Marc : modèles informatiques des émotions et de leurs expressions faciales pour l'interaction Homme-machine affective temps réel

Matthieu Courgeon

Thèse

Présentée pour obtenir le grade de

Docteur de l'Université Paris Sud

Discipline : **Informatique**

Préparée au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
Dans le cadre de l'Ecole Doctorale d'informatique de l'Université Paris Sud (ED427)

Présentée par :

Matthieu COURGEON

Titre :

MARC : Modèles Informatiques des Emotions et de leurs Expressions Faciales pour l'Interaction Homme-Machine Affective Temps Réel

Dirigée par :

Jean-Claude MARTIN, *Professeur à l'université Paris-Sud*
Christian JACQUEMIN, *Professeur à l'université Paris-Sud*

Rapporteurs :

Sylvie PESTY, *Professeur à l'université Pierre Mendès France (Grenoble)*
Elisabeth ANDRÉ, *Professeur à l'université d'Augsburg*

Examinateurs :

Marc CAVAZZA, *Professeur à l'université de Teesside*
Anne VILNAT, *Professeur à l'université Paris-Sud*
Gaspard BRETON, *Ph.D, CEO de la société Dynamixyz*

Résumé

Les émotions et leurs expressions par des agents virtuels sont deux enjeux importants pour les interfaces homme-machine affectives à venir. En effet, les évolutions récentes des travaux en psychologie des émotions, ainsi que la progression des techniques de l'informatique graphique, permettent aujourd'hui d'animer des personnages virtuels réalistes et capables d'exprimer leurs émotions via de nombreuses modalités. Si plusieurs systèmes d'agents virtuels existent, ils restent encore limités par la diversité des modèles d'émotions utilisés, par leur niveau de réalisme, et par leurs capacités d'interaction temps réel.

Dans nos recherches, nous nous intéressons aux agents virtuels capables d'exprimer des émotions via leurs expressions faciales en situation d'interaction avec l'utilisateur. Nos travaux posent de nombreuses questions scientifiques et ouvrent sur les problématiques suivantes : Comment modéliser les émotions en informatique en se basant sur les différentes approches des émotions en psychologie ? Quel niveau de réalisme visuel de l'agent est nécessaire pour permettre une bonne expressivité émotionnelle ? Comment permettre l'interaction temps réel avec un agent virtuel ? Comment évaluer l'impact des émotions exprimées par l'agent virtuel sur l'utilisateur ?

A partir de ces problématiques, nous avons axé nos travaux sur la modélisation informatique des émotions et sur leurs expressions faciales par un personnage virtuel réaliste. En effet, les expressions faciales sont une modalité privilégiée de la communication émotionnelle. Notre objectif principal est de contribuer l'amélioration de l'interaction entre l'utilisateur et un agent virtuel expressif. Nos études ont donc pour objectif de mettre en lumière les avantages et les inconvénients des différentes approches des émotions ainsi que des méthodes graphiques étudiées.

Nous avons travaillé selon deux axes de recherches complémentaires. D'une part, nous avons exploré différentes approches des émotions (catégorielle, dimensionnelle, cognitive, et sociale). Pour chacune de ces approches, nous proposons un modèle informatique et une méthode d'animation faciale temps réel associée.

Notre second axe de recherche porte sur l'apport du réalisme visuel et du niveau de détail graphique à l'expressivité de l'agent. Cet axe est complémentaire au premier, car un plus grand niveau de détail visuel pourrait permettre de mieux refléter la complexité du modèle émotionnel informatique utilisé. Les travaux que nous avons effectués selon ces deux axes ont été évalués par des études perceptives menées sur des utilisateurs.

La combinaison de ces deux axes de recherche est rare dans les systèmes d'agents virtuels expressifs existants. Ainsi, nos travaux ouvrent des perspectives pour l'amélioration de la conception d'agents virtuels expressifs et de la qualité de l'interaction homme machine basée sur les agents virtuels expressifs interactifs.

L'ensemble des logiciels que nous avons conçus forme notre plateforme d'agents virtuels MARC (Multimodal Affective and Reactive Characters). MARC a été utilisée dans des applications de natures diverses : jeu, intelligence ambiante, réalité virtuelle, applications thérapeutiques, performances artistiques, etc.

Mots clés : Interaction homme-machine, agents virtuels expressifs, animation faciale expressive temps-réel, modèles informatiques des émotions, interaction affective, études perceptives

Abstract

Emotions and their expressions by virtual characters are two important issues for future affective human-machine interfaces. Recent advances in psychology of emotions as well as recent progress in computer graphics allow us to animate virtual characters that are capable of expressing emotions in a realistic way through various modalities. Existing virtual agent systems are often limited in terms of underlying emotional models, visual realism, and real-time interaction capabilities.

In our research, we focus on virtual agents capable of expressing emotions through facial expressions while interacting with the user. Our work raises several issues: How can we design computational models of emotions inspired by the different approaches to emotion in Psychology? What is the level of visual realism required for the agent to express emotions? How can we enable real-time interaction with a virtual agent? How can we evaluate the impact on the user of the emotions expressed by the virtual agent?

Our work focuses on computational modeling of emotions inspired by psychological theories of emotion and emotional facial expressions by a realistic virtual character. Facial expressions are known to be a privileged emotional communication modality. Our main goal is to contribute to the improvement of the interaction between a user and an expressive virtual agent. For this purpose, our research highlights the pros and cons of different approaches to emotions and different computer graphics techniques.

We worked in two complementary directions. First, we explored different approaches to emotions (categorical, dimensional, cognitive, and social). For each of these approaches, a computational model has been designed together with a method for real-time facial animation. Our second line of research focuses on the contribution of visual realism and the level of graphic detail of the expressiveness of the agent. This axis is complementary to the first one, because a greater level of visual detail could contribute to a better expression of the complexity of the underlying computational model of emotion. Our work along these two lines was evaluated by several user-based perceptual studies.

The combination of these two lines of research is seldom in existing expressive virtual agents systems. Our work opens future directions for improving human-computer interaction based on expressive and interactive virtual agents.

The software modules that we have designed are integrated into our platform MARC (Multimodal Affective and Reactive Characters). MARC has been used in various kinds of applications: games, ubiquitous intelligence, virtual reality, therapeutic applications, performance art, etc.

Keywords: human computer interaction, expressive virtual agents, expressive facial animation, real-time computational models of emotions, affective interaction, perceptual studies

Remerciements

Je remercie tout d'abord Sylvie Pesty de l'université de Grenoble et Elisabeth André de l'université d'Augsburg d'avoir accepté de rapporter cette thèse et pour leurs commentaires très positifs sur ce manuscrit. Merci également à Marc Cavazza de l'université de Teesside, Anne Vilnat de l'université Paris Sud XI, et Gaspard Breton de la société Dynamixyz, d'avoir accepté de siéger dans mon jury de thèse.

Merci également au LIMSI-CNRS de m'avoir accueilli ainsi qu'à mes deux directeurs de thèse : Merci à Jean-Claude Martin pour avoir su diriger mes recherches sans les limiter, me permettant ainsi de m'approprier mon sujet de thèse, et de m'avoir permis toutes les collaborations interdisciplinaires qui j'espère, contribuent à la qualité de cette thèse. Merci à Christian Jacquemin, d'abord pour m'avoir fait découvrir le LIMSI il y a de ça cinq ans, mais surtout pour m'avoir souvent donné un éclairage différent sur mes problématiques de recherche, me poussant ainsi à pousser plus loin ma réflexion. Merci également pour les travaux art-sciences, et notamment le projet *Réalité Augmentée Mobile*, auxquels j'ai eu le plaisir de participer, qui ont été une source d'enrichissements culturels et intellectuels, et qui n'auraient pas été possibles sans lui.

Merci à toutes les personnes avec qui j'ai pu collaborer au cours de ces années de doctorat. Chronologiquement : Stéphanie Buisine, Céline Clavel, Bertrand Planes, Michel-Ange Amorim, Pascale Barret, Rudi Giot, Marc Rébillat, Brian Katz, Victoria Eyharabide et David Sander. Toutes ces collaborations ont été pour moi un grand enrichissement scientifique qui j'espère transparait dans ce manuscrit.

Mais une thèse ne résume pas simplement à des interactions scientifiques !

C'est d'abord à mes parents que je dois d'avoir fait cette thèse. C'est en effet ma mère qui, alors que je n'avais que treize ans, m'a donné mon premier ordinateur capable d'accueillir une carte graphique, et c'est mon père qui, profitant de l'occasion, m'a transmis le virus de la programmation. Sans ces deux évènements, j'aurais sans doute eut des meilleurs notes au bac, mais je doute d'avoir obtenu un doctorat d'informatique.
Merci à toute ma famille.

Merci également à tous les gens dont la présence a rendu la vie au LIMSI agréable : Rami, Jonathan, Céline, François, Laurent, Fred, Wai, Tifanie, Marc, Gaëtan, Victoria, Julien, Yoren, Tom, etc. Avec deux mentions spéciales, une pour Rami, pour le soutien moral et l'inspiration qu'il m'a apporté au début de ma thèse, et une pour Jonathan, qui a si souvent prêté sa voix, ses gestes et son apparence à mes personnages virtuels que j'ai finalement de lui un double numérique complet (et c'est vrai !).

Merci enfin à Aurore, qui partage ma vie depuis maintenant plus de 7 ans et qui supporte je ne sais comment mes périodes de mauvaise humeur et mes moments de stress. Merci d'avoir eu le courage de relire minutieusement les deux cents pages de ce manuscrit. Merci de m'avoir écouté t'expliquer des centaines de problèmes et concepts informatiques, juste parce que te les expliquer en les vulgarisant m'aidait à y trouver des solutions. Et aussi pour tout le reste… Merci mille fois.

Table des matières

Chapitre 1. Introduction générale

Sommaire du chapitre

1.1 Constat et besoins

1.1.1 L'émergence de l'informatique grand public

Avec l'émergence rapide de l'informatique grand public, l'informatique est de plus en plus omniprésente, rapide et performante. Pourtant, l'interaction homme-machine, c'est-à-dire la manière avec laquelle un utilisateur interagit avec un système informatique, n'évolue pas aussi rapidement (Beaudouin-Lafon, 2004). Elle repose sur des dispositifs d'interactions créés il y a maintenant cinquante ans, comme par exemple, la souris d'ordinateur, créée en 1963 par Douglas Engelbart (Figure 1).

Figure 1 - La souris inventée par Douglas Engelbart (1963)

Ces dispositifs, en étant associés à des interfaces logicielles de type WIMP (W*indow, Icon, Menu, Pointing device*), ont été acceptés par le grand public car ils reposent sur une métaphore du monde réel : un bureau et des documents rangés dans des dossiers. Effectivement, cette métaphore semble bien adaptée à une utilisation professionnelle bureautique de l'informatique.

Depuis, l'informatique a évolué, et avec cette évolution, le cadre strict d'une utilisation bureautique professionnel est dépassé (Jeux vidéo, mondes virtuels, réseaux sociaux, etc.). La métaphore du bureau n'est alors plus forcément la plus adaptée. Il devient nécessaire de concevoir des interfaces plus intuitives, plus ludiques, plus ergonomiques et plus appropriées à une utilisation par un public de plus en plus large, avec des besoins et des enjeux de plus en plus variés.

En effet, l'évolution de l'informatique vers les applications grand public diversifie considérablement les types d'utilisateurs. Certains utilisateurs sont alors susceptibles de manquer d'expérience pour utiliser correctement l'interface proposée. Ils peuvent alors perdre du temps ou se fatiguer inutilement simplement car ils n'ont pas vu ou pas compris une interaction particulière qui leur aurait permis d'être plus efficaces.

Selon Beaudouin-Lafon, (2004), *le paradoxe des interfaces homme-machine est que les meilleures interfaces sont celles qui sont invisibles, celles que l'utilisateur ne remarque pas, et avec lesquelles il trouve l'interaction « naturelle »*. Cependant, Cassell (2001a) fait valoir que pour interagir avec un système, l'utilisateur a besoin d'être capable de situer son « centre d'intelligence ». Selon l'auteur, l'informatique totalement « invisible » pose donc un problème.

1.1.2 Les personnages virtuels comme interface utilisateur

Dans ce contexte, l'utilisation d'agents conversationnels animés (ACA) constitue une direction de recherche qui semble pertinente (Cassell, 2000, Rist, André et Müller, 1997). L'interface informatique prend alors forme humaine et on utilise l'interaction entre deux humains comme métaphore de communication. Le personnage virtuel devient l'entité dans laquelle l'utilisateur situe l'intelligence du système (Cassell, 2001a). Les ACA

semblent donc offrir une interface intuitive et naturelle car ils utilisent les modalités de la communication humaine. De plus, les ACA sont « non invisibles », car si le système informatique réel est masqué, l'affichage d'un personnage virtuel permet d'y situer le « centre d'intelligence » du système.

Nous utiliserons dans la suite de ce manuscrit l'expression « agent virtuel » pour signifier « un personnage virtuel animé dynamiquement en fonction du contexte ou des actions de l'utilisateur ».

Les ACA que l'on trouve aujourd'hui prennent généralement la forme de systèmes de dialogues avec une représentation graphique, soit d'un personnage schématique, soit d'une forme plus ou moins réaliste d'un humain virtuel. Ces systèmes permettent de centraliser différentes tâches, telle que la recherche d'information, l'aide à la navigation, le divertissement, etc. Les ACA sont de plus en plus répandus (Figure 2). On les trouve sur le web (Paiva et Machado, 2002), pour répondre aux questions des visiteurs d'un site, les aider à se diriger ou encore à trouver une information. On les trouve également dans les applications pédagogiques (*e-learning*), les applications liées au monde de l'entreprise (Serious Games) et les applications ludiques (Jeu vidéo). Cependant, leur comportement est en général extrêmement prédéfini, ce qui limite leur interactivité.

Figure 2 - Exemples d'agents virtuels. A Gauche, l'agent web Andrew (société La Cantoche[1]), au centre, l'agent expressif Greta (Pelachaud et al, 2006), à droite, un agent pour la langue des signes (Héloir et al. 2008)

1.1.3 Les agents virtuels comme outils expérimentaux

Les agents virtuels peuvent également servir d'outils pour mener des expérimentations contrôlées en psychologie. Ils permettent ainsi de mieux comprendre la perception humaine et l'interaction entre humains. En retour, cela permet d'améliorer la qualité des agents virtuels. Les personnages virtuels permettent de créer une interaction temps-réel « naturelle » entre un humain et une machine, tout en contrôlant et manipulant précisément les différents paramètres de cette interaction. Les agents virtuels ont ainsi été utilisés pour étudier différents aspects de la perception humaine et de l'interaction *homme-ACA* (Pelachaud, 2005, Rehm et André, 2005, de Melo et Gratch, 2009).

Plus généralement, Wallraven et al. (2005) ont montré l'influence réciproque entre les agents virtuels et la perception humaine. Dans le modèle proposé par les auteurs, les agents virtuels sont utilisés pour étudier la perception humaine, et réciproquement, la perception humaine est utilisée pour étudier et améliorer les modèles et les rendus d'agents virtuels.

1.1.4 Limites des agents virtuels actuels

Les agents virtuels actuels souffrent de limitations importantes, à la fois technologiques et scientifiques. Tout d'abord, il est nécessaire de s'interroger sur l'utilité des agents virtuels lorsque qu'ils ne contribuent pas directement à rendre l'interaction plus fluide, plus simple, ou qu'ils risquent de distraire l'utilisateur de son objectif et de sa tâche en cours (Dehn et Mulken, 2000). Par exemple, dans certaines applications, l'agent virtuel

[1] http://www.cantoche.com/

ne semble pas s'intégrer au reste du système, ni apporter d'information supplémentaire utile à l'utilisateur. C'est par exemple le cas du tristement célèbre « trombone assistant » de Microsoft® Word. On peut émettre l'hypothèse que le manque d'intégration de ce type d'agent vient du mélange entre la métaphore du bureau virtuel et la métaphore de l'agent virtuel. Dans une application où l'utilisateur est focalisé sur une application bureautique, l'agent assistant peut devenir redondant, voire gênant, car il ne s'intègre pas à la métaphore principale. Cet exemple illustre clairement l'importance de concevoir des applications utilisant les agents virtuels de manière appropriée.

L'une des limitations majeures des ACA actuels grand public (présents sur le web par exemple) est qu'ils ignorent généralement une caractéristique fondamentale de la communication humaine : l'émotion. Les émotions sont pourtant largement étudiées en psychologie, et sont reconnues comme une facette fondamentale de la communication humaine. Lorsque des humains communiquent, ils créent une boucle interactive affective dans laquelle chacun s'adapte (de manière consciente ou inconsciente) aux émotions exprimées par l'autre. Il semble donc nécessaire de simuler ce comportement affectif lors d'une interaction avec un humain virtuel.

1.1.5 Agents virtuels et informatique affective

Sloman et Croucher (1981) font apparaitre pour la première fois la nécessité de prendre en compte les émotions dans l'intelligence artificielle. Cette idée est ensuite étendue plus largement par Picard (1997) à l'informatique et en particulier aux interactions homme-machine. L'informatique affective (ou *Affective Computing*) est *l'informatique qui traite, simule, ou influence les émotions ou les autres phénomènes affectifs* (Picard, 1997, 2010). Un agent virtuel capable d'exprimer son état mental interne (ce qui inclut son état émotionnel), peut utiliser différents canaux de communications, verbaux et non verbaux. Par la suite, nous appellerons ce type d'agent « agent virtuel expressif ».

Simuler une interaction humaine affective implique donc : 1) de savoir reconnaître l'état émotionnel de l'utilisateur, 2) de doter l'agent d'une représentation interne d'état émotionnel et des raisonnements associés, et 3) de permettre à l'agent d'exprimer son état émotionnel interne. Chacun de ces aspects pose des problématiques scientifiques complexes. Par exemple, doter l'agent d'un état émotionnel propre nécessite de modéliser un processus émotionnel autonome et dynamique. Pour cela, il est nécessaire d'étudier les travaux en psychologie sur les processus émotionnels de l'être humain. Cependant, en psychologie, aucune définition de ce qu'est une émotion ne fait totalement consensus (Gross et Feldman-Barrett, 2011). Plusieurs théories coexistent et s'influencent (Scherer, 2010). On trouve ainsi différentes approches de l'émotion : catégorielles, dimensionnelles, cognitives, sociales, etc. Dans le cadre d'une application informatique avec des agents virtuels expressifs et interactifs, l'agent doit être doté d'un modèle informatique des émotions. Plusieurs modèles de gestion informatique des émotions ont été proposés (Pelachaud et Bilvi, 2003, Rivière et Pesty, 2010) et évalués (Marsella et Gratch, 2006), mais il reste nécessaire de sélectionner, modéliser et d'implémenter plusieurs approches des émotions afin de les évaluer avec des utilisateurs et de comparer leur impact en situation d'interaction.

De plus, la large variété des émotions chez l'humain reste à explorer. Nous verrons en effet dans la suite de ce document qu'au-delà des émotions dites « primaires » (Ekman et Friesen, 1975), il existe différents états affectifs et émotions « complexes ».

1.1.6 Expressions faciales chez les agents virtuels expressifs

Ekman et Friesen (1975) ont étudié les expressions faciales de certaines émotions appelées émotions de base (Joie, Colère, Peur, Surprise, Dégout, Tristesse). Leurs travaux décrivent en détail les différents indices visibles sur les familles d'expressions pouvant être associées à ces émotions. Ces travaux ont donc logiquement servi de base pour l'animation faciale de nombreux agents virtuels expressifs.

Dans ses premiers travaux sur l'animation faciale (Figure 3), Parke (1974) a mis en évidence qu'un visage virtuel permet de véhiculer un contenu émotionnel. Bien que ces travaux ne traitent pas directement des émotions, ils ouvrent la voie aux travaux sur la simulation des expressions faciales d'émotions.

Figure 3 - Le visage de synthèse animé de Parke (F. I. Parke, 1974)

La qualité graphique des visages virtuels a depuis nettement progressé. Néanmoins, la plupart des agents virtuels expressifs se limitent aux émotions de base proposées par Ekman et peu d'entre eux explorent d'autres émotions et états mentaux. De plus, la plupart de ces agents se limitent seulement à une expression par émotion. Pourtant ces émotions de base apparaissent rarement lors des interactions humaines. Lorsqu'elles apparaissent, elles sont souvent mélangées avec d'autres états mentaux plus complexes (Scherer et Ceschi, 1997, Abrilian et al. 2005). Ainsi, il est nécessaire d'explorer d'autres approches des émotions que les émotions de base, ainsi que les liens entre ces approches et l'animation faciale de personnages virtuels expressifs. De plus, peu de plateformes permettent simultanément d'utiliser et d'étudier un agent virtuel réaliste et interagissant en temps réel avec un utilisateur. Pourtant, l'application temps réelle est souvent la finalité recherchée, et elle nécessite donc d'être étudiée de manière formelle.

1.2 Problématique et objectifs scientifiques

L'objectif principal de cette thèse est de contribuer à l'amélioration de l'interaction entre les agents expressifs et l'utilisateur en se focalisant sur le modèle émotionnel sous-jacent. Nous avons donc proposé plusieurs modèles computationnels (et leurs liens avec l'animation faciale) adaptés des modèles des émotions issus de la psychologie. Proposer plusieurs modèles nous a permis de comparer leurs apports respectifs dans le cadre de l'interaction *Homme-ACA.*

Les travaux effectués dans cette thèse portent donc sur l'interaction affective non verbale avec un visage virtuel expressif réaliste. Ce thème de recherche pose un certain nombre de problématiques dans plusieurs domaines.

1.2.1 Les agents virtuels réalistes

La problématique semble simple : « comment et pourquoi interagir avec un agent virtuel interactif et réaliste ? ». Pourtant, cette question soulève de nombreuses questions de recherche.

MacDorman et al. (2010) ont montré que l'utilisation de personnages plus réalistes visuellement augmente l'engagement de l'utilisateur dans l'interaction avec l'agent virtuel. De plus, Yee et al., (2007) ont réunis les résultats de 22 études comparant l'effet d'un réalisme visuel faible à un réalisme élevé. Sur ces 22 études, 19 suggèrent qu'un rendu plus réaliste a un effet positif sur la perception que les sujets ont de l'agent. Cependant, le *réalisme* est une notion subjective qui nécessite d'être évaluée par des études perceptives effectuées sur des utilisateurs. Dans nos travaux, le terme *réaliste* est utilisé à double sens. D'une part, nous avons proposé et

implémenté différents modèles informatiques (inspirés de différentes approches des émotions issues de la psychologie) pour contrôler l'expressivité du visage virtuel. Nous avons ainsi cherché à obtenir un haut niveau de ***réalisme comportemental*** du visage virtuel. Ce réalisme comportemental s'exprime à la fois en termes de dynamique faciale et de cohérence expressive. D'autre part, nous avons utilisé les techniques récentes de l'informatique graphique pour obtenir un ***réalisme visuel*** en modélisant un agent virtuel animé en temps réel avec un haut niveau de détails graphiques (comparé aux autres agents virtuels expressifs interactifs existants).

1.2.2 Les émotions

Le réalisme d'un agent virtuel repose notamment sur sa capacité à exprimer des émotions durant l'interaction avec l'utilisateur. Pour cela, nous avons choisi de proposer différents modèles computationnels des émotions, en se basant sur les théories et modèles proposées en psychologie.

Cette approche pose deux problématiques principales : Comment choisir, parmi la littérature, les modèles émotionnels les plus pertinents ? Comment concevoir un modèle computationnel fonctionnel à partir d'un modèle théorique, souvent sous-spécifié, issu de la psychologie ? Pour effectuer ces choix, il est nécessaire d'évaluer successivement plusieurs modèles des émotions, afin de déterminer les avantages et les inconvénients de chacun.

1.2.3 Les expressions faciales

Les expressions faciales sont un canal de communication affective non verbal important. Les agents virtuels utilisant les expressions faciales pour exprimer des émotions sont de plus en plus répandus. Pourtant, plusieurs problématiques sont encore en suspens. Par exemple, comment générer des expressions faciales qui soient bien reconnues ? Quels modèles d'émotion utiliser pour générer les expressions faciales ? Comment exprimer des mélanges d'émotions et des états émotionnels subtils ?

En plus de ces problématiques, de nouvelles directions de recherche sont apparues durant ces dernières années de par l'évolution rapide des technologies de rendu graphique temps-réel. En effet, il est aujourd'hui possible d'animer des visages humains de très haute qualité, et utilisant des indices faciaux (par exemple la rougeur de la peau, les rides expressives). Pourtant, les agents virtuels interactifs actuels n'ont pas encore totalement tiré parti de ces nouveaux outils, et beaucoup de ces nouvelles possibilités restent à exploiter et à évaluer dans un contexte interactif. Quelles différences observe-t-on en utilisant un rendu réaliste, des rides d'expressions détaillées, ou des expressions faciales temporaires ? Plus généralement, comment l'utilisation des techniques graphiques récentes est-elle bénéfique pour les agents expressifs interactifs ?

1.2.4 Agents virtuels expressifs interactifs

C'est pour tenter d'apporter des réponses à ces questions que nous avons choisi de centrer cette thèse sur les expressions faciales d'agent virtuel en situation d'interaction.

L'expression « en situation d'interaction » soulève également plusieurs problématiques : comment combiner le réalisme visuel et le réalisme comportemental avec l'interaction temps réel ? Cette combinaison nécessite en effet la mise en place d'architectures dédiées (logicielles et matérielles). De plus, pour pouvoir comparer différents modèles d'émotions et pour être applicables à diverses applications, ces architectures doivent être modulaires et flexibles.

1.2.5 Evaluation de l'impact sur les utilisateurs

Nos travaux sur le réalisme, l'expressivité et le comportement affectif de l'agent partagent le but d'améliorer l'interaction entre l'agent virtuel et l'utilisateur. Comme l'ont montré Wallraven et al. (2005), il est donc nécessaire d'évaluer nos travaux par des expérimentations perceptives.

D'une part, nous devons évaluer le réalisme visuel du visage. La qualité visuelle du visage, l'utilisation de rides d'expression, influent-ils sur la perception émotionnelle de l'utilisateur ? En d'autres termes, l'utilisateur reconnait-il mieux les émotions exprimées par l'agent lorsqu'on augmente la qualité visuelle de l'agent ?

D'autre part, nous devons évaluer le réalisme comportemental. Quelles sont les différences que l'on observe lorsqu'on compare l'utilisation de différents modèles internes d'émotion ? Observons-nous des différences dans la perception que l'utilisateur a de l'agent et de ses expressions ? Observons-nous une modification du comportement de l'utilisateur ou de ses performances lors d'une tâche interactive ?

1.3 Démarche et contributions

1.3.1 Démarche itérative

Les modèles et approches des émotions sont si complexes qu'un seul modèle peut faire l'objet d'une thèse entière. Dans cette thèse nous avons préféré aborder de manière incrémentale plusieurs modèles afin de pouvoir les comparer et mettre en relief leurs complémentarités.

La démarche utilisée durant cette thèse est une approche itérative en trois étapes (Figure 4).

La première étape de cette démarche consiste à sélectionner une théorie d'émotion pertinente pour l'animation faciale et la modélisation du comportement affectif des agents virtuels. Il est pour cela nécessaire d'étudier la littérature en psychologie des émotions et en informatique. La psychologie permet de comprendre les enjeux de la théorie considérée et les différences qu'elle présente par rapport aux autres théories émotionnelles. Les travaux en informatique permettent de comprendre les limites des modèles computationnels existants et basés sur la théorie considérée, aussi bien au niveau modélisation des émotions qu'au niveau de l'animation faciale.

La seconde étape de notre démarche consiste à proposer et implémenter un modèle informatique inspiré du modèle émotionnel étudié. Cette étape implique des choix d'implémentation et de modélisation/formalisation de la théorie afin de l'adapter à des processus computationnels et aux contraintes de l'animation faciale temps réel. L'implémentation de la théorie donne lieu à la fois au développement d'un modèle computationnel dédié, et à des implémentations dédiées dans le système d'animation faciale. Ces développements ont été réalisés avec pour objectif de conserver les modèles implémentés de manière successive, augmentant progressivement les possibilités de l'ensemble des logiciels développés.

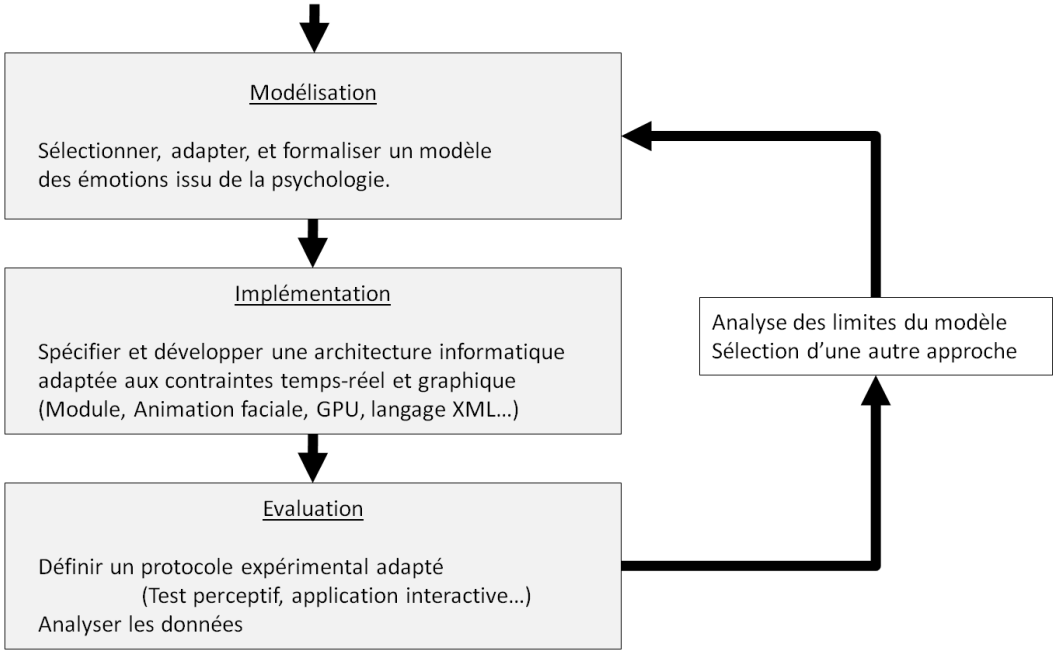


Figure 4 - démarche expérimentale adoptée dans cette thèse.

La troisième étape consiste à mettre en place un protocole expérimental permettant de tester le modèle proposé. L'analyse des données issues de l'expérimentation permet de mettre en lumière ce qu'apporte la théorie, d'expliciter les limites qu'elle impose, ce qu'apporte le modèle informatique proposé et ses limites Il est ensuite possible d'itérer la démarche, soit en élargissant le modèle informatique, soit en sélectionnant une autre approche complémentaire issue de la psychologie permettant de dépasser certaines limites des approches considérées jusqu'alors, et en proposant un nouveau modèle informatique.

Tout au long de cette thèse, cette approche itérative a bénéficié de plusieurs collaborations multidisciplinaires (psychologues et ergonomes) pour la conception des modèles informatiques inspirés des théories des émotions et la mise en place des protocoles expérimentaux. Ainsi, nous avons collaboré avec Céline Clavel (LIMSI-CNRS), Stéphanie Buisine (ENSAM), Michel-Ange Amorim (UFR STAPS, Université Paris Sud XI), et David Sander (Université de Genève).

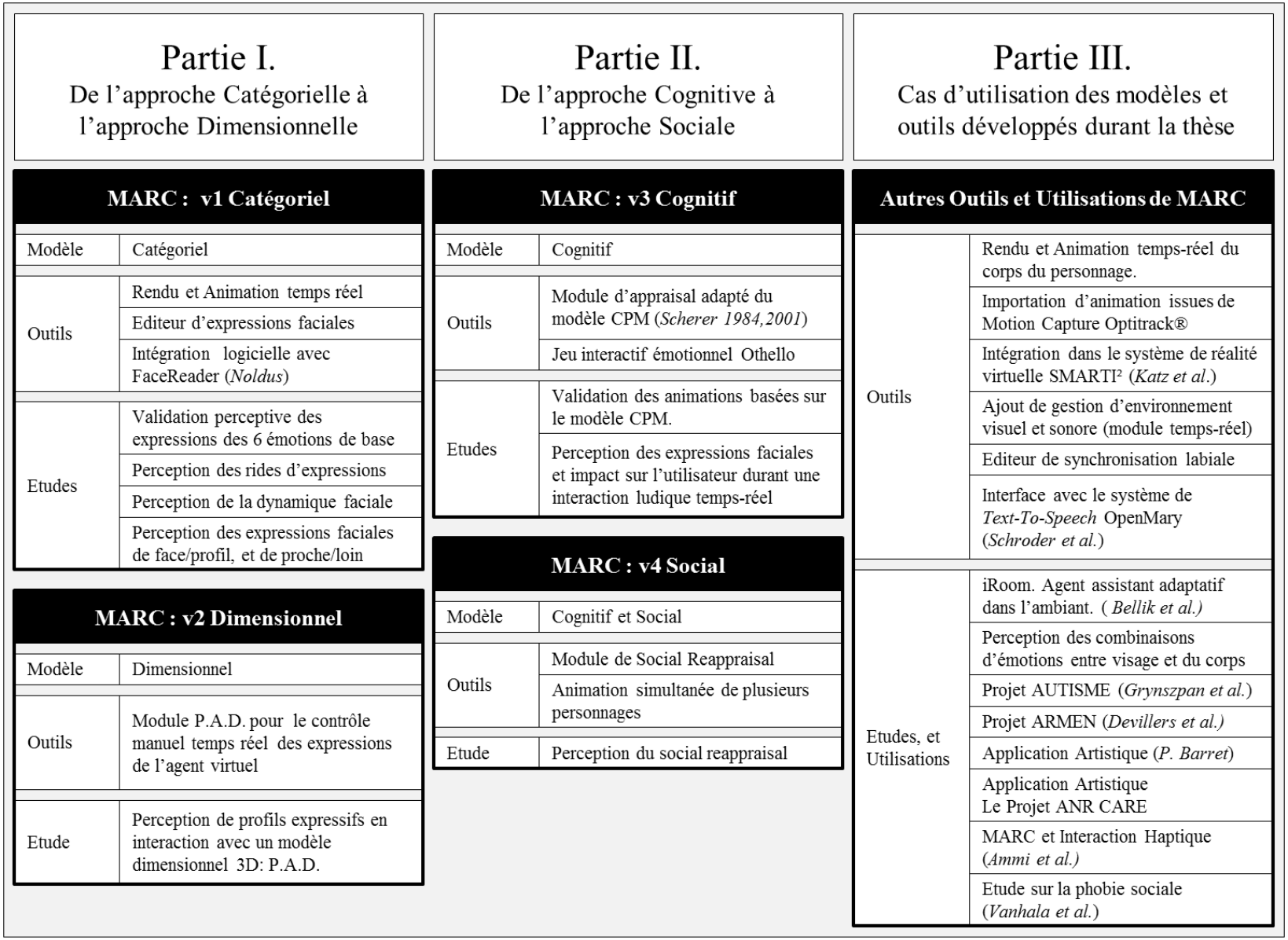
1.3.2 Organisation du mémoire

Cette approche itérative se reflète dans l'organisation du présent manuscrit, schématisé dans la Figure 5. Chaque version de MARC (v1 à v4) correspond à une itération de notre démarche scientifique développée tout au long de cette thèse. Pour chaque itération la Figure 5 reprend les modèles psychologiques considérés, les outils logiciels réalisés, et les études qu'ils ont permis de mener.

1.3.3 Contributions de la thèse

Les différentes études menées durant cette thèse ont été rendues possibles par le développement de logiciels dédiés. De plus, nous avons proposé, spécifié, et implémenté plusieurs modèles computationnels des émotions. A travers ces développements, notre ambition était de proposer des améliorations aux modèles informatiques et aux systèmes existants.

Figure 5 - Plan d'organisation du manuscrit



1.3.3.1 Modèles émotionnels

En utilisant les différentes classifications des modèles émotionnels disponibles dans la littérature (Scherer, 2010, Gross et Feldman-Barrett, 2011) (Cf. section 2.1.2), nous avons sélectionné quatre approches des émotions de complexité croissante et qui semblent pertinentes pour l'animation faciale temps réel :

1) L'approche catégorielle : L'approche catégorielle permet de représenter simplement les émotions par des labels. De plus, la littérature fourni un grand nombre de descriptions des expressions associées à plusieurs de ces labels émotionnels (Ekman et Friesen, 1975).

2) l'approche dimensionnelle : L'approche dimensionnelle situe les émotions dans un espace continu (Russell et Mehrabian, 1977). Par exemple, l'utilisation de la dimension Positif/Négatif permet de dire qu'une émotion est plus positive qu'une autre.

3) l'approche cognitive : L'approche cognitive permet de faire intervenir un raisonnement via une évaluation dynamique de la situation (Lazarus, 1968, Scherer, 1984) et un modèle séquentiel d'expressions faciales d'émotion (Scherer, 2001).

4) l'approche sociale : L'approche sociale consiste à donner de l'importance au contexte social dans lequel a lieu l'événement responsable de l'émotion. Ce contexte social est pertinent pour des interactions impliquant plusieurs utilisateurs ou plusieurs personnages virtuels (Averill, 1985).

Ces approches ont été sélectionnées car elles permettent d'adresser des problématiques complémentaires et des modèles de complexité croissante en psychologie. Elles répondent également à différents besoins en termes d'interaction homme-machine. Ces différents modèles et leurs implémentations seront présentés au fur et à mesure de ce manuscrit, en les introduisant dans le contexte de leur développement.

Dans cette thèse, nous appellerons *modèle* les différentes représentations simplifiées du processus émotionnel qui servirons de base à l'implémentation. Dans la plateforme MARC, les modèles que nous avons adaptés de la psychologie sont concrétisés par des *modules* logiciels émotionnels

Pour évaluer ces différents modèles d'émotions, nous avons mis en place plusieurs types d'interaction, allant du contrôle manuel de l'expression de l'agent, à l'interaction avec un agent autonome. Nous avons ainsi pu mener des études perceptives à travers différents contextes interactifs.

1.3.3.2 Outils logiciels réalisés

L'ensemble des logiciels réalisés dans le cadre de cette thèse s'intègre dans notre plateforme d'agents virtuels expressifs nommée MARC (*Multimodal Affective and Reactive Characters*[2]). La conception de MARC résulte d'un processus itératif. Ainsi, les versions successives de MARC ont contribué à son développement, que nous diviserons en plusieurs catégories. Les développements graphiques, les développements des modèles informatiques émotionnels, et les développements liés à l'interconnexion de MARC avec d'autres systèmes informatiques.

Chaque version de MARC sera donc présentée en détails en séparant distinctement ces trois types de développements.

1.3.3.3 Etudes réalisées

Pour chacune de ces approches, plusieurs expérimentations ont été menées pour évaluer leurs apports et leurs limites. Les résultats principaux des études présentées dans ce manuscrit sont résumés dans la Figure 6.

