

MARC : un personnage virtuel réactif expressif

Matthieu Courgeon

Jean-Claude Martin

Christian Jacquemin

LIMSI-CNRS

BP 133, 91403 Orsay, France

+33.1.69.85.81.04

courgeon@limsi.fr

martin@limsi.fr

christian.jacquemin@limsi.fr

ABSTRACT

La conception d'interfaces de communication affective utilisant des agents virtuels soulève de nombreuses questions. Le système doit notamment être en mesure d'afficher des expressions faciales d'émotions complexes en réaction aux entrées de l'utilisateur, dynamiquement et en temps réel (combinaisons de plusieurs expressions d'états affectifs variés, expressions fines via l'affichage de rides et coloration de la peau). De plus, afin de permettre une interaction variée, différents agents doivent être capables de réagir de différentes manières aux actions de l'utilisateur. Cet article présente un système permettant de contrôler en temps réel un personnage animé capable d'exprimer des expressions faciales émotionnelles. Différents profils expressifs peuvent être spécifiés afin de moduler en temps réel les réactions de l'agent aux entrées de l'utilisateur, simulant ainsi des différences interindividuelles en termes d'expression des émotions. Nos travaux lient l'infographie à la modélisation informatique des émotions et offrent une interface de contrôle adaptable et réutilisable dans différents contextes (agents pédagogiques, arts virtuels, études perceptives).

Categories and Subject Descriptors

H5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies, Interaction styles, User-centered design.

General Terms

Algorithms, Design, Human Factors.

Keywords

Expressive agent, real time interaction, facial expressions.

1. INTRODUCTION

Dans le domaine des interfaces homme-machine multimodales, les agents animés jouent un rôle important dans les univers virtuels, les applications éducatives, ludiques ou artistiques. Ils peuvent être aussi utilisés comme outil lors d'expériences sur la perception des émotions. Pour certaines de ces applications, il peut être utile d'étudier la perception qu'ont les utilisateurs de différents degrés de réalisme et d'expressivité du personnage (en

termes de reconnaissance des émotions).

Celui-ci doit alors pouvoir montrer une animation expressive et un rendu graphique qui permettent à l'utilisateur de percevoir des états affectifs sur une large gamme nuancée. Il doit aussi pouvoir réagir en temps réel aux actions de l'utilisateur, et son comportement doit refléter une personnalité cohérente et réagissant correctement en fonction de l'interaction.

Concevoir un agent virtuel expressif temps réel, réaliste et réagissant de manière automatique aux requêtes de l'utilisateur est cependant un projet à très long terme. Une première étape logique consiste donc à modéliser et construire un dispositif de contrôle manuel. Des tests perceptifs peuvent servir à valider la perception qu'a l'utilisateur des émotions exprimées par l'agent, et à évaluer la boucle perception / action créée par l'interaction continue entre l'utilisateur et le personnage virtuel.

L'objectif de notre travail est d'apporter des éléments de réponse aux limites actuelles de ces agents expressifs en faisant coopérer deux disciplines complémentaires : l'infographie et le traitement informatique des émotions. Aussi bien dans le domaine de l'infographie que dans celui de l'animation, nos objectifs sont de permettre plusieurs degrés de réalismes des images *statiques* (modèle 3D et texture). Nous souhaitons également disposer d'un contrôle sur la *dynamique* de l'expression des émotions et notamment de mélanges d'émotions qui sont souvent observés dans la vie de tous les jours [1-3]. Enfin, en terme d'*interaction*, il s'agit de trouver des correspondances cohérentes entre d'une part les médias d'interface homme-machine mis à la disposition de l'utilisateur (souris, joystick, souris 3D), et les représentations internes des émotions et leurs expressions faciales affichées par l'agent virtuel d'autre part. Le système doit permettre de définir des profils individuels pour ces agents virtuels, servant à moduler en temps réel leur manière d'exprimer des émotions.

Ces différentes contributions s'intègrent dans une plateforme multimodale affective à base d'agents virtuels intitulée MARC (*Multimodal Affective and Reactive Character*) qui a fait l'objet de deux évaluations expérimentales que nous décrivons également ici [4, 5].

Ce travail ouvre des perspectives en termes d'agent expressif, d'animation, d'études sur la perception des émotions et d'applications multimodales affectives.

2. ETUDE DE L'EXISTANT

2.1. L'étude pluridisciplinaire des émotions

Une émotion peut être définie comme une « construction hypothétique indiquant un processus de réaction d'un organisme à des événements significatifs » [6]. Historiquement limitée à la notion de sentiment subjectif (« feeling » en anglais), les émotions sont maintenant reconnues comme faisant intervenir d'autres composantes : un traitement cognitif, des changements physiologiques, des tendances à l'action (par exemple s'enfuir), et une expression motrice. Chacune de ces composantes a différentes fonctions. Darwin avait postulé l'existence d'un nombre fini d'émotions présentes dans toutes les cultures et ayant une fonction d'adaptation [7]. Ce postulat a été par la suite confirmé par Ekman [8] qui a mis en évidence six émotions universelles de base : *Peur*, *Colère*, *Joie*, *Tristesse*, *Dégout*, *Surprise*.

