sensorielle. C'est effectivement le sens des recherches que nous menons dans notre laboratoire. Nos humains virtuels ont maintenant une vision, une audition et un toucher de synthèse. Ils peuvent se diriger vers un bruit déterminé ou éviter la table qu'ils voient

ou enfin réagir même dans l'obscurité en reconnaissant qu'ils ont cogné un mur. Ils peuvent même jouer au tennis en basant leur jeu sur celui de l'adversaire. Le principe repose sur une boucle dite de perception-action. A chaque instant, pour chaque être virtuel, on crée une perception de l'espace virtuel, c'est à dire que l'on simule dans l'ordinateur ce que chaque être peut voir, entendre et toucher à l'aide de sens virtuels [4]. Ainsi l'humain virtuel aura un vrai champ de vision, il pourra entendre les bruits produits par d'autres êtres ou se cogner à une table. Sur la base de ce que ces êtres ressentent, ils sont capables de prendre des décisions simples. Par exemple, un être virtuel, attiré par un son en dehors d'une pièce, pourra chercher la porte et lorsqu'il l'aura aperçue, pourra trouver le chemin jusqu'à cette porte et sortir et ceci sans aucune intervention d'un vrai humain.

2.3.2 Les sens virtuels: vision, audition et toucher

La **vision** est un sous-système perceptuel fondamental pour l'acteur. C'est une approche idéale à lui tout seul pour modéliser de l'animation comportementale. Elle fournit à l'acteur un flot d'information réaliste à partir de l'environnement. Les applications sont la recherche de chemins, l'évitement d'obstacles, les jeux de balle. Le concept de vision de synthèse a été introduit en premier par Renault et al. [5] comme le principal canal d'information entre l'environnement et l'acteur virtuel. Plus récemment, plusieurs auteurs ont adopté cette approche pour simuler le comportement de groupes [6], de poissons [7], ou d'un chien [8]. Dans [5], chaque pixel de la vision d'entrée a l'information sémantique donnant l'objet projeté sur le pixel et l'information numérique donnant la distance à cet objet. Ainsi, il est facile de reconnaître, par exemple, qu'il y a une table juste devant à 3 mètres. L'acteur de synthèse perçoit son environnement selon une petite fenêtre dans laquelle l'environnement est rendu à partir de son point de vue. Comme il peut accéder aux valeurs de profondeur du pixel (z-buffer), la couleur du pixel et sa position propre, il peut situer des objets visibles dans son environnement 3D. Noser et al. [9] ont aussi proposé l'emploi d'une octree comme représentation interne de l'environnement vu par un acteur parce qu'il offre plusieurs caractéristiques intéressantes. L'octree doit représenter la mémoire visuelle d'un acteur dans un environnement 3D avec des objets dynamiques et statiques. Les objets dans cet environnement peuvent croître, rétrécir, se déplacer ou disparaître.

Dans la vie réelle, le comportement des gens et des animaux est souvent très influencé par les sons. Pour recréer une **audition** de synthèse, nous devons dans une première étape, modéliser un environnement sonore où l'acteur de synthèse peut directement accéder à l'information sur la position et la sémantique de la source de l'événement sonore audible. Ainsi, maintenant, nos acteurs virtuels sont capables d'entendre [10]. Le programme de rendu sonore tient compte des contraintes de temps réel. Il est ainsi capable de rendre chaque incrément de temps pour chaque microphone en temps réel et en prenant en compte la vitesse finale de propagation du son, les déplacements des microphones et des sources sonores.