[2] http://marc.limsi.fr

Figure 6 - Récapitulatif des études menées dans le cadre de la thèse

Etudes				
Approche	**Sujet de l'étude**	**Objectifs**	**Résumé du protocole**	**Résultats Principaux**
Catégorielle	Validation perceptive de reconnaissance des six expressions de base	Valider la reconnaissance catégorielle des expressions faciales des six émotions de base créées avec MARC	Présentation des expressions et choix multiple de labels émotionnels	→ Les expressions sont bien reconnues. → Quelque confusions entre Peur et Surprise
	Influence du réalisme des rides d'expressions	Comparer l'influence du réalisme des rides d'expressions sur la perception émotionnelle de l'utilisateur et sur la préférence utilisateur	Utilisation de différents types de rides d'expressions, 'pas de rides', 'rides non réalistes', 'rides réalistes' → Tache de reconnaissance des émotions → Evaluation de la préférence	→ Le type de ride n'influence pas la reconnaissance des émotions, mais la présence de rides augmente l'intensité perçue → Les utilisateurs préfèrent les expressions utilisant des rides d'expression réalistes
	Perception de la dynamique des expressions faciales	Etudier la perception de la dynamique des expressions faciale et des émotions associées lors de changements visuels majeurs (changement de plan de caméra, occlusion…)	Le sujet observe des animations coupées brusquement Une image statique de l'expression coupée est présentée avec un décalage d'intensité (-30% -15% 0% +15% et +30%) Le sujet doit répondre « plus » ou « moins » intense	→ Les sujets anticipent une intensification lorsqu'on coupe une expression peu intense → Les sujets anticipent une réduction d'intensité lorsqu'on coupe une expression très intense → Effet observé sur toutes les catégories d'émotion, mais plus ou moins marqué en fonction des catégories
	Reconnaissance des expressions faciales de face et de profil, de près, et de loin	Etudier l'influence de l'angle de vue de l'observateur sur la reconnaissance des expressions faciales	Les expressions faciales sont présentées soit de face, soit de profil. Choix multiple de labels émotionnels	→ Aucune différence significative de reconnaissance catégorielle → Pour certaines émotions les sujets reportent une plus grande confiance en leurs réponses de face
Dimensionnelle	Contrôle temps réel de l'expression de l'agent doté d'un profil expressif	Etudier la perception du profil expressif de l'agent lors du contrôle temps réel via un espace dimensionnel 3D	Via une souris 3D, l'utilisateur fait évoluer l'état émotionnel de l'agent dans un espace 3D. L'expressivité de l'agent est modulée par son profil expressif que l'utilisateur doit percevoir.	→ Les profils expressifs sont bien reconnus. → Les sujets sont capables de manipuler correctement l'expression de l'agent.
Cognitive	Validation perceptive de séquences expressives émotionnelles basées sur un modèle cognitif	Valider perceptivement les animations générées par le module d'évaluation cognitive	Les animations basées sur un modèle adapté de la théorie du CPM (Scherer 1984, 2001) sont présentées au sujet qui doit reconnaitre l'émotion exprimée.	→ Les profils CPM des émotions exprimées sont reconnues par les sujets (taux de reconnaissance supérieur au seuil de hasard)
	Comparaison de l'approche catégorielle et cognitive via une application ludique temps réel	Etudier l'impact du modèle cognitif pour l'animation faciale dans la perception de l'agent et le comportement de l'utilisateur durant une interaction.	L'utilisateur joue à Othello contre l'agent virtuel expressif. Les expressions faciales de l'agent sont soit neutres, soit catégorielles, soit dynamiquement générées par le processus cognitif interne.	→ Utiliser un modèle cognitif augmente la perception que l'utilisateur a des capacités mentales de l'agent → Utiliser un modèle cognitif modifie le comportement de l'utilisateur au cours du jeu
Sociale	Etude de la perception du concept de Social Appraisal en utilisant deux agents virtuels	Etudier l'impact d'un modèle social sur la perception de deux agents virtuels. Perçoit-on l'influence de l'un des agents sur l'autre ?	Deux agents sont cote à cote, et ils réagissent à un stimuli. Puis, l'un des agents regarde l'autre, et effectue une seconde réaction « sociale ». Plusieurs modèles de réaction sociale sont comparés.	→ L'utilisation du modèle social induit la perception par le sujet d'une influence d'un agent sur l'autre. → Notre modèle de Règles Logique pour le calcul de la réaction sociale augmente l'expressivité perçue.

1.3.3.4 Limites de la thèse

Nos travaux se focalisent uniquement sur les émotions (affects au sens large) et nous ne considérons pas d'autres fonctions de communication des personnages virtuels (par exemple, les rétroactions, ou *backchannels*, qui consistent à fournir un retour non émotionnel à l'utilisateur durant l'interaction (de Sevin et al., 2010)).

En ce qui concerne les techniques d'animation faciale, nous nous sommes limités à l'approche dite « animation paramétrique » inspirée du modèle MPEG-4. Nous n'abordons pas l'animation par simulation physique du visage (Terzopoulos et Waters, 1993). En effet, si cette approche présente certains avantages en termes de réalisme et de dynamique des expressions faciales, elle est peu adaptée à la génération d'animations faciales en temps réel et à l'interaction. Nous ne considérons pas non plus la capture d'expressions faciales. Bien que recevant un intérêt grandissant dans la communauté de l'animation faciale, elles ne concernent pas directement notre approche qui vise à générer algorithmiquement et dynamiquement l'animation faciale en fonction de l'interaction avec l'utilisateur. De plus la capture d'expressions faciales pose d'autres problématiques, telles que les méthodes de capture et le prétraitement des données, que nous n'avons pas pour objectif d'explorer.

Cette thèse n'aborde pas non plus les problématiques liées à la synthèse audiovisuelle de la parole, aux mouvements de tête, aux jeux de regard et l'utilisation de la posture de l'agent virtuel. Nos travaux sont en effet articulés uniquement autour de l'animation faciale interactive. En revanche, la plateforme MARC a été dotée de toutes ces capacités techniques. Ainsi, MARC permet d'explorer des problématiques plus larges que celles abordées dans la thèse présentée ici. Par exemple, un éditeur dédié permettant l'édition des animations posturales, ainsi que l'importation de fichier de capture de mouvements ont été développés pour répondre aux besoins de différentes recherches menées par d'autres membres de notre équipe et partenaires utilisant MARC. Nous avons utilisé ces capacités posturales expressives dans le cadre de cette thèse, pour mener une étude exploratoire sur les combinaisons entre expressions faciales et posturales. Cette étude est décrite en section 7.1.2.

Chapitre 2. Etat de l'art

Sommaire du chapitre

2.1 Approches, théories et modèles des émotions en psychologie et en informatique

Dans cette première partie de l'état de l'art, nous allons nous concentrer sur les différentes théories des émotions issues de la psychologie. Dans un second temps, nous aborderons l'animation faciale de synthèse. Nous relierons ensuite ces deux domaines en étudiant l'état de l'art sur les agents virtuels existants.

2.1.1 Définition d'une émotion

Dans son article *What is an Emotion?,* James (1884) définit les émotions comme des expériences mentales dues à un ensemble de changements physiologiques. James soutient que « nous sommes tristes parce que nous pleurons, nous avons peur parce que nous tremblons » etc. Cette causalité a depuis largement été remise en question. Les théories actuelles des émotions présentent l'émotion comme un phénomène complexe intervenant de manière synchrone à plusieurs niveaux (psychologique, physiologique, subjectif, moteur, etc.) (Scherer 1984, 2001).

En psychologie, il est généralement admis que les émotions sont l'une des facettes centrales de la psyché humaine (Gross et Feldman-Barrett 2011). La plupart des théories des émotions admettent l'existence de plusieurs composantes émotionnelles. Par exemple, les processus cognitifs, les réactions physiologiques périphériques, les changements motivationnels, l'expression motrice, et le sentiment subjectif associé (Scherer 2001). Les théories diffèrent sur leurs hypothèses concernant la façon dont ces composants sont intégrés, et en particulier, de comment différencier les différents états émotionnels à partir de leurs schémas expressifs (Scherer et Peper, 2001). Il n'existe pas de consensus sur la définition des émotions, sur leur nature exacte, ou sur les processus responsables de nos réactions émotionnelles.

A l'instar de Scherer (2005) nous considérerons dans ce manuscrit que l'émotion est un épisode de changements synchronisés des cinq sous-systèmes de notre organisme (Composants cognitifs, composants neurophysiologiques, tendance à l'action, expression motrice, et sentiment subjectif) en réponse à l'évaluation d'un évènement interne ou externe, considéré comme pertinent pour les besoins et les objectifs majeurs de l'organisme.

Notre travail se focalise uniquement sur deux de ces composants. D'une part, nous considérons le composant cognitif, que nous tenterons de modéliser informatiquement. D'autre part, nous considérons l'expression motrice et plus particulièrement l'une de ses manifestations : l'expression faciale.

2.1.2 Classification des différentes théories de l'émotion

Scherer (2010) propose une classification des modèles émotionnels selon deux critères. D'une part, il considère les différents composants émotionnels impliqués, et d'autre part il considère les différentes phases du processus émotionnel considérées par les différentes théories.

Ces différentes approches ne sont pas exclusives, certaines théories peuvent donc relever de plusieurs d'entre elles. La Figure 7 (extraire de Scherer 2010) propose une visualisation des différentes approches en fonction des deux axes proposés par Scherer.

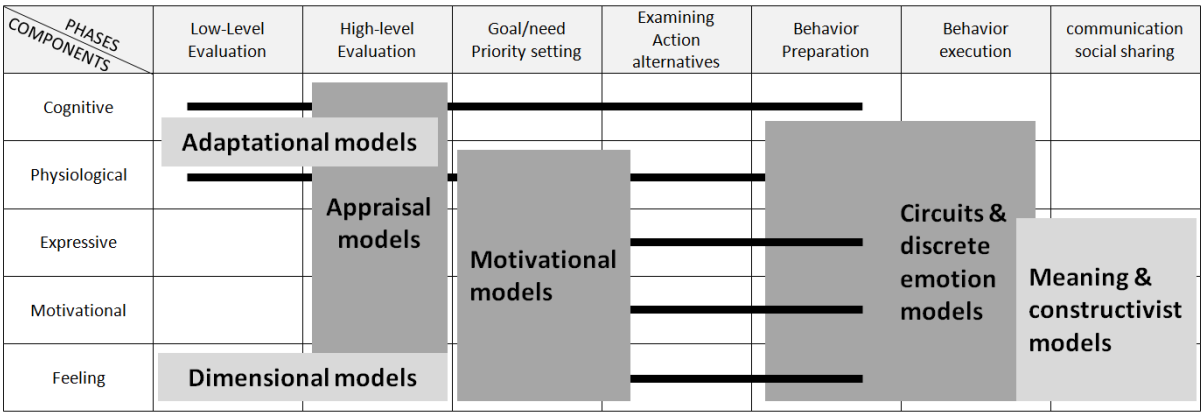


Figure 7 - Les approches des émotions classées selon les dimensions composants émotionnels (lignes) et phases de l'évaluation cognitive (colonnes) (Scherer 2010).

Les différentes approches identifiées par Scherer sont :

1) Approches adaptatives

Scherer regroupe dans cette approche l'ensemble des théories qui considère l'émotion comme une fonction adaptative développée lors de l'évolution comme une capacité à détecter les événements significatifs pour la survie et le bien-être de l'organisme.

2) Approches dimensionnelles

Ces théories considèrent que les émotions sont réparties dans un espace à plusieurs dimensions, telles que la valence (plaisant, déplaisant) ou le niveau d'activation ou d'éveil (actif, passif).

3) Approches cognitivistes

Les théories de l'appraisal supposent que la plupart des émotions sont générées par un processus cognitif multi-componentiel dédié. Une situation est évaluée selon un certain nombre de critères cognitifs. C'est cette évaluation qui est à l'origine des réactions physiologiques et motrices ainsi que du sentiment subjectif.

4) Approches motivationnelles / Tendance à l'action

Cette approche s'inscrit également dans une vision évolutionnaire des émotions. Elles considèrent plus précisément le lien entre les émotions et les motivations de l'individu. Les émotions sont considérées comme un facteur motivationnel de l'individu. Par exemple, la Peur est une émotion motivant une fuite, alors que la Colère est une émotion motivant une préparation au combat.

5) Approche des circuits mentaux

Les théories de cette approche sont influencées par la neuroscience. Elles considèrent qu'il existe un nombre fini d'émotions, causées et différenciées par des circuits neuronaux génétiquement pré-codés.

6) Approches des émotions de bases

Ces théories sont similaires aux théories des circuits mentaux. Elles supposent l'existence d'un certain nombre d'émotions de base, prédéfinies génétiquement. Ces émotions sont différenciées par des programmes neuronaux distincts qui génèrent les réactions physiologiques et expressives prototypiques associées, et en particulier les expressions faciales.

7) Approches lexicales

Les approches lexicales reposent sur la diversité lexicale émotionnelle. En effet, dans la plupart des langues, le vocabulaire lié aux émotions est relativement large. Ainsi, ces théories utilisent une analyse sémantique et lexicale de ce vocabulaire pour déterminer l'organisation et la structure du lexique émotionnel.

8) Approches par constructions sociales

Les théories de constructions sociales considèrent que l'émotion est un produit social, dont les manifestations sont construites selon les règles sociales. Les manifestations physio-biologiques ne sont que secondaires par rapport à la signification de l'émotion dans son contexte social. L'émotion a un rôle communicatif, et le sentiment subjectif associé à l'émotion joue un rôle central.

Gross et Feldman-Barrett (2011) proposent une autre classification des différentes théories de l'émotion. Selon ces auteurs, quatre approches des émotions peuvent être organisées sur un continuum (Figure 8). Ils proposent ainsi une échelle monodimensionnelle sur laquelle ils situent les principales théories des émotions et leurs auteurs. Tout comme la classification de Scherer, les différentes approches se recoupent, et certaines théories relèvent de deux de ces approches simultanément.

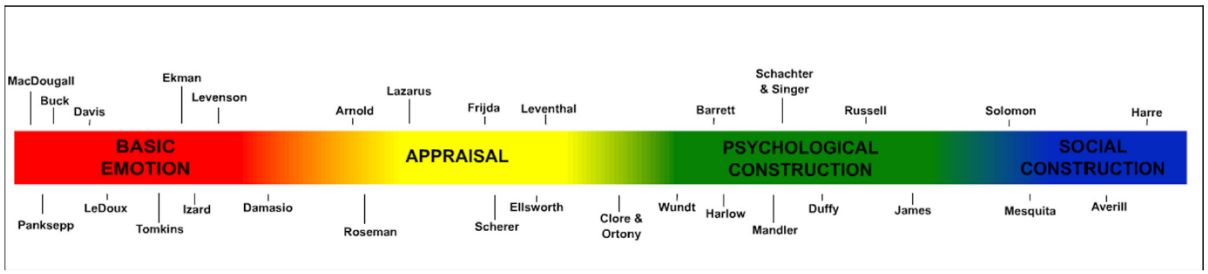


Figure 1. Perspectives on emotion can be loosely arranged along a continuum. We have populated this continuum with representative theorists/researchers drawn from the field of psychology. We distinguish four "zones": (1) basic emotion, in red, e.g., MacDougall (1908/1921), Panksepp (1998), Buck (1999), Davis (1992), LeDoux (2000), Tomkins (1962, 1963), Ekman (1972), Izard (1993), Levenson (1994), and Damasio (1999); (2) appraisal, in yellow, e.g., Arnold (1960a, 1960b), Roseman (1991), Lazarus (1991), Frijda (1986), Scherer (1984), Smith and Ellsworth (1985), Leventhal (1984), and Clore and Ortony (2008); (3) psychological construction, in green, e.g., Wundt (1897/1998), Barrett (2009), Harlow and Stagner (1933), Mandler (1975), Schachter and Singer (1962), Duffy (1941); Russell (2003), and James (1884); (4) social construction, in blue, e.g., Solomon (2003), Mesquita (2010), Averill (1980), and Harré (1986). Given space constraints, as well as the goals of this article, we have limited ourselves to a subset of the many theorists/researchers who might have been included on this continuum (e.g., those who only study one aspect of emotion were not included in this figure).

Figure 8 - Continuum des théories des émotions (Gross et Feldman-Barrett 2011)

Les quatre approches distinguées par Gross et Feldman-Barrett (2011) sont les suivantes :

1) Emotion de base

Ces approches considèrent une liste finie d'émotions universelles ayant chacune un mécanisme mental unique, responsable de l'apparition de l'émotion, et générant des réponses physiologiques et expressives caractéristiques.

2) Appraisals

Dans les théories de l'appraisal, les émotions sont toujours des états mentaux uniques dans leur forme et leurs fonctions, mais il n'existe plus de mécanisme mental dédié à chaque émotion. Certains modèles d'évaluation (Lazarus et Folkman, 1984, Arnold, 1960) considèrent les appraisals comme des causes cognitives de l'émotion, qui donnent du sens aux événements et à leurs contextes. Dans ces modèles, l'évaluation cognitive déclenche des réactions émotionnelles biologiques caractérisées par les motifs expressifs et physiologiques stéréotypés.

3) Construction Psychologique

La troisième approche des émotions, identifiée par Gross et Feldman-Barrett, considère les émotions comme une construction psychologique. Les émotions ne sont plus considérées comme des états mentaux spéciaux, avec une forme unique, une fonction et une cause. Selon cette approche, les émotions ne possèdent pas de mécanismes dédiés. Les états mentaux émotionnels émergent d'un processus cognitif constructif en permanence modifié. Ce processus se compose d'un ensemble de composants cognitifs basiques, mais non spécifique aux émotions.

4) Construction Sociale

La dernière approche identifiée par Gross et Feldman-Barrett considère l'émotion comme une construction sociale. Les émotions sont considérées comme des phénomènes sociaux, ou des réactions culturellement prescrites. Elles sont constituées par des facteurs socioculturels, et contraintes par les rôles et statuts des participants ainsi que par le contexte social. Certains modèles de construction sociale traitent la situation sociale comme le déclencheur des réactions émotionnelles, de la même manière que certaines théories des appraisals considèrent les appraisals comme déclencheurs des réactions émotionnelles. Toutefois, d'autres modèles considèrent les émotions comme des produits socioculturels prescrits et construits par des gens, plutôt que par la nature. Les émotions sont vues comme les composantes de la culture plutôt que comme des états internes humains. Dans la mesure où les processus cognitifs impliqués portent des aspects culturels, les émotions sont considérées comme apprises plutôt que données par la nature. Les émotions varient donc d'une culture à l'autre.

Nos travaux portent sur quatre approches différentes des émotions (catégorielle, dimensionnelle, cognitive, et sociale) que nous avons jugées pertinentes pour l'animation faciale et l'interaction temps-réel. Nous allons donc présenter chacune de ces quatre approches en détail, et pour chacune d'entre elles, mettre en lumière les différentes utilisations qui en ont été proposées en informatique.

2.1.3 L'approche catégorielle des émotions en psychologie

Selon l'approche catégorielle des émotions, il existe un ensemble d'émotions de base, par exemple la joie, la surprise, la peur, la colère, la tristesse et le dégoût (Ekman et Friesen, 1986). Selon Ekman, plusieurs caractéristiques distinguent les émotions de base les unes des autres, ainsi que des autres phénomènes affectifs (Ekman, 1999). Ces émotions de base sont censées être déclenchées sur les conditions spécifiques par des programmes internes universels. Selon Ekman (2003) chaque émotion se caractérise ainsi par un « circuit » spécifique. Ces programmes neuromoteurs sont exécutés automatiquement et sont relativement résistants aux changements. Ils peuvent cependant être modifiés par la culture ou l'apprentissage. Ekman et Friesen (1975) ont établi des listes d'indices visibles distincts et des familles d'expressions faciales de surprise, de peur, de dégoût, de colère, de joie et de tristesse. Selon les théories catégorielles, une émotion de base n'est pas considérée comme un seul état affectif, mais plutôt comme une famille d'états liés. Divers chercheurs (Tomkins, 1984, Izard, 1977, Plutchik, 1980, Ekman et Friesen, 1975) considèrent différentes listes des émotions fondamentales, incluant par exemple, la Honte, la Culpabilité, la Fierté (Izard, 1977).

En plus des émotions de base, l'approche catégorielle couvre aussi d'autres états mentaux (Baron-Cohen, 2007), dont des émotions complexes. Selon Damasio (1994), les émotions de base seraient innées, et les émotions complexes se construiraient durant le développement de l'individu. Cependant, dans un contexte réel, il est rare d'observer des émotions basiques seules (Scherer et Ceschi, 1997, Abrilian et al., 2005). Souvent, l'état émotionnel d'une personne est un mélange de plusieurs états mentaux.

Selon l'approche catégorielle, l'intensité se définit généralement de deux façons. La première approche consiste à associer une échelle d'intensité aux labels émotionnels. On parlera donc de joie intense, de légère surprise, etc. La seconde approche consiste à définir plusieurs labels émotionnels pour chaque famille d'émotion. Ainsi, le chagrin sera une variante moins intense de la tristesse. La Rage sera une variante plus intense de la Colère, etc (Ekman 2003).

2.1.4 L'approche catégorielle des émotions en informatique

L'approche catégorielle des émotions est généralement utilisée de manière implicite en informatique. Cette approche permet la représentation directe des émotions comme des états indépendants. En effet, l'approche catégorielle ne permet pas de structurer les émotions entre elles. Ainsi, les modèles informatiques utilisent généralement un ensemble fixé (et généralement réduit) de catégories d'émotions, sur lesquelles s'effectuent des raisonnements pour chaque émotion sans tenir compte des autres. Par exemple, les logiciels de détection d'émotion se concentrent souvent sur les signes spécifiques aux six émotions de base (FaceReader, Noldus) ou parfois sur un ensemble plus restreint (Busso et al. 2004).

En ce qui concerne la génération, l'approche catégorielle est souvent utilisée pour représenter les paramètres expressifs de l'agent virtuel (ex : exprimer un mélange de Surprise et de Joie). Comme nous le verrons dans la suite, les raisonnements de haut niveau sont souvent gérés par des modèles d'émotion plus complexes (Ruttkay et al., 2003, Becker-Asano et Wachsmuth, 2008, Gratch et Marsella, 2006, Rivière el al., 2011, etc.).

L'approche catégorielle laisse en suspens un certain nombre de problématiques concernant la dynamique des expressions. Ainsi, différents types de dynamique ont été proposés (Dariouch et al, 2004). Le modèle *onset-apex-onset* (ou *attack-sustain-decay*) propose une animation en trois étapes (Figure 9 gauche). Une phase d'augmentation d'intensité jusqu'à un apex, un maintien de l'apex, puis une décroissante progressive. Le modèle *attack-decay-sustain-release* ajoute en plus un pic expressif après l'*onset* du premier modèle (Figure 9 droite). Ces dynamiques peuvent être appliquées directement sur l'intensité des émotions catégorielles, ou bien à une échelle plus locale, sur les paramètres utilisés par la génération expressive (ex : l'expression faciale de l'agent virtuel).

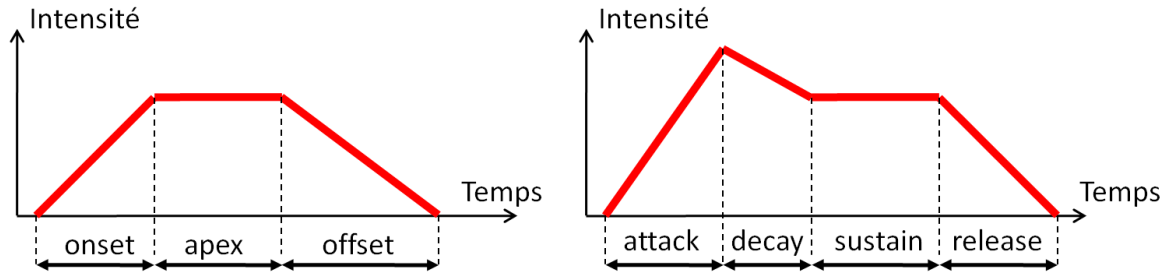


Figure 9 - Dynamiques des émotions applicables à l'approche catégorielle. A gauche le modèle onset-apex-onset, à droite, le modèle attack-decay-sustain-release

L'approche catégorielle permet donc de représenter différents états émotionnels chez un agent virtuel, ainsi que de les exprimer et de les mélanger. Différents aspects sont cependant sous spécifiés dans la littérature en psychologie relative aux émotions. Principalement, l'effet des mélanges d'émotions sur l'expressivité, et la dynamique émotionnelle. Pourtant, l'approche catégorielle semble être une approche pertinente pour l'animation d'agents virtuels interactifs. Le chapitre 3 présente donc les travaux que nous avons effectués en nous inspirant de cette approche.

2.1.5 L'approche dimensionnelle des émotions en psychologie

Contrairement aux approches catégorielles des émotions, les approches dimensionnelles considèrent que les émotions ne sont pas des états indépendants les uns des autres, mais qu'elles font partie d'un espace continu à plusieurs dimensions. Il est donc possible de définir des relations entre les émotions. Par exemple, dans un espace utilisant une dimension *Positive/Négative*, il est possible de définir que la Joie est plus positive que la Colère.

Plutchik (1980) propose un espace conique (ou circumplexe) en 3 dimensions pour représenter les émotions. La Figure 10 présente la représentation graphique de cet espace, ainsi qu'une version « à plat » de sa surface. Dans cet espace, la dimension verticale représente l'intensité. Ainsi, plus on se déplace vers le haut du circomplexe, plus on trouve des émotions intenses, telles que la Rage, la Terreur, ou l'Extase.

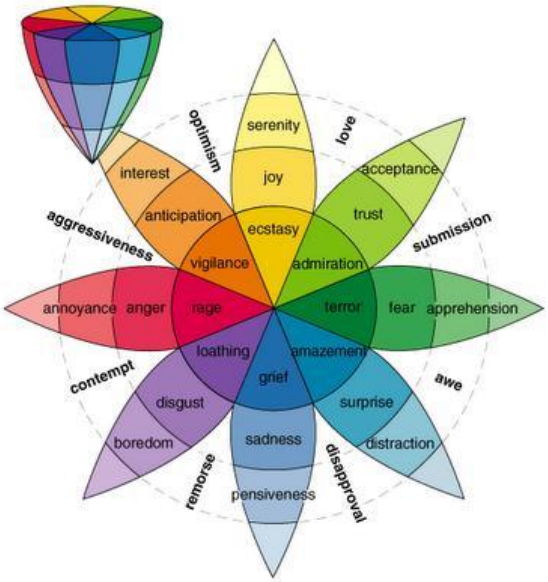


Figure 10 - Le Circomplexe de Plutchik

Russell et Mehrabian (1977) proposent un espace selon 3 dimensions continues : Plaisir, Activation, Dominance (*Pleasure, Arousal, Dominance*) pour représenter le tempérament et les émotions. Dans cet espace, chaque dimension varie entre -1,0 et 1,0. En utilisant une approche lexicale, Russell et Mehrabian ont mené des études sur plusieurs centaines de personnes en leur demandant de situer un certain nombre de labels émotionnels à l'intérieur de cet espace 3D. Ils suggèrent ainsi de manière empirique l'emplacement de plusieurs états affectifs au sein de cet espace en 3 dimensions. Contrairement au *Circomplex* de Plutchik, l'intensité n'est pas représentée dans cet espace 3D. Il est effectivement nécessaire de distinguer la dimension Activation de l'intensité. Par exemple, une tristesse intense est très passive (faible Activation), et a une intensité forte.

L'intensité émotionnelle est considérée de manière variable selon les différentes théories dimensionnelles de l'émotion (Reisenzein, 1996). Un certain nombre de ces théories ne spécifient pas comment l'intensité interagit avec les dimensions (Osgood, 1966, Russell et Mehrabian, 1977, Smith et Ellworth, 1985). Certaines théories incluent l'intensité émotionnelle comme une dimension supplémentaire et indépendante (Gerrig et al., 1991). Dans d'autres théories (par exemple, Russell 1989), l'intensité est considérée comme proportionnelle à la norme du vecteur de l'émotion dans l'espace dimensionnel. Par exemple, une émotion très Active, ou très Négative, sera considérée comme plus intense qu'une émotion plus neutre (plus proche du point de neutralité de l'espace dimensionnel).

2.1.6 L'approche dimensionnelle des émotions en informatique

Ruttkay et al. (2003) proposent un modèle à deux dimensions dans lequel les six émotions de base (selon Ekman) sont distribuées sur un disque (Figure 11). Le centre du disque représente l'état émotionnel neutre. Chaque point du disque représente un mélange des expressions faciales des deux plus proches émotions de base, la distance au centre du disque étant l'intensité de l'état émotionnel.

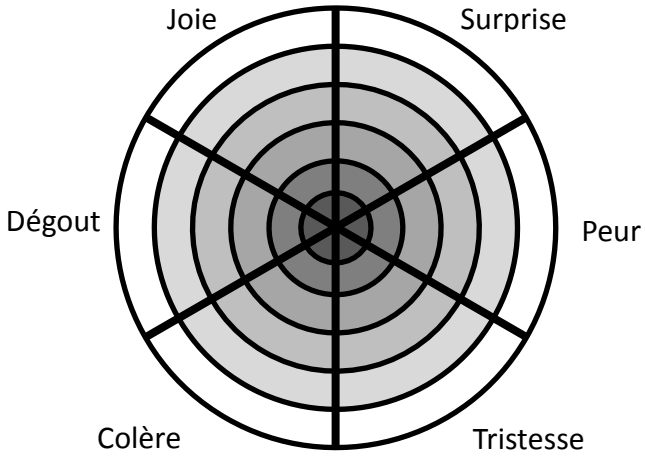


Figure 11 - Le disque des émotions proposé par Ruttkay et al. (2003)

Bien que ce modèle puisse être rapproché du modèle théorique de Plutchik (1980) dans sa forme, il n'en respecte pas l'organisation (Le circomplex de Plutchik est constitué de huit quadrants, contre 6 pour le modèle de Ruttkay), et il se focalise uniquement sur les six émotions de base.

Ce modèle a été utilisé pour animer un visage cartoon. L'utilisateur déplace la souris à l'intérieur du disque (Figure 11), et le visage adopte l'expression de l'émotion sélectionnée. La progression de la souris permet ainsi de définir une dynamique faciale dans un espace continu, et de créer des mélanges entre deux émotions adjacentes.

Fonctionnant sur un principe d'interaction similaire, l'outil d'annotation Feeltrace (Cowie et al., 2000) utilise une approche dimensionnelle à deux dimensions (Pleasure / Arousal) comme espace d'annotation d'émotion en temps réel. L'annotateur peut déplacer également son curseur de manière continue dans l'espace à deux dimensions. Il peut ainsi annoter sa perception dynamique d'une vidéo ou d'un extrait musical. Ainsi, la dynamique émotionnelle perçue par l'annotateur est enregistrée. La Figure 12 montre l'interface du logiciel Feeltrace.

Dans ce logiciel, l'intensité émotionnelle est considérée comme étant la norme du vecteur 2D formé entre le centre du repère dimensionnel, et le curseur.

L'outil Feeltrace a été évalué pour l'annotation de l'état émotionnel de deux personnes, à partir de l'enregistrement audio de leur conversation. Il a également été évalué pour l'annotation des émotions perçues dans un extrait de musique. Si l'outil semble montrer une capacité d'annotation intéressante, il est cependant limité, par exemple, par son incapacité à discriminer la Peur de la Colère (Négative et Active) dans sa version à deux dimensions. Le point fort de l'outil Feeltrace étant sa capacité à annoter précisément la dynamique émotionnelle et sa continuité.

Ce type d'interface pourrait également être utilisé pour contrôler un agent virtuel. En utilisant un paradigme d'interaction proposé par Ruttkay et al. (2003) mais en utilisant un espace à deux dimension (comme Feeltrace), voire trois dimensions (par exemple, P.A.D, Russell et Mehrabian, 1977).

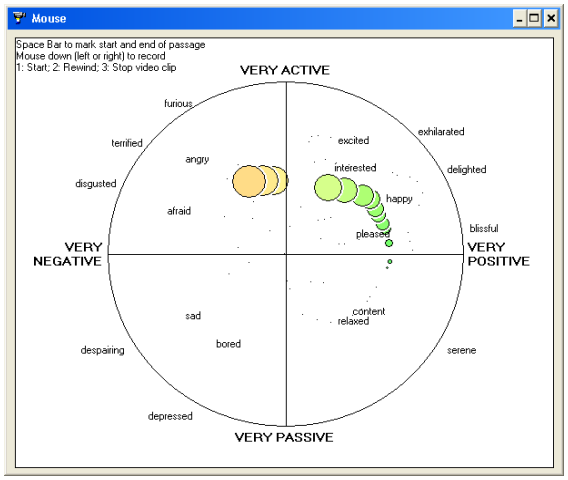


Figure 12 - Interface du logiciel Feeltrace (Cowie et al., 2000)

Une autre application des modèles dimensionnels a été proposée par Becker-Asano et Wachsmuth (2008) dans la conception du modèle émotionnel de l'architecture WASABI. WASABI est une architecture dédiée à la simulation émotionnelle. Le modèle contient une représentation des émotions, des capacités de raisonnements cognitifs, et une représentation visuelle sous forme d'agent virtuel (agent MAX). L'état émotionnel courant de ce système est représenté comme une localisation dans l'espace 3D P.A.D. Comme nous le verrons plus loin, cet état émotionnel est ensuite catégorisé pour être appliqué à l'agent expressif (Becker-Asano et Wachsmuth, 2008).

L'approche dimensionnelle apporte donc un certain nombre d'avantages par rapport à l'approche catégorielle des émotions. D'une part, les émotions font partie d'un espace continu qui les met en relation les unes avec les autres. Ainsi les émotions ne sont plus des états indépendants. En ce qui concerne les modèles informatiques, cela permet d'appliquer un certain nombre de règles supplémentaires dans les calculs d'état émotionnel. Ces règles ont donc un impact sur l'animation faciale qui en découle. Ainsi, nous pensons qu'il est pertinent d'explorer l'approche dimensionnelle des émotions pour l'animation faciale temps réel. Notre chapitre 4 présente donc les travaux que nous avons menés en nous inspirant de cette approche.

2.1.7 Approches cognitives des émotions en psychologie

Selon les théories cognitives multi-componentielles (ou théories de l'*appraisal*), les émotions résultent d'un processus d'évaluation de la situation et son rapport à l'individu (Arnold, 1960, Frijda, 1986, Scherer, 1984, Ellsworth et Scherer, 2003).

Les théories cognitives postulent que 1) la plupart des émotions (mais pas toutes) sont provoquées par une évaluation cognitive (mais pas nécessairement consciente ou contrôlée) de la situation et des événements antécédents, et que 2) les différentes réactions motrices sont provoquées par le résultat de ce processus d'évaluation. Les théories des processus multi-componentielles supposent donc qu'il y a autant d'états différents émotionnels que de résultats différents possibles de l'évaluation cognitive.

Depuis les années soixante, les théories de l'appraisal connaissent un intérêt croissant grâce aux travaux précurseurs de Arnold et Lazarus. L'un des objectifs de ces théoriciens est d'expliquer pourquoi différentes personnes réagissent de manières différentes à des évènements similaires. En effet, contrairement à ce que la théorie des émotions de base soutient, les gens réagissent de façons partiellement différentes à une même

situation, et ce en fonction de leur perception/évaluation de cette situation. Les différences qui apparaissent dans la manière de percevoir une situation provoquent des émotions diverses qui sont spécifiques à chaque personne.

Les travaux sur les théories de l'appraisal trouvent leurs origines dans les années 1950. Arnold concentre ses recherches sur le processus cognitif à l'origine des émotions. Elle développe sa "théorie cognitive", qui précise que la première étape d'une émotion est la phase d'évaluation de la situation. Selon Arnold, l'évaluation cognitive est à l'origine de la séquence émotionnelle et génère à la fois les actions appropriées et l'expérience émotionnelle en elle-même.

Lazarus (1984) identifie deux catégories d'évaluations cognitives 1) l'évaluation primaire, qui détermine le sens et les implications de l'événement pour l'organisme, et 2) l'évaluation secondaire, visant à l'évaluation de la capacité de l'organisme à faire face aux conséquences de l'événement. Ces deux catégories d'appraisal sont complémentaires et séquentielles. Lazarus distingue ensuite deux types de réactions à l'évènement : 1) les actions directes, visant à modifier les conséquences de l'évènement, et 2) des processus de réévaluation cognitive, visant à s'adapter à l'évènement sans le modifier.

Suite à ces travaux précurseurs, plusieurs théories de l'appraisal ont été proposées. Toutes sont composées d'un certain nombre de critères d'évaluation cognitive. Cet ensemble de critères varie selon les théories.

Frijda propose un certain nombre de critères d'évaluation, appelée *lois* (Frijda, 1988), pour décomposer le processus d'évaluation cognitive en règles unitaires. Frijda soutient que l'existence de ces lois est difficile à démontrer avec précision, car leur mesure repose sur une verbalisation de l'évaluation cognitive faite par le sujet. En effet, les sujets sont souvent questionnés sur leur expérience émotionnelle a posteriori, ce qui introduit une altération subjective de leur évaluation cognitive. Cependant, Frijda (1988) soutient une forte corrélation entre les mesures subjectives effectuées et les résultats prédits par ses *lois*. Selon lui, cela suggère l'existence des mécanismes cognitifs qu'il défend.

La théorie cognitive proposée par Scherer (1984, 2001) fait valoir que les émotions sont générées par des cycles d'évaluations multi-componentielles des événements. Scherer propose une description plus fine que celle proposée par Lazarus, en détaillant plus précisément les étapes du processus cognitif. Le processus cognitif n'est plus séparé en deux catégories (primaire/secondaire), mais en une liste de critères d'évaluations. Ces critères d'évaluations sont appelés *Stimulus Evaluation Checks (SECs).* A l'instar de Scherer (Scherer et Sangsue, 2004), nous utiliserons l'appellation « *checks* » dans la suite de ce document pour faire référence à ces critères d'évaluation. Scherer regroupe ces checks en quatre groupes (Scherer et Sangsue, 2004). Ces quatre aspects peuvent s'illustrer par les questions suivantes, concernant l'évaluation d'un événement émotionnel :

1) ***La pertinence***
   Cet évènement est-il nouveau ou pertinent pour moi ? Est-ce qu'il m'affecte directement (ou mon groupe social) ?

2) ***Les rapports aux buts***
   Quelles sont les implications ou les conséquences de cet évènement et à quel point vont-elles affecter mon bien-être, ou mes buts à court et long terme ?

3) ***Le potentiel de maitrise***
   A quel point suis-je capable de faire face à ces conséquences, en les modifiant ou en m'adaptant ?

4) ***L'accord avec les standards***
   Comment cet évènement se situe par rapport à mes convictions personnelles ainsi que face aux normes et valeurs sociales ?

Les groupes 1 et 2 correspondent aux appraisals primaires de Lazarus, le groupe 3 correspond aux appraisals secondaires de Lazarus, et le groupe 4 ne trouve pas d'équivalence directe.

Chacun de ces groupes est composé de plusieurs checks. Scherer a, au cours de ses travaux, défini plusieurs listes de checks, et ses études se concentrent le plus souvent sur 5 à 7 d'entre eux (Scherer, 1999). Le Tableau 1 donne une définition de certains checks fréquemment utilisés.

Check	Groupe	Définition (Adaptée de Scherer, 1999)
Nouveauté	1	Avais-je anticipé que la situation allait se produire?
Agrément intrinsèque	1	Ai-je trouvé l'évènement plaisant ou déplaisant?
Rapport aux buts	2	Cet évènement m'a-t-il empêché d'atteindre mes buts ou de suivre mes plans ?
Causalité externe	2	Une personne tierce était-elle à l'origine de l'évènement ?
Potentiel de maitrise	3	Comment ai-je évalué ma capacité à modifier ou à m'adapter aux conséquences de cet évènement ?
Standards externes	4	Si l'évènement a été causé par mon comportement ou celui d'un autre, ai-je jugé ce comportement comme non adapté ou immoral ?
Standards Internes	4	De quelle façon cet évènement a-t-il affecté ma vision de moi-même, par exemple mon estime de moi ou ma confiance en moi ?

Tableau 1 - Huit critères d'évaluation issus de la théorie de Scherer

L'ensemble de ces checks font partie du *Componential Process Model* (Scherer, 2001). A l'instar des premiers travaux sur les théories de l'appraisal, ce modèle tente d'expliquer la différenciation des états émotionnels comme le résultat d'une séquence d'évaluation, mais il fait également des prédictions concernant les réactions physiologiques et expressives qui en découlent.

Les différents checks peuvent avoir des effets multiples et simultanés sur l'expression faciale, la posture du corps, la voix et le système nerveux (Scherer, 2001). La temporalité et l'ordre de ces checks sont fondamentaux. Par exemple, une étude a constaté que l'agrément intrinsèque est évalué avant le rapport aux buts (Lanctôt et Hess, 2007). Cette étude a également mesuré les délais de réaction des muscles faciaux pour ces deux checks (environ 400ms pour l'agrément et 800ms pour le rapport aux buts) par rapport au début du stimulus. Une autre étude a suggéré que certains checks (nouveauté, agrément, et rapport aux buts) se produisent dans le cerveau en 250ms (Grandjean et Scherer, 2008). Les checks de plus haut niveau, tels que le potentiel de maitrise et les standards, peuvent induire des délais plus longs, mais peu de données sont encore disponibles dans la littérature à ce sujet.

Des travaux en neuropsychologie suggèrent également que certaines parties du cerveau seraient dédiées à certains critères d'évaluation. Par exemple, Sander et al. (2003) soutiennent que l'amygdale, longtemps considérée comme le siège de la Peur et d'autres émotions négatives, serait en fait un circuit neuronal dédié à la détection de la pertinence des stimuli.