Les émotions sont limitées dans le temps et focalisées sur un événement émotionnel. On parle plus généralement d'états affectifs lorsqu'on englobe des émotions secondaires ou des phénomènes affectifs à plus long terme (par exemple l'humeur) [9]. Dans la suite de ce texte, nous utiliserons cependant le terme « émotion » avec le sens plus général d'état affectif incluant les émotions primaires, mais aussi d'autres états affectifs (reproche, soulagement, admiration).

Les émotions peuvent être représentées de deux manières complémentaires. Soit par des catégories discrètes représentées par une étiquette verbale (par exemple « Colère »), soit par une position, ou un ensemble de positions, dans un espace défini par des dimensions continues. La dimension « valence » permet ainsi de décrire l'axe « négatif – positif ». La dimension « activation » permet de décrire l'axe « peu actif – très actif ». L'axe « dominance » sert à représenter le sentiment de contrôle que la personne ressentant l'émotion a sur les événements ou non. Ces 3 dimensions ont été réunies dans un espace intitulé PAD (Pleasure, Arousal, Dominance) initialement défini pour décrire des tempéraments [10].

Enfin, certaines différences interindividuelles ont été observées en termes de perception des expressions des émotions, par exemple différences homme-femme [11] notamment en ce qui concerne la perception des mélanges d'émotions dans des photos d'expressions faciales [12].

2.2. Les agents virtuels expressifs

La capacité des agents virtuels à exprimer ou plus généralement à gérer les émotions fait actuellement l'objet de recherches dans le domaine du traitement informatique des émotions (« *Affective Computing* ») [13]. Les agents expressifs actuels sont limités. Ils ne permettent généralement pas d'interagir en temps réel avec l'utilisateur. Quand ils sont interactifs, cela se fait aux dépens du réalisme et de la finesse du rendu qui deviennent alors insuffisants pour certaines applications (pas de rides, pas de modification de la coloration de peau en fonction de l'état émotionnel). Plusieurs études ont comparé différentes apparences et degrés de réalisme de personnages virtuels mais sans se focaliser sur l'expression d'états affectifs [14-21]. Beaucoup d'agents virtuels sont limités à quelques émotions de base et ne permettent pas de réagir

différemment aux émotions en fonction de profils d'agents qui seraient individuels. Certaines études ont exploré la perception d'expressions d'émotions complexes affichées par un agent virtuel mais rarement dans le cadre d'une interaction continue avec l'utilisateur [22]. Enfin, plusieurs chercheurs ont exploré différentes correspondances entre les actions de l'utilisateur et les expressions d'états affectifs mais sans se focaliser sur des signes fins d'expressions (modèle de tête 3D détaillé, rides, coloration et rendu de la peau, ombres) [23, 24] qui sont connus comme étant pertinents pour l'expression et la perception d'états affectifs [1].

3. MARC – DESCRIPTION DU SYSTEME

3.1 Objectifs

L'interaction affective entre un utilisateur et un agent virtuel dépend de plusieurs paramètres comme les capacités de l'agent à 1) réagir en temps réel aux actions de l'utilisateur, 2) afficher des signes fins d'expressions et 3) exprimer des différences individuelles en termes de réaction émotionnelle. La combinaison de ces trois paramètres reste actuellement une perspective long-terme.

Nous avons conçu MARC (*Multimodal Affective and Reactive Character*) afin de mener des études perceptives sur l'impact de ces trois paramètres sur la perception des émotions et développer des applications de traitement affectif (agents pédagogiques, arts virtuels).

3.2 Principes généraux

MARC est axé sur l'interaction affective temps réel d'une part, et sur le réalisme du rendu visuel et de l'animation d'autre part. La combinaison de ces deux domaines impose l'utilisation des techniques récentes de rendu GPU telles que le *Shadow Mapping/BSSRDF* [25]. L'implémentation de ces techniques ne sera pas abordée dans cet article, mais signalons que ce type de rendu requière une grande proportion du temps de calcul GPU afin de garantir le temps réel. Pour cette raison, on ne peut réaliser de calculs complexes physiques pour l'animation et conserver le temps réel.

Dans MARC, l'animation du visage est donc réalisée par déplacement de points clés selon le principe d'animation faciale MPEG-4 [26]. Cette technique ne permet pas de générer de rides d'expressions, ou de garantir la cohérence physique du visage déformé. Pour compenser ce manque, notre logiciel d'édition précalcule un certain nombre d'informations par des systèmes physiques, telles que des cartes de rides, qui seront utilisées ensuite en interpolation durant la phase d'animation temps réel. On peut donc afficher des rides basées sur des calculs physiques d'élasticité à coût moindre qu'avec celui induit par un système physique temps réel.

3.3 MARC – FaceEditor : un outil d'édition d'expressions faciales

Lors de la phase de rendu temps réel, l'animation est faite par interpolation d'expressions statiques prédéfinies par le concepteur. Pour le moment, à chaque émotion correspond une seule expression faciale éditée manuellement via notre logiciel d'édition, et stockée dans une table XML sous forme de déplacements de points clés dans l'espace du visage.

L'interface d'édition permettant de créer ces expressions statiques est une interface de type WYSIWYG développée spécifiquement

pour notre système, et qui permet de manipuler directement les points clés dans l'espace pour déformer le visage et générer ainsi une expression pour chacune des émotions que l'on veut pouvoir exprimer durant l'interaction temps réel. Il permet de 1) spécifier l'influence des points clés sur le maillage, et 2) de positionner les points clés pour chaque expression.