Notre objectif est de bâtir un modèle comportemental basé sur la contribution sensorielle tactile reçue de l'environnement au niveau de la peau. Cette information sensorielle peut être employée dans des tâches comme **toucher** des objets, presser des boutons ou donner des coups de pieds à des objets. Un cas intéressant est le contact entre les mains et les objets durant la préhension. Notre approche [11] est basée sur des capteurs sphériques multiples. Ceux-ci sont considérés comme un groupe d'objets

attachés à la figure articulée. Un capteur est activé pour toute collision avec d'autres objets ou capteurs. Chaque capteur sphérique est ajusté à sa forme conjointe associée avec des rayons différents. Cette configuration est importante dans notre méthode parce que quand un capteur est activé dans un doigt, seules les articulations situées au-dessus doivent être bloquées, les autres pouvant encore bouger. En suivant cette méthode, tous les doigts sont finalement positionnés naturellement autour de l'objet. Ces capteurs multiples ont été intégrés dans une méthodologie générale pour la saisie automatique d'objets par des acteurs de synthèse.

2.4 Les acteurs perceptifs et interactifs

Nous définissons un acteur perceptif et interactif comme un acteur conscient d'autres acteurs et des gens réels. Un tel acteur est aussi supposé être autonome bien sûr. De plus, il est capable de communiquer interactivement avec les autres acteurs quelque soit leur type et avec les gens réels.

2.4.1 Intercommunication entre acteurs de synthèse

Les comportements peuvent être aussi dépendant de l'état émotionnel de l'acteur. Nous avons ainsi développé un modèle de communication non-verbale [12]. La crédibilité des acteurs virtuels est améliorée par leur aptitude à interpréter et employer un langage non-verbal. Une communication non-verbale est basée sur les postures et l'information sur ce que les gens ressentent. Les postures sont des moyens de communiquer et sont définies par des positions spécifiques des bras et des jambes et certaines orientations du corps. Habituellement, les gens n'emploient pas consciemment une communication non-verbale, mais ils la comprennent instinctivement de manière étonnante et sont capables d'y répondre sans raisonnement explicite. Cette communication non-verbale est essentielle pour diriger l'interaction entre des gens avec ou sans contact (Figure 4). Dans notre cas, nous avons récemment développé des simulations d'intercommunication non-verbale entre des êtres de synthèse. Le principe est le suivant, pour chaque être virtuel, on définit son état émotionnel qui peut aller de 0 (triste et aucun désir de communiquer) à 1 (joyeux et désirant vivement entrer en communication). On définit également pour chaque paire d'êtres leur intensité relationnelle allant de 0 (se détestent) à 1 (s'adorent). On laisse alors une dizaine d'êtres virtuels se déplacer au hasard dans un parc public. Lorsque deux êtres se rencontrent, ils vont s'ignorer ou communiquer en fonction de leurs états émotionnels respectifs et leur intensité relationnelle. Les états émotionnels et les intensités relationnelles ne sont pas constants et évoluent en fonction de la communication. Ainsi, un être joyeux qui rencontre un être qu'il n'aime pas verra une diminution de son état émotionnel à cause de la rencontre qu'il a fait. On notera également que l'état émotionnel a une influence directe sur la manière de se comporter, de marcher par exemple (Figure 5).



Figure 4. Interaction entre deux acteurs



Figure 5. Marche avec état émotionnel triste et joyeux

2.4.2 Communication entre personnes réelles et virtuelles

Les gens réels sont bien sûr aisément conscients des actions des acteurs de synthèse grâce aux outils de la Réalité Virtuelle tels que les casques de visualisation. Le problème majeur à résoudre est de rendre les acteurs virtuels conscients du comportement des gens réels. Les acteurs virtuels devraient percevoir les vraies personnes ou leurs avatars par leur sens virtuels. Les acteurs perceptifs et autonomes peuvent avoir des degrés différents d'autonomie et des canaux différents pour ressentir l'environnement. Pour les sens visuel et auditif, nous pouvons considérer le cas d'un avatar jouant au tennis avec un acteur perceptif interactif. Concernant le sens tactile, nous pouvons considérer l'exemple suivant: une lutte entre une personne réelle et un acteur autonome. Le mouvement de la personne réelle est capturée avec un Flock of Birds. Les gestes sont reconnus par le système [13] et l'information est transmise à l'acteur virtuel qui est capable de réagir aux gestes et décider quelle attitude adopter. La Figure 6 montre un exemple.