Les différentes théories cognitives proposent plusieurs segmentations du processus cognitif émotionnel, mais il est possible de mettre en relation les différents ensembles de critères d'évaluation. Le Tableau 2 (Ellsworth et Scherer, 2004) propose une mise en correspondance des critères proposés par plusieurs théories cognitives.

Tableau 2 - Comparatif des théories de l'évaluation cognitive (extrait d'Ellsworth et Scherer 2004)

	Frijda (1986)	Roseman (1984)	Scherer (1984)	Smith et Ellsworth (1985)
Novelty	Changes Familiarity		Novelty suddenness familiarity	Attentionnal Activity
Valence	Valence Focality	Appetitive / Aversive motives	Intrinsic pleasantness	Pleasantness Importance
Goals/Needs	Certainty	Certainty	Goal Significance Concern relevance Outcome probability	Certainty
Agency	Intent / Self-other	Agency	Cause: agent Cause: motive	Human Agency
Norms/values	Value Relevance		Compatibility with standards External Internal	Legitimacy

Contrairement aux autres théories cognitives, le modèle cognitif OCC (Ortony, Clore, et Collins, 1988) distingue différents types de processus cognitifs, sélectionnés selon la nature du stimulus évalué. On trouve 1) le processus lié aux conséquences d'un événement, 2) le processus lié à l'action de soi ou d'un tiers, et 3) le processus lié à l'aspect d'un objet. La Figure 13 donne la structure globale des différentes évaluations cognitives selon la théorie OCC.

Certains rapprochements peuvent être effectués entre la théorie OCC et les théories cognitives telles que le modèle CPM. Par exemple, l'évaluation « *Consequence for other/self* » du modèle OCC semble se référer au même concept que le check « pertinence par rapport aux buts » du modèle CPM. Néanmoins, le fait de considérer plusieurs processus cognitifs en fonction de la nature du stimulus distingue le modèle OCC des autres théories cognitives, dans lesquelles le processus cognitif est automatique, unique, et inconscient, quel que soit le type de stimulus.

Dans le modèle OCC, l'intensité des émotions est définie par un ensemble de variables associées à chaque émotion. Par exemple, l'intensité de déception (Figure 13 : *disappointment*) sera proportionnelle à l'évaluation des conséquences positives que l'événement aurait pu avoir. Dans la plupart des autres théories cognitives, l'intensité émotionnelle n'est pas clairement explicitée, mais est implicitement reliée à l'intensité des différents checks. Par exemple, plus un événement sera obstructif pour les buts de l'individu, ou plus il manquera de ressources pour faire face aux conséquences, plus le sentiment de tristesse sera intense.

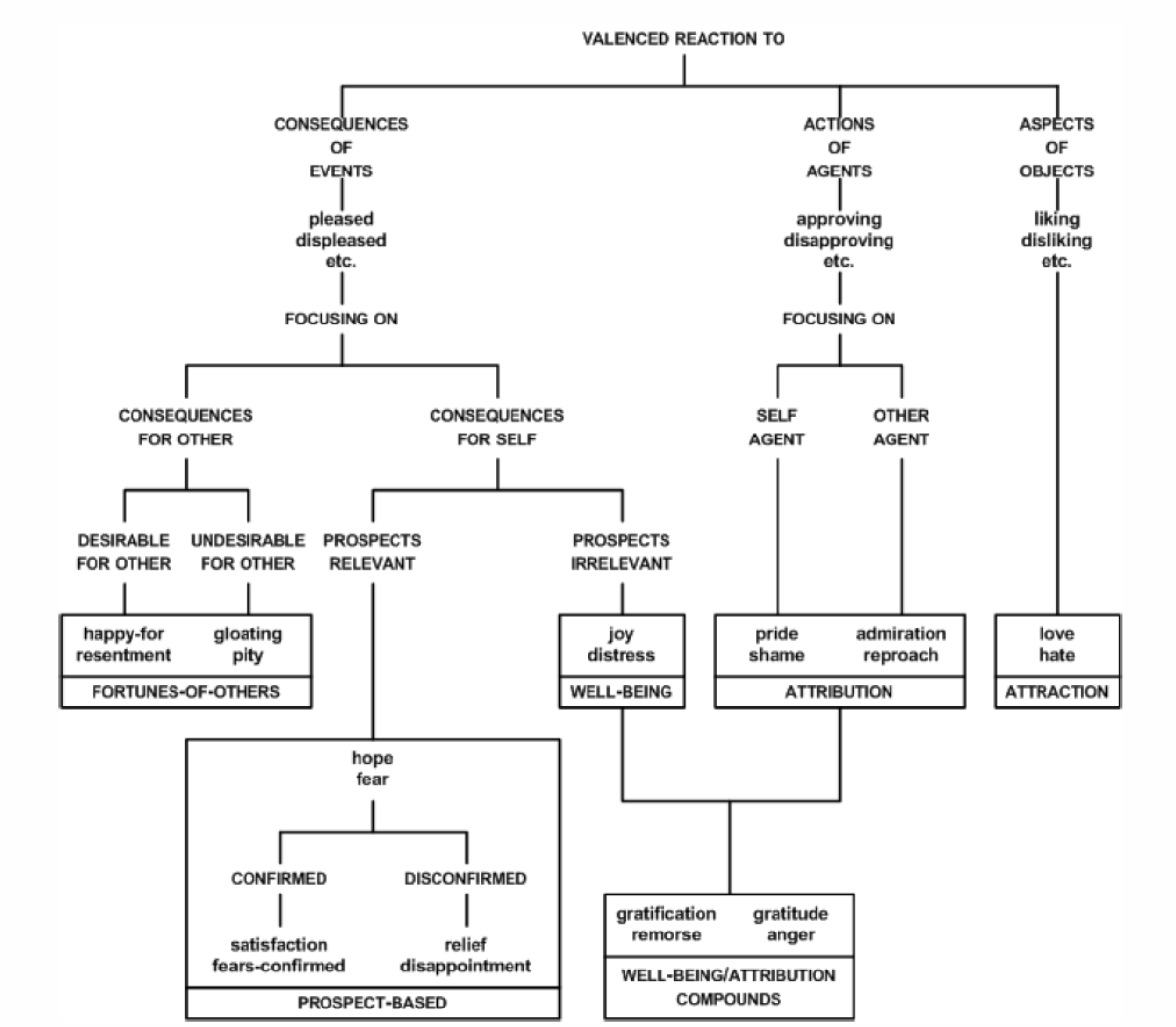


Figure 13 - Structure globale du processus émotionnel selon la théorie OCC (extraite de Ortony, Clore, Collins, 1988)

2.1.8 Approches cognitives des émotions en informatique

La diversité des définitions et des approches cognitives des émotions en psychologie a mené au développement d'une grande variété de modèles informatiques des émotions. Toutes les théories cognitives comprennent plusieurs composantes et plusieurs étapes dans le processus émotionnel. Ce fonctionnement par étapes fait des théories cognitives des candidates idéales pour la modélisation informatique car ces étapes sont très adaptées aux contraintes des modèles computationnels.

Plusieurs implémentations ont été proposées en se basant sur ces différentes théories. L'un des premiers modèles informatiques a été créé en se basant sur les *lois* de Frijda. Frijda et Swagerman (1987) proposent l'implémentation ACRES, capable, étant donné une description de l'événement soumis à l'évaluation cognitive, de donner une liste des émotions résultantes plausibles. Ce système montre de bonnes performances, de l'ordre de 31% de réponses identiques aux réponses des sujets humain, et 71% de réponses cohérentes (dans les cinq premiers labels donnés par les sujets).

D'autres modèles computationnels ont depuis été proposés. Marsella (2010) propose une classification des modèles existants en fonction des différents modèles psychologiques sous-jacents. La Figure 14 résume cette classification. Les blocs gris (à gauche sur le schéma) représentent les théories de la psychologie, et les flèches

représentent l'influence des modèles les uns sur les autres. Certains modèles informatiques sont inspirés à la fois par des modèles issus de la psychologie, et par d'autres modèles informatiques.

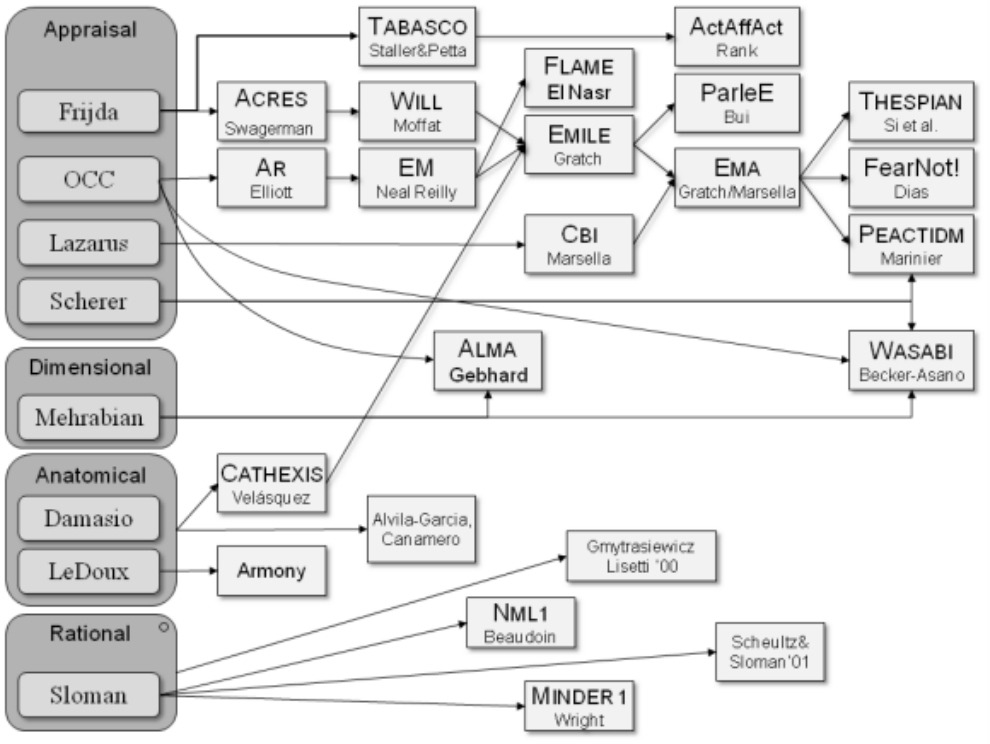


Figure 14 - Classification de certains modèles informatiques issus de l'approche cognitive selon Marsella (2010)

Le *Componential Process Model (CPM)* (Scherer, 2001) (*Scherer* dans la Figure 14) suppose l'existence d'autant d'états émotionnels possibles que de résultats possibles de l'évaluation cognitive de la situation. Sander et al. (2005) ont proposé un modèle informatique basé sur les réseaux de neurones, et ayant pour but de prédire un label émotionnel en fonction des checks du modèle CPM mis en entrée. Avec ce système, ils ont pu valider la cohérence de la théorie du CPM et des prédictions émotionnelles associées. De plus, ils ont pu étudier le comportement du système en faisant varier des paramètres tels que l'ordre, ou la simultanéité des entrées du système. En revanche, ce système ne traite pas le problème de comment simuler, à partir d'un événement, le traitement cognitif associé.

Pour simuler ce traitement, le modèle OCC a été largement exploité dans les recherches en informatique. La plupart des modèles computationnels récents sont en effet fondés sur cette théorie (Becker-Asano et Wachsmuth, 2008). En effet, sa structure en arbre de décision facilite son implémentation informatique. Par exemple, l'agent pédagogique Pat, proposé par Jaques et al. (2009), utilise le modèle OCC pour estimer l'état émotionnel d'un étudiant à partir de ses actions dans une application d'e-learning. Pat est ainsi capable de détecter plusieurs émotions (la Joie, la Détresse, la Satisfaction, la Déception, la Gratitude, la Colère, et la Honte) en simulant le processus cognitif de l'étudiant au fur et à mesure de ses action). A partir des inférences sur l'état émotionnel de l'étudiant, l'agent virtuel présente des comportements ciblés, visant notamment à augmenter la motivation et la confiance en soi de l'étudiant.

Néanmoins, le modèle OCC impose des limites. En effet, il ne considère que peu de critères de l'évaluation cognitive, par exemple, le check « capacité d'adaptation » n'est pas pris en compte (Ochs et al., 2008). De plus, il ne considère que vingt émotions (Voir Figure 13) et limite leur contexte. Pour finir, certaines des émotions de base ne sont pas considérées. Si le dégout peut être considéré comme intégré dans le critère « disliking » de l'évaluation « Aspect d'un objet », la Surprise est en revanche absente du modèle OCC. Pourtant, le modèle

OCC est souvent considéré comme trop complexe pour être intégré dans les personnages virtuels, et est souvent simplifié (Bartneck, 2002).

Le système EMA (Marsella et Gratch, 2006), également basé sur des théories d'évaluation cognitive, tente de modéliser les réactions émotionnelles consécutives à un événement. Ce modèle distingue les réactions rapides de type appraisal, et les réactions à plus long terme, basées sur les inférences, c'est-à-dire, un processus de réflexion plus approfondi, et plus long. Dans le modèle EMA, la différenciation des émotions est effectuée par un système inspiré du modèle OCC. Ce modèle est augmenté par un processus de calcul de stratégies d'adaptation, telles que définies par la théorie cognitive de Lazarus. Différentes stratégies peuvent être mises en place, telles que l'action, le déni, la recherche d'information, la recherche de moyen d'agir, ou la procrastination (attendre qu'un autre événement modifie la situation).

Le framework FeelMe (Broekens et DeGroot, 2004, 2008) intègre un modèle cognitif dans un cadre plus large, permettant de modéliser l'environnement de l'agent, de simuler l'évaluation cognitive, et de générer les comportements émotionnels associés. Ce modèle fonctionne sur un modèle de traitement du signal, où chaque dimension d'appraisal évolue comme un signal continu, et peut être modulée de plusieurs manières par des modulateurs externes. Ce modèle a été appliqué par les auteurs au jeu vidéo PacMan. L'état émotionnel du personnage virtuel (PacMan) est calculé dynamiquement en fonction des événements du jeu. Les dimensions d'appraisal utilisées dans cet exemple sont les dimensions *Pleasure Arousal et Dominance* (Russell et Mehrabian 1977) (En français, Valence, Activation, Dominance). Ces dimensions sont plus réductrices que les théories d'appraisal modernes.

L'une des difficultés des implémentations de modèles cognitifs est de représenter le contexte de l'évaluation sous forme computationnelle. Par exemple, évaluer si un événement est inattendu pour un personnage virtuel nécessite de savoir quels événements étaient prévisibles (de son point de vue). De même, évaluer si un événement va à l'encontre des buts d'un agent nécessite de connaitre ses objectifs. Pour cela, plusieurs approches ont été proposées. Les modèles BDI (Belief, Desire, Intention), modélisant les croyances, les désirs, et les intentions de l'agent, permettent de modéliser une partie de la théorie de l'esprit de Bratman (1987). Plusieurs implémentations informatiques ont été proposées, notamment dans le cadre d'utilisation d'agents virtuels expressifs (Ochs et al, 2005, Rivière et Pesty, 2010, Rivière et al., 2011). La plupart des modèles BDI classiques ne traitent cependant pas directement les émotions comme un paramètre de raisonnement. Certain modèles ont cependant été proposés pour tenir compte de « l'état émotionnel » courant du système sur le fonctionnement et les raisonnements effectués par le modèle DBI (Adam, 2007, Pereira et al., 2008a), créant ainsi un lien entre la simulation des émotions et la simulation des processus cognitifs.

L'approche cognitive tente donc de modéliser les processus cognitifs à l'origine des émotions. Ainsi, il semble pertinent d'explorer cette approche pour la gestion des émotions et de l'expressivité des agents virtuels. Nos travaux présentés dans le chapitre 5 proposent un modèle computationnel inspiré du modèle CPM (Scherer, 1984, 2001), ainsi que l'animation faciale associée.

2.1.9 Approches sociales des émotions en psychologie

Les causes et fonctions sociales des émotions ont été identifiées et reconnues dans plusieurs travaux en psychologie (Averill, 1985, Parkinson, 1996). Selon Rimé et al. (1992), les émotions sont généralement partagées socialement rapidement après leur apparition. Elles ont ainsi un rôle de communication sociale. En effet, Shaver et al. (1992) ont récolté et étudié plus de six cent descriptions d'épisodes émotionnels. Ils ont observé que trois quart de ces descriptions impliquent la présence d'une autre personne.

Parkinson (1996) défend l'idée que l'apparition des émotions est le plus souvent liée au contexte social. Plus précisément, il identifie les autres personnes comme l'une des causes principales des émotions. Selon l'auteur, si

les autres personnes sont l'une des cibles des expressions émotionnelles, elles en sont également l'une des principales sources. Le rôle de l'environnement social pour les émotions est donc bidirectionnel. Les émotions servent ainsi à maintenir nos relations sociales. L'empathie est un bon exemple de cette bidirectionnalité. Smith (1759) définit en effet l'empathie comme une réaction émotionnelle rapide et involontaire à l'expérience émotionnelle d'autrui. La personne pour qui on éprouve de l'empathie est donc à la fois la source et la cible de l'émotion.

L'une des autres facettes sociales des émotions est leur impact sur notre évaluation de la situation. Manstead et al. (2001) définissent le *social appraisal* (Evaluation Cognitive Sociale) comme une évaluation des pensées, sentiments et actions des autres personnes en réponse à un événement émotionnel. Ainsi, contrairement aux théories cognitives, les autres personnes ne sont plus uniquement partie intégrante du contexte. Leur processus d'évaluation cognitive est pris en considération. Manstead et al. (2001) distinguent deux rôles du social appraisal. Premièrement, son rôle dans l'expérience émotionnelle. L'évaluation d'un événement émotionnel faite par l'individu est influencée par les évaluations que les autres font du même événement. Deuxièmement, le Social Appraisal joue un rôle dans l'expression émotionnelle : la façon dont les gens expriment leurs émotions sera influencée par les implications sociales de ces expressions

Ces deux facettes du Social Appraisal renvoient au concept plus général de régulation des émotions. En effet, Gross (1998) définit la régulation des émotions comme « le processus par lequel un individu influence sa propre expérience émotionnelle, au moment même où elle survient, ainsi que sa manière d'exprimer ses émotions. ». Selon Gross, le processus de régulation émotionnelle serait à la fois automatique et contrôlé, conscient et inconscient, et pourrait influencer plusieurs étapes du processus de génération émotionnelle. Le Social Appraisal pourrait donc être interprété comme une forme de régulation émotionnelle liée aux relations sociales de l'individu avec les autres personnes en présence. Cependant, nous n'aborderons pas directement la problématique de la régulation émotionnelle dans les agents expressifs, car elle n'est pas directement liée à nos objectifs de recherche.

2.1.10 Approches sociales des émotions en informatique

Le phénomène social de l'empathie est celui qui a reçu la plus grande attention dans la communauté informatique. Plusieurs modèle ont été proposés (McQuiggan et Lester, 2006, Ochs et al., 2008, Leite et al. 2010) pour permettre à l'agent virtuel de montrer de l'empathie envers l'utilisateur. Cependant, ces modèles sont généralement limités par approximation de l'état émotionnel de l'utilisateur. Et effet, l'état émotionnel de l'utilisateur est souvent déduit en fonction de paramètres tels que l'historique de l'interaction et les inférences effectuées sur les actions de l'utilisateur. Pourtant, Ochs et al. (2008) ont mis en relief qu'un agent empathique était mieux perçu par l'utilisateur qu'un agent non empathique. De même, Leite et al. (2010) ont montré que leur robot interactif est plus perçu par les sujets comme un compagnon de jeu lorsqu'il présente un comportement empathique. Si l'empathie n'est que l'un des aspects sociaux liés aux émotions, ce résultat suggère néanmoins l'importance de la prise en compte du contexte social dans la simulation des émotions.

En ce qui concerne le Social Appraisal, aucun modèle informatique n'a, à notre connaissance, été proposé à ce jour. Pourtant, Mumenthaler et Sander (2009) ont montré que le concept de Social Appraisal est applicable à des agents virtuels expressifs. En utilisant deux visages virtuels, les auteurs ont étudié l'influence du « social appraisal » sur la catégorisation d'expressions faciales émotionnelles. Les deux visages sont disposés côte à côte, l'un au centre de l'écran et l'autre en périphérie. Le sujet doit reconnaitre la catégorie émotionnelle de l'expression de l'agent central. L'agent exprime de la Joie, de la Colère ou de la Peur. Simultanément, l'agent en périphérie exprime une expression faciale, également de Joie, de Colère ou de Peur. Pour finir, la direction du regard du personnage en périphérie est manipulée. L'étude montre que lorsque l'agent en périphérie regarde l'agent central et exprime une émotion, cela modifie la perception que l'utilisateur a de l'expression de l'agent central. En revanche, lorsque l'agent en périphérie ne regarde pas l'agent central, aucun effet n'est observé. Ces résultats montrent donc que le phénomène du social appraisal est applicable dans le cadre d'applications utilisant des agents virtuels.

Si plusieurs modèles informatiques ont été proposés pour la simulation de l'empathie, aucun modèle n'a été proposé pour la modélisation du social appraisal. Pourtant, le social appraisal semble relier les modèles cognitifs avec l'aspect social des émotions. Il semble donc qu'il soit pertinent d'explorer le social appraisal dans le cadre des agents virtuels expressifs. Dans notre chapitre 6, nous proposons donc un modèle informatique du social appraisal, basé sur notre modèle cognitif inspiré du CPM, et étendu aux aspects sociaux.

2.1.11 Conclusion sur les émotions et leur traitement en informatique

En psychologie, la définition de ce qu'est une émotion, leur nature exacte, ou le processus responsable de nos réactions émotionnelles, ne fait pas consensus. Différentes approches coexistent et s'influencent et plusieurs théories des émotions tentent de formaliser ce qu'est une émotion. Ces approches ne sont pas totalement exclusives, ainsi certaines théories relèvent de plusieurs approches. La plupart des théories supposent l'existence d'un processus cognitif sous-jacent aux émotions. Pourtant, certaines théories considèrent l'existence d'un processus cognitif dédié à chaque émotion (Ekman et Friesen, 1975, Tomkins, 1984), alors que d'autres considèrent un processus cognitif unique, commun à toutes les émotions (Scherer, 1984).

Le traitement informatique des émotions se heurte donc au problème de devoir sélectionner une ou plusieurs théories parmi l'ensemble des théories formulées en psychologie. Cette sélection doit être effectuée en fonction de l'application ciblée. Plusieurs types d'application ont été explorés (contrôle expressif, annotation émotionnelle, simulation affective, etc.) pour plusieurs de ces approches.

L'animation faciale de personnages virtuels expressifs peut être effectuée à partir de plusieurs de ces approches des émotions (catégorielle, dimensionnelle, cognitive, sociale, etc.). Il est donc nécessaire de concevoir différents modèles computationnels à partir de chacune des approches possibles et de concevoir les méthodes d'animation faciale associées. Les différents modèles conçus doivent également être évalués par des études perceptives afin de comparer les différentes approches possibles dans le cadre d'applications interactives incluant des agents virtuels expressifs.

2.2 Expressions faciales des émotions

2.2.1 Expressions faciales émotionnelles chez l'humain

Ekman a réalisé un grand nombre d'observations d'expressions faciales émotionnelles en se basant sur des individus de plusieurs cultures. Ainsi, il soutient l'existence de six émotions de base : Joie, Colère, Tristesse, Dégout, Surprise, et Peur. Chacune de ces émotions serait associée à un processus cognitif spécifique et inné. De plus, les expressions faciales de ces émotions (Figure 15) seraient universelles, car observées dans la totalité des cultures étudiées. Ekman décrit les différents signes faciaux saillants représentatifs de ces six émotions (Ekman et Friesen, 1975). Il définit ainsi des familles d'expressions faciales pour chaque émotion de base. Les expressions décrites par Ekman sont prototypiques, et il est difficile de les observer dans un contexte usuel, mais elles semblent, d'après les études menées par Ekman, être universellement reconnues et exprimées.

Figure 15 - Certaines expressions faciales des émotions de bases présentées par Ekman (dans l'ordre Colère, Peur, Dégout, Surprise, Joie, et Tristesse)

Ekman a également étudié les mélanges d'expressions faciales d'émotions. Il a ainsi mis en relief l'importance de zones particulières du visage pour certaines émotions. Par exemple, le froncement des sourcils semble être le seul indice fondamental pour l'expression de Colère.

Pour les autres états mentaux que ces six émotions, peu de descriptions précises sont disponibles. Baron-Cohen (2007) a créé le corpus vidéo MindReading. Il contient des vidéos de séquences émotionnelles actées (six vidéos par état mental, par exemple, intéressé, incertain, etc.). Cette base de données a notamment été utilisée dans un logiciel pédagogique pour personnes autistes, afin de les entrainer à reconnaitre les expressions faciales de ces différents états mentaux.

Si les expressions faciales d'émotions sont décrites de manière précise, les processus à l'origine de ces expressions sont plus controversés. Les approches catégorielles et les théories cognitives proposent des mécanismes sous-jacents de génération d'expressions faciales différents. L'approche catégorielle suppose l'existence d'un programme pré-écrit pour chaque émotion. Ainsi, Tomkins suggère que le mécanisme expressif est un ensemble de programmes neuromoteurs. Ces programmes prédisent que, suite à l'élicitation d'une émotion, le modèle expressif prototypique sera produit (Tomkins, 1984). Sur le visage plus particulièrement, un modèle prototypique de l'expression faciale est sélectionné parmi une famille d'expressions (correspondant à l'émotion élicitée) et affichée.

Au contraire, les théories cognitives estiment que le processus cognitif de l'émotion fait apparaitre successivement plusieurs éléments de l'expression qui se combinent dans le temps. Le mécanisme d'expression de l'émotion est donc très différent de celui décrit par l'approche catégorielle. L'événement suscitant l'émotion est évalué selon un ensemble de critères. Selon le modèle CPM (Scherer, 2001), chaque critère produirait une réaction expressive modulée. Ces réactions expressives se combineraient de façon dynamique, produisant une grande variété d'expressions faciales différentes. L'expression faciale produite par chaque évaluation s'ajoute de manière cumulative aux expressions faciales des évaluations précédentes (Wehrle, et al 2000). Selon Smith et Scott (1997) chacun de ces composants individuels de l'expression faciale contribuent à l'expression du visage et contient une signification propre.

Ainsi, les théories catégorielles et cognitives diffèrent selon le nombre et la proto-typicalité des expressions du visage. Selon le modèle CPM, il existe différentes composantes expressives associées aux différents checks de l'évaluation cognitive. Ces composantes expressives peuvent ainsi être partagées par plusieurs émotions.

Une étude a comparé l'expression faciale d'émotion en termes de catégorie et d'évaluation liées à l'expressivité du visage (Scherer et Ellgring, 2007). Douze acteurs ont été invités à exécuter des scénarios couvrant 14 émotions. Aucune expression prototypique complète n'a été observée pour les émotions de base, contrairement aux prédictions faites sur la base de l'approche catégorielle. L'apparition d'expression incomplète ou partielle a

été relativement rare, apparaissant dans seulement environ un tiers des vidéos. La rareté des expressions faciales prototypiques dans ces enregistrements peut également être interprétée comme une preuve contre l'existence de programmes expressifs prédéfinis (donc de l'approche catégorielle). De plus les dynamiques expressives montrent une variabilité beaucoup plus importante que celle que l'on attendrait dans le cas de l'approche catégorielle des émotions (Scherer 2010).

2.2.2 Perception humaine des expressions faciales d'émotions

De nombreuses études ont évalué la reconnaissance humaine des émotions dans les expressions faciales prototypiques (Russell, 1994). Souvent, ces études montrent des forts taux de reconnaissance catégorielle. Cependant ces études se limitent à un nombre restreint d'émotions, et très souvent, aux émotions de base uniquement.

De plus ces études laissent généralement de côté l'effet du visage en lui-même. En effet, certains paramètres sont importants dans la perception d'un visage. La familiarité, les proportions du visage, etc., doivent être contrebalancés lors de la passation de tests perceptifs. En effet, Gamond et al. (2011) ont montré que certains circuits neuronaux sont spécifiquement activés lorsque le sujet regarde un visage familier. Lorsqu'on présente un visage possédant au moins 35% de traits communs avec un visage connu, alors les mêmes circuits neuronaux sont activés. Selon les auteurs, les visages observés seraient associés de manière automatique avec certains traits de personnalités.

Hess et al. (2004) ont également mis en relief l'importance du genre (visage féminin/masculin) et de l'apparence du visage sur la perception des émotions. Par exemple, un visage masculin serait lié à une dominance plus élevée qu'un visage féminin. Pourtant, dans cette étude, la colère exprimée par les visages féminins a été perçue plus intense que les expressions faciales de colère des visages masculins. Cette étude met en relief les interactions complexes genre/personnalité et les biais associés dans la perception des expressions faciales d'émotions. Ces différentes études montrent donc la nécessité de considérer la forme du visage (et la personnalité associée à ce visage par le sujet) lors de l'analyse de sa perception émotionnelle.

Barkhuysen et al. (2010) ont étudié les aspects dynamiques de la perception des expressions faciales. Ils ont présenté à leurs sujets des expressions faciales d'émotion contenues dans des stimuli de durées variables (160ms, 320ms, 480ms, 640ms, 800ms, et 960ms). Cette technique est nommée « *gate paradigm* » (Grosjean, 1996). Pour chaque stimulus, les sujets avaient pour objectif de reconnaitre l'émotion exprimée, mais n'étaient pas obligés de répondre en cas de doute. Les résultats montrent les durées les plus courtes (160ms, et 320ms), le nombre de sujets donnant une réponse est nettement plus bas (environ 65% des sujets) que pour les stimuli plus longs (environ 95% des sujets). Cette étude met donc en relief que la perception humaine des expressions faciales est un processus progressif et qui nécessite un temps minimum pour être précis, qui semble se situer entre 400ms et 500ms.

2.2.3 Rendu et animation de visage expressif

Pour animer un visage virtuel expressif interactif et réaliste, il est nécessaire d'exploiter les technologies d'animations qui permettent d'exprimer les émotions données par le modèle émotionnel et leurs subtilités expressives. Nos travaux s'articulant autour de l'animation faciale temps réel, cette section présente les différents concepts importants liés à ce domaine.

2.2.3.1 Problématiques

Depuis le début des années 70, la simulation du visage humain représente un axe important de l'informatique graphique. Le premier visage de synthèse a été présenté au début des années 1970 dans les travaux de Parke (1974). Il y présente un visage virtuel animé par un nombre restreint de paramètres et crée une animation en s'appuyant sur une interpolation linéaire entre deux expressions clés.

La simulation de visage se heurte à deux problématiques (Magnenat-Thalmann et al., 1988). La première problématique est la qualité du rendu de visage réaliste statique. Dans les années 80, les capacités graphiques des ordinateurs sont encore trop limitées pour rendre la peau humaine de manière photoréaliste. La seconde problématique porte sur le réalisme de l'animation issue de la modélisation de la dynamique faciale. Le visage humain est extrêmement complexe, car composé d'un nombre important de muscles qui interagissent avec une structure osseuse et avec la peau.

Pourtant, au milieu des années 90, l'augmentation des performances des ordinateurs et l'émergence des processeurs graphiques programmables (*Graphic Processing Unit* ou *GPU*) donne un nouvel élan à plusieurs aspects de la simulation pour le visage.

2.2.3.2 Informatique graphique : définitions

L'informatique graphique est l'informatique liée à la génération (ou rendu) d'images de synthèse. On distingue les méthodes de rendu temps-réel, et non temps-réel, dont les objectifs sont différents. Les méthodes non temps réel, tels que le *ray tracing*, sont par exemple utilisées dans l'industrie du cinéma, et permettent aujourd'hui un niveau de réalisme visuel très élevé (Figure 16).

Figure 16 - Exemple de rendu réaliste en ray tracing (Source : Wikipédia). Cette image illustre la simulation lumineuse complexe des moteurs de ray tracing, gérant la diffraction, la réflexion, et la profondeur de champ (flou des objets en arrière-plan)

Les travaux présentés dans cette thèse reposent uniquement sur des techniques temps réel. En effet, le temps réel est une contrainte imposée par l'interaction homme machine.

Les méthodes de rendu temps réel tirent parti des capacités matérielles des cartes graphiques (GPU). Les GPU sont des composants matériels dédiés spécifiquement à la génération d'image de synthèse. Ces méthodes sont celles utilisées notamment par l'industrie du jeu vidéo. Les images obtenues sont généralement moins réalistes mais sont générées en moins de 40 millisecondes, permettant un rendu minimum de 25 images par secondes. Pour obtenir cette fréquence, ces méthodes exécutent des programmes spécifiques (nommés *Shaders*) dédiés aux processeurs graphiques. Il existe différents types de shaders, exécutés de manière séquentielle. C'est ce qu'on appelle le *Pipeline Graphique* (Figure 17).

Dans la partie vectorielle du pipeline, les shaders sont dédiés au traitement de la géométrie 3D. Le *Vertex Shader* est responsable des changements de référentiels, généralement pour placer les sommets dans l'espace de la caméra. Le *Tesselation Shader*, séparé en deux sous étapes, permet de raffiner localement un maillage, et ainsi de générer dynamiquement des détails. Ces shaders ne sont cependant pas disponibles sur tous les matériels

graphiques actuels. Le *Geometry Shader* permet lui de créer dynamiquement tout type de géométrie. L'une des applications usuelles est la duplication des triangles du maillage pour créer des ombres projetées. Finalement, le *Fragment Shader* est dédié au traitement des couleurs, c'est-à-dire le remplissage des pixels.

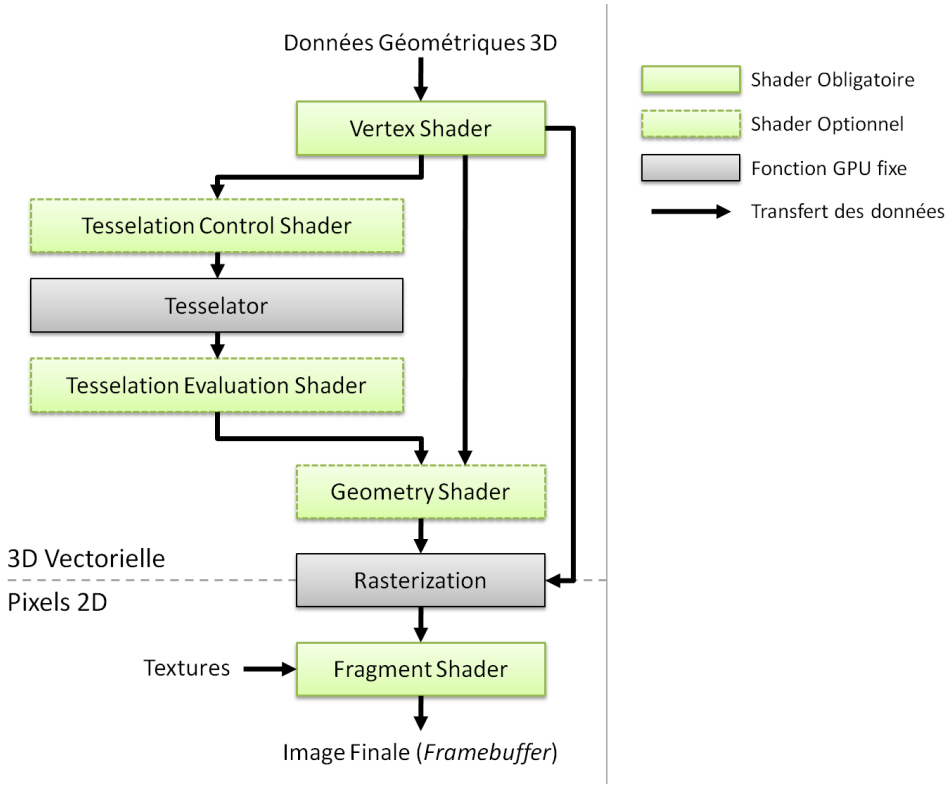


Figure 17 - Schéma simplifié du pipeline graphique des cartes actuelles (OpenGL 4.1 / DirectX 11)

Certains de ces shaders sont « optionnels » : le pipeline graphique peut fonctionner sans que ce type de shader soit implémenté. Actuellement, le Vertex Shader et le Fragment Shader sont les deux seules étapes indispensables, et sont donc requis pour générer une image de synthèse.

L'avantage du GPU sur le CPU est sa capacité à traiter ***en parallèle*** un très grand nombre d'opérations sur les nombres réels, permettant ainsi de traiter un grand nombre d'informations de manière simultanée. Contrairement à la programmation CPU classique, utilisée notamment par les méthodes non temps réel telles que le *ray-tracing*, les shaders sont donc écrits pour être exécutés avec un haut niveau de parallélisme.

Les méthodes temps réel et non temps réel semblent vouées à fusionner, car les matériels graphiques, de plus en plus puissants, utilisent aujourd'hui certains algorithmes développés pour le rendu non temps réel. Cependant, à l'heure actuelle, l'architecture matérielle des cartes graphiques (Figure 17) ne se prête pas bien à l'implémentation de ces algorithmes, ainsi la fusion des deux approches pourrait nécessiter des changements importants dans la conception des matériels graphiques, ce qui repousse son échéance.

2.2.3.3 Qualité de rendu

La qualité d'une image de synthèse repose sur 2 aspects fondamentaux : 1) La qualité des données, en particulier les structures géométriques 3D (ou maillages) et les textures (images plaquées sur les surfaces, propriétés lumineuses), et 2) la qualité du système d'illumination (simulation lumineuse responsable des calculs d'éclairage).

Le maillage est une structure géométrique constituée de sommets (appelés Vertex), reliés ensemble pour former des surfaces (généralement des triangles). Animer un objet consiste donc à déplacer les vertex qui le composent dans l'espace 3D pour le déformer. Cette opération est effectuée par les shaders dédiés à la géométrie 3D (généralement par le *Vertex Shader*).

Une fois déformé, le maillage est projeté en 2D sur l'écran, découpé en pixels (rastérisé), puis la couleur de chaque pixel peut être calculée de manière indépendante (donc en parallèle). Cette opération repose sur les propriétés lumineuses de l'objet et sur la texture que l'on souhaite plaquer dessus (qualité des données), et sur les méthodes de calcul lumineux programmées en shader (qualité du système d'illumination). Chaque type de surface requiert son propre algorithme de simulation lumineuse. Par exemple, le shader de rendu de la peau sera très différent du shader de rendu de surfaces minérales, telles que le marbre.

2.2.3.4 Rendu réaliste de la peau

Le rendu réaliste de la peau humaine a longtemps été un enjeu pour l'informatique graphique. Ce n'est que récemment que 1) la puissance de calcul et les algorithmes existants ont permis d'obtenir des résultats satisfaisants en termes de systèmes d'illumination. 2) Les dispositifs de capture de données ont permis d'acquérir des mesures précises de spéculométrie de la peau du visage humain (comportement de la lumière dans la structure de la peau).

Ces nouvelles méthodes de calcul et d'acquisition de données étant très récentes, nous verrons plus loin que le rendu de la peau des agents virtuels est souvent peu élaboré, et donne un effet « plastique » qui peut faire obstacle à la crédibilité de l'agent.

2.2.3.5 Système d'illumination pour la peau du visage humain

La peau humaine est une surface irrégulière, transparente et composée de plusieurs couches superposées (épiderme, derme, muscle, os) dont l'irrigation sanguine varie avec la physiologie et l'état émotionnel. La lumière s'y comporte donc de manière complexe. Les algorithmes de rendu de la peau doivent tenir compte de cette complexité physiologique. Les rayons lumineux sont diffusés et réfléchis plusieurs fois à l'intérieur des couches de la peau et ne sont presque jamais réémis de l'endroit où ils sont entrés. De plus, leur trajectoire change en fonction de leur longueur d'onde et de leur énergie. Ce phénomène est illustré par la Figure 18 ci-dessous.

Pour simuler ce comportement de la lumière, différent modèles ont été proposés. Certains modèles prennent par exemple en compte l'influence de la composition chimique de la peau (mélanine, hémoglobine) sur la couleur de la surface de la peau, mais sans tenir compte de la diffusion lumineuse en surface. En effet, ces deux composés chimiques naturellement présents dans la peau ont une influence directe sur sa couleur, car au cours de sa trajectoire à l'intérieur de la surface de la peau, une partie de la lumière est absorbée. En fonction des caractéristiques chimiques de la peau, les différentes longueurs d'ondes lumineuses sont absorbées différemment. Modifier la composition chimique de la peau modifie donc sa couleur. Augmenter la mélanine rend un effet plus bronzé. Augmenter/diminuer l'hémoglobine permet de faire rougir/blêmir la peau.