La figure 1 montre l'interface d'édition. Les points clé MPEG-4 (en bleu), peuvent être sélectionnés (en jaune) et déplacés par déplacement de la souris pour éditer l'expression faciale. Notre but dans le développement en interne de cet outil est de nous permettre d'effectuer des recherches long terme en fonction de nos besoins logiciels spécifiques avec des modèles et fonctions extensibles selon nos besoins. Ainsi notre éditeur manipule des modèles au format natif du moteur de rendu graphique que nous utilisons (VirChor, décrit au paragraphe suivant). Il permet également la génération de *normal map* pour le rendu de ride en utilisant la technique GPU nommé *Bump Mapping*, ainsi que des textures de *Displacement Mapping* pour les modifications physiques du visage non accessibles par la technique des déplacements de points clés. La génération de ces « cartes de rides » utilise un système masse ressort prenant en compte l'épaisseur et l'élasticité de la peau. La figure 2 résume les sorties du logiciel d'édition et leur intégration dans la phase d'animation.

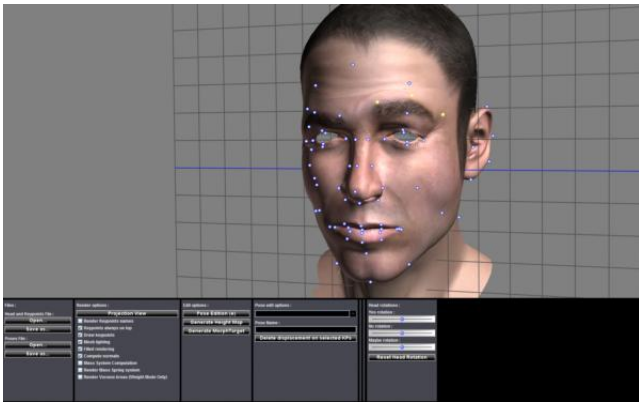


Figure 1 – Logiciel d'édition d'expression faciale dédiés aux besoins de MARC.

3.4 VirChor : rendu et animation temps réel

Comme le montre la figure 2, l'animation temps réel est pilotée par le logiciel nommé *Virtual Choreographer* (ou VirChor <http://virchor.sf.net>). Ce logiciel est un moteur de rendu et de gestion d'animation capable de faire le lien entre une scène graphique animée et des logiciels de contrôle de plus haut niveau servant à contrôler l'animation de manière interactive.

VirChor est en charge de gérer les différentes étapes du rendu graphique via l'utilisation d'OpenGL et du contrôle de l'exécution des Shaders de rendu en carte graphique, qui servent à générer l'animation et le rendu réaliste de la peau (Figure 3). Il gère également les mélanges entre les différentes expressions statiques. Comme nous l'avons dit, les expressions sont stockées dans des tables de déplacements de points clés MPEG-4. VirChor est en charge de faire un mélange pondéré de ces tables, et de générer ainsi l'expression complexe qu'il transmet à la carte graphique pour créer l'image résultante. Les valeurs de pondérations utilisées par VirChor pour mélanger les expressions

lui sont fournies par le module d'interaction décrit dans la section suivante.

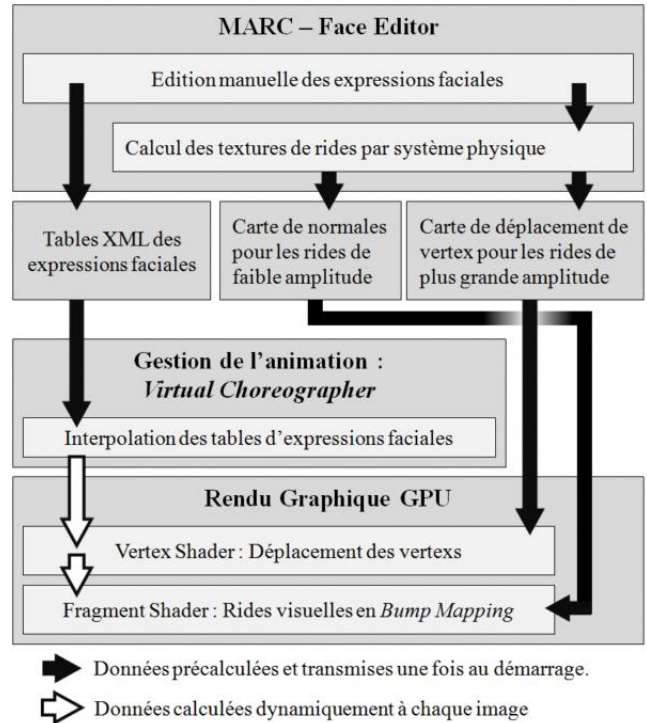


Figure 2. Intégration des informations issues de l'éditeur d'expressions dans le rendu de l'animation temps réel.



Figure 3 – Différents modèles de visage et différents rendu.

3.5 MARC – Gestion des interactions affectives

Nous décrivons dans ce paragraphe le module qui gère l'interaction et contrôle le moteur de rendu (Figure 4).