Figure 6. Combat entre avatar et acteur perceptif interactif

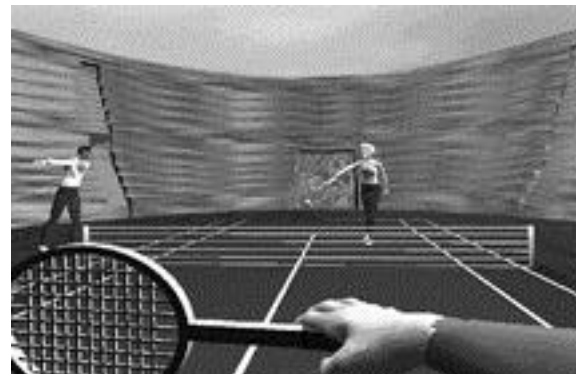


Figure 7. Jeu de tennis avec avatars et acteurs perceptifs interactifs temps-réel

2.4.3 Les Humains Virtuels dans les Environnements Distribués

Le système VLNET[8] (Virtual Life NETwork) supporte un environnement virtuel distribué et partagé permettant à des utilisateurs multiples d'interagir avec chacun et leur environnement en temps réel. Les utilisateurs sont représentés par des acteurs humains virtuels 3D, des acteurs guidés. Ils ont des comportements et une apparence similaire aux humains réels, et aident à soutenir le sens de présence des utilisateurs dans l'environnement. En plus d'acteurs guidés, l'environnement peut aussi inclure des acteurs perceptifs interactifs et autonomes qui servent surtout comme interface amicale pour les utilisateurs désirant accéder à certains services. Les humains virtuels peuvent aussi être employés pour représenter les partenaires actuellement indisponibles, permettant ainsi une coopération asynchrone entre des partenaires distants.

2.4.4 Jouer au tennis avec Marilyn

Comme application, nous sélectionnons un jeu virtuel de tennis avec un acteur guidé, l'actrice perceptive interactive Marilyn, et un arbitre perceptif interactif, les deux avec une vision virtuelle. La Figure 7 montre une séance interactive. Les caractéristiques géométriques du court de tennis font partie de la connaissance des joueurs. Pour la simulation de la dynamique de la balle, la gravité, le filet, le sol et la raquette, nous employons des calculs basés sur la physique. Le dépistage de la balle par le système de vision est contrôlé par un automate spécial qui suit la balle, estime le temps de collision et le point de collision de la balle avec la raquette et effectue le coup avec la force donnée et dans la direction voulue. L'arbitre juge le jeu en suivant la balle avec son système de vision. Marilyn peut aussi entendre des événements sonores et obéit aux décisions de l'arbitre.

3. Conclusion

Notre objectif ultime est de créer des acteurs de synthèse crédibles et réalistes, des humains virtuels intelligents, autonomes avec faculté d'adaptation, perception et mémoire. Ces acteurs devraient être capables d'agir librement et avec émotion. Idéalement, ils devraient être conscients et imprévisibles. Mais, sommes-nous loin d'une situation aussi idéale ? Nos acteurs perceptifs interactifs sont capables de percevoir le monde virtuel, les gens vivant dans ce monde et dans le monde réel. Ils peuvent agir en se basant sur leur perception de façon autonome. Leur intelligence est contrainte et limitée aux résultats obtenus dans le développement de nouvelles méthodes d'Intelligence Artificielle. Les acteurs intelligents sont capables d'apprendre ou de comprendre des situations très simples. La mémoire est généralement définie comme le pouvoir de reproduire ou de se rappeler ce qui a été appris et retenu surtout par des mécanismes d'association.