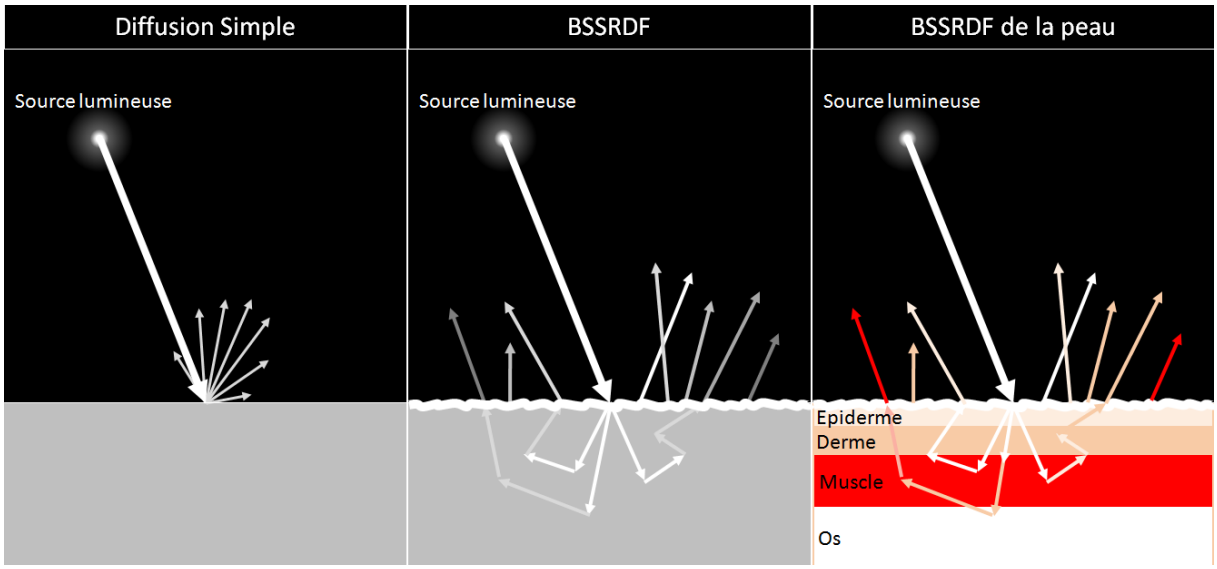


Figure 18 - Trajectoire de la lumière dans une surface partiellement transparente

Le phénomène de rougissement/blêmissement étant un facteur expressif (De Melo et al, 2009), il est important de simuler la composition chimique de la peau dans le système d'illumination.

Les techniques de rendu modernes les plus réalistes sont les modèles de BSSRDF (*Bidirectional Subsurface Scattering Reflectance Distribution Function*). D'Eon et al. (2007) proposent une technique de calcul d'un modèle de BSSRDF temps réel sur GPU, inspiré du modèle non temps réel de Donner et Jensen (2005). Ce modèle suit les lois physiques telles que la conservation de l'énergie. Ce rendu temps réel de la peau est dit multi passes. Il pré-calcule successivement plusieurs images hors écran, pour les utiliser dans le calcul de l'image qui sera affichée à l'écran. Le rendu temps réel obtenu par D'Eon et al. (Figure 19 à droite) est difficile à différencier du rendu non temps réel (Figure 19 à gauche).

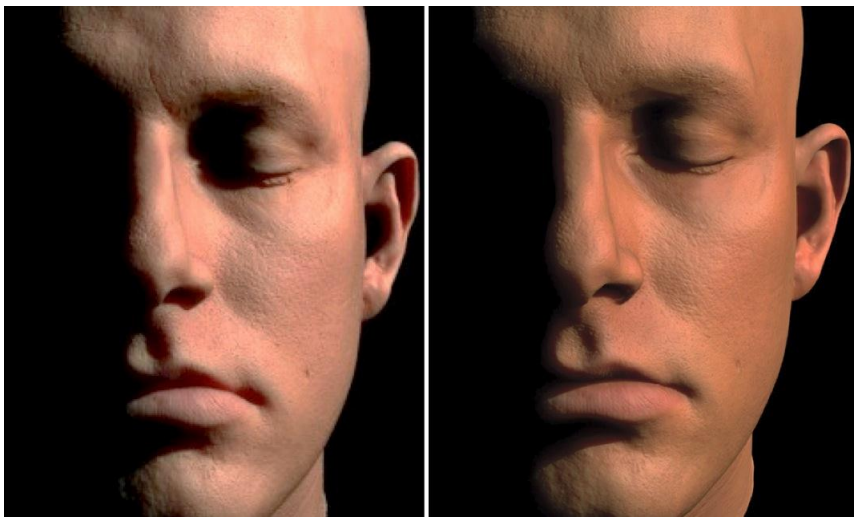


Figure 19 - Rendu de la peau. A gauche : Modèle non temps-réel CPU de Donner et Jensen (2005), A droite : Modèle temps-réel GPU de D'Eon et al. (2007).

2.2.3.6 Données spéculométriques et dispositifs de capture du visage humain

Quel que soit le modèle de rendu utilisé, il est nécessaire d'utiliser des données précises sur les propriétés lumineuses et chimiques de la peau. De plus, ces données sont différentes en fonction de la zone du visage considérée. En effet, la peau du visage n'est pas uniforme. Certaines parties sont plus brillantes, plus mattes, plus fines, plus épaisses, etc.

La qualité des données associées au visage (textures, spéculométrie) est fondamentale pour obtenir un rendu de qualité. En utilisant des appareils de mesures spécifiques pour enregistrer le comportement lumineux de la peau, il est possible aujourd'hui d'acquérir les données spécifiques à chaque région du visage, et ainsi de reproduire le comportement de la lumière de manière fidèle sur l'ensemble du visage. Ces données capturées sont donc aussi importantes que les algorithmes d'illumination utilisés.

Les dispositifs actuels considèrent généralement plusieurs types de comportements lumineux. Le système LightStage (décrit ci-après), considère quatre types de comportements lumineux : La lumière réfléchie, la lumière diffuse directe, la lumière diffusée en surface de la peau, et la lumière diffusée par les couches profondes de la peau.

Plusieurs jeux de données sont donc capturés, pour être ensuite recombinés par le système d'illumination lors du rendu final. La Figure 20 montre les différentes composantes lumineuses et leur combinaison, comparée à la photographie du modèle humain.

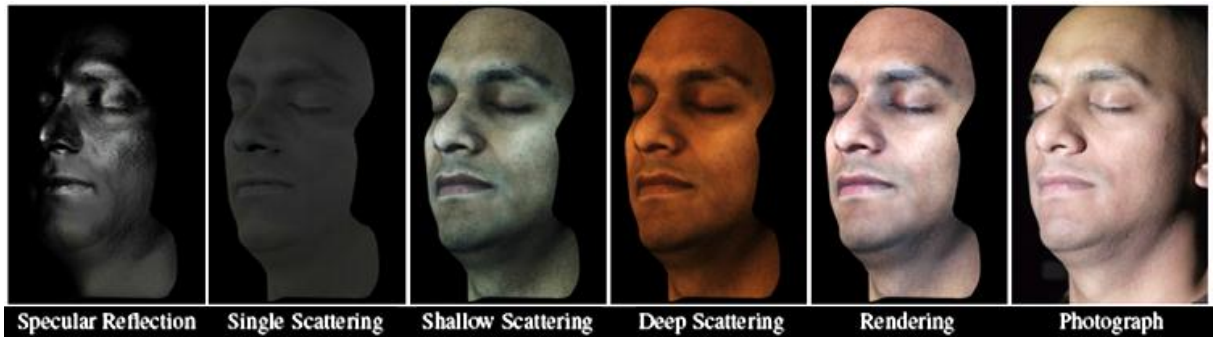


Figure 20 - Composantes lumineuses séparées, combinées, et photo de référence (Ma et al., 2008).

Ces données, sont généralement acquises depuis plusieurs angles de vue, et en considérant plusieurs angles d'incidence de la lumière. L'ensemble de captures lumineuses représente donc une grande quantité de données.

Diverses approches de captures dédiées au visage humain ont été développées au cours des dernières années. Elles ont toutes en commun l'utilisation d'un grand nombre de lampes et de caméras, parfois basées sur des rayons de lumière structurée.

Debevec et al. (2000) ont conçu l'un des premiers systèmes de capture dédiés au visage humain. La première version du système LightStage est composée de plusieurs caméras et sources lumineuses montées sur un axe rotatif permettant de capturer le visage selon plusieurs angles (Figure 21).

Ce système a ensuite été modifié et composé d'une structure rigide (Figure 22), utilisant de la lumière structurée, et permettant ainsi de capturer la géométrie faciale du sujet (en plus de mesurer les réflexions lumineuse de la peau). Cette technologie (Alexander et al., 2010), est aujourd'hui utilisée dans de nombreux films pour créer des effets spéciaux numériques (James Cameron's Avatar, Benjamin Button, Spiderman 2, etc.).

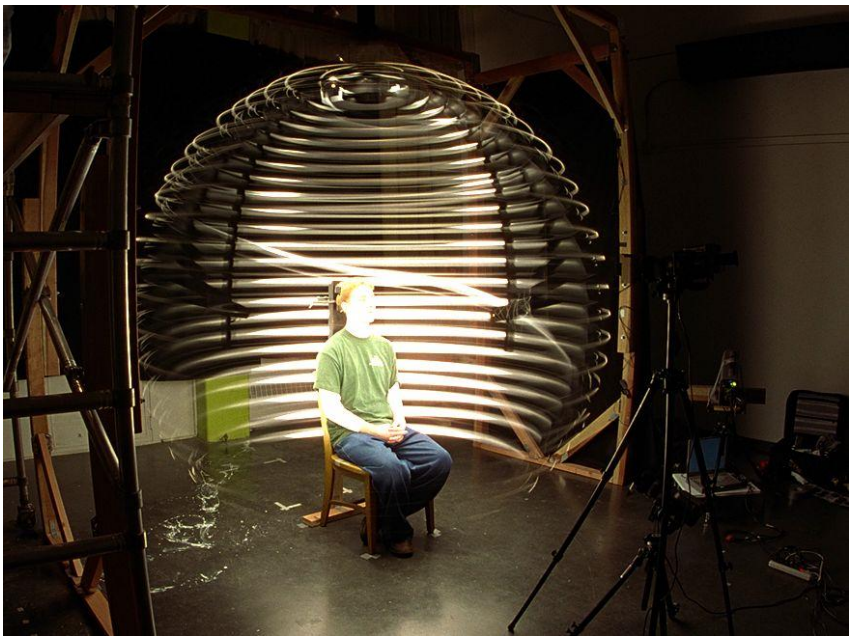


Figure 21 - Première version du système LightStage, la structure de caméra et de lumière est ici en rotation (effet de trainées lumineuse) (Debevec et al., 2000)

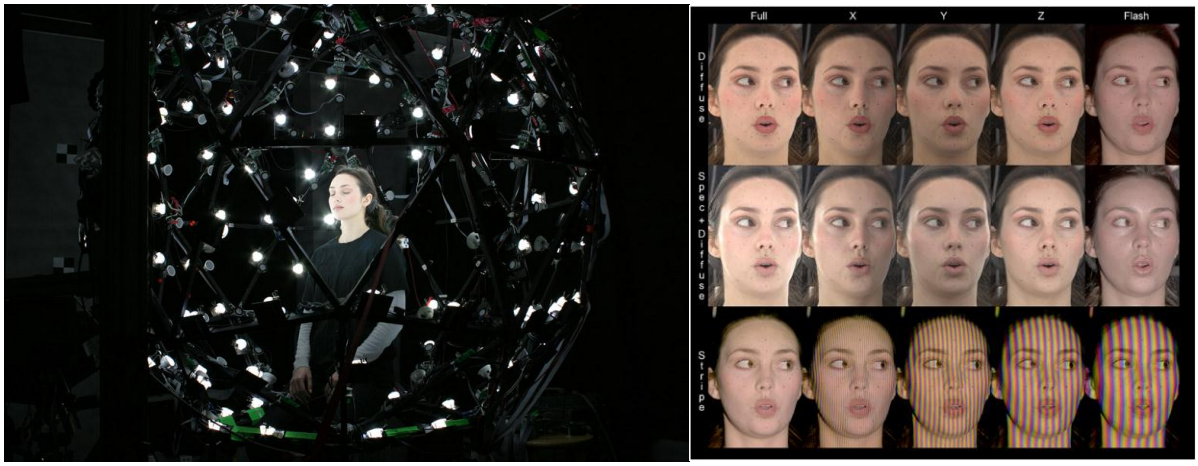


Figure 22 - LightStage Version 6. Structure fixe (gauche) et lumière structurée (droite) (Alexander et al., 2010)

Weyrich et al (2006) ont conçu un système de capture similaire se composant de deux appareils. Le premier est un dispositif de caméra pour enregistrer les propriétés réflectives de la peau selon différents angles. Le dispositif, dont la structure est similaire à celle de LightStage, permet la capture depuis 16 angles de vue différents, avec pour chacun, 14 angles lumineux. Soit un total de 150 prises de vue (Figure 23).

Le second dispositif (Figure 24) se compose d'un capteur de lumière diffusée. Il mesure, via un ensemble de fibres optiques, la quantité et la variation colorimétrique de lumière diffusée sous la surface de la peau. Il permet de mesurer le comportement lumineux de la peau sur une portion précise du visage, à une échelle plus locale que la structure de 150 caméras. Dans la partie droite de la Figure 24, on peut visualiser la diffusion intra-cutanée de la lumière. Plus on s'éloigne de la source, plus la lumière est rougie et faible. Ainsi, utiliser ce second dispositif sur plusieurs parties du visage permet d'obtenir des informations fines sur l'influence des couches internes du visage en différents points.

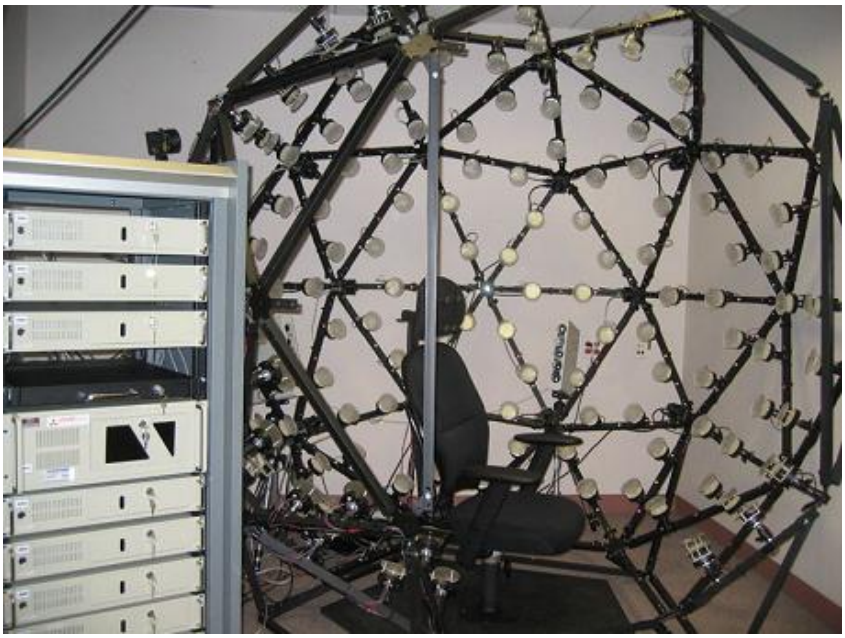


Figure 23 - Dispositif de capture de réflectance de Weyrich et al. (2006)

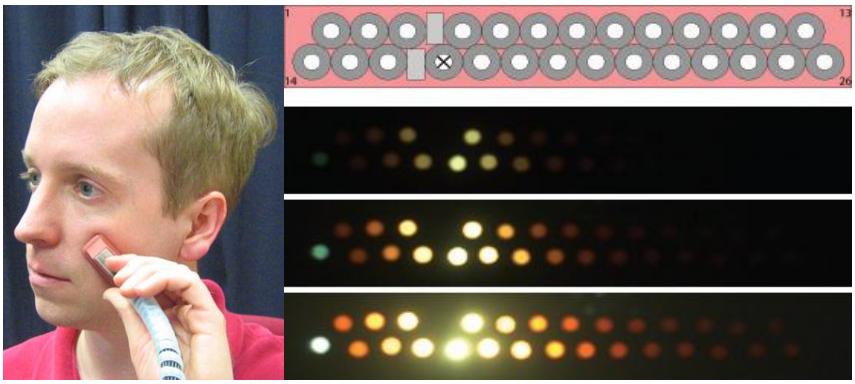


Figure 24 - Outils de mesure de diffusion lumineuse intra-cutanée de Weyrich et al. (2006)

Il est important de noter que ces données sont indépendantes du type de rendu utilisé pour générer un visage de synthèse. Les algorithmes temps réel peuvent profiter de la qualité de ces données de la même manière que les algorithmes plus complexes et non temps réel utilisés, par exemple, dans l'industrie du cinéma. Quels que soient les dispositifs de capture utilisés, les données brutes nécessitent en revanche un prétraitement important pour pouvoir être utilisées pour l'animation faciale.

2.2.3.7 Techniques d'animation faciale

Comme l'a souligné Magnenat-Thalmann (1988), le rendu de la peau n'est que l'une des problématiques de la simulation de visage. L'autre point important porte sur l'animation. En effet, la plupart des modèles photo-réalistes proposés ne sont pas animés (par exemple, D'Eon et al, 2007). Pour ceux qui le sont, l'animation est en général moins réaliste, et même si il est difficile de distinguer le virtuel du réel sur des images statiques, la différence devient visuellement évidente lorsque ces visages sont animés.

Aujourd'hui, plusieurs techniques d'animation faciale coexistent. Pour les présenter, nous avons choisi de les répartir en trois catégories distinctes (Figure 25) : l'animation paramétrique, l'animation par simulation physique, et l'animation par corpus. Ces approches se distinguent selon deux axes : la quantité de données requise, et la quantité de calculs requise.

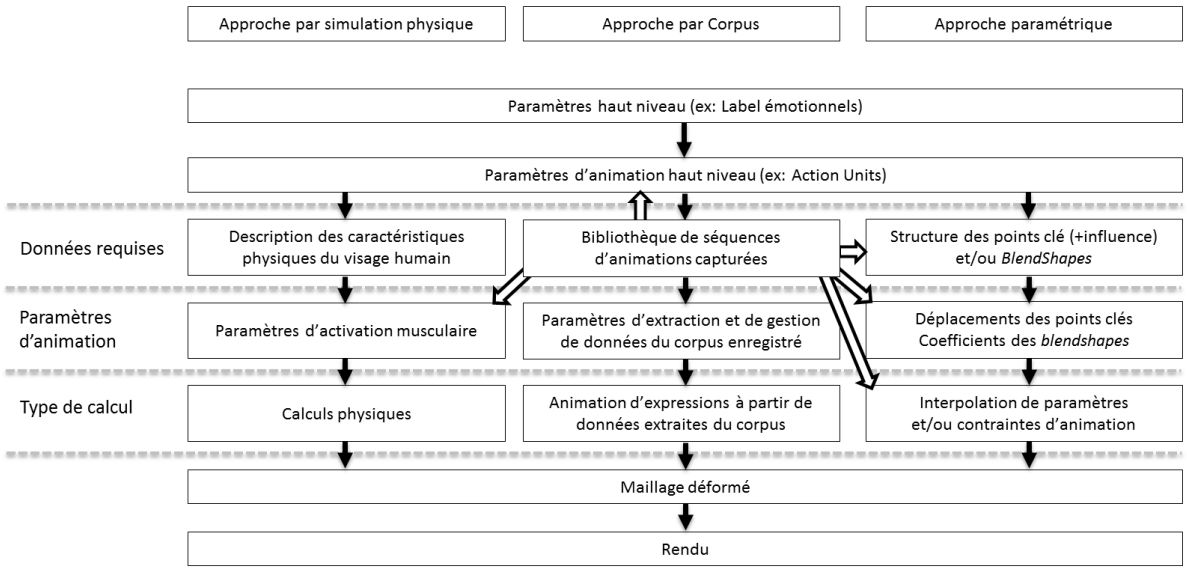


Figure 25 - Différentes approches de l'animation faciale et relations entre elles.

- ***Approche paramétrique***

L'animation paramétrique est l'approche la moins consommatrice de puissance de calcul et elle ne nécessite que peu de données en comparaison des deux autres approches.

Cette méthode a pour principe que la génération d'animations faciales dynamiques et temps réel requiert un niveau d'abstraction. L'animation doit être décrite par un nombre de paramètres expressifs réduits qui détermineront la forme du visage.

Le visage présenté par Parke était un ensemble réduit de polygones dont l'animation était obtenue par interpolation linéaire des sommets entre deux positions clés. Bien que la complexité des modèles 3D ait évolué depuis, cette méthode est encore utilisée par plusieurs systèmes d'animation faciale car elle permet de définir dynamiquement la séquence d'expressions faciales. L'interpolation peut être faite soit entre deux formes du visage (technique nommée *blendshapes*), soit par interpolation entre deux configurations de points-clés qui serviront de paramètres à la déformation du maillage (technique nommée *skinning*).

La méthode du *skinning* consiste à placer chaque point clé à une position précise sur le visage. Pour chacun d'entre eux, on associe une région d'influence sur la surface du visage. Ainsi, le déplacement de chaque point clé déplace une zone du visage. Ces régions d'influences sont pondérées. Ainsi, les sommets les plus éloignés seront moins impactés par le point clé que les vertices très proches du point clé considéré (Figure 26).

Plusieurs configurations de point clés ont été proposées. Par exemple, le modèle de point clés MPEG-4 (Pandzic & Forchheimer, 2003) (Figure 27) est largement utilisé dans la conception de personnages virtuels expressifs et interactifs.

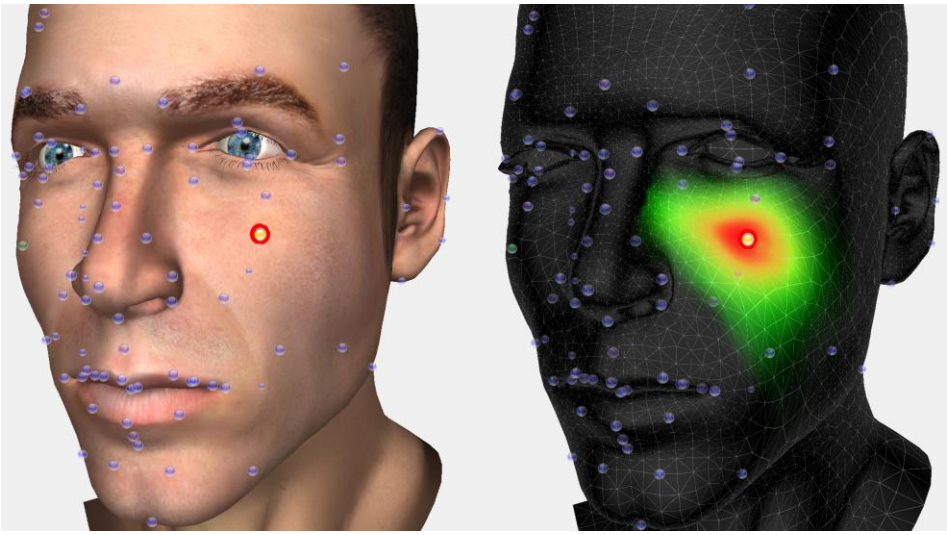


Figure 26 - Exemple de paramètres de Skinning. A Gauche : Positions des points-clés. A Droite : Influence de l'un des points-clés (Rouge : très fort, Vert : Moyen, Noir : Aucune influence)

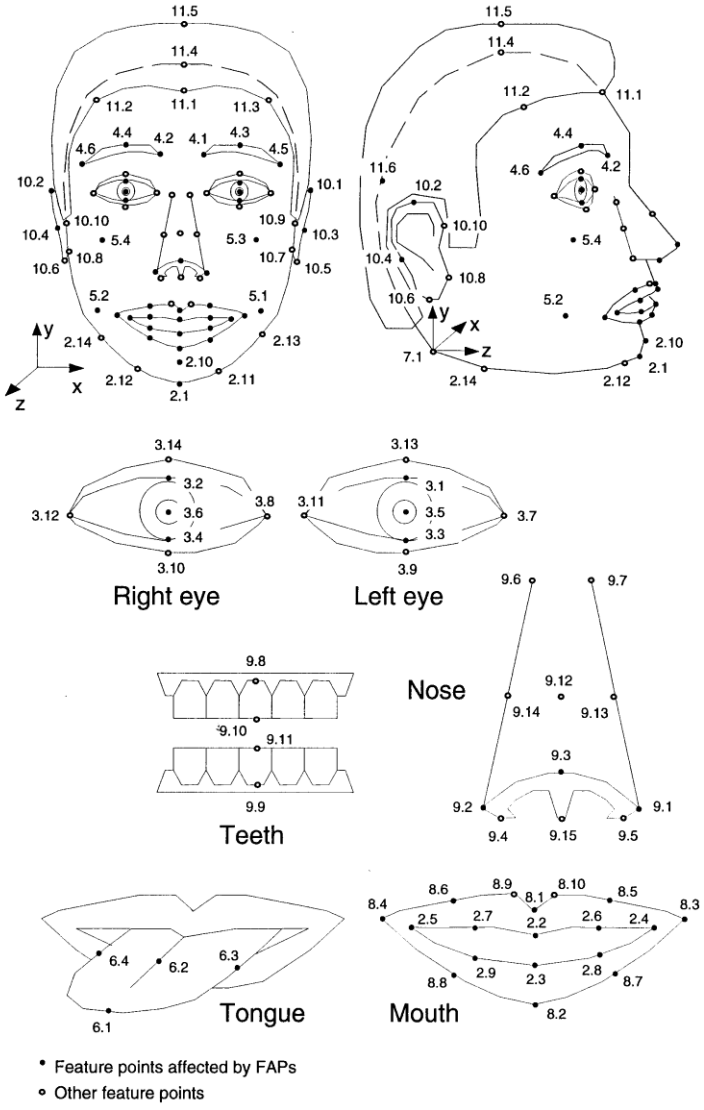


Figure 27 - Schéma des points-clés de la norme MPEG-4 (Pandzic & Forchheimer, 2003)

Une autre approche paramétrique classique est la technique des *blendshapes.* Cette méthode consiste à interpoler le visage entier entre deux formes prédéfinies de visage, généralement obtenues à l'aide d'éditeur ou de dispositif de capture. L'animation est ensuite réalisée par interpolation entre une forme de visage et une autre. Cette technique peut être considérée comme une variante de l'animation à base de points-clés. Chaque sommet du visage est alors considéré comme un point clé, n'influençant que lui-même. De même que pour les points-clés, il est possible de définir plusieurs zones du visage comportant chacune un ensemble de blendshapes. La combinaison de plusieurs blendshapes permet alors la création d'expressions faciales à partir de sous expressions unitaires.

L'approche paramétrique, et en particulier l'approche par points-clés, permet de reporter les expressions faciales d'un personnage virtuel sur un autre. C'est le principe du *retargeting* (Chuand et Bregler, 2002). En effet, la structure de points-clés permet une abstraction de la géométrie. Ainsi si deux maillages différents, mais de géométries similaires, partagent le même jeu de points-clés, il est possible de reporter les animations des points-clés d'un personnage sur un autre. En revanche, lorsque la structure des maillages diffère de manière plus importante, diverses problématiques apparaissent. Par exemple, si on veut appliquer la même expression faciale entre un humain et un animal cartoon, la structure du visage varie de manière trop marquée pour appliquer directement les paramètres de déplacement des points-clés. Le retargeting permet alors de modifier les paramètres pour les transposer sur différents modèles 3D. Plusieurs méthodes de retargeting ont été proposées (Pighin et Lewis, 2006). Ces diverses méthodes permettent de transférer des expressions et des animations sur des personnages dont la morphologie varie largement (Figure 28) (Song et al. 2011).

Figure 28 - Retargeting d'une expression faciale sur divers types de personnages (Song et al. 2011)

Le *retargeting* permet également l'animation par capture de l'utilisateur (*Performance Driven Facial Animation*), dans lesquels la dynamique faciale de l'utilisateur est captée, puis transformée pour être appliquée sur un personnage virtuel (Dutrève et al. 2008).

Cependant, les problématiques du retargeting n'étant pas directement liées à nos travaux, nous ne décrirons pas plus en détails ces techniques.

L'approche paramétrique manque généralement de réalisme. Deux problématiques ont été identifiées : 1) l'animation synchronisée sur l'ensemble du visage, et 2) l'interpolation linéaire (Parke, 1982).

Les points-clés sont généralement déplacés tous simultanément, créant une animation faciale parfaitement synchronisée sur l'ensemble du visage, et donc très peu réaliste. Considérant ce problème, certains travaux ont divisé le visage en plusieurs zones, et ont réalisé des animations locales à des vitesses différentes selon les zones par interpolation linéaire (Niewiadomski et al., 2009).

En raffinant ce principe, on aboutit à un visage divisé en unité d'animation les plus élémentaires possibles. Par exemple, la liste des Unités d'Action (*Action Units*) (Ekman et al., 2002). Chaque Action Unit peut être considérée comme une composante de base de l'animation du visage. Il devient alors plus complexe d'obtenir une animation cohérente sur l'ensemble du visage, et la multiplication des unités d'animation élémentaires multiplie également le nombre de paramètres nécessaires pour contrôler le visage.

Les animations obtenues restent cependant basées sur le principe d'interpolation. Le principe d'interpolation linéaire est le second obstacle au réalisme de l'animation (Parke, 1982, Cosker et al., 2010). Certains modèles dynamiques utilisent des interpolations sigmoïdes, ou encore des modèles de dynamique plus complexes pour générer des animations non linéaires sur le visage virtuel.

L'étude perceptive menée par Cosker et al. (2010) repose sur l'utilisation d'un scanner haute définition 3D enregistrant le mouvement de 30.000 points sur un visage humain à une fréquence de 60Hz. Cet outil a permis aux auteurs d'observer que lorsqu'un visage humain passe d'une expression neutre à une expression d'émotion, la plupart des points de la surface du visage suivent une courbe et non une ligne droite (Figure 29).

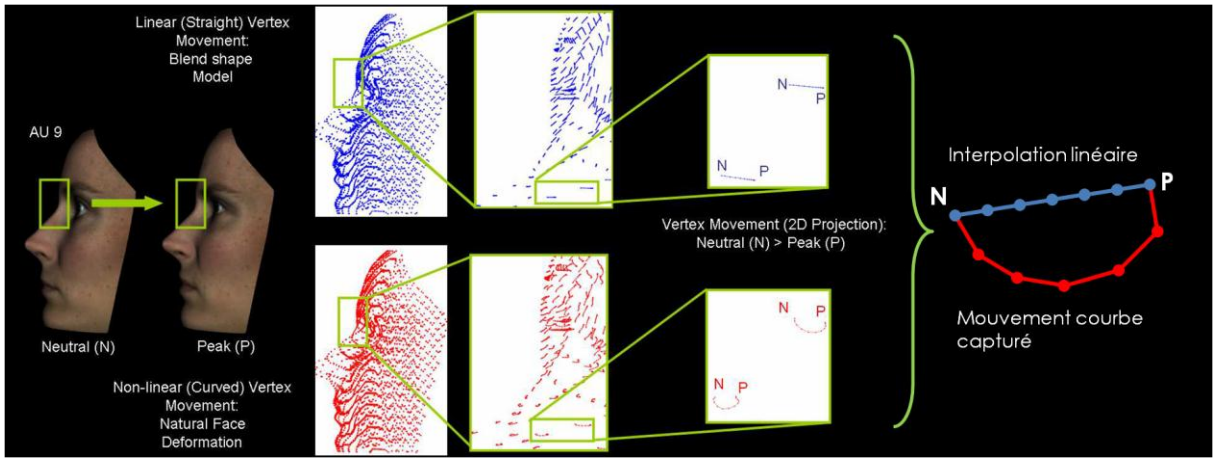


Figure 29 - Capture de mouvements faciaux VS interpolation linéaire (A partir de Cosker et al. 2010).

Après avoir enregistré un corpus d'animations faciales, les auteurs ont effectué une étude perceptive sur un groupe de sujets, en comparant les mouvements linéaires à des mouvements suivant les courbes enregistrées par leur système de capture. Les auteurs montrent que les animations basées sur les mouvements enregistrés sont significativement préférées aux mouvements linéaires et sont jugées comme étant plus naturels.

Pourtant, la reproduction de ce type de mouvements complexes à base de mouvements courbes, en utilisant un modèle paramétrique, reste un défi. Seules les deux autres approches permettent actuellement d'obtenir ce genre d'animations réalistes.

- ***Approche par corpus de capture de mouvement***

L'animation faciale par corpus est actuellement l'unique méthode temps réel qui génère des animations dont la dynamique est fidèle à celle du visage humain. Elle consiste effectivement à reproduire les mouvements d'un visage humain à partir d'enregistrements 3D. En contrepartie, cette approche nécessite l'enregistrement d'une quantité de données importante. En effet, chaque configuration du visage doit être enregistrée en 3D. Il faut donc enregistrer au minimum 30 structures 3D pour chaque seconde de film.

Par exemple : Soit une animation de 3 secondes durant laquelle le visage virtuel composé de 5000 vertex passe de l'expression neutre à l'expression de colère, en affichant 30 images par seconde.

En partant du principe que la position ou le déplacement d'un point dans l'espace se représente en informatique sous forme de trois valeurs flottantes sur 32bits (x, y, et z), soit 12 octets, on obtient (Figure 30) :

- Animation par 90 point clés
    - Structure du visage : 5000 vertex = 60 kilo octets
    - Points-clés de départ : 90 positions = 1,08 kilo octets
    - Points-clés d'arrivée : 90 positions = 1,08 kilo octets
    - Total : 62.16 kilo octets

- Animation par blendshapes :
    - Structure du visage de départ : 5000 vertex = 60 kilo octets
    - Structure du visage d'arrivée : 5000 vertex = 60 kilo octets
    - Total : 120 kilo octets

- Animation par capture de mouvement :
    - Total : 5000 vertex * 3 secondes * 30 images par seconde = 5.4 Méga Octets.

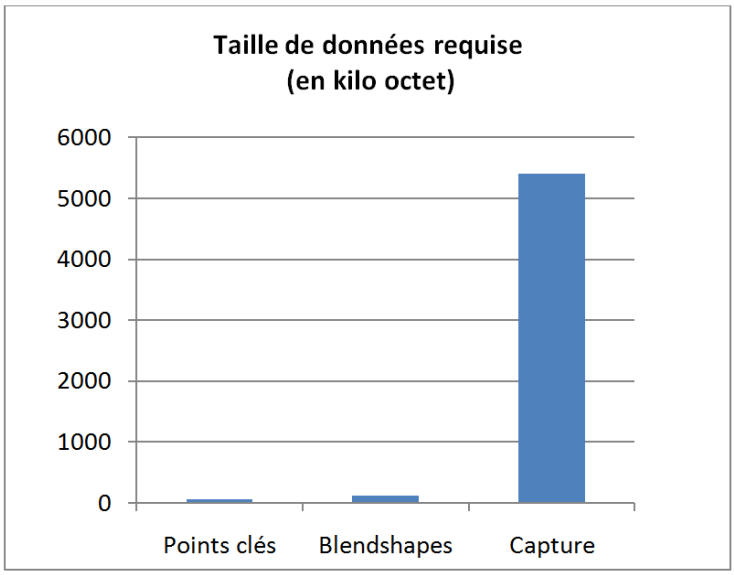


Figure 30 - Comparaison des données requises entre l'approche par points-clés, blendshapes, et capture de visage pour une animation de 3 secondes

Bien entendu, la qualité de l'animation générée est largement supérieure dans le cas de la capture de mouvement. En effet, une interpolation linéaire synchronisée sur l'ensemble du visage, entre deux expressions clé, et d'une durée de trois secondes, n'est pas une animation crédible (Cosker et al. 2010).

Cependant, la dynamique faciale générée par l'animation à base de capture de mouvement est extrêmement liée au modèle 3D de personnage utilisé. Il est donc difficile de transposer les expressions d'un visage virtuel à un autre. De plus, cette technique ne permet pas de générer des animations dynamiquement. Toute animation doit donc être enregistrée au préalable à partir d'un visage réel.

Néanmoins, cette technique est applicable dans un domaine aussi exigeant que celui du jeu vidéo. Les studios Rockstar ont récemment utilisé cette approche dans la conception du jeu « L.A. Noire »[3] (Studio Rockstar).

[3] http://www.rockstargames.com/lanoire/

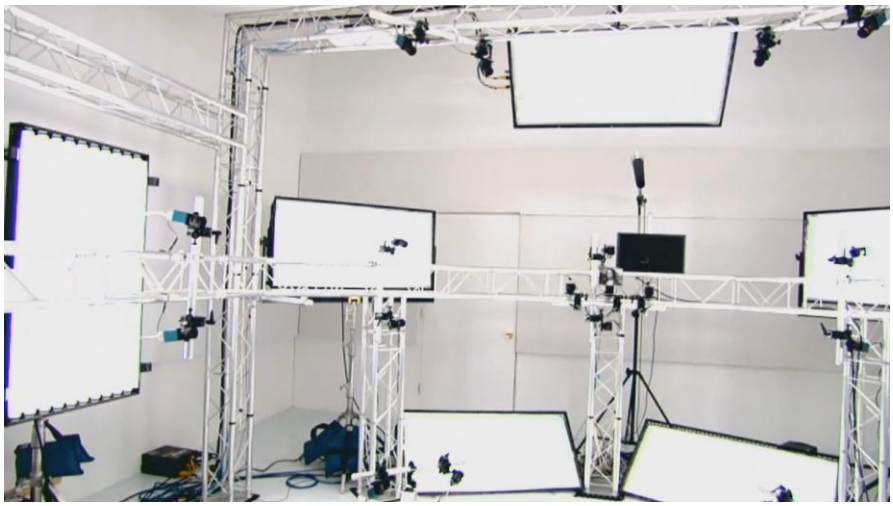


Figure 31 - Dispositif de capture d'expressions faciales utilisé par Rockstar pour la production du jeu LA. Noire (Studio Rockstar)

Figure 32 - Un acteur dans le dispositif de capture sans marqueur utilisé par les studios Rockstar (gauche). Exemples d'images obtenues à partir des données capturées (droite)

Figure 33 - Exemple de rendus d'expressions faciales temps réelles du moteur de jeu de LA. Noire, basés sur un corpus obtenu par motion capture (extraits de la vidéo Making of L.A. Noire)

Le dispositif de capture utilisé est un dispositif sans marqueurs faciaux, et donc moins intrusif, basé sur l'utilisation de caméras haute définition. La géométrie du visage est reconstruite en 3D, image par image. Ainsi, la forme du visage 3D est extrêmement proche de celle de l'acteur. On peut alors reconnaitre le visage de l'acteur, son jeu, et la dynamique de son visage. Cette technique est notamment efficace pour capturer les détails dynamiques du visage de l'acteur, tels que les rides permanentes, les rides d'expressions et les mouvements subtils. La Figure 33 montre trois images générées par le moteur de rendu du jeu L.A. Noire en utilisant les données capturées sur des acteurs.

- ***Approche par simulation physique***

L'animation par simulation physique consiste à considérer l'ensemble du visage humain, et non pas uniquement sa surface. Les expressions ne sont pas dues à une simple déformation de la peau, mais à une simulation complexe des os, des muscles, des tendons et de la peau.

Plusieurs systèmes d'animation faciale basés sur la simulation physiologique du visage humain ont été conçus pour éviter l'interpolation linéaire de paramètres. Waters (1987) a proposé le premier modèle musculaire simulant le visage humain. Ce modèle a ensuite été étendu pour prendre en compte les dynamiques physiologiques dues aux propriétés viscoélastiques de la peau (Lee, 1993). Sifakis et al. (2006) proposent une simulation physique complète du visage humain et utilisent des paramètres d'activation musculaire issu d'un corpus capturé. Ainsi la dynamique faciale est calculée par une simulation physique exacte, et à partir de données de contractions musculaires issues d'un visage humain réel (Figure 34). Toutefois, ces modèles physiques et physiologiques exigent beaucoup plus de temps de calcul que les modèles paramétriques. De fait, ils ne sont pas adaptés pour l'animation en temps réel.