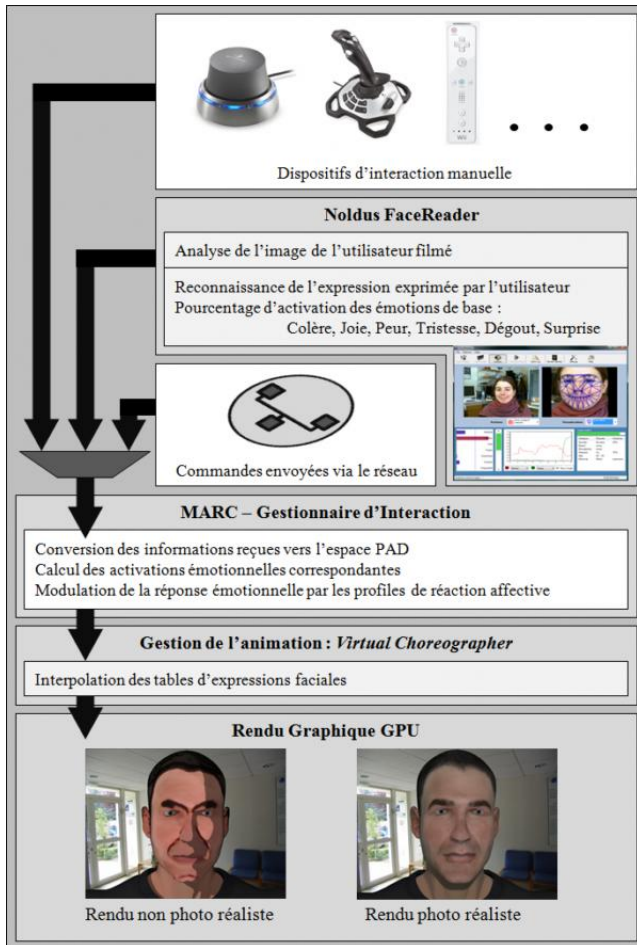


Figure 4. Schéma de l'interaction affective multimodale.

Nous avons choisi d'évaluer deux modes d'entrée vis à vis de leur pertinence pour l'interaction affective : le geste 3D et la reconnaissance des expressions faciales de l'utilisateur par traitement d'images. Ces deux modes peuvent être utilisés de manière indépendante ou complémentaire. Une fois l'interaction intégrée dans un espace continu des émotions, un module de gestion des profils de réaction affective permet de gérer l'évolution de l'état émotionnel de l'avatar, en fonction de son profil de réaction affective et de l'entrée utilisateur (Gestuelle et/ou Image). Ce module est donc indépendant de la méthode d'entrée.

3.5.1 Représentation interne des états affectifs

Nos besoins en termes de représentations internes des états affectifs étaient les suivants : pouvoir représenter une large variété d'états affectifs pertinents pour des expériences en perception des émotions (transitions entre émotions, dynamique des expressions) et en applications affectives, permettre de représenter ces états à l'aide de dimensions continues qui sont adaptées aux dispositifs d'interaction homme-machine, être suffisamment intuitives pour pouvoir être manipulées éventuellement par interaction directe de l'utilisateur.

Nous avons choisi l'espace PAD que nous avons introduit plus haut [10, 27]. PAD est un espace en trois dimensions (Pleasure, Arousal, Dominance) utilisé en Psychologie pour l'étude des émotions et utilisé dans plusieurs agents virtuels expressifs [28].

Cet espace peut être représenté par un cube d'arêtes de taille 2, chaque axe variant de -1 à 1. Il présente l'intérêt d'être à la fois simple à visualiser, car c'est un cube, et d'offrir une représentation continue de l'espace des émotions. Il permet une représentation de l'espace des émotions dans un espace 3D qui nous a semblé adapté aux dispositifs d'interaction 3D actuels (joystick, souris 3D...).

3.5.2 Entrée gestuelle et l'espace PAD

Le premier mode d'interaction est le geste, à travers un dispositif d'interaction 3D. Cette méthode utilise un module JAVA gérant le périphérique d'interface utilisateur, ainsi que l'interface entre ce périphérique et la correspondance des interactions avec l'espace PAD.

Nous avons expérimenté plusieurs dispositifs d'entrée. Au départ, nous utilisons un joystick pour manipuler notre curseur dans l'espace PAD. Le joystick contient effectivement 3 axes, mais l'un d'entre eux est une rotation sur l'axe vertical et s'est révélé peu intuitif à manipuler. Le deuxième dispositif testé est une souris 3D possédant six degrés de liberté [29]. Le déplacement dans le cube se fait de manière plus intuitive.

L'utilisateur peut contrôler soit la position, soit la vitesse du curseur à l'intérieur du cube, et définir ainsi des trajectoires de l'état affectif de l'agent dans l'espace PAD. Le système calcule ensuite l'activation relative à la position du curseur pour chaque émotion disposée aux 8 coins du cube. La figure 5 montre les axes de PAD et les 8 émotions positionnées, dans une première approche, à chaque coin du cube.

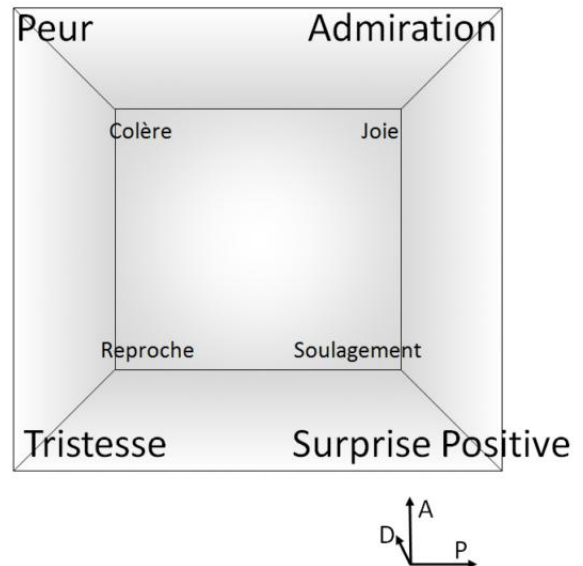


Figure 5. Disposition des émotions dans l'espace PAD.