Nous avons vu que des aspects émotionnels peuvent être importants dans l'intercommunication non-verbale. Les émotions sont aussi essentielles dans l'animation faciale. Cependant, une émotion réelle devrait être considérée comme un état de ressentir, une réaction physique et psychique subjectivement ressentie comme une sensation forte et entraînant physiologiquement des changements qui préparent le corps pour une action vigoureuse immédiate. Dans ce cas, nous sommes loin de réaliser des acteurs véritablement émotionnels.

Enfin, les acteurs dans l'avenir devraient être flexibles, conscients et libres. Un acteur est flexible aussi longtemps qu'il peut "survivre" dans des environnements plus ou moins imprévisibles et dangereux. Selon Alexander [14], un acteur conscient devrait être conscient surtout de quelque chose en lui-même ou caractérisé par une sensation, une émotion, ou une pensée. Un acteur peut être considéré comme libre si son comportement futur est imprévisible aux autres.

Remerciements

L'auteur est reconnaissant aux gens qui ont contribué à ce travail, en particulier Pascal Bécheiraz, Ronan Boulic, Tolga Çapin, Amaury Aubel, Tom Molet, Luc Emering et Hansrudi Noser. La recherche a été financée par le Fonds National Suisse de la Recherche Scientifique et l'Office Fédéral de l'Éducation et de la Science dans le cadre des projets Européens Humanoid-2 et Coven.

References

1. J.Shen, D.Thalmann, Interactive Shape Design Using Metaballs and Splines, Proc. Eurographics Workshop on Implicit Surfaces '95 , Grenoble, pp.187-196.
2. D.Thalmann, A New Generation of Synthetic Actors: the Interactive Perceptive Actors, Proc. Pacific Graphics '96 Taipei, Taiwan, 1996, pp.200-219.
3. D.Thalmann, Using Virtual Reality Techniques in the Animation Process in: Virtual Reality Systems (Earnshaw et al. eds), Academic Press, pp.143-159.
4. D. Thalmann, Virtual Sensors: A Key Tool for the Artificial Life of Virtual Actors, Proc. Pacific Graphics '95, Seoul, Korea, 1995, pp.22-40.
5. O.Renault, N.Magnenat Thalmann, D.Thalmann, A Vision-based Approach to Behavioural Animation, The Journal of Visualization and Computer Animation, Vol 1, No 1, 1990, pp.18-21.
6. C.W.Reynolds An Evolved, Vision-Based Behavioral Model of Coordinated Group Motion, Proc. 2nd International Conf. on Simulation of Adaptive Behavior, MIT Press, 1993, pp.384-392.
7. X.Tu, D.Terzopoulos, Artificial Fishes: Physics, Locomotion, Perception, Behavior, Proc. SIGGRAPH '94, Computer Graphics, 1994, pp.42-48.
8. B.M.Blumberg, T.A.Galyean, Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments, Proc. SIGGRAPH 95, 1995, pp.47-54.
9. H.Noser, O.Renault, D.Thalmann, N.Magnenat Thalmann, Navigation for Digital Actors based on Synthetic Vision, Memory and Learning, Computers and Graphics, Vol.19, No1, 1995, pp.7-19.
10. H.Noser, D.Thalmann, Synthetic Vision and Audition for Digital Actors, Proc.Eurographics'95, 1995, pp.325-336.
11. Z.Huang, R.Boulic, N.Magnenat Thalmann, D.Thalmann, A Multi-sensor Approach for Grasping and 3D Interaction, Proc. CGI '95, Academic Press, 1995, pp.235-254.
12. P.Bécheiraz, D.Thalmann, A Model of Nonverbal Communication and Interpersonal Relationship between Virtual Actors, Proc.Computer Animation '96, IEEE Computer Society Press, 1996.
13. L.Emering, R.Boulic, D.Thalmann, Interacting with Virtual Humans through Body Actions, IEEE Comp. Graphics & Applic., 1998, Vol.18, No1, pp.8-11.
14. I.Aleksander, Artificial Consciousness, in: (Magnenat Thalmann and Thalmann, eds) Artificial Life and Virtual Reality, John Wiley, Chichester, 1994, pp.73-81.