Figure 34 - Simulation de l'ouverture de la mâchoire par le système biomécanique non temps réel de Sifakis et al. (2006)

Kähler et al. (2001) ont proposé une méthode basée sur un système physique dit « masse-ressort » pour simuler les muscles et l'élasticité de la surface de la peau sur un maillage simple. Ce type de modèle simplifié permet une rapidité de calcul adaptée à l'animation temps réel, au détriment de la fidélité physiologique du visage et de la qualité des maillages utilisés.

Les systèmes physiques permettent également le retargeting d'animation faciale, bien que les algorithmes soient différents, car il s'agit d'adapter les activations musculaires d'un modèle à un autre (Choe et al. 2001).

Puisque nous souhaitons utiliser des modèles avec une géométrie faciale très détaillée et animés en temps réel, les travaux présentés dans la suite de ce manuscrit n'utilisent pas de modèle par simulation physique. Nous ne détaillerons donc pas plus les travaux liés à ce champ de recherche.

- ***Partage de données et approches hybrides***

Ces différentes approches partagent un certain nombre de caractéristiques. Les données capturées pour réaliser de l'animation à base de corpus peuvent par exemple être utilisées pour extraire des paramètres d'animation haut-niveau, tel que les Action Units ou les paramètres d'activations musculaires (Sifakis et al. 2005) pour être appliqués ensuite à une autre technique d'animation. Il est également possible de calculer les paramètres d'animation d'un modèle à base de points-clés en se basant sur des données capturées sur un visage humain.

De plus, les paramètres de haut niveau sont indépendants de la technique d'animation utilisée. Par exemple, le déclenchement d'une Action Unit particulière peut être appliqué à tous les modèles d'animation. Le résultat obtenu doit être « similaire » quel que soit la technique d'animation utilisée, ce qui permet, par exemple, de les comparer entre elles.

Certains modèles ont également cherché à utiliser des données issues de capture de mouvements faciaux pour améliorer la qualité de l'animation paramétrique. Par exemple, le modèle de Stoiber et al., (2010) permet de répliquer la dynamique viscoélastique des différentes parties du visage humain sur un modèle paramétrique à partir d'analyses de capture de mouvement. Ainsi, les auteurs obtiennent un système paramétrique non basé sur l'interpolation linéaire, mais qui simule au contraire un comportement plus réaliste.

2.2.3.8 Visages expressifs virtuels

Plusieurs systèmes d'animation faciale sont conçus pour exprimer des émotions. Par exemple, la boîte à outils XFace (Balci et al., 2005) ou Greta (Niewiadomski et al., 2009). Ces systèmes utilisent une animation paramétrique basée sur MPEG4 pour modéliser des expressions faciales d'émotions. Dans la vie réelle, il est rare de voir apparaitre des expressions faciales d'une émotion de base seule. Au lieu de cela, nous éprouverions plutôt des mélanges d'émotions (Scherer et Ceschi, 1997, Abrilian et al., 2005). Ainsi, certains de ces systèmes (par exemple Greta) ont été également dotés de la capacité d'afficher des mélanges de plusieurs expressions d'émotions (Albrecht et al., 2005, Ochs et al., 2005). La plupart des modèles de mélange d'expressions utilisent uniquement des interpolations entre les expressions prédéfinies des émotions de base, affichées simultanément sur l'ensemble du visage. Certains modèles décomposent le visage en régions distinctes animées de manière indépendante ou des modèles d'expression séquencée (Niewiadomski et al., 2010).

Niewiadomski et Pelachaud (2010) ont proposé un modèle combinatoire reposant sur une méthode floue qui permet de générer des expressions faciales complexes d'émotions. Ce modèle a été appliqué à l'agent virtuel Greta. Les expressions faciales complexes affichées résultent de la combinaison de plusieurs expressions faciales simples (par exemple la superposition de deux expressions) ou qui sont volontairement modifiées par le système de rendu (par exemple, inhiber ou masquer). Cette méthode permet le masquage de l'émotion selon des règles de politesse.

Ces systèmes d'animation faciale permettent d'exprimer une grande variété d'émotions. Pourtant, la plupart des applications se concentrent sur les émotions de base : Joie, Tristesse, Colère, Surprise, Dégout, Peur (Ekman et Friesen, 1975). Becker-Asano et Wachsmuth (2008) ont observé que les agents exprimant à la fois des émotions simples et complexes ont été perçus comme étant plus âgés que les agents n'exprimant que des émotions de base.

2.2.3.9 Les rides d'expression et les rides permanentes

1. Chez l'être humain

En plus du rendu de la peau et de la dynamique du visage, d'autres informations visuelles sont fondamentales sur un visage humain. Par exemple, les expressions faciales engendrent des rides d'expression. Ces rides

apportent des informations expressives importantes et sont susceptibles d'augmenter significativement le réalisme du visage simulé.

En nous basant sur les distinctions proposées par Ekman et Friesen (1975), nous considérons séparément les rides rapides de l'expression émotionnelle (temporairement produites par l'activité des muscles du visage) et les rides dues à l'âge (rides permanentes qui émergent au cours de la vie).

Les propriétés de la peau sont impliquées dans les deux types de rides. La peau adhère étroitement aux tissus sous-jacents dans un certain nombre de régions du corps; ailleurs, elle glisse plus librement et, dans certains endroits, elle peut provoquer des rides (Wu et al., 1997). La forme triangulaire de la microstructure de la peau définit l'emplacement des sillons ou micro lignes, endroits auxquels la peau peut plisser et former des rides.

Au-delà des descriptions d'Ekman des rides dans l'expression des émotions de base (Ekman et Friesen 1975), les rides temporaires sont parfois mentionnées. Par exemple les pattes d'oie sont impliquées dans le sourire sincère de Duchenne (Ekman et al., 1990). Le manuel FACS (Ekman et al., 2002) présente une description détaillée des changements de texture (sillons, rides, etc.) par rapport aux Action Units, mais sans faire de lien avec les émotions.

Le rôle des rides permanentes est parfois discuté dans les travaux sur la relation entre les émotions et le vieillissement. L'expression neutre du visage de sujets âgés est perçue comme plus intense que celle des sujets plus jeunes. Cet effet serait dû aux rides permanentes et à des changements morphologiques du visage en fonction de l'âge. En outre, ces rides permanentes, qui restent visibles dans les expressions neutres, peuvent véhiculer des informations sur la personnalité (Malatesta et al, 1987). Par exemple, la colère en tant que trait de personnalité tend à laisser une empreinte permanente sur le visage. Suite à ces résultats, nous pouvons émettre l'hypothèse que la simulation des rides d'expression devrait augmenter l'expressivité du visage.

2. Application aux agents virtuels

Plusieurs techniques ont été proposées pour simuler cet effet sur un visage virtuel. Mais la problématique des rides n'est pas spécifique à l'animation faciale. Ces recherches s'appliquent aussi à simulation de vêtements et autres tissus déformables, qui répondent aux mêmes principes de tissus viscoélastiques. En effet, la simulation physique de tissu se base sur un ensemble de « couches » d'éléments solides ponctuels. Cet ensemble d'éléments est liés par des forces élastiques. Si le nombre de couches et les caractéristiques élastiques des forces varient entre un système de simulation de vêtements et un système de simulation de peau, le principe du modèle physique reste le même.

La génération de rides sur un tissu peut être divisée en deux approches : 1) l'approche paramétrique, et 2) l'approche générative. L'approche paramétrique nécessite de prédéfinir les rides en les éditant manuellement ou bien en les capturant (à partir de photos par exemple). Ces rides prédéfinies sont alors déclenchées localement lors de l'animation (Hadap et al., 1999). Afin de déterminer en temps réel quelles rides doivent être rendues visibles, plusieurs approches existent. Larboulette et Cani (2004) proposent un algorithme de détection de compression appliqué au maillage (la structure géométrique 3D) pour déclencher progressivement les rides en certaines zones. Les rides apparaissent automatiquement et progressivement au fur et à mesure que le maillage est comprimé.

L'approche générative consiste à générer les rides dynamiquement. Par exemple, en utilisant une simulation physique du tissu. Cette approche ne nécessite donc pas de rides prédéfinies. Cette approche est généralement beaucoup plus complexe et utilise plus de temps de calcul. Toutefois, les rides qui en résultent sont générées automatiquement, sans aucune édition manuelle préalable (Bridson et al. 2003). Certains modèles physiques ont été développés spécifiquement pour la génération de rides (Wu et al., 1995) à partir de la modélisation du squelette, des propriétés physiques des muscles et de la peau. D'autres modèles génèrent les rides sur la surface comme un effet secondaire. Par exemple, un système d'animation faciale par simulation physique fera apparaitre les rides dynamiquement de par sa simulation de structure de la peau.

Comme les modèles génératifs requièrent plus de temps de calcul, certains d'entre eux utilisent le GPU pour générer dynamiquement des rides (Loviscach, 2006) par une méthode similaire à celle de Larboulette et Cani (2004).. Cependant, cette approche utilise une grande quantité de la capacité de GPU, que les systèmes d'animation faciale utilisent habituellement pour le rendu et l'animation. Ainsi, cette technique ne peut être combinée avec un rendu réaliste du visage, par manque de ressources computationnelles GPU (sans considérer le cas spécial des machines multi-GPU).

Quelle que soit l'approche utilisée, il faut différencier le calcul des rides (forme et position) de leur affichage. Calculer qu'une zone est compressée, et qu'elle doit donc être ridée, n'est qu'une première étape. Il faut ensuite déterminer comment ces rides seront effectivement affichées sur le visage virtuel.

L'utilisation de la technique du *bump-mapping* sur GPU est aujourd'hui largement utilisée pour le rendu des rides sur les visages virtuels. Cette technique est relativement récente, puisqu'elle repose sur les nouvelles générations de cartes graphiques programmables. Le *bump-mapping* présente l'avantage de pouvoir afficher des rides fines sur un maillage de basse résolution. De plus, les rides prédéfinies sont enregistrées sous forme d'image, ce qui facilite leur édition. Cependant, le *bump-mapping* ne modifie pas la structure du maillage, et donc ne modifie pas le profil de la surface. Il ne crée qu'un effet visuel local, destiné à être vu de face. Finalement, plusieurs zones de rides peuvent être facilement définies en utilisant plusieurs images (Dutrève et al, 2009). La qualité de rendu des rides dépend donc de la qualité des images qui contiennent les rides prédéfinies. Dutrève et al (2011) ont proposé une méthode pour effectuer du retargeting des rides d'expression. En identifiant manuellement les zones de rides d'expression sur une photo, l'algorithme proposé extrait les rides en niveau de gris, génère dynamiquement les cartes de rides (sous forme de *normal-map* nécessaires au *bump-mapping*). Une fois générées, ces cartes de rides sont appliquées en temps réel sur le visage virtuel animé. Ainsi, le système proposé par Dutrève et al (2011) effectue le retargeting à la fois des rides d'expressions et de l'animation faciale. Une approche similaire a été utilisée par De Melo et al. (2009). A partir de photographie de rides, les cartes de rides sont manuellement extraites, puis utilisées en temps réel pour créer des rides d'expression virtuelles.

2.2.3.10 L'évaluation des rides d'expressions

Bien que plusieurs méthodes existent pour créer des animations faciales et des rides d'expression, l'évaluation de ces méthodes est souvent limitée à des critères techniques (par exemple : vitesse de rendu, complexité algorithmique, ou la notion floue de «réalisme»). Dans un contexte dans lequel des visages virtuels sont utilisés pour véhiculer un contenu affectif, il est nécessaire d'évaluer perceptivement l'apport des rides d'expression.

Wallraven et al. (2005) ont mené une expérience en utilisant des expressions actées. Ils ont comparé un rendu de visage réaliste et avec rides d'expressions et un rendu flouté, dans lequel les rides n'étaient plus visibles. Leurs principaux résultats sont les suivants: 1) le flou du visage augmente le temps nécessaire pour reconnaître les expressions faciales, et 2) réduire l'amplitude des expressions du visage réduit l'intensité perçue de l'expression.

Une étude récente utilisant des images statiques (De Melo et al., 2009) a observé que la combinaison de signes multiple (rides, rougissement et larmes) augmente l'intensité perçue d'émotions et de l'expressivité perçue de l'agent. Toutefois, la reconnaissance de la catégorie d'émotion n'a pas été abordée dans cette étude. Dans le protocole expérimental utilisé par les auteurs, la catégorie émotionnelle était en effet donnée aux participants. De plus, les trois axes de rendu (rides, rougissement et larmes) ne varient pas séparément. Il n'est donc pas possible de discuter de leurs impacts respectifs.

Pour résumer, peu d'agents virtuels expressifs réalistes intègrent des mécanismes de génération de rides. La plupart des agents interactifs temps-réel utilisent la norme MPEG-4 qui ne permet pas explicitement la génération de rides. En conséquence, aucune étude détaillée n'a été menée pour évaluer l'impact spécifique des différentes caractéristiques des rides, telles que leur réalisme visuel et leur dynamique.

2.3 Humains virtuels expressifs

Les technologies utilisées pour l'affichage de personnages virtuels sont en perpétuelle évolution, mais pour rendre ces personnages crédibles, un travail non graphique est également nécessaire. En effet, pour qu'un agent virtuel soit complet, il doit savoir générer un comportement adapté à la situation d'interaction, dynamique, et réactif. Pour cela, plusieurs modèles ont été proposés couvrant les expressions faciales, mais aussi d'autres modalités, telles que la parole ou les gestes.

2.3.1 Architectures et modèles de comportements affectifs

2.3.1.1 Génération comportementale multimodale : Le système BEAT

Le système BEAT (Cassell et al. 2001b) (*Behavior Expression Animation Toolkit*) utilise une entrée textuelle que l'on souhaite faire prononcer par un agent virtuel. Il génère des paramètres expressifs non verbaux cohérents avec le contenu sémantique du texte.

Le modèle présenté consiste en quatre étapes. La première étape consiste à analyser, et annoter automatiquement le texte donné en entrée. L'annotation est réalisée par analyse linguistique, sémantique, et en utilisant les informations contextuelles disponibles. Cette phase va par exemple permettre d'identifier les mots importants à mettre en relief. La seconde phase consiste à identifier, et suggérer un ensemble de comportements cohérents, pouvant correspondre à l'annotation du texte faite dans la phase. La troisième phase consiste alors à sélectionner, parmi les comportements proposés en phase 2, un ensemble de comportements qui seront effectivement joués par le module d'animation. En effet, certains comportements peuvent être conflictuels, c'est-à-dire que deux (ou plus) des comportements proposés ne peuvent pas avoir lieu en même temps. Il est donc important de vérifier la cohérence globale des comportements proposés. Finalement, une fois les comportements sélectionnés, la quatrième phase consiste à planifier les animations et les synchroniser pour qu'elles soient affichées par le personnage virtuel.

Le système BEAT met en relief la nécessité de séparer les différents niveaux logiques sous-tendant l'expression. Ici, la sémantique est séparée des comportements, eux même séparés de la réalisation. Cette approche modulaire permet de raisonner à plusieurs niveaux, et ainsi de mieux contrôler les différentes étapes.

2.3.1.2 Génération comportementale incrémentale

Kopp et al. (2007) ont proposé un système de génération incrémentale de comportement d'écoute. L'agent MAX est ainsi capable de montrer des comportements cohérents lorsqu'il « écoute » ce que dit l'utilisateur. L'agent possède plusieurs états internes, comme par exemple « en recherche de contact », « à l'écoute », « compréhensif » (manifeste qu'il comprend ce que l'utilisateur explique), ou « poli ».

Chacun de ses états internes sont associés à une valeur continue entre 0 et 1 qui représente leur activation. Plusieurs états peuvent donc être actifs en même temps. Les changements de valeur des états internes se font progressivement en fonction des entrées du système. C'est l'aspect incrémental : le système maintient un état interne que les entrées viennent altérer progressivement.

Pour chaque état un certain nombre de comportements peuvent être déclenchés. Ces comportements sont décidés par un système de règles de décision, associé à un système probabilistique. Ces deux systèmes vont permettre de décider quels comportements doivent être activés.

Les comportements décidés sont ensuite envoyés vers un module d'animation chargé de les animer de manière multimodale. L'intérêt majeur de cette approche incrémentale est d'éviter que deux comportements successifs puissent être contradictoires. En effets, l'évolution progressive d'un état interne stable permet d'éviter de générer de telle situation.

2.3.1.3 L'initiative SAIBA

SAIBA (Vilhjalmsson et al. 2007, Bevacqua et al, 2008) est une initiative de recherche internationale visant à spécifier une architecture logicielle modulaire pour la planification et la génération de comportements d'agents virtuels. L'architecture SAIBA est composée de trois modules, mis en relation par deux types de protocoles, basés sur des langages XML (Figure 35).

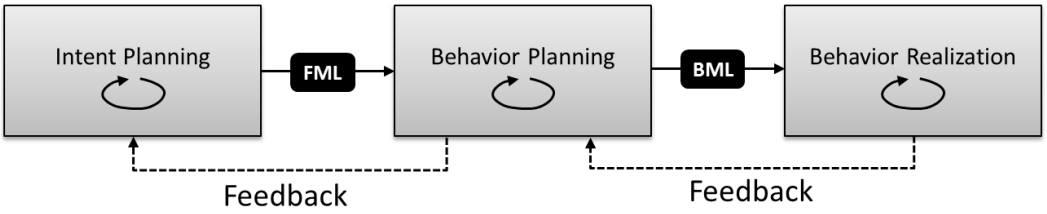


Figure 35 - Architecture SAIBA

Le premier module est appelé *Intent Planner*. C'est dans ce module que sont définies les intentions communicatives de l'agent, par exemple, l'émotion qu'il doit exprimer. Ce module haut niveau est généralement dépendant de l'application interactive dans laquelle évolue l'agent. Les intentions communicatives générées par ce module sont exprimées en langage FML (*Function Markup Language*) et transmises au second module. Plusieurs types de langages FML existent. Par exemple, le langage EmotionML est spécifiquement dédié aux émotions.

Le second module est appelé *Behavior Planner*. Son but est de convertir les intentions communicatives, reçues en FML, en comportements multimodaux (par exemple : gestes, expressions faciales, parole, locomotion). Ces modalités sont donc verbales et non verbales. Une fois les comportements définis, ils sont exprimés en langage BML (*Behavior Markup Language*), et transmis au dernier module.

Le troisième et dernier module, nommé *Behavior Realizer* est chargé de calculer l'animation exprimant l'intention communicative définie par le premier module, en utilisant les paramètres d'animation définis par le second module. Ce troisième module est généralement un moteur d'animation et de rendu d'agents virtuels (2D ou 3D) ou de gestion robotique dans le cas de robot expressifs.

La force de l'architecture SAIBA réside dans sa généricité et dans l'interopérabilité. En théorie, changer de logiciel d'agent virtuel ne nécessite pas de changement dans les modules de haut niveau. Dans la pratique, les plateformes d'animation acceptant le format BML n'implémentent qu'une sous partie du format. Ainsi, la généricité n'est que théorique, car une partie de l'intension communicative sera potentiellement ignorée par le *behavior realizer* lors de l'animation. Pour gérer ce genre de cas, l'architecture SAIBA prévoit un système de feedback entre les différents modules, par exemple pour que le *Behavior Realizer* puisse informer le *Behavior Planner* qu'il ne supporte pas un certain type de comportement. Néanmoins, l'objectif de SAIBA est d'uniformiser les langages entre les différentes plateformes d'agents expressifs, et de permettre de partager plus rapidement des ressources et d'interconnecter les systèmes.

2.3.1.4 Les modèles d'interaction

Afin de concevoir un modèle d'interaction complet, contenant notamment une boucle d'interaction affective, il est possible de s'inspirer de l'interaction humaine en la formalisant.

Thòrisson (1999) a proposé le modèle YMIR, formalisation à trois niveaux. Le modèle Ymir est un modèle génératif de dialogue non spécifique aux émotions qui relie la perception multimodale, la décision et l'action multimodale dans un cadre cohérent. YMIR est un modèle modulaire qui peut être utilisé pour créer des

personnages autonomes, dotés d'une perception multimodale et de la génération d'action. Les trois niveaux du modèle YMIR (Figure 36) sont :

1) Le niveau réactif (Figure 36, niveau inférieur). Ce niveau sert à caractériser les événements de l'interaction. Par exemple, le locuteur tend un bras pour montrer un objet. Le rôle du niveau réactif est de caractériser le geste, sans lui donner de sens. Les informations factuelles ainsi générées par le niveau réactif serviront de données d'entrée aux niveaux supérieurs.

2) Le niveau de contrôle (Figure 36, niveau intermédiaire). Ce niveau a pour but de créer des blocs de données structurées. Par exemple, il gère la reconstruction des structures de phrase à partir des mots reconnus par le niveau réactif, et fait le lien avec les gestes associés. Par exemple, si l'utilisateur montre un objet et demande « Quel est cet objet ? », le niveau de contrôle doit associer ces deux événements.

3) Le niveau sémantique (Figure 36, niveau supérieur). Ce niveau reconstruit le sens des différentes commandes données par l'utilisateur et se charge de générer les réponses appropriées. Ce niveau repose sur les informations structurées, générées par le niveau de contrôle.

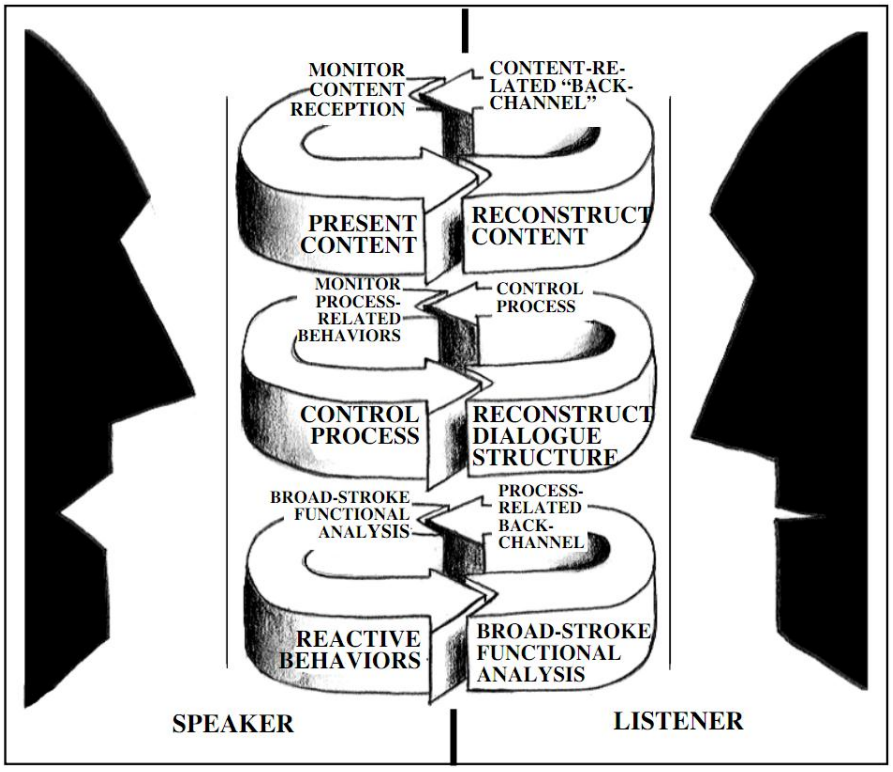


Figure 36 - Formalisation de l'interaction de dialogue du modèle YMIR (Thòrisson, 1999)

2.3.2 Agents virtuels exprimant des émotions par expressions faciales

La plupart des systèmes d'animation faciale estiment que les expressions du visage apparaissent comme des mélanges d'expressions faciales prototypiques, correspondant aux émotions générées par le système émotionnel. Toutefois, l'humain est capable de simuler ou de masquer une émotion afin de respecter certaines normes sociales liées au contexte. Par exemple, il est préférable d'exprimer de la joie en face de journaliste pour ne pas montrer sa déception suite à un échec à des élections (Abrillian et al. 2005). Pour simuler ce type de comportement, Niewiadomski et al. (2010) ont proposé un modèle permettant d'inclure des règles de masquage et de superposition d'expressions faciales à l'agent Greta. Ainsi l'agent serait capable d'afficher un

comportement social, et pourrait s'adapter à différents rôles (ex : agent Tuteur dominant, agent assistant conciliant, agent commercial poli, etc.).

De plus, les travaux sur l'analyse vidéo suggèrent que les expressions émotionnelles ne sont pas des expressions prototypiques, mais plutôt des séquences dynamiques de signaux apparaissant dans un ordre spécifique (Bänzinger et Scherer, 2007). Niewiadomski et al. (2009) ont souligné que ces signaux présentent également des contraintes temporelles. Par exemple, certains signaux apparaissent toujours suivis par un autre. En se basant sur une approche corpus, Niewiadomski et al. (2009) ont conçu des animations faciales pour l'agent virtuel Greta basé sur des séquences d'expressions faciales.

L'agent Alfred (Bee et al. 2009) est un agent permettant le contrôle temps réel par l'utilisateur, et possédant un niveau de détails géométrique élevé et un système de rides d'expressions appliquées par bump mapping. L'animation de l'agent Alfred est basée sur le principe des *blendshapes*. Une *blendshape* a été définie pour chaque Action Unit (Ekman et Friesen, 2002). L'animation d'Alfred est donc basée sur le système FACS, ce qui lui permet une grande variété d'expressions faciales. Cependant aucun modèle émotionnel n'est utilisé, ainsi, l'animation d'Alfred nécessite de manipuler un nombre important de paramètres.

L'agent MAX, (Boukricha et al., 2007) permet l'affichage d'expressions faciales d'émotions de base et d'autres états affectifs, basé sur l'architecture WASABI, un modèle computationnel de raisonnement cognitif. Ce modèle mélange l'approche cognitive des émotions avec l'approche dimensionnelle. MAX est donc capable d'exprimer, via ses expressions faciales, des émotions complexes issues d'un processus émotionnel dynamique.

Un système d'animation faciale utilisant des travaux Scherer (2001) a été mis en œuvre par Paleari et Lisetti (2006). Le système affiche des expressions temporaires lors des différentes phases de l'évaluation cognitive. Malatesta et al. (2007) ont également abordé la dynamique des expressions faciales en se basant sur la théorie cognitive de Scherer (2001) en utilisant l'agent Greta (Niewiadomski et al., 2009) et ont proposé deux modes d'animation. Dans une première approche, ils effectuent une accumulation successive des expressions des différentes évaluations cognitives : chaque expression est ajoutée à l'expression du visage actuel. Dans une seconde approche chaque expression remplace les expressions faciales des évaluations cognitives précédentes. Chaque expression disparaît lorsque l'expression de l'évaluation suivante apparait. Ils ont comparé ces deux approches dans une étude perceptive exploratoire. Les sujets de l'étude devaient indiquer quelle émotion ils avaient reconnu dans chacune des animations. Selon cette étude, la méthode additive montrerait des taux de reconnaissance supérieurs au hasard, alors que la méthode séquentielle donnerait des taux de reconnaissance moins élevés, très légèrement supérieurs au hasard. Toutefois, cette première étude met en relief certaines limites de la méthode additive. Par exemple, «*dans le cas de la colère, selon les paramètres des tables de prédiction d'expression faciale de la théorie du CPM (Scherer, 2001), l'expression de l'évaluation « événement nouveau » inclut un haussement des sourcils. Mais, l'évaluation suivante (rapport aux buts) induit le froncement des sourcils, ce qui entre en conflit avec le haussement de sourcils précédent. Ce conflit rend problématique l'animation additive. Le résultat de l'animation séquentielle cumulative correspondante est source de confusion* » (Malatesta et al., 2007).

La conception d'un système d'évaluation des événements affectifs au cours d'une interaction temps réel avec un utilisateur reste un défi. De plus, en dehors de la génération et de l'évaluation des événements, la mise en place d'une application temps réel implique de gérer des séquences dynamiques de l'émotion, c'est à dire la façon dont les différentes évaluations cognitives de plusieurs événements séquentiels se combinent dans le temps (Marsella et al., 2010).

Les systèmes ne considèrent que des animations faciales exprimant l'évaluation cognitive en dehors de tout contexte d'interaction. Les sujets visualisent des animations et doivent ensuite répondre à des questions évaluant leur perception. Aucun de ces systèmes ne fait vraiment d'évaluation des événements survenus dans le contexte dynamique d'une application interactive.

2.3.3 Evaluation de la perception humaine des agents animés

Comme l'ont soutenu Deng et Ma (2008), la perception humaine est l'un des outils les plus efficaces pour mesurer l'expressivité d'un agent virtuel. Différentes applications interactives utilisant des modèles informatiques émotionnels ont été conçues pour créer des installations expérimentales. Ces expérimentations fournissent des informations pour la conception des agents virtuels. Par exemple, certaines études montrent que l'affichage d'émotions complexes est susceptible d'altérer la perception des utilisateurs. Ainsi Becker-Asano et al. (2008) ont observé que les agents exprimant des émotions complexes ont été perçus comme étant plus âgés que les agents n'exprimant que des émotions de base.

Wallraven et al (2005) proposent une modélisation de l'influence réciproque entre les agents virtuels et la perception humaine. Dans ce modèle, les agents virtuels sont utilisés pour étudier la perception humaine, et réciproquement, la perception humaine est utilisée pour étudier et améliorer les modèles et les rendus d'agents virtuels.

Généralement, les études perceptives utilisent des questionnaires post-expérimentaux pour recueillir le sentiment subjectif et a postériori des utilisateurs. Prendinger et al. (2006) ont proposé une méthode plus directe pour évaluer les agents virtuels. L'objectif de leur expérimentation est d'évaluer l'utilisation de données physiologiques dans l'évaluation des agents virtuels, et plus particulièrement la *conductivité de la peau* (mesure d'activation émotionnelle) du sujet et son *électromyographie* (mesure de valence émotionnelle). Dans cette étude, l'agent MAX, doté du système émotionnel WASABI, a été présenté dans une application de jeu de carte interactif : Le Skip-Bo. Dans cette installation, le joueur est face à l'agent, qui est son adversaire. La variable manipulée est le module comportemental et émotionnel de MAX, qui varie entre plusieurs conditions. Les données physiologiques recueillies mettent en évidence des différences physiologiques entre les mesures des différentes conditions expérimentales, ce qui semble valider que les capteurs physiologiques peuvent être utilisés pour évaluer les agents virtuels. Cependant, l'utilisation de ce type de dispositif implique la mise en place de capteurs intrusifs et nécessite des matériels très spécifiques et l'expertise requise pour les mettre en œuvre.

Les études menées sur les agents virtuels ont également mis en lumière des phénomènes perceptifs particuliers, comme l'*uncanny valley*, ou le *personna effect*.

2.3.3.1 L'uncanny valley

Lors de la conception d'agent virtuel interactif, il est difficile de faire correspondre le niveau de réalisme de l'apparence du personnage virtuel, de la qualité de son animation, de son interactivité, et de son comportement. Si l'un de ces aspects est moins abouti que les autres (MacDorman et al., 2009), cela résulte dans un phénomène perceptif connu sous le nom d'*Uncanny Valley*.

L'*uncanny valley* a été défini par Mori (1970) comme un phénomène de répulsion à l'égard des humanoïdes, robotiques ou virtuels, lorsqu'ils atteignent un niveau de réalisme très élevé sans pour autant être parfaitement similaire à l'humain. Ce phénomène intervient en particulier lorsque les rendus visuels et comportementaux d'un agent n'ont pas le même niveau de finesse/réalisme (Walters et al, 2008).

En effet, le degré de réalisme visuel de l'agent doit correspondre à son degré de réalisme comportemental (Vinayagamoorthy et al., 2005). Si le rendu visuel statique est supérieur en qualité à la dynamique des mouvements de l'agent, alors on obtient une impression négative. L'*Uncanny Valley* affecte également d'autres modalités que la vision. En effet, il a été observé (Mitchell et al., 2011) qu'un trop grand écart de qualité entre la voix et l'aspect visuel provoque également un effet perceptif négatif.

Cet effet concerne également le comportement de l'agent en général. En effet, une apparence très humaine provoque des attentes importantes sur les capacités de raisonnement de l'agent et sur son comportement social. Ces attentes seront moins élevées si l'agent a un aspect plus robotique (Goetz et al., 2003).

La Figure 37 montre la représentation graphique de l'évolution de la perception qu'un humain a d'un humanoïde en fonction de sa similarité avec l'humain. On distingue la vallée qui apparait entre le robot humanoïde et l'humain. Cette figure met également en relief l'importance de la dynamique, puisque l'*uncanny valley* est plus importante pour les humanoïdes animés que pour les humanoïdes statiques.

L'*uncanny valley* est donc un phénomène perceptif complexe, mais fourni des recommandations importantes pour la conception d'agents virtuels et robotiques. C'est à cause de ce phénomène que les jeux vidéo et certains films d'animation ont tendance à ne pas respecter les proportions humaines du visage et du corps des personnages virtuels. Ainsi, en diminuant le réalisme, les designers graphiques améliorent la perception des personnages virtuels.

Cependant, peu d'études expérimentales évaluent l'impact de la qualité de rendu visuel des agents expressifs sur la perception des utilisateurs, car le réalisme des agents est généralement limité (Brenton et al., 2005). Les rides d'expression, les simulations complexes de la structure peau, la simulation physiologique du visage, sont autant de phénomènes dont la contribution à éviter l'*uncanny valley* n'a pas été évaluée.

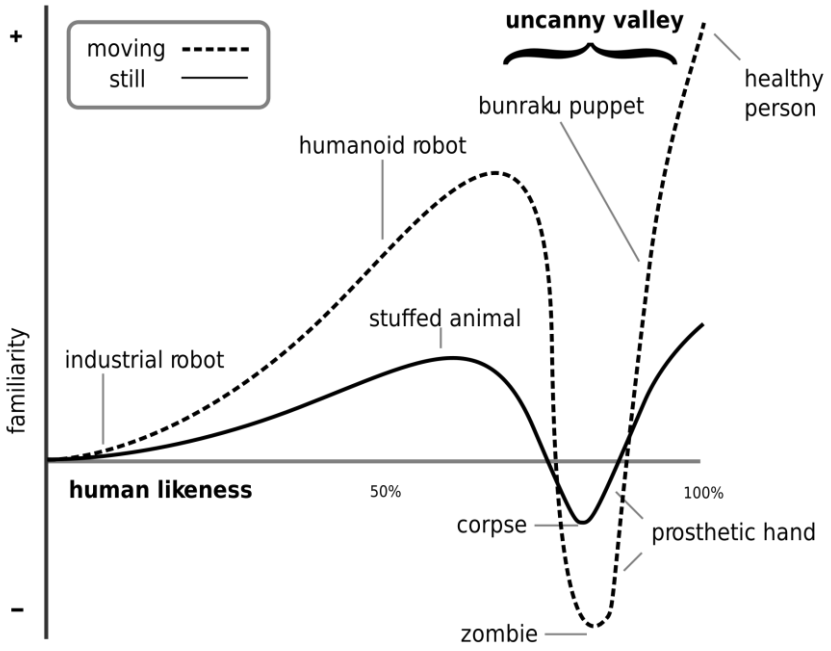


Figure 37 - Représentation Graphique de l'uncanny valley (Mori, 1970)

2.3.3.2 Le persona effect

Le *persona effect*, se rapporte au fait que *la seule présence d'un agent virtuel (y compris non expressif) dans une application interactive d'e-learning a un impact positif sur la perception que les étudiants ont de leur expérience d'apprentissage* (Lester et al. 1997).

Cet impact positif a été cependant nuancé par divers travaux. Moundridou et Virvou (2002) ont montré que si l'expérience subjective des élèves est améliorée, la présence d'un agent n'améliore pas les performances d'apprentissage. Dehn et Van Mulken (2000) soutiennent même que dans certaines applications, la présence d'un agent peut être une distraction et gêner l'apprentissage, ainsi l'agent aurait une influence négative sur l'objectif premier : apprendre.

2.3.4 Dispositifs de contrôle temps réel d'agents virtuels

Dans certaines applications, telles que les réseaux sociaux ou les mondes virtuels, il peut être nécessaire de permettre à l'utilisateur de contrôler directement les expressions d'un agent virtuel, qui sert dans ce cas d'avatar à l'utilisateur. Pour permettre un tel contrôle, l'une des approches consiste à mettre en place un ensemble de curseurs glissants pour chaque paramètre ou émotion disponible. Cependant, une telle interface est rapidement limitée. D'une part, elle ne permet de contrôler qu'un paramètre à la fois, et d'autre part, la correspondance entre un curseur et son paramètre d'animation est difficile à appréhender (Bee et al., 2009).

Différents dispositifs ont donc été testés pour permettre une manipulation plus intuitive et un contrôle plus fin de l'expressivité de l'agent virtuel. Le dispositif Pogany[4] (Jacquemin, 2007) est un dispositif anthropomorphique en forme de visage inspiré de l'œuvre de Constantin Brancusi intitulée « Mademoiselle Pogany ». Le dispositif d'interaction est équipé d'une caméra interne qui capte les variations lumineuses à travers de petits trous percés sur le « visage » du dispositif. Ainsi, le contrôle du visage virtuel se fait par manipulation directe du dispositif physique. Par exemple, en touchant les sourcils du visage du dispositif, on obscurcit la zone correspondante dans l'image de la camera interne, ce qui déclenche l'animation des sourcils du visage virtuel (Figure 38).

Figure 38 - L'interface Pogany (Jacquemin, 2007) et un exemple d'utilisateur la manipulant.

D'autres dispositifs, non spécifiquement créés pour le contrôle d'expressions faciales, ont été étudiés. Bee et ses collaborateurs ont notamment exploré deux dispositifs innovants pour contrôler l'agent Alfred : les gants de données et les contrôleurs de jeux vidéo (Figure 39).

Figure 39 - A gauche, le gant de données, à droite un contrôleur de Xbox 360. (Images extraites de Bee et al. 2009)

[4] http://perso.limsi.fr/jacquemi/Artsciedu/pogany.html

Les gants de données sont des dispositifs permettant de capter l'orientation de la main dans l'espace (3 rotations) et la flexion des cinq doigts. La manette de Xbox 360 possède quant à elle deux joysticks à deux dimensions, plus quinze boutons. L'enjeu pour l'utilisation de ces différents dispositifs est de trouver une correspondance intuitive avec les paramètres d'animation du visage virtuel. En effet, dans le cas de Pogany, l'interface anthropomorphique permet une correspondance physique entre le dispositif et le visage virtuel. Pour les dispositifs comme les contrôleurs de jeux vidéo, plusieurs correspondances sont possibles. Bee et ses collaborateurs ont donc mené différentes études perceptives avec ces dispositifs pour identifier les correspondances les plus intuitives. Dans leurs travaux, ils ont montré que ces dispositifs, comparés aux interfaces utilisant des curseurs glissants, permettent un gain de temps dans la création d'expressions faciales, sans réduire la qualité des expressions obtenues (Bee et al., 2009).

2.3.5 Applications des agents virtuels expressifs

Si les agents virtuels expressifs sont largement utilisés dans les travaux de recherche pour effectuer des études perceptives, peu d'applications interactives ont été mises en place. Charles et al. (2009) ont proposé une application interactive de *story telling* interactive inspirée de l'œuvre de Flaubert « Madame Bovary », et dans laquelle l'utilisateur prend part à un dialogue affectif avec un personnage virtuel. Cette application utilise un système de détection dans la voix de l'utilisateur (Cavazza et al., 2009), et exprime les émotions de l'agent virtuel par différente modalités (voix, visage, etc.). Cependant, dans un système de *story telling*, l'interaction est très scriptée, ce qui limite l'interactivité.

Niewiadomski et al. (2009) ont proposé un système interactif de dialogue dans lequel l'agent émet des rétroactions (ou *backchannels*) à partir du dialogue avec l'utilisateur. Dans cette application, les réactions de l'agent Greta sont limitées à un ensemble d'éléments communicatifs réduit, tels que l'accord (mouvement vertical de la tête), ou le désaccord (mouvement horizontal de la tête).

André et Rist (2000) ont proposé une application dans laquelle plusieurs agents sont présentés simultanément dans l'interface et dialoguent entre eux dans une application de vente de voiture. Ici, l'utilisateur assiste à l'interaction des agents sans y prendre part. L'application ne propose donc pas une interaction « avec les agents virtuels » au sens propre, mais explore l'utilisation d'un groupe d'agents virtuel pour présenter de l'information à l'utilisateur de différentes manières.

Ces applications interactives restent cependant limitées à une utilisation en laboratoire. Les applications grand public des agents virtuels expressifs sont en effet relativement rares.