3.5.3 Entrée vidéo et mime des expressions faciales

Notre second mode d'interaction avec l'avatar passe par l'utilisation du logiciel de reconnaissance d'expression faciale par

traitement d'image FaceReader qui analyse l'image de l'utilisateur pour tenter de détecter les émotions qu'il exprime [30]. Ce logiciel est utilisé pour différentes recherches en interaction homme-machine [31, 32] et présentait pour nous l'avantage d'être intégrable rapidement comme un module indépendant.

Ce logiciel calcule un taux de reconnaissance pour les six émotions de base : Peur, Colère, Joie, Tristesse, Surprise, Dégout. En interfaçant FaceReader et MARC, nous avons obtenu un miroir émotif nous permettant d'obtenir une interaction continue temps-réel [13, 23]. L'intégration se fait en utilisant les valeurs données par FaceReader pour calculer une position cible à l'intérieur de l'espace PAD. Le curseur est alors progressivement attiré par cette cible, elle-même déplacée en temps réel par l'analyse de FaceReader. Le reste de l'interaction est alors similaire à l'interaction gestuelle. La figure 6 montre l'interface du logiciel FaceReader et l'animation générée.



Figure 6 - Miroir émotif : FaceReader et MARC

Le logiciel FaceReader montre cependant quelques inconvénients. L'éclairage de l'utilisateur doit être vif et venir du haut. Les valeurs produites en sortie par le logiciel sont fluctuantes, même lorsque l'utilisateur garde une expression fixe. Pour éviter que l'expression de l'avatar semble « vibrer », nous avons donc dû appliquer une fonction de lissage sur les valeurs envoyées par FaceReader. Ceci a introduit une certaine inertie à l'état émotionnel de l'avatar, qui ne réagit plus instantanément aux changements d'expressions de l'utilisateur.

Un premier test informel a donné des retours positifs avec quelques sujets qui doivent être confirmés par une évaluation.

3.5.4 Modulation des entrées utilisateur par des profils de réaction affective

Nous avons développé un système de Profils de Réaction Affective (PRA), permettant de donner une individualité à l'agent virtuel en modulant ses réactions aux actions de l'utilisateur.

Un profil de réaction affective définit la réponse de l'agent aux commandes bas niveau de l'utilisateur (gestes ou image) pour une émotion donnée à l'aide 1) d'une courbe de réponse (Figure 7), 2) d'une vitesse de réaction (vitesse maximale de transition positive vers l'émotion) et 3) d'une vitesse de déclin (vitesse maximale de décroissance de l'émotion).

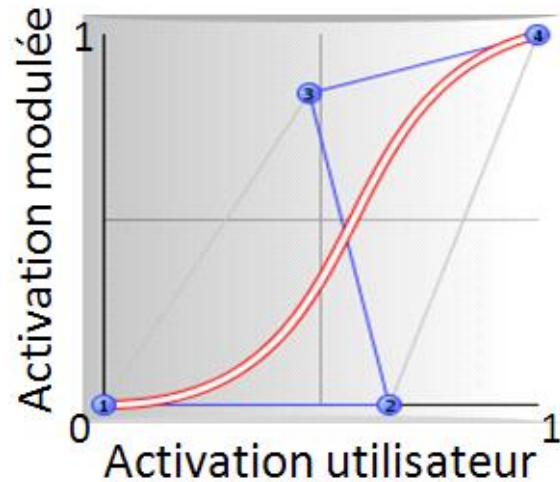


Figure 7 - Courbe de réponse

La figure 8 montre l'influence des deux vitesses de réaction. Sur chaque courbe, la courbe fine et rouge montre un exemple d'activation d'une émotion par l'utilisateur. La courbe en gras montre l'activation réelle envoyée au système d'animation, correspondant au signal utilisateur modulé temporellement par les deux vitesses (augmentation et diminution).

La courbe du haut de la Figure 8 montre l'effet de deux vitesses élevée donnant une bonne réactivité temporelle. La courbe du milieu montre l'effet de deux vitesses faibles donnant une inertie importante à la réaction de l'agent. La courbe du bas montre l'effet de la dissymétrie des deux vitesses (augmentation et diminution). L'ensemble des huit réponses émotionnelles (une pour chaque coin du cube PAD) forme un Profil de Réaction Affective individuel à un agent.

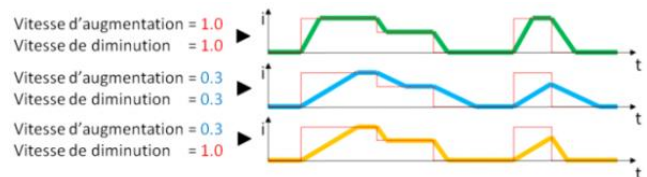


Figure 8 - Influences des vitesses de réaction de l'agent aux commandes de l'utilisateur.

3.5.5 Spécification et interaction

L'interface de spécification est intégrée à l'interface de contrôle du personnage, pour faciliter l'utilisation des profils durant l'interaction. Elle présente une série de contrôles graphiques à gauche de l'interface (Figure 9) permettant de régler les paramètres des profils de réaction affective.