Depuis 2004, l'agent virtuel MAX (Jörding et Wachsmuth, 1997, Kopp et Jung, 2000) est utilisé comme guide virtuel au Musée *Heinz Nixdorf MuseumForum* (Paderborn, Allemagne) dédié aux sciences informatiques. L'agent y est régulièrement mis à jour, ainsi que sa base de connaissance. Ces mises à jour régulières lui permettent de rester fonctionnel et attractif pour le public visitant le musée, en fonctionnant dix heures par jour, six jours par semaine, depuis sept ans. Scientifiquement, cette application grand public a permis à l'équipe de MAX d'obtenir une grande quantité de données sur la perception que le public a de l'agent. Il a été observé (Pfeiffer et al. 2011) que visiteurs adultes semblent relativement réservés vis-à-vis de l'agent. Les informations recueillies suggèrent qu'ils éprouvent de l'inquiétude sur l'évolution des machines et de l'intelligence artificielle. Au contraire, les visiteurs enfants approchent l'agent avec moins de retenue que les visiteurs adultes. Ils lui parlent en langage naturel, alors que le dialogue avec l'agent s'effectue via une console et un clavier. De plus, les jeunes visiteurs semblent avoir tendance à « tester » l'agent en l'insultant plus ou moins directement, ce que les auteurs interprètent comme un besoin de tester les limites du système émotionnel de l'agent.

Depuis 2009, le musée des sciences de Boston a également mis en place un système de guide virtuel (Swartout et al., 2010). Dans cette installation, deux agents féminins sont présentés, Ada et Grace (Figure 40), et cherchent à engager le public dans l'interaction. L'utilisation de deux personnages virtuels est, selon les auteurs, propice à l'engagement des utilisateurs. Elle en effet permet aux agents d'interagir entre eux et ainsi d'enrichir le dialogue. Cette application montre un certain nombre d'autres différences avec celle présentée en Allemagne. Le rendu

graphique des deux agents a été conçu en utilisant le système LightStage (Debevec et al., 2000) et un rendu temps réel tirant parti du moteur de jeu vidéo *Gamebryo*. Ainsi, le réalisme visuel des agents est de très haute qualité. L'interaction avec les agents se fait par langage naturel, et non par clavier et console. Ainsi, l'utilisateur est plus facilement engagé dans l'interaction. En revanche, contrairement à l'agent MAX, le système présenté ne possède pas de gestion des émotions. L'expressivité émotionnelle des deux agents est donc limitée. Pourtant, l'installation semble engager les utilisateurs avec succès dans l'interaction.

Figure 40 - Utilisatrice interagissant avec Ada et Grace au musée des sciences de Boston. (Swartout et al. 2010)

Ces applications montrent que les agents virtuels interactifs sont aujourd'hui acceptés par le grand public. Pourtant, ils souffrent encore de limitations. L'agent MAX par exemple, nécessite l'utilisation d'un clavier pour communiquer, son rendu visuel lui donne l'aspect d'un pantin articulé, et son système de dialogue souffre de nombreuses limitations. Par exemple, les phrases écrites par le sujet doivent être formulées avec une orthographe et une grammaire parfaite (Pfeiffer et al. 2011), ce qui ralenti la communication et provoque des incompréhensions. Cependant, l'exemple des insultes faites à MAX met également en relief que l'utilisation d'un modèle émotionnel crée des attentes importantes de la part des utilisateurs humains. Cela met en relief l'importance du modèle computationnel des émotions.

2.3.6 Robots exprimant des émotions par expressions faciales

Les personnages virtuels présentent l'inconvénient de ne pas avoir de réalité physique tangible. Dans certains cas, l'utilisation de robots expressifs peut donc être préférable. Plusieurs robots humanoïdes ont été développés. Comme nous allons le voir, ces robots présentent en revanche des limites importantes en termes d'expressivité.

L'iCat, développé par Philips©, est un robot en forme de chat humanisé (Figure 41) et doté de capacité expressives. Il possède 11 moteurs dont 9 dédiés aux expressions faciales. L'iCat ne possède donc que peu de degrés de liberté utilisables pour animer ses expressions, en comparaisons avec les agents virtuels dotés de plusieurs dizaines de degrés de liberté.

Figure 41 – L'iCat de Philips

Leite et al. (2009) utilisent l'iCat comme adversaire dans un jeu d'échec. Ces travaux font suite à une précédente version dans laquelle l'agent ne prenait pas en compte l'état émotionnel de l'utilisateur (Pereira et al., 2008b). Dans cette nouvelle version, le système *emotivector* (Leite et al. 2008) permet la prise compte de l'état émotionnel de l'utilisateur et l'utilisation d'une mémoire autobiographique pour d'influencer l'état émotionnel de l'agent représenté par l'iCat. Ainsi, l'iCat peut être doté de capacités empathiques (Leite et al. 2010). Dans cette étude, les auteurs montrent que l'iCat est plus perçu par les sujets comme un compagnon de jeu lorsqu'il présente un comportement empathique.

Figure 42 - Le robot expressif Geminoid F et le modèle humain ayant inspiré sa conception

La technologie Geminoid F. est actuellement le stade le plus avancé, en termes de réalisme, du robot humanoïde expressif (Figure 42). Cependant, d'un point de vue animation faciale, le robot souffre de larges limitations. En effet, son visage de comporte que 12 dégrées de liberté, dont 7 dédiés aux rotations des yeux et de la tête (Figure 43).

Becker-Asano et Ishiguro (2011) ont mené une étude pour comparer la perception des utilisateurs des expressions faciales du robot à celle de son modèle humain à l'aide d'images statiques présentées dans un questionnaire web. Cette étude a montré des taux de reconnaissance catégorielles élevés et équivalents entre les expressions de l'humain et celle du robot. Seules deux expressions n'ont pas été reconnues sur le robot : la Peur, confondue avec la Surprise, et la Colère, confondue avec la Tristesse. Si la littérature fourni divers exemples dans lesquelles la Peur et la Surprise sont confondues, la confusion entre Colère et Tristesse est plus inattendue. Ce résultat s'explique par l'absence de moteur pour simuler le froncement de sourcils. Ainsi l'un des indices

faciaux nécessaire à l'expression de Colère ne peut être répliqué par le robot. Ceci met en relief les sévères limitations des robots humanoïdes actuels.

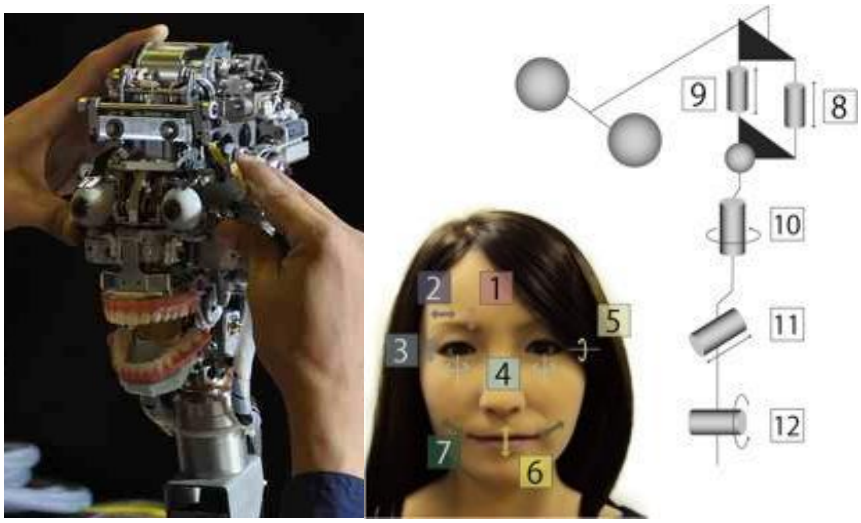


Figure 43 - Mécanismes d'animation faciale du robot Geminoid F.

Pereira et al. (2008b) ont également utilisé le robot iCat pour mener une étude comparant l'utilisation d'un robot à celle d'un agent virtuel 3D. Dans cette étude, l'iCat est à nouveau utilisé comme adversaire de l'utilisateur dans le cadre d'une partie d'échec. Le but de cette installation est d'évaluer l'impact de la présence physique du robot sur le plaisir qu'à l'utilisateur à jouer la partie d'échec. Les auteurs ont comparé l'utilisation du robot lui-même avec l'utilisation d'un iCat virtuel, affiché sur un écran d'ordinateur. Cette étude a montré que l'utilisation d'un robot, plutôt que d'un agent virtuel, augmente la sensation d'immersion et d'interaction sociale.

Figure 44 - Robot et agent virtuel utilisé comme démonstrateur de l'architecture CMION (Kriegel et al., 2011)

L'utilisation d'architectures génériques telle que SAIBA permet une convergence progressive des architectures logicielles utilisées par les systèmes d'agents virtuels et celles utilisées par les robots humanoïdes (Kriegel et al., 2011). En effet, si la manière de réaliser les animations (*behavior realizer*) est très différente, les modules de plus haut niveau, tels que les modules d'intention (*intent planner*) et de comportement (*behavior planner*)

peuvent être adaptés aux deux types de systèmes. Ainsi, la migration des systèmes d'agents virtuels vers des systèmes robotiques (et inversement) est facilitée (Leite et al. 2010). Certaines architectures logicielles telles que CMION (Kriegel et al., 2011), ont ainsi été proposés pour permettre à la fois l'utilisation de robots et d'agent virtuels expressifs. L'architecture CMION a été appliquée à un robot et au personnage virtuel ayant la même apparence (Figure 44).

Cependant, les agents virtuels permettent actuellement une meilleure finesse expressive que les robots. L'étude de Pereira et al. (2008b), ainsi que les travaux de Kriegel et al. (2011), utilisent en effet un agent virtuel extrêmement limité car contraint aux mêmes limites expressives que le robot de forme identique. Ainsi, les études comparatives entre robots et agents virtuels sont limitées, car il faudrait comparer l'utilisation du robot expressif à celle d'un agent expressif plus élaboré, tel que Greta ou Alfred.

2.4 Conclusions de l'état de l'art

Notre état de l'art couvre plusieurs domaines de recherches. Tous ces domaines sont importants pour comprendre les enjeux actuels des agents virtuels expressifs. D'une part, nous avons vu que les travaux en psychologie des émotions sont variés, et que plusieurs approches sont possibles pour définir ce qu'est une émotion. Divers modèles émotionnels ont ainsi été proposés, Ainsi, lorsqu'il s'agit de développer un modèle computationnel utilisable pour manipuler les émotions d'un agent virtuel et contrôler son expressivité, des choix doivent être effectués. C'est pour cette raison qu'il nous est nécessaire d'explorer plusieurs approches proposés en psychologie. Pour plusieurs de ces approches, nous proposerons donc un modèle computationnel et nous l'évaluerons. Ces différents modèles psychologiques n'ont pas les mêmes objectifs, ainsi il n'est pas toujours possible d'en effectuer des comparaisons deux à deux. Nous verrons cependant dans la suite de ce manuscrit les différents protocoles que nous avons proposé pour les comparer partiellement.

Dans le domaine de l'informatique affective, plusieurs modèles computationnels ont été proposés pour augmenter le réalisme comportemental des agents virtuels. Cependant, les agents virtuels utilisés sont généralement limités en termes de réalisme visuel, car ils utilisent des techniques de rendu et d'animation ne tirant pas parti du potentiel des matériels informatiques actuels.

En effet, les agents virtuels expressifs et interactifs peuvent également bénéficier des progrès récents de l'informatique graphique. Comme nous l'avons vu, dans le domaine de la 3D industrielle (en particulier dans l'industrie du cinéma et des jeux vidéo), les techniques d'animations, de rendu, et de capture sont en perpétuelle évolution, et offrent ainsi des niveaux de réalisme visuel de plus en plus fins. L'un des objectifs de ces travaux est d'augmenter l'immersion des joueurs en proposant un rendu visuel toujours plus réaliste. La flexibilité, c'est-à-dire la capacité à générer dynamiquement des animations adaptées à la situation, est fondamentale dans la recherche sur les agents virtuels expressifs. Elle n'est en revanche pas une priorité pour les sociétés de développement de jeux vidéo, et encore moins pour le cinéma. Certains jeux (ex : le jeu L.A. Noire, Rockstar, 2011) prennent donc le parti de pré-scripter l'intégralité des dialogues et animations, et ainsi obtenir des expressions faciales parfaitement contrôlées, puisque qu'enregistrées intégralement. Cette approche, en plus d'être très couteuse en quantité de données, n'est pas applicable à la génération d'animations en temps réel durant une interaction dynamique entre un utilisateur et un agent virtuel expressif. Néanmoins, certaines des avancées techniques apportées par ces travaux peuvent bénéficier aux agents virtuels. Il convient donc de les étudier pour évaluer leur apport à l'interaction affective dynamique.

En résumé, afin de déterminer comment améliorer la qualité de l'interaction temps réel entre l'utilisateur et l'agent virtuel, il est nécessaire de modéliser, implémenter et évaluer perceptivement plusieurs modèles informatiques des émotions, ainsi que différentes techniques récentes pour le rendu graphique et l'animation faciale.

PARTIE I

De l'approche Catégorielle à l'approche Dimensionnelle

Chapitre 3. Approche catégorielle des émotions pour l'animation faciale

Sommaire du chapitre

Publications associées

- M. Courgeon, J-C. Martin, C. Jacquemin, (2008) *MARC: a Multimodal Affective and Reactive Character* In: Proceedings of the ICMI workshop on Affective Computing (AFFINE), Chania, Greece, 20-22 octobre 2008
- M. Courgeon, J-C. Martin, C. Jacquemin, (2008) *MARC : Un Personnage Virtuel Réactif Expressif*, In: Proceedings of the 2nd Workshop on Animated Conversational Agents, Paris, France, 25-26 novembre 2010
- M. Courgeon, S. Buisine, J-C. Martin (2009) *Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions*, In: Proceedings of the 9th International Conference on Intelligent Virtual Agents, pp 201-214, Amsterdam, The Netherlands, 10-12 septembre 2009
- M. Courgeon, M-A. Amorim, C. Giroux, J-C. Martin (2010), *Do Users Anticipate Emotion Dynamics in Facial Expressions of a Virtual Character?,* in: Proceedings of the 23rd International Conference on Computer Animation and Social Agents, Saint Malo, France, 31mai-2juin 2010
- M. Courgeon, C. Clavel, N. Tan, J.C. Martin, (2011) *Front View vs. Side View of Facial and Postural Expressions of Emotions in a Virtual Character*, In: LNCS Transactions on Edutainment, 6, pp 132-143
- M. Courgeon, S. Buisine, J-C. Martin (2011) *Expressive Wrinkles in a Virtual Character's Facial Expressions: Impact on Emotion Recognition, Perceived Expressiveness and Users' Preferences*, (Soumis à la revue IEEE Transactions on Affective Computing)

3.1 Objectifs

Ce chapitre présente la première architecture de la plateforme MARC (*Multimodal Affective and Reactive Character*) que nous avons conçue et développée durant cette thèse. La première version de MARC est inspirée de l'approche catégorielle des émotions. En effet, l'utilisation de l'approche catégorielle est fréquente dans les agents virtuels. Cependant, notre objectif est de combiner cette approche avec un rendu graphique détaillé, pour créer un agent virtuel temps réel, réaliste, et interactif. Ainsi, le choix d'utiliser l'approche catégorielle nous permet de nous appuyer sur les travaux effectués sur les agents virtuels, tout en intégrant des problématiques différentes en termes de perception et de réalisme.

Cette première version de l'architecture pose les principes de base de MARC, tels que l'objectif d'obtenir un personnage « réaliste » et temps-réel, permettant d'exprimer des mélanges d'émotions, et sa modularité (séparation entre « édition offline » et « rendu temps réel », ainsi qu'entre le module d'animation et les modules émotionnels).

Cette version de la plateforme nous a servi de point de départ pour établir plus précisément les objectifs de nos travaux, et ainsi, permettre le développement des versions successives présentées dans la suite de ce manuscrit.

La seconde partie du chapitre présente les différentes études que nous avons réalisées avec cette première version de notre plateforme.

3.2 Architecture logicielle de MARC v1 (approche catégorielle)

Dans la première version de MARC, la plateforme se compose de trois modules. La Figure 45 montre l'architecture globale de la plateforme. Les trois modules sont les suivants : 1) Le logiciel d'animation 3D temps réel interactif, qui réalise l'animation du visage expressif lors de l'interaction avec l'utilisateur. 2) L'éditeur d'expressions faciales « offline », qui permet aux concepteurs/graphistes de créer et de modifier la bibliothèque d'expressions faciales utilisée par le module d'animation temps réel. 3) Le module émotionel catégoriel.

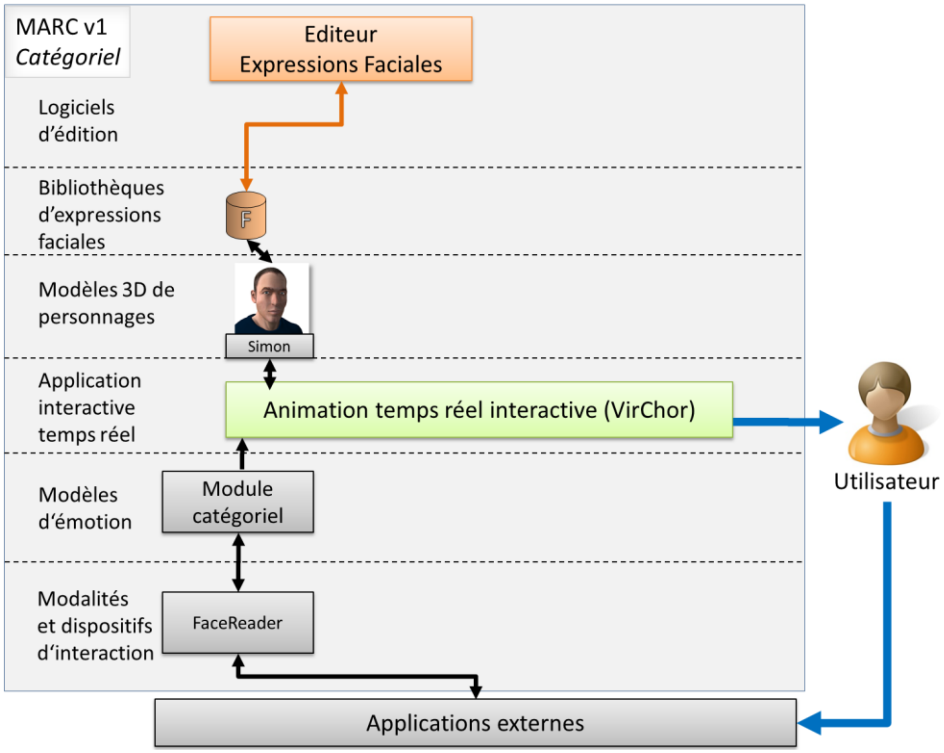


Figure 45 - Architecture de MARC V1 : Modèle Catégoriel

Les modules liés à l'animation (1 et 2) utilisent et éditent une base de donnée extensible d'expressions faciales prédéfinies. Le module de rendu 3D temps-réel interactif s'intègre dans une boucle interactive avec l'utilisateur, les dispositifs d'interaction, et les applications externes. Ces applications externes peuvent piloter l'expressivité du personnage directement par des commandes réseaux, en utilisant le module d'émotion catégoriel, qui permet de générer des expressions faciales en se basant sur les expressions de la bibliothèque.

3.2.1 Développements logiciels : modèle émotionnel catégoriel

Dans le module émotionnel catégoriel, les émotions sont représentées comme des états émotionnels discrets, associés chacun à un label émotionnel (ex : Joie, Peur, Colère, etc.). Le module émotionnel que nous avons implémenté permet donc la création d'un ensemble d'émotions « disponibles » pour l'animation, lesquelles sont associées à un nom (label), et à une seule expression faciale. Les expressions faciales sont définies selon la norme MPEG-4 (Cf. section 3.2.2.1). Chaque expression est donc codée sous forme d'une liste de déplacements de points-clés.

Notre modèle informatique catégoriel permet également de mélanger les expressions faciales de plusieurs émotions. Pour cela, le module émotionnel catégoriel mélange les expressions faciales en utilisant un coefficient indépendant pour chaque émotion.

Chaque émotion peut être activée de manière indépendante. Afin de garantir une évolution continue des émotions, et ainsi éviter des « ruptures » visuelles dans l'animation résultante, nous avons implémenté un système de gestion de la dynamique de type *onset-apex-offset* (décrit dans l'état de l'art). L'activation d'une émotion requière donc quatre paramètres : 1) L'intensité cible (apex), 2) la durée totale de l'émotion, 3) la durée de l'interpolation incrémentale (*onset*) 4) la durée de l'interpolation en sortie d'animation (*offset*).

Dans le cas où une seconde activation de la même émotion est déclenchée avant la fin de la durée de la première activation, la première activation est interrompue. L'intensité courante est alors considérée comme point de départ de la seconde animation. Ainsi, la continuité de dynamique de l'activation de l'émotion sera conservée, comme illustrée par la Figure 46.

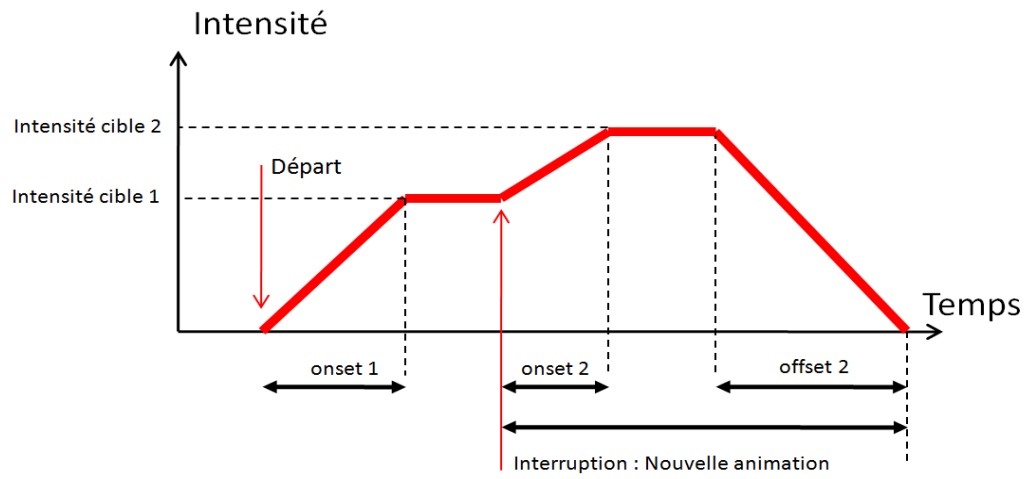


Figure 46 - Courbe d'activation d'une émotion avec interruption

3.2.2 Développements logiciels : partie graphique

Dans notre première version, MARC se compose d'un modèle 3D unique, exporté depuis le logiciel Poser® (*SmithMicro*). Le maillage de ce modèle 3D comporte environ 100.000 polygones (dont 20.000 pour le visage). La texture du visage, plaquée sur le modèle 3D, est un bitmap de 4096x4096 pixels.

3.2.2.1 Le système d'animation faciale de MARC

Dans la première version de MARC, l'animation repose sur le logiciel Virtual Choreographer (Jacquemin, 2004), développé par Christian Jacquemin. Le Virtual Choreographer (VirChor) est basé sur l'utilisation du GPU par l'intermédiaire d'OpenGL et du langage nVidia CG. Le langage XML de VirChor permet de décrire une scène virtuelle ainsi que les comportements et animations associés aux objets de la scène. Ainsi, VirChor permet un rendu audiovisuel 3D dynamique, et contrôlable depuis des applications externes. De plus, il fonctionne à la fois sous Linux et sous Windows. Pour finir, un ensemble de scripts Python d'import/export ont été développés pour permettre l'interconnexion entre le logiciel libre d'édition 3D Blender[5] et VirChor.

La première étape de la conception de MARC a été d'extraire le modèle géométrique du logiciel Poser, puis de l'éditer. Le modèle a en effet requis quelques modifications, effectuées en utilisant le logiciel Blender, pour permettre son intégration dans VirChor et MARC.

Le système d'animation faciale de MARC est un système de skinning. Un ensemble de points-clés sont définis sur le visage, et chaque point clé possède une influence sur la géométrie du visage. Nous avons choisi d'utiliser l'ensemble des points-clés défini par la norme MPEG-4, fréquemment utilisée dans la conception des agents virtuels actuels. Les points MPEG-4 sont donc répartis sur le visage conformément aux spécifications du format.

Le mélange d'expressions faciales, c'est-à-dire le déplacement des points-clés (décrit dans la section 3.2.1) est effectué par VirChor. L'ensemble des déformations géométriques du maillage 3D (associées aux déplacements des points-clés) est effectué en GPU (Vertex Shader) permettant ainsi un rendu dynamique avec une fréquence minimum de 30 images par secondes.

Le module d'animation se base sur les données issues du module émotionnel pour générer des séquences d'expressions faciales. Le déplacement de chaque point clé est donc calculé en effectuant une somme pondérée des vecteurs de déplacements des expressions faciales associées aux émotions à mélanger.

Soit $\boldsymbol{KP_e}$ la liste des déplacements de chaque point clé pour une émotion $\boldsymbol{e}$. $\boldsymbol{KP_e}$ est donc la représentation numérique de l'expression faciale associée à l'émotion $\boldsymbol{e}$. $\boldsymbol{KP_e\ [p]}$ correspond donc au vecteur 3D de déplacement du point clé d'indice $\boldsymbol{p}$.

Soient $\boldsymbol{(\ KP_1\ ..\ KP_e\ )}$, l'ensemble des expressions faciales des $\boldsymbol{n}$ émotions à mélanger, et $\boldsymbol{(\ c_1\ ..\ c_n\ )}$ les coefficients d'intensité respectifs associés à ces émotions. $\boldsymbol{KP_i}$ doit donc être exprimé avec un coefficient $\boldsymbol{c_1}$, etc.

L'expression $\boldsymbol{KP_{res}}$ résultante du mélange d'émotions est donc calculée selon la formule suivante :

Pour tout point clé d'indice $\boldsymbol{p}$:

$$KP_{res}\,[\mathrm{p}] = \sum_{i=1}^{n} ((\ KP_i\ )\,[\mathrm{p}] \quad \mathrm{x} \quad c_i \quad)$$

La continuité de l'évolution des coefficients d'intensité (issus du module émotionnel) permet de garantir une trajectoire continue des points-clés, et ainsi d'éviter les « ruptures » dans l'animation du visage virtuel.

Cette méthode de mélange d'expression est une méthode « simple ». En effet, certaines études montrent que les émotions ne se mélangent pas de manière uniforme sur toutes les parties du visage, mais que certaines émotions ont une influence plus importante sur certaines parties du visage que sur d'autres (Ochs et al. 2006). Néanmoins, notre approche permet la création d'expressions variées, et permet d'utiliser une liste d'émotions extensible. En effet, l'utilisation de règles de mélange d'expressions plus sophistiquées nécessite la mise en place d'une procédure dédiée pour chaque émotion. Il devient alors plus complexe d'ajouter des émotions dans le système, car il faut alors définir les règles de mélange de chaque émotion avec chaque autre.

[5] www.blender.org

3.2.2.2 Le système de rendu du visage de MARC

Notre objectif étant d'obtenir un haut réalisme visuel tout en restant compatible avec l'animation temps réel, le rendu temps réel du visage virtuel de MARC est effectué en utilisant des techniques de rendu graphique récentes (D'eon et el. 2007) implémentées sur processeur graphique (GPU).

La principale difficulté pour rendre la peau humaine est due au comportement complexe de la lumière à l'intérieur des tissus. MARC utilise une technique dite « multi-passes » dans le processeur graphique afin de simuler ce comportement. Le modèle physique de réflectance lumineuse utilisé, nommé Bidirectional Subsurface Scattering Reflection Distribution Function (BSSRDF), permet la simulation de trois comportements lumineux distincts : 1) les ombres auto-projetées, 2) la pénétration de la lumière dans la surface, et 3) la diffusion de la lumière à l'intérieur d'une surface.

Pour effectuer ce type rendu en temps réel, notre algorithme de BSSRDF a été inspiré de D'eon et Luebke (2007) et consiste en 4 étapes (Figure 47). Notre algorithme, comme celui de D'eon et Luebke, est une approximation du modèle de BSSRDF complet. En effet, le principe complet du BSSRDF ne peut être implémenté uniquement qu'en utilisant la méthode du *ray-tracing*.

La première étape de notre méthode de rendu (Figure 47, A) consiste à enregistrer, depuis la position de la lumière, une texture de profondeur. Ainsi, il est possible de déterminer si une surface du visage est directement éclairée, ou si elle est ombrée par une autre surface.

A B C D E

Figure 47 - Etapes du rendu BSSRDF dans MARC

La seconde étape (Figure 47, B) consiste à calculer l'illumination de la surface par la méthode de Phong, c'est-à-dire, calculer le coefficient d'illumination de la surface en fonction de son inclinaison par rapport à la source lumineuse. Contrairement à la méthode de Phong classique utilisée en rendu graphique, cette phase ne considère pas la composante spéculaire de la peau (effet brillant), qui sera ajoutée lors du rendu final du visage.

Cette illumination est ensuite atténuée par le calcul des ombres obtenues en utilisant la texture de profondeur calculée en première étape (Figure 47, C). On observe, dans notre exemple, que la partie gauche du visage (à droite sur l'image) est principalement dans l'ombre, car la lumière se trouve sur la droite du personnage. Durant cette phase, l'algorithme tient compte de la pénétration lumineuse dans les couches internes de la peau. Ainsi, la face arrière de l'oreille droite du visage virtuel apparait comme illuminée en rouge (Voir Figure 48 pour le résultat final correspondant).

La dernière phase consiste ensuite à appliquer un flou gaussien à la texture ainsi obtenue. (Figure 47, D) Le flou gaussien permet de simuler la diffusion de la lumière en surface de la peau. Il convient donc de diffuser plus largement le canal de couleur rouge que les deux autres (car la peau absorbe le bleu et le vert en plus grande proportion que le rouge).

La texture d'illumination obtenue ainsi peut finalement être appliquée sur le visage virtuel qui semble alors diffuser la lumière de manière comparable à la peau humaine (Figure 47, E). Cet effet est le plus visible au niveau des oreilles (Figure 48) et des narines.

Figure 48 - Rendu de l'arrière de l'oreille de MARC. A gauche, sans BSSRDF, à droite, avec BSSRDF

3.2.2.3 Les rides d'expressions

En plus de l'animation des points-clés MPEG-4, nous utilisons leur structure pour calculer la compression de certaines parties du visage sur lesquelles des rides d'expression seront affichées par la technique du *bump-mapping* présentée dans l'état de l'art.

Les zones de rides utilisées dans MARC ont été définies en utilisant les descriptions d'Ekman (Ekman et Friesen 1975). Chaque type de rides cité dans l'un des descriptions des émotions de base est associé à deux ou trois points-clés qui permettront l'activation du type de rides correspondant.

La Figure 49 donne la localisation des axes de détection de la compression. Notre modèle implémente les rides nasolabial (A), les "pattes d'oies" (B), les rides du front (C and D axes), et les rides nasales (E,F) pour les six émotions de base.

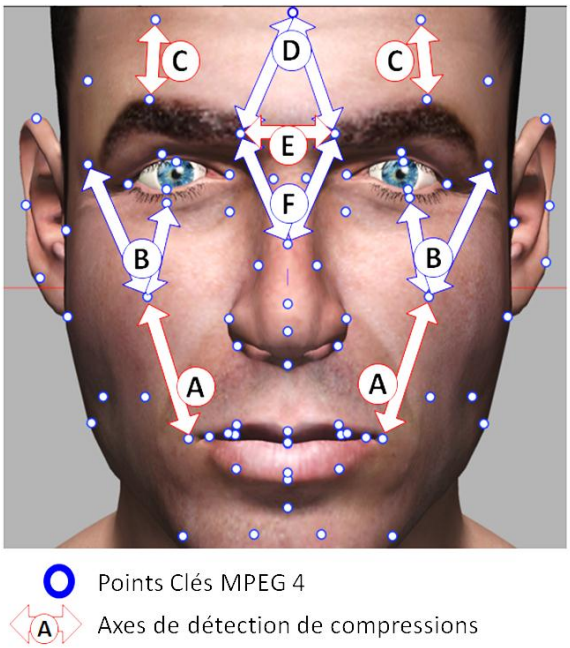


Figure 49 - Axes de détection de compression utilisés dans notre modèle.

Les rides sont déclenchées à l'aide d'une adaptation de la technique de plissage de tissu (Larboulette et Cani, 2004), appliquée à l'animation faciale au format MPEG-4. Lorsque la distance entre ces points-clés est réduite par la déformation du visage due à l'expression faciale, alors les rides associées sont proportionnellement activées. Cette technique est pertinente pour l'animation faciale paramétrique, car l'apparition des rides d'expressions est liée aux mêmes paramètres que ceux qui animent le visage. Ainsi, les rides d'expressions sont directement synchronisées, à la fois temporellement et en intensité, avec le reste de l'animation faciale.

3.2.2.4 Editeur d'expressions faciales dédié

Les données géométriques du visage (la structure de son maillage) peuvent être éditées grâce à des logiciels 3D tels que Blender. En revanche, les paramètres d'animation, tels que les positions et les influences des points clés, ou encore la bibliothèque d'expressions faciales, sont dépendants de notre logiciel d'animation. Pour cette raison, nous avons conçu un logiciel dédié à l'édition des données liées à l'animation.

Comme nous allons le voir, les différentes interactions homme-machine de ce logiciel d'édition nécessitent un accès bas niveau aux données du visage virtuel et au pipeline graphique (que nous ne détaillerons pas). L'utilisation du logiciel VirChor ne nous aurait pas permis ce type de programmation bas niveau. Ainsi, nous avons dû utiliser directement la bibliothèque OpenGL au lieu de l'utiliser indirectement par le Virtual Choreographer. Nous avons donc programmé cet éditeur en JAVA. L'accès à la bibliothèque OpenGL depuis JAVA s'effectue à travers la bibliothèque open source *Light Weight Java Game Library* (LWJGL).

A partir de la structure géométrique du visage, il est tout d'abord nécessaire de placer les points-clés en respectant les positions spécifiées par la norme MPEG4. Notre logiciel permet de placer ces points clés par manipulation directe (par la technique du « drag & drop »). Le déplacement des points clés s'effectue dans le référentiel de la caméra. Ainsi, le point est fixé dans un plan lors de tout déplacement. Pour éditer la profondeur, il faut donc pivoter la caméra autour du visage virtuel. Notre logiciel prévoit pour cela différentes méthodes de rotation de la caméra selon tous les axes possibles du visage virtuel.

Le logiciel d'édition permet ensuite de contrôler l'influence des points clés sur le maillage. Cette édition manuelle est permise par un pinceau 3D, que l'on utilise pour peindre les influences directement sur la structure 3D du visage. Cependant, avec une édition entièrement manuelle, il est à la fois long et difficile d'obtenir un résultat satisfaisant. Nous avons donc conçu un algorithme de pré-calcul des influences des points clés. Le résultat obtenu requiert généralement des rectifications manuelles de certaines influences, mais permet d'accélérer le processus de mise en place des paramètres d'animation.

L'algorithme fonctionne selon les étapes suivantes :

1- Pour chaque point clé, on identifie le vertex le plus proche en distance euclidienne 3D.

2- A partir de ces vertex, on calcule des régions de Voronoï (une région par point clé).

3- En analysant les frontières des régions de Voronoï, on en déduit la matrice de voisinage des points clés. Les points clé K1 et K2 sont dit « voisins », s'ils partagent une frontière commune.

4- Pour chaque point clé, on calcule la distance euclidienne moyenne de ses voisins. Grâce à cette distance, nous pouvons alors calculer la courbe de décroissance de l'influence du point clé. En effet, plus les voisins d'un point-clé sont proches, plus son influence va décroitre rapidement en fonction de la distance.

5- On applique ce coefficient de décroissance pour calculer l'influence du point clé sur chaque vertex appartenant soit à sa région de Voronoï, soit à la région d'un point clé voisin. Chaque vertex sera donc influencé par son point clé « principal » et l'ensemble de ses voisins.

6- Finalement, on applique une passe de flou gaussien sur le maillage pour « lisser » le résultat. Chaque vertex va donc transmettre une partie de ces paramètres aux vertex qui lui sont connexes.

La Figure 50 montre l'influence calculée pour l'un des points-clés du sourcil gauche du visage virtuel. Une fois les influences manuellement ajustées avec le pinceau 3D, il devient possible de créer des expressions faciales.

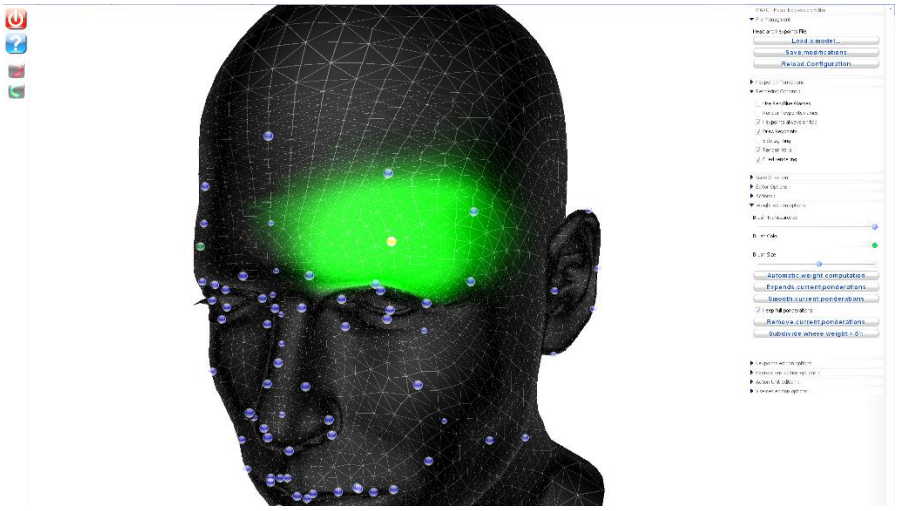


Figure 50 - Editeur d'expressions faciales. Interface de l'éditeur d'influence des points-clés. Le niveau de couleur verte est proportionnel à l'influence du point clé sélectionné (en jaune)

Il est difficile de valider formellement l'influence des points-clés sur le maillage. Une part de l'édition est en effet dépendante du savoir-faire artistique de la personne qui effectue l'édition. Cependant, nous pouvons considérer que la validation des expressions faciales par une étude perceptive valide simultanément les paramètres sous –jacents. Ainsi, les études présentées dans la suite de ce chapitre tiendront lieu de validation de nos paramètres d'animation.

En mode édition d'expressions, le logiciel permet ensuite de créer de nouvelles expressions et de modifier les expressions existantes. Chaque expression est modifiée en déplaçant des points-clés par rapport à leur position d'origine. Le visage est déformé en temps réel en tenant compte de l'influence des points clés. Ainsi, l'utilisateur visualise un aperçu de l'expression en temps réel. Une expression est représentée de manière interne comme un ensemble de déplacements relatif des points clé par rapport à leur position d'origine. Les différentes expressions éditées grâce à ce logiciel peuvent ensuite être utilisées par le module d'animation temps réel.

En plus de l'édition des expressions faciales, l'éditeur intègre une partie dédiée à la spécification des rides d'expressions qui seront utilisées lors de l'affichage des rides par la technique du bump-mapping.

L'édition des rides s'effectue en traçant les lignes de plissement directement sur la structure 3D. Sur la Figure 48, les cercles noirs représentent les points MPEG4 utilisés pour créer les expressions faciales. Les lignes bleues, éditées manuellement par le graphiste, représentent les lignes de plissement des rides d'expressions. Une fois éditées, ces lignes sont converties en textures de relief (« bump-map »), puis en normal-map qui seront utilisées lors du rendu temps réel.

La conversion des lignes vers une texture est effectuée en projetant les courbes de Bézier dans le référentiel de la texture du visage (c'est à dire tangentiellement à la surface du visage). Lors de cette projection, les extrémités des lignes de rides sont atténuées, afin de les faire disparaitre plus progressivement.

Une fois les lignes projetées dans un espace 2D, la texture de relief ainsi obtenue est lissée par un flou gaussien. Ce flou gaussien permet de créer des rides plus progressives lors du rendu final. On évite ainsi que les rides ressemblent à des « fissures » trop nettes. La texture est ensuite utilisée pour calculer une normal-map de chaque zone de rides, en utilisant un calcul classique de dérivée selon les axes horizontal et vertical.