La partie droite de l'interface permet de visualiser le cube PAD de l'espace de représentation des émotions.

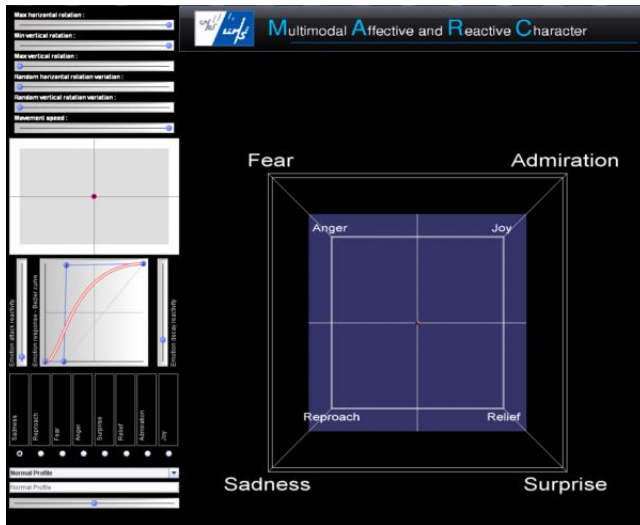


Figure 9. Interface de visualisation de l'espace PAD (à gauche) et de contrôle des Profils de Réaction Affective (à droite)

4. EVALUATIONS

Deux expérimentations ont été menées. La première expérimentation vise à évaluer les expressions *statiques* définies pour chaque émotion. Ces expressions ont été spécifiées pour cinq d'entre elles d'après Ekman [1] (Colère, Peur, Joie, Tristesse, Surprise positive). Les trois dernières expressions (Admiration, Soulagement, et Reproche) permettent de couvrir les 8 coins de l'espace PAD et ont été définies de manière empirique. On pouvait donc s'attendre à de moins bons résultats de reconnaissance sur ces trois expressions. La seconde évaluation cherche à évaluer la perception *dynamique et interactive* des différents profils de réaction affective par les utilisateurs.

4.1 Première évaluation : reconnaissance des expressions faciales statiques

4.1.1 Protocole

L'utilisateur est mis face à l'agent virtuel MARC affichant des expressions faciales statiques. Le formulaire à remplir contient neuf cases à cocher (huit émotions de base, plus *Neutre*). L'utilisateur peut cocher une ou plusieurs cases. Quatorze expressions faciales sont présentées à la suite. Huit émotions « simples », et six émotions « composées » d'un mélange de deux expressions à parts égales. La même expression faciale ne peut apparaître deux fois à l'utilisateur. L'ordre des émotions est tiré au hasard ainsi que les six mélanges parmi les douze mélanges possibles (l'expression *Neutre* ne peut être mélangée avec une autre, puisque cela serait équivalent à afficher une émotion simple à plus faible intensité).

L'expérimentation a été menée sur deux groupes de huit personnes. Le second groupe avait la possibilité de revoir chaque expression une deuxième fois, et éventuellement changer les réponses données. Ce second groupe devait permettre d'évaluer un possible effet d'apprentissage.

4.1.2 Résultats

Comme le montre la figure 10, *Tristesse*, *Peur* et *Colère* sont reconnues lorsqu'elles sont seules par 100% des sujets. *Joie* est reconnue à 85%, et *Reproche* à 79%. *Admiration* est reconnue à 46% et *Soulagement* à 18%. De plus, les trois émotions les plus reconnues sont en moyenne regardées de 10 à 15 secondes par les utilisateurs avant de faire leur choix, alors que *Admiration* est regardée en moyenne 20s, *Reproche* 25s, et *Soulagement* (l'émotion la moins reconnue) 36s.

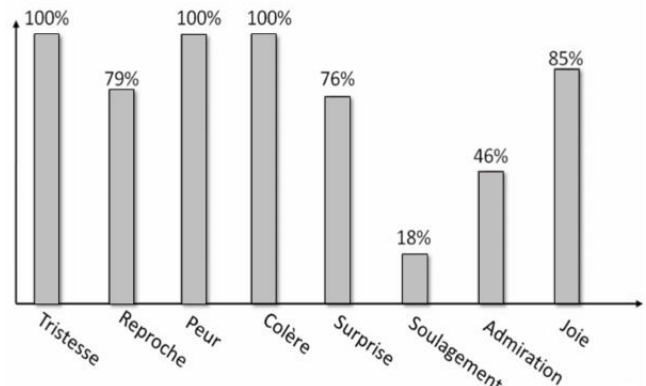


Figure 10. Taux de reconnaissance moyen des émotions non mélangées.

Les mélanges d'émotions sont moins bien perçus. Avec un taux de 40% à 60%, sauf la *Surprise*, perçue à 72% quand elle est combinée avec une autre émotion. Le soulagement et l'admiration sont à nouveau moins bien reconnus que les autres émotions lorsqu'ils sont combinés avec une autre émotion, (respectivement 14% et 22%).

On constate des résultats similaires à ceux de Hall et Matsumoto [12] sur les différences hommes/femmes dans la détection d'émotions composées. Leurs expérimentations montraient que les femmes perçoivent mieux et plus rapidement les expressions faciales d'émotions mélangées. Dans notre expérience, nous avons observé que sur les mélanges, les femmes ont un score moyen supérieur de 12% à celui des hommes.

4.1.3 Conclusions de l'évaluation

Les expressions d'émotions de base conçues en utilisant les descriptions issues de la littérature sont bien reconnues, comme on s'y attendait. Les trois autres expressions sont moins bien perçues.

4.2 Seconde évaluation : perception dynamique des Profils de Réaction Affective lors de l'interaction

4.2.1 Protocole

Dans cette évaluation l'utilisateur a la possibilité d'agir sur l'état affectif de l'agent à l'aide de l'entrée gestuelle. Différents Profils de Réaction Affective ont été définis pour cette expérimentation: *Positif*, *Négatif*, *Rapide*, *Lent*, *Très expressif*, *Peu expressif*. Par exemple, le profil *Positif* réagira très rapidement à l'activation des émotions placées du côté « Valence Positive » du cube PAD, et très lentement aux émotions placées du côté « Valence Négative ».

Ces différents profils sont proposés à l'utilisateur, sans identification, et dans un ordre aléatoire. Pour chaque profil, l'utilisateur manipule et observe les réactions de l'agent. Le sujet remplit ensuite un formulaire proposant des adjectifs utilisés en différences individuelles telles que Expressif / Peu expressif, Spontané/Non spontané, Honnête/Manipulateur, Emotif/Peu émotif, Chaleureux/Froid, Curieux/Fermé...

4.2.2 Résultats

Les résultats de cette expérimentation ont montré que les profils *Très expressif* et *Négatif* sont bien reconnus. La figure 11 montre les résultats pour le profil *Très expressif*. On constate que ce profil est décrit comme expressif, émotif, plutôt social mais avec peu de *self-control*. Les autres caractéristiques sont autour de la moyenne et montrent des différences dans la perception des sujets.

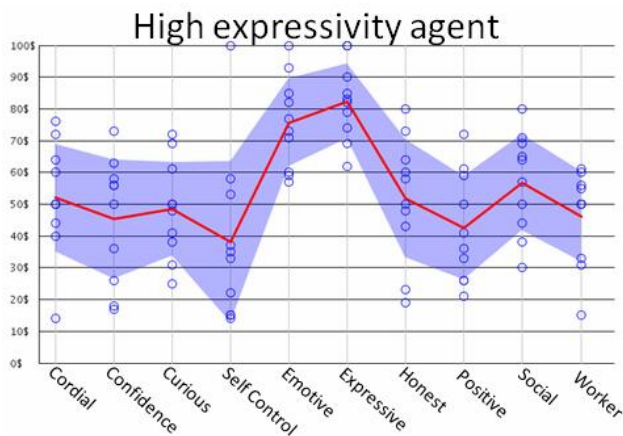


Figure 11. Perception du profil « Très expressif ».

5. CONCLUSIONS ET PERSPECTIVES

Nous avons présenté un outil d'interaction temps-réel avec un agent virtuel expressif, axé d'une part sur le réalisme des expressions exprimées, et d'autre part sur l'exploration de différences interindividuelles des agents virtuels dans l'expression des émotions. Le réalisme passe par l'utilisation de techniques récentes en matière de rendu graphique, et par l'étude approfondie de la littérature en matière de perception des expressions faciales.

Le système d'interaction, grâce à sa modularité, est utilisable avec de nombreux périphériques d'entrée, et le moteur de profils expressifs est adaptable en fonction des besoins. Enfin, il est possible de changer le mode rendu graphique pour utiliser un autre type de rendu, et varier le degré de réalisme visuel.

Nous l'avons initialement présenté lors de la conférence AAMAS 2008. Un outil similaire (« Alfred », <http://alfred.bernhardfalk.com/>) a fait ensuite l'objet d'une démonstration à GALA 2008.

Comme le montre la figure 12, nous avons également filmé MARC pour faire analyser ses expressions faciales par *FaceReader*. Le visage semble bien reconnu par *FaceReader* (ce qui confirme en partie la qualité du rendu de MARC). Ceci nous permet d'envisager de concevoir un protocole expérimental utilisant le logiciel *FaceReader* en parallèle d'évaluations perceptives réalisées avec des sujets humains, pour évaluer

l'expressivité de notre avatar. Nous pourrions ainsi comparer les résultats obtenus avec des personnes, et ceux obtenus par *FaceReader*, qui pourraient être plus objectifs, en particulier sur les mélanges d'émotions.



Figure 12 - Miroir émotif : FaceReader et MARC

De nombreuses perspectives sont actuellement à l'étude, notamment l'utilisation des systèmes physiques de type Muscle/Squelette/Peau qui devront être développés, puis évalués en terme de gain perceptif pour l'expression des émotions.

Les applications possibles de ce système sont variées, par exemple outils d'aide pour les animateurs. Il est actuellement utilisé dans un projet ANR artistique nommée CARE (*Cultural experience : Augmented Reality and Emotion*). L'agent y est utilisé pour enrichir la perception des spectateurs d'un ballet augmenté. L'agent est également utilisé pour une application d'entraînement social pour des enfants autistes [33].