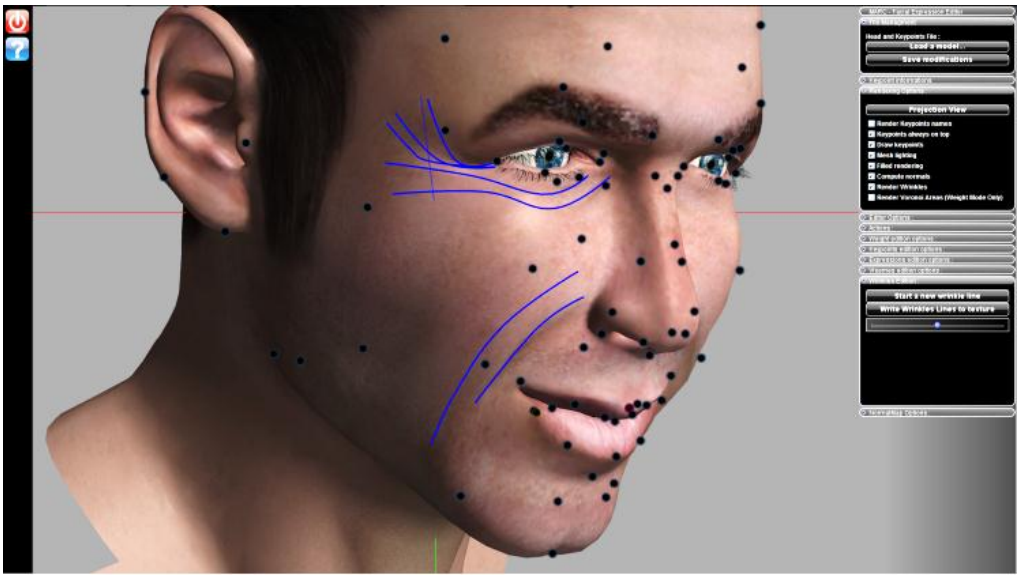


Figure 51 - Editeur d'expression faciale. Mode création de rides d'expressions

3.2.3 MARC v1 catégoriel : conclusions

La conception de cette première version de MARC est axée à la fois sur les aspects visuels et émotionnels. Ces deux aspects doivent donc été évalués.

Cette version de MARC correspond à l'état de l'art des techniques de rendu de visages virtuels. Peu de systèmes d'agents virtuels utilisent ces techniques graphiques récentes. De ce fait, MARC nous permet d'étudier l'influence d'un rendu graphique réaliste sur la perception émotionnelle que les utilisateurs ont de l'agent.

Notre module émotionnel, basé sur le modèle catégoriel des émotions, nous permet de réaliser l'animation faciale d'émotions de base et d'émotions complexes. Le mélange des expressions faciales est réalisé par combinaison linéaire. Malgré les limitations qu'impose cette première approche en termes d'animation, nous avons pu réaliser un certain nombre d'études. Nous verrons dans les versions suivantes de MARC les différentes techniques que nous avons mises en place pour réaliser une animation dynamique plus complexe.

Les outils informatiques et le modèle computationnel d'émotion catégoriel présentés ci-dessus nous ont donc permis d'effectuer plusieurs expérimentations perceptives. Ces expérimentations, présentées dans les sections suivantes, nous ont permis d'obtenir différents résultats et d'améliorer progressivement nos modèles informatiques.

3.3 Evaluation perceptive des expressions statiques des 6 émotions de base

L'interface d'édition de MARC nous a permis de créer une expression faciale pour chaque émotion de base, en utilisant les descriptions des expressions humaines effectuées par Ekman et Friesen (1975). Les expressions ainsi créées peuvent être animées et mélangées par le module d'animation temps réel. Néanmoins, il est nécessaire de les valider perceptivement pour nous permettre de les utiliser dans nos expérimentations.

3.3.1 Objectif

L'objectif de notre première expérimentation est donc de valider les expressions faciales de notre personnage virtuel en comparant les taux de reconnaissance des expressions faciales des six émotions de base (Colère, Peur, Tristesse, Dégout, Surprise et Joie) avec les taux de reconnaissance de visages humains réels issus des travaux en psychologie expérimentale (Russell, 1994).

3.3.2 Hypothèse

Notre hypothèse est la suivante :

H1 : Les taux de reconnaissance des expressions faciales de MARC doivent être comparables aux taux de reconnaissance obtenus dans les expérimentations équivalentes menées avec des photos d'expressions humaines des mêmes émotions de base.

3.3.3 Protocole

Nous avons conçu une expression faciale de MARC pour chacune des six émotions de base: colère, peur, surprise, joie, tristesse et dégoût. Leur conception a été inspirée par les descriptions écrites et picturales de Ekman et Friesen (1975). La Figure 52 montre les expressions des six émotions de base et le visage neutre.

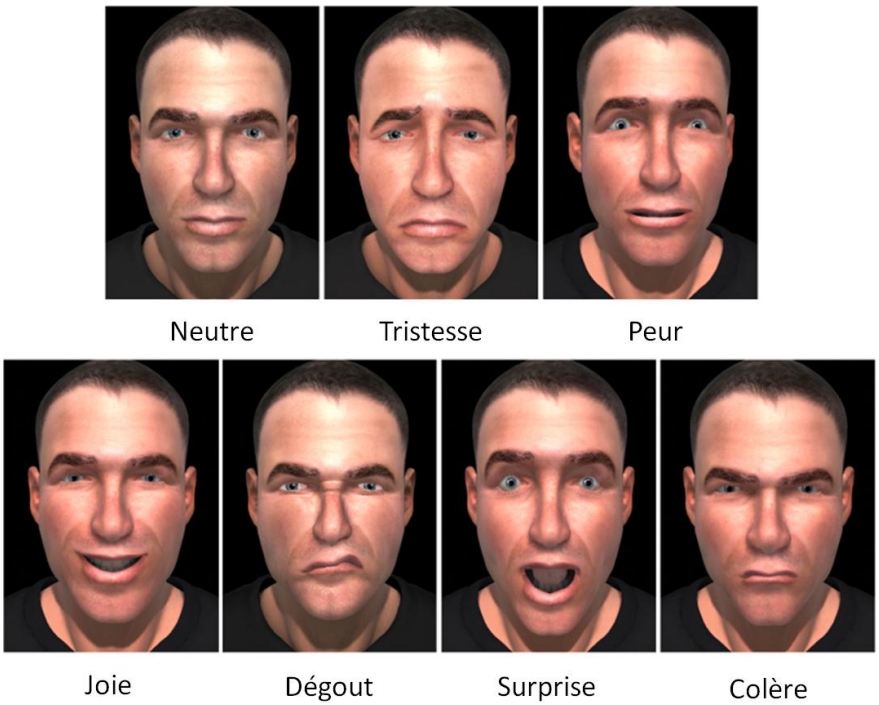


Figure 52 - L'expression neutre de MARC et les expressions des six émotions de base.

Lors des questionnaires du test perceptif, nous avons utilisé un paradigme de choix multiple semblable à celui utilisé dans les études examinées par Russell (1994).

L'expérience a été menée sur 53 sujets (18 femmes, 35 hommes), âgés de 24 à 55 ans (moyenne 31,6). Les sujets ont d'abord dû répondre à quelques questions personnelles. L'expression neutre de l'agent virtuel a été affichée afin que les sujets puissent avoir un point de référence. Ensuite, les expressions des six émotions de base ont été affichées dans un ordre aléatoire, trois fois chacune.

Pour chaque image, les sujets devaient choisir l'un des sept labels émotionnels (choix forcé entre 7 labels: neutre + six émotions de base).

3.3.4 Résultats

Les taux de reconnaissance obtenus sont de: 98.74% pour la Joie, 98.11% pour la Tristesse, 84.91% pour la Colère, 83.65% pour la Surprise, 57.86% pour le Dégout, et 40.24% pour la Peur.

Selon Ekman, dans ce type d'étude, le niveau de hasard doit tenir compte de la valence. Le niveau de hasard pour la Joie est de 50%, car c'est la seule émotion positive. Le niveau de hasard pour la Surprise est de 33.3%, et de 25% pour les émotions négatives (Peur, Dégout, Tristesse, Colère).

Les résultats que nous avons obtenus (Figure 53) montrent à la fois des similarités et des différences avec les études menées avec des photos de visage humain (Russell, 1994).

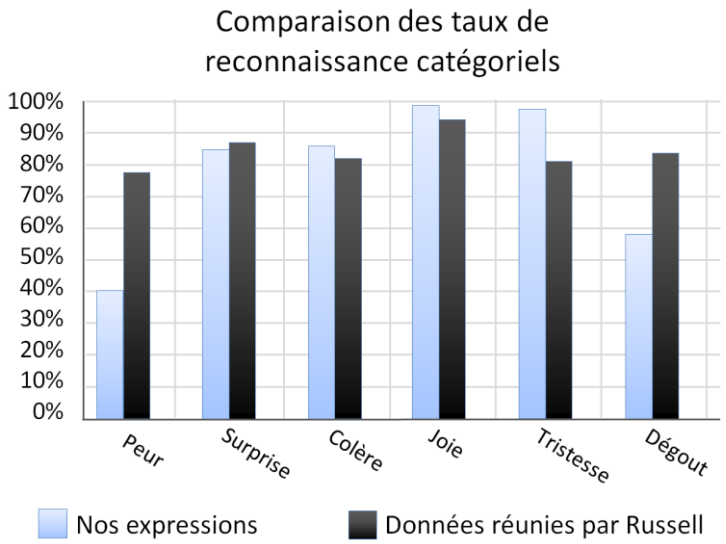


Figure 53 - Taux de reconnaissance des 6 émotions de base, comparés aux études de la littérature (Russell, 1994)

Nos résultats sont globalement cohérents avec les études précédentes: La Joie est l'émotion la plus reconnue, et nous avons les taux de reconnaissance similaires à ceux des études antérieures. Le Dégoût a été confondu avec Colère, en abaissant son taux de reconnaissance. Enfin, l'expression de Tristesse affiche un taux de reconnaissance catégoriel supérieur aux études de référence.

Toutefois, nous avons mesuré une confusion importante pour l'expression de Peur. Plus de 30% des images de Peur ont été perçues comme de la Surprise. Ainsi, le taux de reconnaissance de la Peur est plus faible que ceux des autres émotions. Cependant, cette confusion entre peur et la surprise a été observée précédemment dans d'autres études. Par exemple, des taux similaires ont été obtenus en utilisant un visage virtuel basé sur des expressions actées (Rapcsak et al., 2000). Nous pouvons émettre l'hypothèse que notre expression de Peur n'était peut-être pas assez intense. L'utilisation d'une expression de peur panique pourrait contribuer à améliorer la reconnaissance de cette émotion. Par exemple, les lèvres pourraient être plus tendues.

Les différences entre nos études et les études de référence peuvent être expliquées par les différences entre les deux types de stimuli. Nous avons utilisé un visage virtuel au lieu de photographies de visages réels. Notre visage virtuel n'étant pas contraint physiquement, certaines expressions peuvent être légèrement non réalistes, ce qui pourrait justifier les écarts de taux de reconnaissance en notre défaveur.

D'autre part, nous avons utilisé des images en couleur au lieu d'images en noir et blanc, comme c'est souvent le cas dans les études de référence. Nos sujets en revanche sont plus habitués à observer des visages en couleur qu'en noir en blanc. Cela pourrait donc expliquer les taux de reconnaissance supérieurs aux études précédentes, notamment pour les émotions Joie, Colère et Tristesse. De plus, nous supposons également que la démocratisation des visages virtuels (ex : cinéma, jeux vidéo, films d'animation, etc.) a contribué à habituer les sujets à observer des visages virtuels.

3.4 Impact des rides d'expression sur la reconnaissance des expressions dynamiques d'émotions catégorielles

La perception d'un visage, en plus de sa dynamique, passe également par un ensemble de caractéristiques visuelles, telles que les rides d'expressions, la coloration de la peau, la moiteur, les larmes, etc. Cependant, peu d'agents virtuels implémentent ces indices visuels expressifs, et peu de travaux évaluent leur impact sur la perception des agents virtuel par les sujets humains.

Dans cette section, nous présentons une expérimentation focalisée sur l'impact des rides d'expressions et leur réalisme sur la perception d'un agent virtuel et la reconnaissance catégorielle les émotions. Cette étude a été menée en collaboration avec Stéphanie Buisine de l'ENSAM.

MARC permet l'utilisation de différents modes de rendu des rides. Le mode "Sans-rides" ne rend pas les rides (seuls les mouvements des sourcils, des lèvres, etc. sont affichés). Le mode "Rides-Réalistes " rend des rides avec un effet de relief sur la peau (une texture de *bump mapping* contenant les rides créées avec l'éditeur décrit précédemment). Le mode "Rides-Symboliques" rend des lignes noires au lieu des rides réalistes, activées par le même principe et à la même intensité que les rides réalistes. Ce mode de rendu est destiné à simuler une animation de type cartoon (film d'animation ou bande dessinée), susceptible de permettre une communication plus efficace. En effet, les animations stylisées ou caricaturées, avec comportements stéréotypés, peuvent-être plus efficaces que les méthodes plus réalistes ou naturelles (Gratch et Marsella, 2004). Par exemple, des études par Calder et al. (2000) ont montré que caricaturer les expressions du visage diminue le réalisme perçu, mais augmente en contrepartie le taux de reconnaissance des émotions par les sujets (avec temps de réaction plus court). Un de nos objectifs est d'étudier la différence entre les rides réaliste et symbolique sur les performances de reconnaissance catégorielle. Enfin, le mode « Rides-Seules » affiche les rides réalistes, mais sans aucun mouvement sur le visage (la forme du visage reste inchangée, mais les rides apparaissent). Ce mode est non crédible, mais il servira de condition contrôle dans notre expérimentation.

3.4.1 Objectifs de l'étude sur les rides d'expression

Cette étude vise à évaluer l'impact de la présence ou de l'absence de rides d'expression, et l'impact du réalisme des rides d'expression dans la perception de stimuli dynamiques. En complément de l'étude par de Melo et Gratch (2009) (Cf. état de l'art), nous avons décidé de 1) nous concentrer uniquement sur les rides comme seul signe émotionnel. Ainsi, nous pourrons isoler sa contribution individuelle sur la perception des utilisateurs. 2) Nous utiliserons des stimuli dynamiques plutôt que des images statiques en raison de l'importance de la dynamique du visage dans la perception des émotions. Comme le montre notre état de l'art, Wehrle et al. (2000) ont en effet observé que les expressions faciales dynamiques sont mieux reconnues que des images statiques.

Il est important de considérer plus d'état mentaux que les six émotions de base (Baron-Cohen, 2007, Russell et Mehrabian, 1977) Les expressions faciales des six émotions de bases exprimées par MARC ont été validées dans

l'étude présentée en section 3.3. Le but de l'étude présentée ici est d'évaluer l'impact des rides. Nous avons donc décidé d'utiliser des expressions d'autres états mentaux que celles validées précédemment. Nous avons donc inclus d'autres états mentaux, tels que la Culpabilité et la Fascination.

Pour finir, nous avons utilisé plusieurs intensités expressives. L'intensité maximale, et une intensité divisée par deux. L'expression faciale de moindre intensité est obtenue en atténuant de moitié les déplacements des points-clés MPEG-4 des expressions faciales.

3.4.2 Hypothèses

H1 : Les émotions de base seront mieux reconnues que les états mentaux complexes

En effet, les émotions de base ont été observées comme étant universellement reconnues (Ekman et Friesen, 1975). Nous pouvons donc supposer qu'elles seront mieux reconnues que les états mentaux complexes, moins documentés.

H2 : Les différents modes de rendu des rides engendreront des différences dans les taux de reconnaissance émotionnels et la perception que l'utilisateur aura de l'expressivité du visage virtuel.

Selon nous, nous devrions observer la relation suivante pour les taux de reconnaissance :

Rides Seules < Sans rides < Rides Symboliques < Rides Réalistes

En effet, nous supposons que les rides d'expression augmentent la reconnaissance des expressions faciales. Cependant, le rendu non réaliste des rides symboliques est susceptible de gêner la reconnaissance des émotions, ainsi, les rides réalistes devraient donner de meilleurs taux de reconnaissance que les rides symboliques.

H3 : A des intensités expressives plus élevées, les catégories d'émotions seront mieux reconnues.

Cette hypothèse se base sur les taux de reconnaissance peu élevés que nous avons observé pour l'expression de Peur. Selon nous, cet effet est dû à un manque d'intensité, qui atténue la reconnaissance des émotions. Nous pensons donc observer un effet similaire sur les autres catégories d'émotions.

3.4.3 Protocole

Participants : 32 sujets (10 femmes, 22 hommes), âgés de 16 à 40 (25 ans en moyenne, écart type = 4,6) ont été recrutés parmi les étudiants et le personnel de recherche du LIMSI et de l'ENSAM. 20 d'entre eux n'étaient pas familiers avec les agents virtuels, 9 ont déclaré être des utilisateurs occasionnels des agents virtuels, et 3 d'entre eux étaient des utilisateurs réguliers.

Matériel : Les catégories émotionnelles ont été sélectionnées au sein de l'intersection entre les états mentaux situés dans l'espace PAD par Russell et Mehrabian (1977) et l'ensemble des états mentaux de la base de données vidéo MindReading de Baron-Cohen (2007). Cette méthode de sélection a été utilisée afin de pouvoir disposer, pour chaque émotion, un point de référence dans l'espace PAD et d'un corpus vidéo d'expressions faciales associées. L'emplacement dans l'espace PAD a été arrondi à l'angle le plus proche du cube PAD. Afin de limiter systématiquement le nombre de stimuli, et d'éviter les confusions possibles par des sujets entre activation et l'intensité (Devillers et al. 2006), seules les émotions d'activation positive ont été sélectionnées. Nous n'avons donc considéré que quatre zones de l'espace PAD. Une émotion de base et une émotion complexe ont été sélectionnées sur chaque quart de cet espace Valence/Dominance. Nous avons gardé les quatre émotions de base: la Joie, la Colère, la Peur, et la Surprise. Nous avons également conservé les quatre expressions faciales correspondantes qui ont été conçues et testées dans la validation décrite précédemment. La tristesse n'a pas été sélectionnée dans cette seconde étude, car son activation est négative. Le dégoût n'a pas été conservé, car il est

situé dans le même quadrant de l'espace PAD que la colère, et que cette émotion a été moins bien reconnue dans notre étude de validation.

Quatre émotions complexes ont été choisies en raison des mêmes critères: l'Intérêt, le Mépris, la Culpabilité, et la Fascination. Les expressions faciales de ces quatre émotions complexes sélectionnées ont été inspirées par la base de données MindReading (Baron-Cohen, 2007) où chaque état mental est interprété par six acteurs. Pour cela, nous avons commencé par extraire les caractéristiques des expressions faciales (ex : froncement de sourcil) apparaissant dans au moins la moitié de ces six vidéos. Ensuite, nous avons utilisé ces caractéristiques pour concevoir les expressions faciales de ces émotions complexes en utilisant notre plate-forme MARC.

Ces expressions d'émotions complexes n'ont pas été validées au préalable (par exemple, selon le protocole utilisé dans notre validation des expressions faciales des émotions de base décrit dans la section précédente). La littérature est rare sur le taux de reconnaissance des expressions faciales des émotions complexes et il aurait été difficile d'interpréter les résultats.

Chaque expression a été conçue comme un ensemble de mouvements du visage (par exemple, joue, sourcils, lèvres, paupières, etc.). Les rides d'expressions sont activées automatiquement à partir des mouvements faciaux (selon la méthode décrite précédemment).

Chacune des expressions des 8 émotions sélectionnées a été rendue en utilisant les 4 modes de rides (Sans-Rides, Rides-Réalistes, Rides-Symboliques, Rides-Seules) avec 2 intensités différentes, pour un total de 64 animations différentes. Le rôle des rides dans l'expression des émotions peut être évalué en effectuant des comparaisons par paires entre le mode Sans-Rides, le mode Rides-Réalistes et le mode Rides-Symboliques. En outre, le mode Rides-Seules nous permet d'évaluer le rôle des mouvements du visage sur la reconnaissance des émotions, en comparant le mode Rides-Réalistes au mode Rides-Seules. Le dispositif expérimental est résumé dans la Figure 54. La Figure 55 montre les expressions de Surprise et la Culpabilité des quatre modes de rides.

L'intensité maximale d'une expression a été définie au point au-delà duquel l'expression semble exagérée. Pour les émotions de base, ces expressions sont celles utilisées lors de la validation en section 3.3.

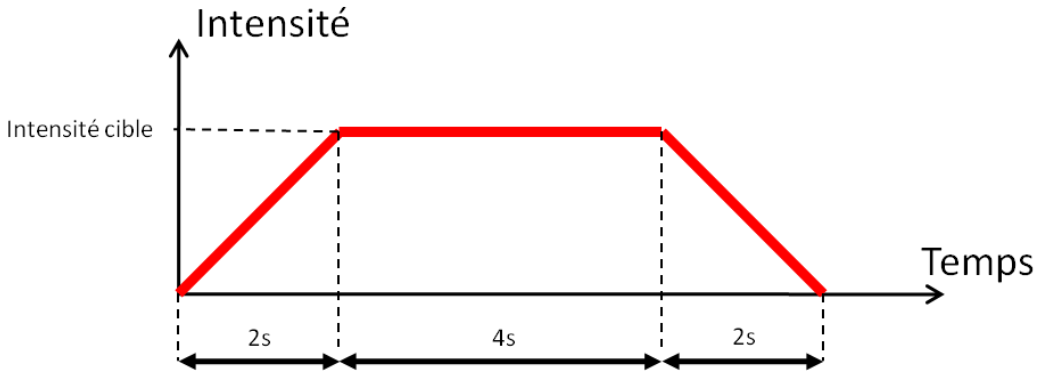
L'intensité 50% est définie en divisant par deux l'intensité des paramètres expressifs. Le mouvement global du visage est ainsi divisé par deux. En effet, Wallraven et al. (2005) soutiennent que la réduction des mouvements faciaux réduit l'intensité perçue. Enfin, nous faisons l'hypothèse que les différences entre les modes de rendu de rides seront moins importantes avec des intensités inférieures.

		Mouvements du visage dans l'expression des émotions	
		Avec mouvements	Sans mouvement
Rides dans l'expression des émotions	Sans Rides	Condition Sans-Rides	
	Rides Symboliques	Condition Rides-Symboliques	
	Rides Réalistes	Condition Rides-Réalistes	Condition Rides-Seules

Figure 54 - Récapitulatif des conditions expérimentales de l'étude sur les rides d'expressions

Chaque animation commence avec une expression neutre, puis le visage adopte progressivement l'émotion cible, soutenue pendant 4 secondes, et retourne progressivement à une expression neutre. Comme les rides sont

directement actionnées par le déplacement des point-clés MPEG-4, leurs animations sont automatiquement synchronisées (dans le GPU). La courbe d'activation est donc la suivante :



Les animations ont été rendues en temps réel durant le test, en utilisant 2 cartes graphiques nVidia 8800GT (en SLI), et affichées en plein écran sur un moniteur LCD 24 pouces (environ 60cm de diagonale) avec une résolution de 1920x1200 pixels. L'utilisateur se situait à environ 60cm de l'écran.

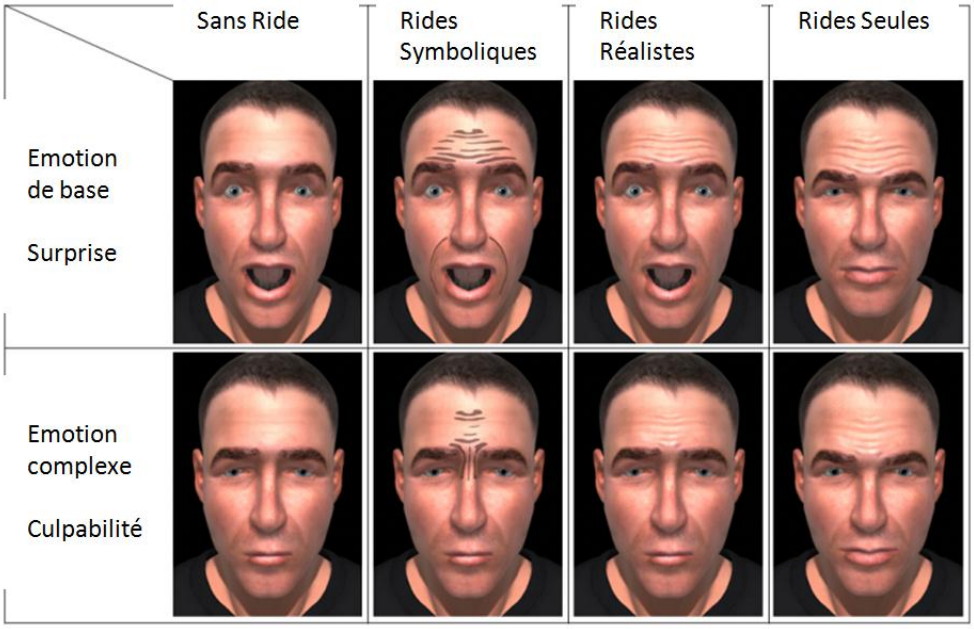


Figure 55 - Expressions faciales de Surprise et de Culpabilité selon les quatre types de rides

Procédure : Les sujets ont été invités à fournir certains renseignements personnels (âge, sexe, profession). L'expérience a ensuite été divisée en 2 étapes. La première étape a consisté à afficher successivement les 64 animations. Pour chaque animation, les sujets devaient choisir une seule étiquette dans un ensemble de 16 descripteurs émotionnels et le taux de l'intensité perçue sur une échelle analogique de 0 à 100%. La Figure 56 montre les 16 descripteurs que nous avons choisis comme réponses possibles: les 8 émotions affichées par les stimuli, plus 8 distracteurs (4 sélectionnés en neutralisant les axes Valence et Dominance, et 4 autres, sélectionnés en utilisant une Activation négative).

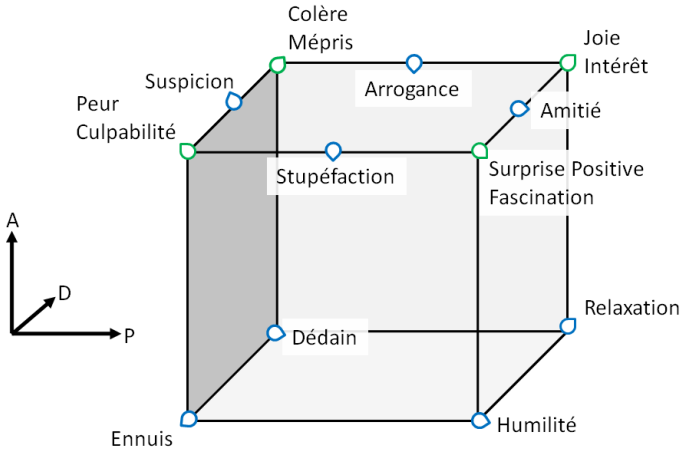


Figure 56 - Les émotions sélectionnées (en vert) et les distracteurs du questionnaire (en bleu). Les localisations des émotions dans l'espace PAD sont approximées.

Les distracteurs sont destinés à accroître la difficulté de la tâche en ajoutant d'autres labels émotionnels que ceux présentés aux sujets de l'expérimentation. En effet, les tests de validation, menés en utilisant les expressions du visage sans rides d'expression, ont montré des taux de reconnaissance élevés (par exemple 98,74% pour la Joie). Nous étions préoccupés par un effet plafond potentiel qui aurait limité l'impact de nos modes de rendu. L'introduction d'éléments de distraction rend le choix forcé beaucoup plus difficile pour plusieurs raisons. Tout d'abord, les distracteurs fournissent des alternatives plausibles aux émotions. De fait, dans le paradigme classique des études listées par Russell (1994), il n'y a pas d'alternative plausible à la Joie, puisque c'est la seule émotion positive à la disposition des sujets. En outre, parce que nos éléments de distraction sont incorrects tout au long du test, ils peuvent entraver une éventuelle stratégie du sujet consistant à deviner l'émotion attendue en fonction de celle qu'il a déjà vue. En effet, dans le paradigme classique, les réponses données par le sujet lui donnent des indices pour les émotions suivantes, car tous les labels sont corrects tour à tour. Ce design expérimental avec distracteurs est un principe de base en tests psychométriques et psychologiques (Kline, P., 1987, Kline, J.B., 2005).

Quatre animations ont été utilisées en tant que session d'entrainement, et sans limite de temps de réponse. A partir de la cinquième animation, les sujets ne disposaient que de 30 secondes pour choisir un label émotionnel et l'intensité émotionnelle perçue. Nous avons choisi cette procédure en temps limité en vue d'assurer la spontanéité des réponses et pour limiter la durée totale de l'expérience. L'ordre de présentation des 64 stimuli a été tiré aléatoirement pour chaque sujet. Les participants ont été autorisés à prendre une pause entre 2 animations en cliquant sur un bouton de pause.

Dans la deuxième étape de l'expérimentation, pour chaque émotion, les participants ont visualisé les images statiques de l'expression correspondant à son intensité maximale. Chaque page du questionnaire (Figure 57) était composée de 4 images (4 modes de rendu graphique) disposées côte à côte avec des positions aléatoires. Les participants ont eu à classer ces modes de rendu en fonction du niveau d'expressivité perçu (**++**, **+**, **−**, et **−−**). Ils ont également eu à choisir leur mode de rendu préféré pour chaque émotion. L'expérience a duré entre 30 et 40 minutes par sujet.

Données collectées : La reconnaissance des émotions catégorielles a été collectée comme une variable binaire (bonne réponse = 1, mauvaise réponse = 0), et la perception d'intensité, comme une valeur numérique comprise entre 0.0 et 1.0. Le classement de l'expressivité (1er rang représentant le stimulus perçu comme le plus expressif) a été converti en scores d'expressivité (le 1er rang est converti en un score de 3 points, et le dernier rang, en un score de 0 point) et les préférences des utilisateurs ont été considérées comme une variable binaire (rendu favoris = 1, autres = 0).

Figure 57 - Formulaire de la seconde partie de l'étude sur les rides d'expression

3.4.4 Résultats

Performances de reconnaissance : L'intensité perçue des émotions, les scores d'expressivité et les scores de préférence ont été analysés par analyses de variance avec la catégorie émotionnelle, le rendu graphique et l'intensité comme variables intra-sujet. Le test LSD de Fisher a été utilisé pour les comparaisons post-hoc par paires. Toutes les analyses ont été effectuées avec le logiciel SPSS.

Performances de reconnaissance : Parmi les 64 stimuli × 32 utilisateurs (2048 items), nous avons eu 56 *timeout*, ce qui correspond à 2,7% des données. Ces items ont été analysés comme étant des réponses de reconnaissance fausse. Le temps de réponse moyen a été de 15,6 secondes par animation (Ecart type = 1,99s). Le taux de reconnaissance global des émotions catégorielles est de 26,6%. Les sujets de cette expérience ayant 16 choix possibles, le niveau de hasard est de 1 / 16 (6,25%). Toutefois, nous avons choisi de suivre la stratégie conservatrice d'Ekman (1994) et de considérer la valence émotionnelle pour calculer le niveau de hasard. Ainsi, pour les expressions positives, nous avons considéré que les sujets avaient 9 choix (Joie, Intérêt, Amical, Surprise Positive, Fascination, Humble, et Relaxé, plus Etonné et Arrogant qui sont de Valence neutre). De même les sujets avaient 9 choix pour chaque expression négative (Colère, Mépris, Peur, Culpabilité, Scepticisme, Dédain, et Ennui, plus Etonné et Arrogant). C'est pourquoi nous avons considéré un niveau de hasard de 1 sur 9 (soit 11.11%).

L'effet principal de la catégorie émotionnelle s'est avéré être significatif ($F(7 / 630)$ $p = 16,24$, $p < 0,001$) (Figure 58). Le test T-Student a montré que le taux de reconnaissance était significativement plus élevé que le niveau de hasard, à l'exception de Culpabilité et de Fascination (9,3% et 9,8%) qui n'étaient pas significativement différents par rapport au niveau du hasard ($t(31) = -0,69$ et $t(31) = -0,6$ respectivement). En outre, les émotions de base ont été mieux reconnues (39%) que les émotions complexes (14,6%, $p < 0,05$). La colère est l'émotion la plus reconnue (52,7%) parmi les émotions de base (tendanciellement plus que la joie, $p = 0,059$, et de manière significative que toutes les autres émotions, $p < 0,002$).

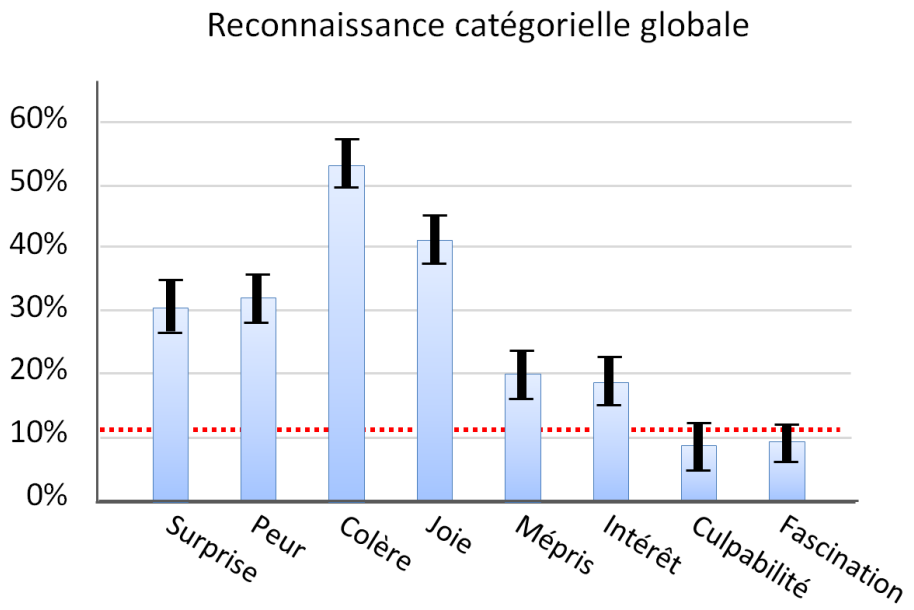


Figure 58 - Reconnaissance catégorielle par émotion. La ligne pointillée représente le niveau de hasard (11.11%)

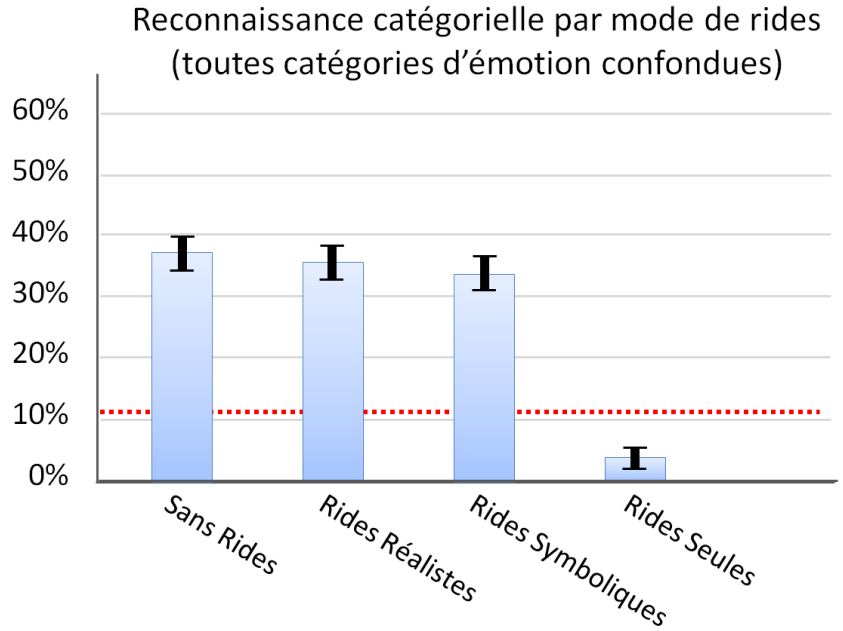


Figure 59 - Taux de reconnaissance pour chaque mode de rendu graphique. La ligne pointillée représente le niveau de hasard (11,11%).

L'effet principal du mode de rendu graphique a également été significatif (F (3 / 630) p = 59,39, p <0,001) (Figure 59). Les modes Sans-Rides, Rides-Réalistes et Rides-Symboliques ont tous donné des taux de reconnaissance équivalents, et significativement plus élevés que le niveau de hasard (34,5%, p <0,001). Toutefois, le taux de reconnaissance du mode Rides-Seules (3,7%) ne diffère pas du niveau de hasard.

Nous avons également observé une interaction entre la catégorie émotionnelle et le rendu graphique (F (21/630) p = 5,27, p <0,001) (Figure 60), montrant que l'effet du modèle de rides (Sans-Rides, Rides-Réalistes, et Rides-Symboliques équivalents et meilleurs que le mode Rides-Seules) n'était pas vrai pour toutes les catégories émotionnelles. Pour l'Intérêt et la Culpabilité, nous n'observons aucun effet du rendu graphique. Pour la Joie le mode Rides-Symboliques donne lieu à des performances légèrement supérieures au mode Sans-Rides (p = 0,078). Enfin, pour la Peur, le mode Rides-Symboliques montre des taux de reconnaissance légèrement inférieurs à ceux du mode Sans-Rides (p = 0,065) et du mode Rides-Réaliste (p = 0,073).

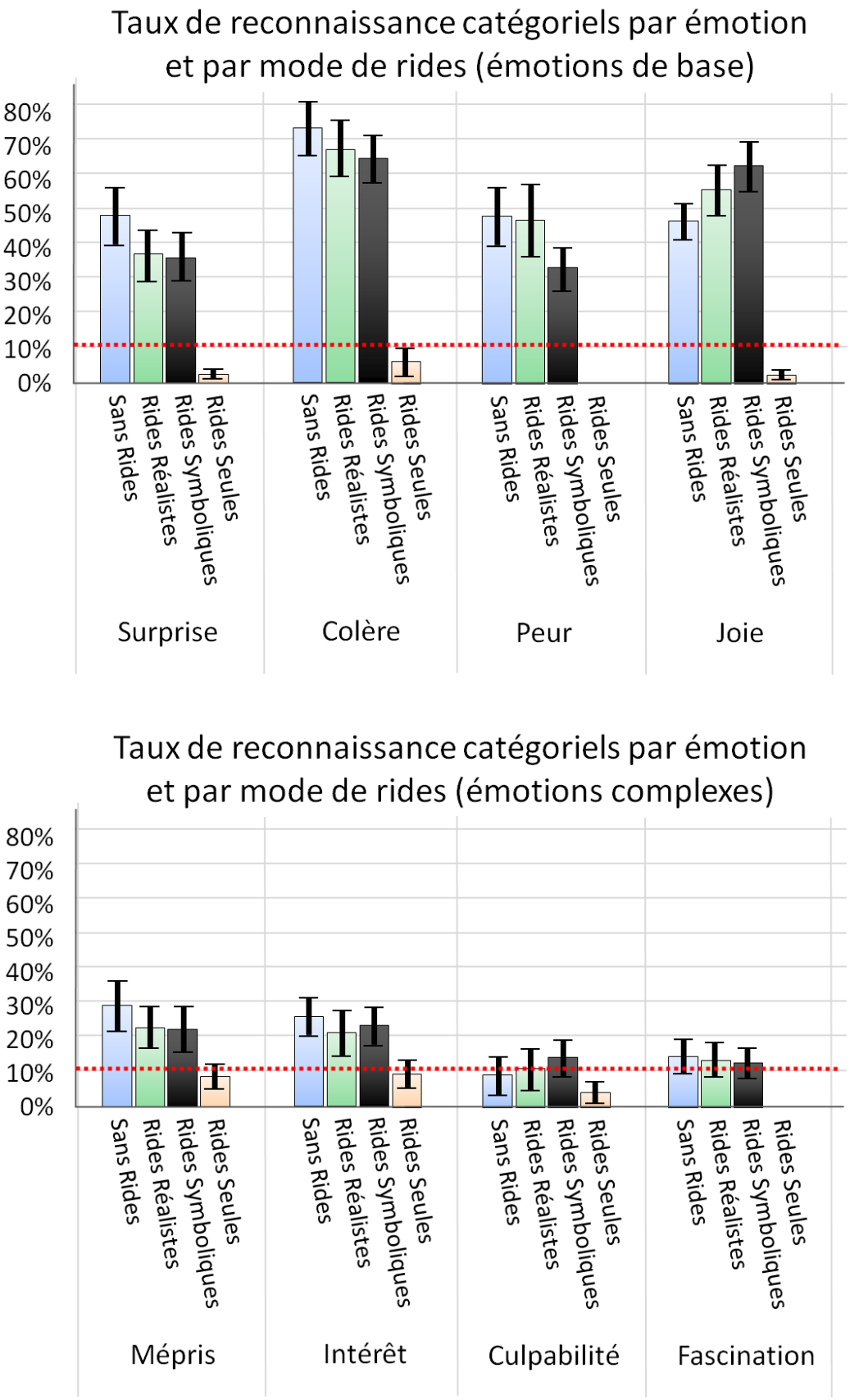


Figure 60 - Taux de reconnaissance pour chaque émotion et pour chaque mode de rendu. La ligne pointillée représente le niveau de hasard (11.11%)

L'intensité des expressions affichées a également eu une influence significative sur les taux de reconnaissance (F (1 / 630) = 12,01, p = 0,002): les expressions affichées à intensité maximale ont été mieux reconnues (30%) que les expressions affichées à intensité moyenne (23,7%). Toutefois, l'interaction entre catégorie émotionnelle et intensité (F (7 / 630) = 2,95, p = 0,006) montre que cet effet n'est vrai que pour la Peur (p <0,001) et la Colère (p = 0,014). Pour toutes les autres émotions, l'intensité des stimuli n'a eu aucune influence sur les performances de reconnaissance.