6. REFERENCES

- [1] Ekman, P. and Friesen, W. V. Unmasking the face. A guide to recognizing emotions from facial clues. Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.
- [2] Scherer, K. R. Analyzing Emotion Blends. Fischer, A., City, 1998.
- [3] Abrilian, S. Représentation de Comportements Emotionnels Multimodaux Spontanés : Perception, Annotation et Synthèse. Directeurs: J.-C. Martin, L. Devillers, J. Mariani., Paris XI, 2007.
- [4] Courgeon, M., Martin, J.-C. and Jacquemin, C. 2008. User's Gestural Exploration of Different Virtual Agents' Expressive Profiles. In Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (Estoril, Portugal, 12-16 May, 2008). ACM.
- [5] Courgeon, M., Martin, J. C. and Jacquemin, C. 2008. MARC: a Multimodal Affective and Reactive Character. In Proceedings of the AFFINE workshop on Affective Interaction in Natural Environments held at the 10th International Conference on Multimodal Interfaces (ICMI 2008) (Chania, Crete, 24th of October 2008).
- [6] Scherer, K. R. Emotion. Introduction to Social Psychology: A European perspective. Oxford: Blackwell, 151-191, 2000.
- [7] Darwin, C. The expression of emotion in man and animal. University of Chicago Press (reprinted in 1965), Chicago, 1872.

- [8] Ekman, P. Basic emotions. *Handbook of Cognition & Emotion*. John Wiley, 301–320, New York, 1999.
- [9] Scherer, K. R. What are emotions? And how can they be measured? . *Social Science Information*, 44, 4 (2005), 693-727.
- [10] Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14 (1996), 261-292.
- [11] Knapp, M. L. and Hall, J. A. *Nonverbal communication in human interaction*. Sixth edition. Thomson Wadsworth, 2006.
- [12] Hall, J. A. and Matsumoto, D. Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4, 2 (2004), 201-206.
- [13] Picard, R. *Affective Computing*. MIT Press, 1997.
- [14] Van Vugt, H. C., Konijn, E. A., Hoorn, J. F., Keur, I. and Eliëns, A. Realism is not all! User engagement with task-related interface characters. *Interacting with Computers*, 19, 2 (2006), 267-280.
- [15] Baylor, A. L. and Kim, Y. *Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role*. City, 2004.
- [16] Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A. and Sasse, M. A. 2003. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. . In *Proceedings of the SIG-CHI conference on Human factors in computing systems* (Fort Lauderdale, FL, USA, April 5-10, 2003).
- [17] Dehn, D. M. and van Mulken, S. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, 52 (2000), 1-22.
- [18] Beun, R. J. *Embodied conversational agents: Effects on memory performance and anthropomorphisation*. City, 2003.
- [19] Koda, T. and Maes, P. *Agents with faces: the effects of personification of agents*. The British HCI Group, City, 1996.
- [20] Mc Breen, H. and Jack, M. Evaluating humanoid synthetic agents in e-retail applications. *IEEE Transactions on Systems, Man and Cybernetics*, 31, 5 (2001), 394-405.
- [21] Wonisch, D. and Cooper, G. *Interface Agents: preferred appearance characteristics based upon context*. City, 2002.
- [22] Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L. and Pelachaud, C. 2006. Perception of Blended Emotions: from Video Corpus to Expressive Agent. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA'2006)* (Marina del Rey, USA, 21-23 august, 2006). Springer.
- [23] Melder, W., A., , Truong, K. P., Uyl, M. D., Van Leeuwen, D. A., Neerincx, M. A., Loos, L. R. and Stock Plum, B. 2007. Affective multimodal mirror: sensing and eliciting laughter. In *Proceedings of the Proceedings of the international workshop on Human-centered multimedia* (Augsburg, Bavaria, Germany, 2007). ACM.
- [24] D'Alessandro, C., D'Alessandro, N., Le Beux, S., Simko, J., Cetin, F. and Pirker, H. 2005. The speech conductor : gestural control of speech synthesis. In *Proceedings of the eINTERFACE 2005. The SIMILAR NoE Summer Workshop on Multimodal Interfaces* (Mons, Belgium, July 18-August 12, 2005).
- [25] D'Eon, E. and Leubke, D. Advanced techniques for realistic realtime skin rendering. *nVidia GPU GEMS 3* 293-347, 2007.
- [26] Pandzic, I. S. and Forchheimer, R. *MPEG-4 Facial Animation. The Standard, Implementation and Applications*. John Wiley & Sons, LTD, 2002.
- [27] Russell, J. A. and Mehrabian, A. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11 (1977), 273-294.
- [28] Gebhard, P. 2005. ALMA - A Layered Model of Affect In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)* (Utrecht, 2005).
- [29] 3D mouse: SpaceNavigator, 3DConnexion, (Logitech). City.
- [30] van Kuilenburg, H., Wiering, M. and Den Uyl, M. 2005. A model based method for automatic facial expression recognition. In *Proceedings of the European Conference on Machine Learning (ECML' 05)*, (2005).
- [31] Zaman, B. and Shrimpton-Smith, T. *The FaceReader: Measuring instant fun of use*. ACM International Conference Proceeding Series; Vol. 189 City, 2006.
- [32] Truong, K. P., van Leeuwen, D. A. and Neerincx, M. A. *Unobtrusive Multimodal Emotion Detection in Adaptive Interfaces: Speech and Facial Expressions*. *Foundations of Augmented Cognition*. Springer, 354-363, Berlin / Heidelberg, 2007.
- [33] Grynszpan, O., Martin, J.-C. and Nadel, J. *Multimedia interfaces for users with high functioning autism: an empirical investigation*. *Int. J. Human-Computer Studies* (2008).