Enfin, il n'y a pas d'effet du genre de l'utilisateur sur les performances de reconnaissance.

Perception de l'intensité : On observe un effet de la catégorie émotionnelle sur l'intensité perçue (F (7 / 630) = 29,82, p <0,001). La Surprise, la Colère, la Joie et la Culpabilité suscitent des niveaux significativement plus élevés d'intensité perçue (p <0,006) que ceux de la Peur, le Mépris, l'Intérêt et Fascination.

On observe également un effet du mode de ride sur l'intensité perçue (F (3 / 630) p = 61,58, p <0,001) (Figure 61) révélant que les Rides-Réalistes et les Rides-Symboliques génèrent une perception de l'intensité supérieure au mode Sans-Ride (p <0,021), qui, lui, montre une intensité perçue supérieure au mode Rides-Seules (p <0,001). La différence entre les modes Rides-Réalistes et les Rides-Symboliques n'est pas significative.

La différence entre les deux niveaux d'intensité a été perçue de manière significative (F (1 / 630) p = 135,87, p <0,001). Ceci confirme que les stimuli d'intensité moyenne ont été perçus comme étant moins intenses (m = 0,35) que les stimuli affichés avec une intensité élevée (m = 0,53). L'interaction entre l'intensité et le rendu graphique (F (3 / 630) = 24,47, p <0,001) montre que le mode de rides n'influe pas sur la capacité des sujets à décoder l'intensité. Cependant, le mode Rides-Seules n'a pas permis aux sujets de décoder l'intensité perçue. Les modes Rides-Réalistes, Rides-Symboliques, et Sans-Rides sont eux équivalents.

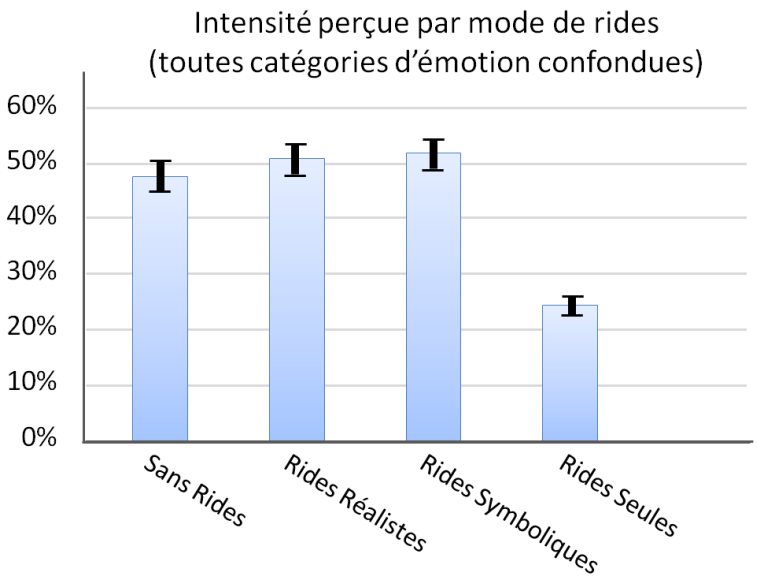


Figure 61 - Intensité perçue en fonction du mode de rendu

Expressivité : En ce qui concerne les évaluations subjectives de l'expressivité, le rendu graphique s'est avéré avoir un effet principal significatif (F (3 / 630) = 97,39, p <0,001) (Figure 62).

Le niveau de hasard est de 1,5 car chaque classement de quatre images correspond à l'attribution de 6 points (3+2+1+0), soit 1,5 par image en moyenne. Le mode Rides-Réalistes est jugé significativement plus expressif (2.4/3) que les trois autres modes (p<0.001). Les scores du mode Sans-Rides et Rides-Symboliques (1,6 et 1,7)

n'étaient pas significativement différents l'un de l'autre alors que le rendu Rides-Seules est jugé significativement moins expressif que les autres (0,2, p <0,001).

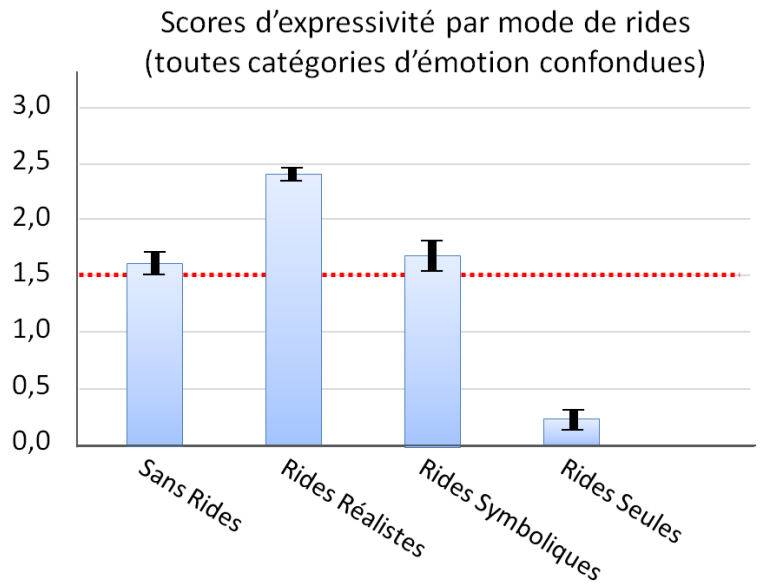


Figure 62 - Expressivité perçue pour chaque mode de rides. La ligne horizontale pointillée représente le niveau de hasard (score de 1.5)

L'étude de l'interaction catégorie émotionnelle × rendu graphique (F (21/630) p = 5,96, p <0,001) montre plusieurs effets différents en fonction des catégories expressives:

Pour la Surprise, la Culpabilité et la Fascination: le principal effet est vérifié :

Rides-Réalistes > Sans-Rides = Rides-Symboliques > Rides-Seules.

Pour la Peur et la Colère, on observe :

Rides-Réalistes > Rides-Symboliques > Sans-Rides > Rides-Seules.

Pour la Joie et le Mépris, on observe :

Rides-Réalistes = Sans-Rides = Rides-Symboliques > Rides-Seules

Pour l'Intérêt, on observe :

Rides-Réalistes > Sans-Rides > Rides-Symboliques = Rides-Seules

Il n'y a aucun effet du genre des sujets sur la perception de l'expressivité.

Préférences : Le mode de rendu des rides a eu un effet significatif sur les scores de préférence (F (3 / 630) = 25,80, p <0,001) (Figure 63). Le mode Rides-Réalistes a été préféré à l'ensemble des 3 autres rendus (p <0,001) avec un score de préférence de 0,51 / 1. Le mode Sans-Rides (0,27) a été préféré au mode Rides-Symboliques (0,09) et au mode Rides-Seules (0,13, p <0,001). Les modes Rides-Symboliques et Rides-Seules montrent des scores équivalents.

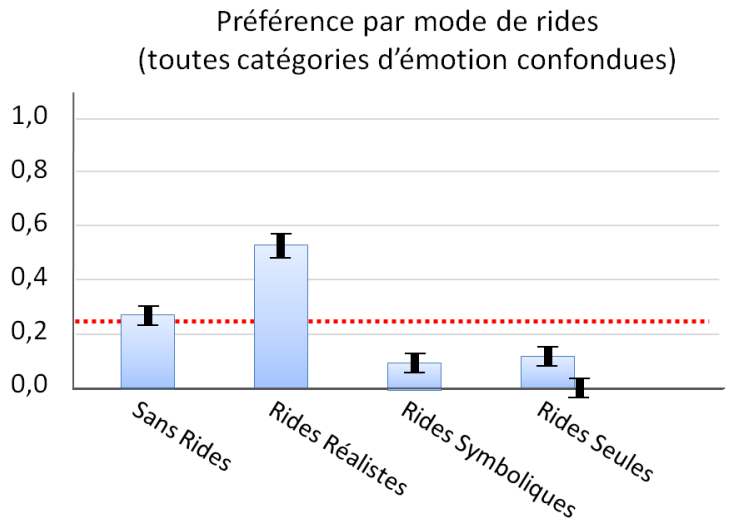


Figure 63 - Préférence pour chaque mode de rendu des rides. La ligne horizontale pointillée représente le niveau de hasard (0.25)

L'interaction catégorie émotionnelle × rendu graphique (F (21/630) p = 2,58, p <0,001) montre 4 effets de la préférence en fonction de la catégorie émotionnelle:

Pour la Surprise, la Peur et la Colère: le mode Rides-Réalistes est préféré (p <0,037) tandis que les 3 autres modes sont équivalents.

Pour la Joie, la Culpabilité et la Fascination: le mode Sans-Rides est équivalent au mode Rides-Réalistes, et ces deux modes sont supérieurs aux deux autres.

Pour l'Intérêt: le mode Rides-Seules est équivalent au mode Rides-Réalistes, et ces deux modes sont supérieurs aux deux autres.

Pour le Mépris: L'effet du mode de ride n'est pas significatif, ce qui signifie qu'aucun des modes n'a été systématiquement préféré.

3.4.5 Conclusions de l'étude sur les rides d'expression

La peur est à nouveau confondue avec la Surprise, ce qui est cohérent puisque que nous avons utilisé les deux mêmes expressions lors des deux expériences. L'un des résultats importants de notre étude sur les rides d'expression est que l'utilisation de rides ne réduit pas cette confusion. Plus généralement, l'utilisation de rides d'expression ne semble pas avoir d'impact sur la reconnaissance catégorielle des émotions. Cela donne des indications quant à la validité de nos stimuli expérimentaux. Le fait que les taux de reconnaissance aient été plus élevés que le niveau de hasard tend à valider la conception de nos animations, sauf dans les cas des expressions de Culpabilité et de Fascination qui ne sont pas mieux reconnues que le niveau de hasard. Par conséquent, nous ne discuterons pas plus les résultats de ces deux expressions d'émotions complexes. Lorsque les données de reconnaissance de Culpabilité et de Fascination sont exclues, les taux de reconnaissance sur les émotions restantes s'élève à 32,6%, ce qui est plus faible que les études listées par Russell (1994). Cependant, ce score de reconnaissance relativement faible doit être interprété avec prudence pour plusieurs raisons. Tout d'abord, il comprend des données du mode Rides-Seules qui ont font chuter le taux de reconnaissance de façon spectaculaire. Deuxièmement, cette étude utilise des stimuli d'intensité modérée et intense alors que les études classiques se limitent à l'utilisation de stimuli intenses seulement. Troisièmement, le taux de reconnaissance global comprend les taux de reconnaissance de deux émotions complexes (Mépris et Intérêt), alors que les études classiques se concentrent sur les émotions de base.

En outre, nous n'avons considéré uniquement que les réponses rigoureusement exactes dans nos analyses. Les erreurs légères ne sont pas distinguées des erreurs majeures. Les deux sont considérées comme fausses. Par exemple, la confusion entre le mépris et l'arrogance a été considérée comme une mauvaise réponse, bien que certaines personnes puissent ignorer la différence entre les expressions de ces émotions. Nous avons adopté cette règle conservatrice afin de maximiser la probabilité d'observer des différences entre nos modes de rides. Pour finir, la tâche proposée dans notre étude a été beaucoup plus difficile que celles des études classiques empiriques utilisant des photos d'humains d'émotions de base statiques à haute intensité et un nombre limité d'options (Russell 1994). En effet, nos sujets ont dû faire face à des expressions d'émotions complexes, des labels émotionnels complexes et des distracteurs.

La seule valeur qui peut être comparée aux études classiques est le taux de reconnaissance des quatre émotions de base (surprise, peur, colère, joie) à haute intensité et sans considérer les données du mode Rides-Seules. Nous obtenons ainsi un score de reconnaissance à 58%. Ce score est plus bas que les taux de reconnaissance obtenus avec paradigme classique, et aussi plus bas que ceux rapportés dans un paradigme à réponses libres au lieu du format à choix forcé (Ekman, 1994). Notre résultat illustre l'impact énorme que le format expérimental peut exercer sur les données collectées. Par exemple, introduire des distracteurs dans les options de réponse peut diminuer de manière significative la reconnaissance des émotions de base.

En ce qui concerne l'intensité, les stimuli que nous avons conçus comme très intenses ont été effectivement jugés plus intenses que les stimuli conçus comme modérément intenses. L'effet global de l'intensité prévue sur les taux de reconnaissance tend également à valider nos stimuli. Les intensités élevées conduisent à des meilleurs taux de reconnaissance. Toutefois, cet effet n'est significatif que pour les expressions de Peur et de Colère. Ceci est cohérent avec notre hypothèse selon laquelle l'expression de Peur utilisée dans notre étude n'est peut-être pas assez intense. La reconnaissance des autres émotions n'a pas été sensiblement améliorée grâce à des stimuli intenses. Nous émettons l'hypothèse que notre façon de manipuler l'intensité explique ce résultat. À haute intensité, une émotion ne peut être exprimée uniquement que par une plus large expression, mais également en utilisant différentes expressions du visage pour une seule émotion et d'autres modalités, telles que les mouvements de tête, du regard et coloration de la peau. D'autres études doivent être menées pour valider cette hypothèse.

Un des principaux objectifs de cette expérience était de comparer plusieurs modes de rendu graphique des rides. Le principal résultat de cette étude, original et inattendu, est l'effet des modes de rides sur la reconnaissance des émotions. La présence de rides n'a pas amélioré les performances de reconnaissance: ni les rides réalistes, ni les rides symboliques, ne fournissent davantage d'informations que le rendu sans rides en termes de catégorie d'émotion. Ainsi le facteur clé pour la reconnaissance catégorielle des expressions faciales de nos stimuli semble être des mouvements du visage plutôt que de rides, car l'absence de mouvements (mode rides seules) fait tomber les performances de reconnaissance de façon spectaculaire au niveau de hasard. Cependant, nos analyses ont également montré que les rides sont susceptibles d'augmenter l'intensité perçue de l'expression émotionnelle, ce qui est cohérent avec les résultats antérieurs obtenus avec des photos de visage réel (Borod et al., 2004). Une telle augmentation est subtile, mais néanmoins significative dans nos données, et est générée de manière similaire par les deux modes Ride-Réalistes et Rides-Symboliques.

D'un point de vue plus subjectif, le rendu Rides -Réalistes a été considéré comme le plus expressif et le mode de rendu favori de nos sujets. Un tel résultat est suffisant pour établir l'utilité d'inclure les rides réalistes dans la conception de personnages virtuels. En ce qui concerne l'expressivité, on peut se demander pourquoi nos participants ont jugé le mode Rides-Réalistes comme plus expressif même s'il ne conduit pas à une amélioration effective des taux de reconnaissance, ni à un accroissement de l'intensité perçue. Nous émettons l'hypothèse que ces évaluations de l'expressivité ont été influencées par les préférences des utilisateurs, qui peuvent les avoir amenés à choisir ce rendu en premier dans la tâche classement de l'expressivité.

Les émotions de base ont été mieux reconnues que les émotions complexes. Ceci peut être expliqué par l'absence d'une description détaillée des expressions faciales des émotions complexes dans la littérature. Nous avons également l'hypothèse que les définitions de ces émotions sont plus subjectives. Par conséquent, la sélection d'un adjectif pour ces émotions est une tâche plus floue que pour les émotions de base.

L'étude de Melo et Gratch (2009) a évalué l'impact de la combinaison des rides, du rougissement et des larmes sur la perception des émotions, en utilisant des images statiques. Ils ont observé que l'utilisation de ces indices augmente l'expressivité et l'intensité émotionnelle perçue. Nos résultats, en ce qui concerne les rides, sont cohérents avec leurs résultats. En outre, nous avons observé que l'utilisation de rides réalistes améliore la préférence des utilisateurs. En conséquence, les rides pourraient aider les utilisateurs à accepter les agents virtuels et à s'engager dans l'interaction avec eux.

3.5 Rides géométriques 3D et perception des émotions

Les rides réalistes considérées dans l'étude précédente souffrent d'une limitation importante : la technique utilisée est limitée par l'absence de déformation réelle du visage lors de l'apparition des rides en raison de la technique de *bump-mapping* utilisée qui n'impacte que le rendu lumineux. Ainsi, de profil, les rides sont problématiques, car la géométrie, vue de côté, n'est pas déformée. Suite à cette première étude sur les rides d'expressions, nous avons donc choisi de proposer une nouvelle approche pour afficher des rides d'expression en 3D.

3.5.1 Animation paramétrique de rides géométriques

Afin d'étudier la perception des expressions faciales 3D de l'émotion à partir de vues de face et de côté, nous avons étendu le système MARC avec un système de rides en 3D. Afin de créer des rides réalistes en 3D, nous avons ajouté au système une deuxième série de points-clés dédiée aux rides. Les zones d'influence des points-clés sont éditées manuellement (comme pour les points-clés dédiés à l'animation faciale). La Figure 64 montre la zone d'influence d'un point-clé dédié aux rides, et la déformation géométrique résultante.

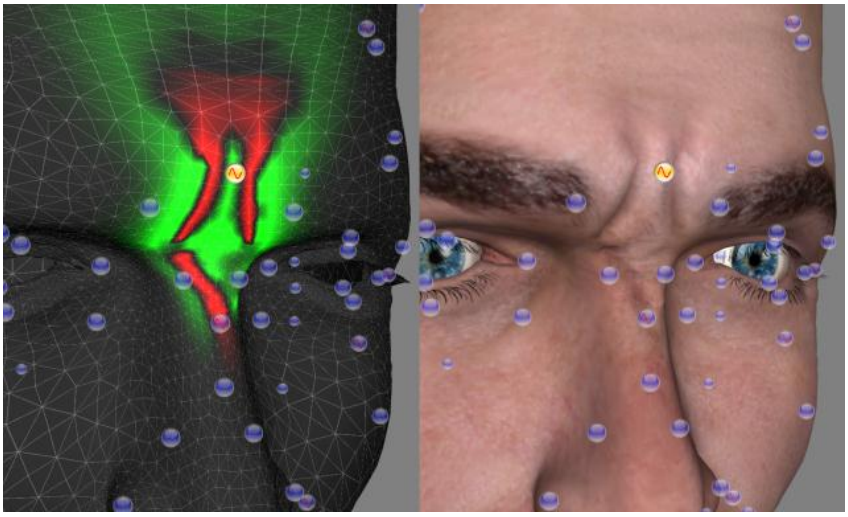


Figure 64 - Influence d'un point clé dédié aux rides (gauche). Le rouge indique une influence négative, le vert, une influence positive. La partie de droite montre la déformation géométrique résultante.

Les emplacements des points-clés dédiés aux rides ont été définis sur la base des descriptions d'Ekman des rides associées aux expressions faciales (Ekman et Friesen, 1975). Trois points clés sont définis sur le front (gauche, centre, droite). Deux points clés sont définis sur le nez (rides de froncement des sourcils et rides de froncement du nez). Deux points clés sont définis pour les pattes d'oie (gauche et droite). Deux points clés sont définis pour la gestion des rides nasolabiales (gauche et droite). Enfin, deux points clés sont définis pour rendre les lignes des rides sous les yeux (gauche et droite).

Les rides 3D géométriques permettent un rendu plus réaliste que le bump mapping. Elles sont visibles à la fois dans une vue de profil et dans une vue de face. La Figure 65 montre une vue de côté de rides géométriques en utilisant la technique 3D (droite), comparée aux mêmes rides rendues en utilisant la technique 2D dite de bump mapping (gauche) que nous avons utilisée dans les expériences décrites précédemment.

Figure 65 - Rides du front. 2D bump mapping (gauche) versus 3D géométrique (droite)

La déformation géométrique 3D est possible grâce aux maillages de haute résolution que nous utilisons: 20.000 polygones pour le visage seul. Utiliser des maillages de basse résolution ne permettrait pas ce rendu de rides géométriques sans utiliser de technique de raffinement de maillage à la volée, très coûteuses en temps de calcul. De plus, cette méthode de rendu de rides 3D n'augmente pas le temps de calcul nécessaire au GPU pour déformer la géométrie. En effet, nous utilisons des points-clés pour les déformations faciales et les rides d'expression. Ainsi, les programmes GPU d'animation GPU (vertex shaders) ne font aucune distinction entre les points-clés MPEG4 et points-clés dédiés aux rides.

Dans notre approche, les vecteurs normaux du maillage déformé (pour créer des effets d'ombrage appropriés) sont calculés dynamiquement. Contrairement aux techniques 2D de type *bump mapping*, notre méthode de rendu de rides 3D permet de visualiser les rides à partir de différents points de vue, sans nécessiter la complexité d'un modèle physique (Terzopoulos et Waters, 1993). En contrepartie, elles demandent un temps de saisie manuelle dont la durée dépend du nombre de zones de rides choisies et de la finesse de leur contrôle.

Les expressions faciales présentées dans les sections suivantes ont été conçues en utilisant ce système. Elles ont été rendues en utilisant le module de rendu temps réel à partir duquel nous avons extrait les images statiques utilisées dans les expérimentations décrits par la suite.

3.5.2 Rides géométriques versus « bump mapping »

Pour évaluer l'apport de cette nouvelle technique de rides en 3D sur la perception des émotions et la préférence utilisateur, une étude comparant rides 2D et rides 3D a été menée en collaboration avec l'ENSC de Bordeaux, dans le cadre du stage de trois étudiants (Camille Barrière, Charlotte Jacobé de Naurois et Laure Mathieu). Cette section présente donc rapidement les hypothèses et les résultats de cette étude.

3.5.2.1 Hypothèse

Dans notre précédente étude, nous avons montré que les rides d'expressions n'augmentent pas la reconnaissance catégorielle des émotions, mais impactent la préférence des utilisateurs. Ainsi l'hypothèse de cette étude est la suivante :

H1. Nous supposons que les émotions sont aussi bien reconnues avec les deux modes de rides mais que le mode 3D est préféré par les participants.

3.5.2.2 Protocole

Les étudiants de l'ENSC ont donc conçu un ensemble d'images statiques présentant le visage de MARC exprimant sept émotions différentes (Dégout, Peur, Surprise, Colère, Joie, Bouleversement, Touché) depuis trois angles différents (Face, 45°, Profil), et en utilisant les deux modes de rides (*bump mapping* 2D, et 3D géométrique). L'étude a été menée sur 35 élèves de l'ENSC.

3.5.2.3 Résultats et discussion

En termes de préférence, nous n'obtenons pas de résultats significatifs. 18 sujets ont préféré le mode géométrique 3D contre 17 sujets préférant le mode *bump mapping* 2D. Les justifications sont variées, mais aucun sujet n'a rapporté avoir remarqué la différence entre les deux types de rides. Le réalisme atteint par la technique 2D du *bump mapping* semble donc suffisant pour faire plafonner la préférence utilisateur.

Les mesures perceptives effectuées suggèrent en revanche que le mode de rides fait une différence dans la reconnaissance catégorielle des émotions. En effet, 65,4% des expressions présentées avec les rides géométriques 3D ont été reconnues en termes de catégorie émotionnelle, contre 53,1% des expressions présentées avec des rides affichées en *bump mapping 2D*.

Notre hypothèse est donc réfutée.

Néanmoins, la différence observée entre les deux modes de rides semble contredire les résultats de notre précédente étude, dans laquelle les rides d'expression ne modifiaient pas les taux de reconnaissance. D'autre part, l'effet observé semble plus important pour la Surprise que pour les autres émotions. Nous obtenons en effet un taux de reconnaissance catégoriel de 94% pour les rides 3D, contre 69% pour les rides 2D.

Pour l'expression de Surprise, nous pouvons émettre l'hypothèse que la présence de rides marquées au niveau du front est responsable de cet effet. En effet, un tiers des expressions étaient présentées de profils. Or, les rides présentes sur le front de l'agent virtuel dans l'expression de Surprise sont nettement moins visibles de profil en mode 2D qu'en mode 3D. Nous pouvons donc supposer que les rides géométriques 3D apportent une information supplémentaire lorsque l'agent est observé de profil augmentant ainsi le taux de reconnaissance catégorielle de l'émotion.

La plupart des expérimentations perceptives explorent la perception de visage vu de face et en gros plan. Cependant, comme le suggère cette étude, il est important d'étudier la perception des visages dans des situations moins optimales pour la perception. Par exemple, dans une application où l'agent peut se déplacer dans une scène, par exemple avec d'autres personnages virtuels. Ainsi, son visage peut ne pas toujours être affiché ni en gros plan, ni de face.

3.6 Perception des expressions vues de face, de profil, de près et de loin

3.6.1 Objectifs

L'expérience présentée dans cette section vise à évaluer l'impact du point de vue de l'observateur sur la perception émotionnelle des expressions faciales. L'unique variable que nous avons choisi de manipuler dans cette expérimentation est le point de vue d'observateur, pour lequel on considère deux facteurs: l'angle de vue (face VS profil) et la distance (gros plan VS plan éloigné).

3.6.2 Protocole

Dans cette étude, nous avons sélectionné 4 émotions: Joie, Colère, Tristesse, et Relaxation. Ces émotions ont été choisies car elles sont équilibrées en termes de Valence et d'Activation (Russell et Mehrabian, 1977). La Joie est une émotion positive avec activation élevée. La relaxation est une émotion positive avec activation faible. La colère est une émotion négative avec activation élevée. La tristesse est une émotion négative avec activation faible.

Une fois de plus, la spécification des expressions faciales pour chacune des trois émotions de base (Joie, Colère et Tristesse) a été inspirée par les travaux d'Ekman (1975). La spécification de l'expression faciale de la Relaxation a été inspirée par la base de données vidéo MindReading (Baron-Cohen, 1997).

Nous avons utilisé 16 images (Figure 66):

8 vues de face + 8 vues de profil

8 gros plans + 8 vues éloignées

36 sujets ont participé à l'expérience (12 femmes, 24 hommes, âge moyen 27 ans). L'ordre de présentation des 56 images a été tiré au hasard. Chaque image a été affichée avec une taille de 560x650 pixels. Outre l'image, un questionnaire a été affiché avec un ensemble de huit curseurs. Pour chaque image, les utilisateurs devaient évaluer les niveaux de Valence, d'Activation et de Dominance qu'ils percevaient. Ils devaient également évaluer dans quelle mesure ils percevaient chacune des quatre émotions et estimer la confiance dans leurs évaluations. Chaque réponse a été donnée en utilisant une échelle de Likert de cinq points. Afin de détecter les mélanges d'émotions, nous avons utilisé l'échelle de Likert plutôt qu'un choix forcé. Ainsi, les sujets ont été en mesure de modérer leurs réponses et de rapporter des mélanges d'émotions.

3.6.1 Hypothèses

H1: Le gros plan du visage doit être mieux reconnu que le point de vue plus éloigné.

En effet, on suppose que les détails visibles en gros plan permettront de mieux détecter l'émotion exprimée par le visage.

H2: Le point de vue (face-profil) devrait avoir un impact sur la reconnaissance des émotions.

Cette hypothèse est issue de l'étude menée en collaboration avec l'ENSC, et qui suggère que les rides 3D augmentent les taux de reconnaissance catégoriels vue de profils.

H3: la confiance que les sujets ont dans leurs réponses devrait être plus faible pour des vues de côté que pour les vues de face.

En effet, outre le fait que la vue de profil masque une partie du visage de l'agent, les interactions humaines sont généralement effectuées face à face. Ainsi, nous supposons que la perception des sujets devrait être altérée en vue de profil. Ainsi, les sujets devraient avoir moins confiance en leur perception qu'avec les vues de face.

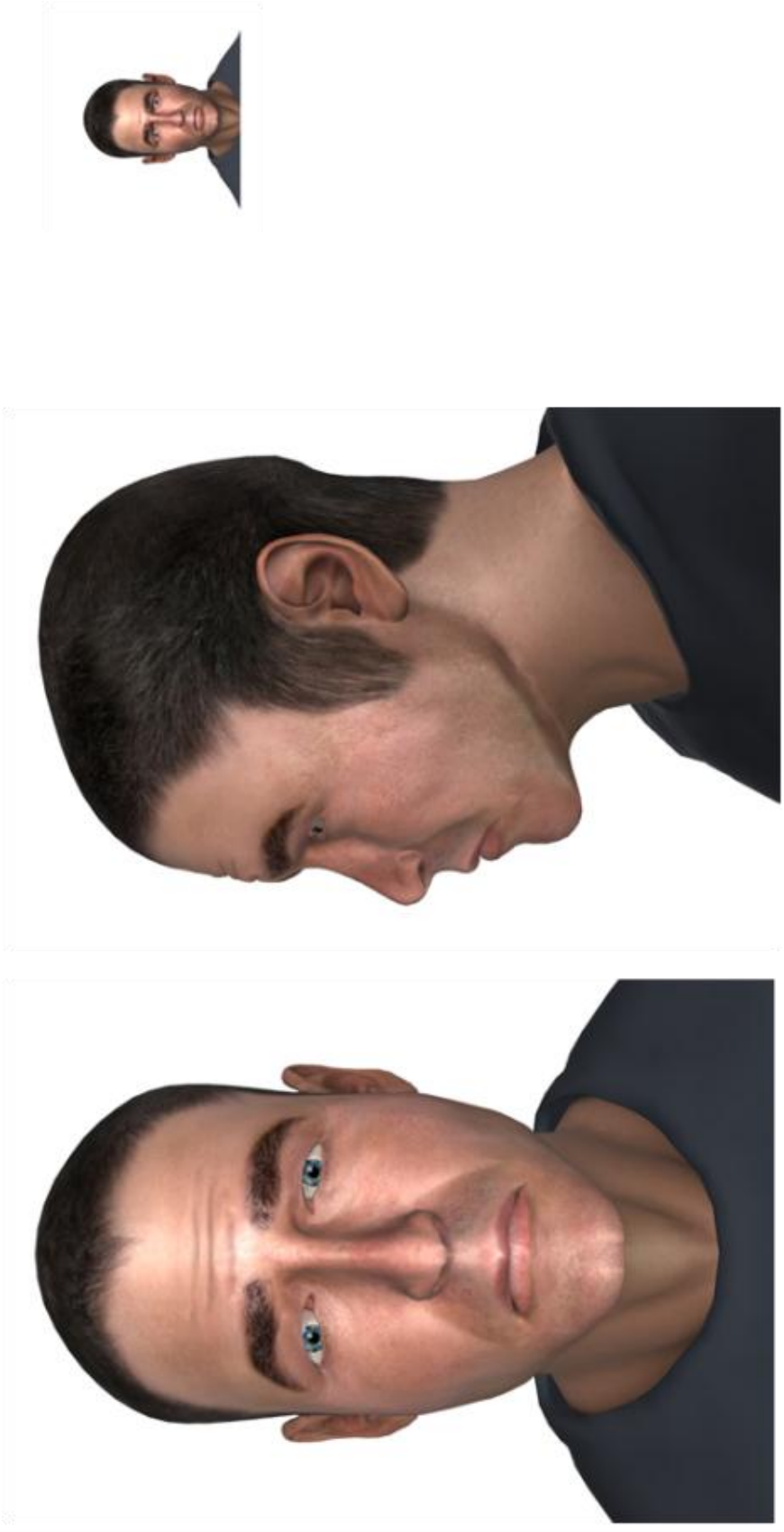


Figure 66 – images de la Tristesse. Gros plan de face (à gauche), Gros plan de profil (au centre) et vue distante de face (à droite)

3.6.2 Résultats

3.6.2.1 Données de reconnaissance catégorielle

Nous avons calculé une ANOVA sur des mesures répétées afin d'estimer les taux de reconnaissance des émotions (catégoriels et dimensionnels) dans chaque modalité.

La Colère a été l'émotion la plus perçue dans nos images de Colère quel que soit le point de vue (près / loin, face / profil). La Tristesse était l'émotion la plus perçue dans nos images de Tristesse quel que soit le point de vue (près / loin, face / profil). La Joie a été l'émotion la plus perçue dans nos images de Joie quel que soit le point de vue (près / loin, face / profil).

Les images de Relaxation étaient perçues comme des mélanges composés à égalité de Joie et de Relaxation quel que soit le point de vue (près / loin, face / profil). Les images conçues pour exprimer la Relaxation ont souvent été perçue comme de la Joie, également une émotion positive.

3.6.2.2 Retour sur les hypothèses

H1. Nous observons que les images en gros plan du visage ne sont pas mieux reconnues que des vues éloignées. Ainsi, nous pouvons émettre l'hypothèse que l'utilisation d'expressions prototypiques d'émotions permet d'afficher les personnages virtuels à distance sans pour autant réduire la reconnaissance des émotions. L'hypothèse H1 est donc réfutée.

Ce résultat est important, il indique en effet que l'utilisation de personnages virtuels de taille réduite permet une expression émotionnelle de qualité comparable aux visages en gros plan. L'utilisation de personnages virtuels ne requiert donc pas nécessairement l'utilisation d'une surface importante de l'écran, et peuvent donc être combinés avec une application présentée sur le même écran. Cette observation concorde avec les travaux de Clavel et al. (2009) qui suggèrent que le visage permet une bonne reconnaissance catégorielle même lorsqu'il est affiché de petite taille.

H2. Nous avons observé que le point de vue influe sur la reconnaissance des émotions. En effet, la perception de la Colère était équivalente quel que soit le point de vue. Mais la Joie a été mieux perçue en vue de profil que de face. De plus, la vue de face de l'expression de Relaxation est mieux reconnue que la vue de profil. Enfin, la tristesse était mieux reconnue en vue de face qu'en vue de profil. Ainsi, le point de vue a une influence sur la reconnaissance des émotions. L'hypothèse H2 est confirmée.

Nous n'avions cependant pas émis d'hypothèse sur la nature de l'influence. Nous observons que cette influence varie en fonction de la catégorie émotionnelle. Il est donc nécessaire de modérer ce résultat. En effet, nous n'observons pas de schéma d'influence précis, et donc, l'influence observée peut être due à nos stimuli.

H3. La confiance était significativement moins élevée sur les images vues de profil que pour les images vues de face. Cela confirme l'hypothèse H3.

Nous observons également un effet d'interaction entre nos deux facteurs (face/profil, loin/proche). Si on considère les quatre conditions séparément (face/loin, face/proche, profil/loin, profil/proche), nous observons que la confiance rapportée par les sujets est équivalente pour trois d'entre eux. Seul le mode *vue de loin et de profil* donne lieu à une chute significative de la confiance reportée.

Ainsi, nos résultats montrent que si les taux de reconnaissance catégoriel sont globalement comparables quel que soit le point de vue de l'observateur, la combinaison des facteurs donne une information intéressante sur le niveau de confiance des sujets dans leurs réponses.

3.6.3 Conclusions

En conclusion, nous avons observé des taux élevés de reconnaissance émotionnelle quel que soit le point de vue et la distance. Pourtant, ces facteurs ont un impact sur la confiance rapportée par les utilisateurs.

Dans cette expérience, nous n'avons utilisé que deux angles (face et profil). L'utilisation d'angles intermédiaires pourrait fournir des résultats intéressants sur la limite à laquelle l'angle de vue devient problématique pour la perception et conduit ainsi à une diminution de la confiance rapportée.

Plusieurs études ont souligné l'impact de la dynamique par rapport à des stimuli statiques sur la perception des expressions émotionnelles (de Melo et al. 2009). Il serait donc pertinent de mener une étude similaire en utilisant des stimuli dynamiques plutôt que des images statiques. Cependant, la dynamique des expressions faciales ne devrait pas être limitée à l'interpolation d'expressions prototypiques.

Des études similaires doivent être menées en utilisant des photos d'expressions de visages humains, en utilisant des simulations physiques biomécaniques du visage, et en utilisant d'autres modèles de visage. Cela permettrait d'évaluer l'impact des modèles paramétriques de rides (tel que celui proposé dans cette étude) sur la perception des expressions faciales des émotions. Cependant, l'utilisation des visages humains introduirait un nouveau défi: comment contrôler précisément tous les paramètres du visage.

3.7 Perception de la dynamique des expressions faciales

Dans l'étude que nous avons présentée dans la section précédente, nous avons évalué comment les sujets percevaient les expressions faciales statiques d'émotion lorsqu'on introduit une difficulté visuelle en présentant le visage de plus loin, ou de profil.

L'objectif général de cette thèse est de nous orienter vers un système interactif. Il est donc important d'étudier la perception de la dynamique des expressions faciales. Par exemple, Bänziger et al. (2009) ont montré que l'expression de Colère est moins bien reconnue dans des images statiques que dans des vidéos dynamiques. L'expérimentation présentée dans cette section porte sur cet aspect de la perception humaine de la dynamique des expressions faciales.

Cette expérimentation a été conçue et mise en place avec Michel-Ange Amorim et Caroline Giroux de l'UFR STAPS de l'Université Paris-Sud XI. Les passations ont été effectuées par Caroline Giroux dans le cadre de son stage de Master. Nous ne décrirons donc pas en détails le protocole et les résultats de ces travaux. Cette section sera focalisée sur les objectifs et les conclusions de ceux-ci. Les détails de l'expérimentation peuvent être trouvés dans l'article publié à la conférence internationale « Computer Animation and Social Agent 2010 ».

3.7.1 Objectifs

Plusieurs études ont été menées sur la perception humaine de la dynamique, et en particulier sur la perception d'objets en mouvement. Ces études ont mené à la définition du concept de *moment représentationnel* (Freyd et Finke, 1985). Le moment représentationnel est un terme désignant le phénomène selon lequel la représentation mentale de la position finale mémorisée d'un objet en mouvement ayant disparu subitement se trouve déplacée dans le sens du mouvement. Par exemple, lorsqu'un utilisateur observe une balle en mouvement, et que la balle est subitement masquée (par exemple, en éteignant la lumière), l'utilisateur se « souviendra » avoir vu la balle plus loin sur sa trajectoire qu'il n'a réellement pu l'observer.

Le concept de *moment représentationnel* reflète donc notre tendance naturelle à anticiper et extrapoler mentalement les mouvements d'une cible. Le moment représentationnel est un sous-produit du processus d'émulation qui aide notre système de perception temps réel (Wilson et Knoblich, 2005). Il préserve la continuité des événements, et ce en dépit de profonds changements visuels (changement de plan (Jarraya et al., 2005)), occlusion (Graf et al., 2007)).

L'existence d'un tel phénomène pour la perception de la dynamique des expressions faciales émotionnelles, que nous appellerons *moment émotionnel*, n'est pas claire dans les travaux en psychologie perceptive. Une étude rapporte une tendance à percevoir une intensification des expressions faciales (Yoshikawa et Sato, 2008). Une autre étude indique que les observateurs anticiperaient plutôt un retour à une expression plus neutre (Thornton, 1998), ce qui serait donc un moment émotionnel inverse. Étonnamment, même si la catégorie émotionnelle (colère, dégoût, peur, joie, tristesse, et surprise) a été manipulée dans ces premières études, les catégories ne sont pas considérées séparément dans l'analyse des données. Notre expérimentation vise donc à évaluer ce *moment émotionnel* en considérant séparément les catégories d'émotions.

3.7.2 Protocole

Les stimuli utilisés dans cette étude étaient conçus en quatre étapes sur le principe suivant : D'abord, on présente au sujet un signal de démarrage durant 2 secondes. Ce signal est ensuite suivi d'une animation de MARC composé d'une interpolation linéaire d'environ 1 seconde entre l'expression neutre et l'expression de l'émotion cible à une intensité « Cible » variable.

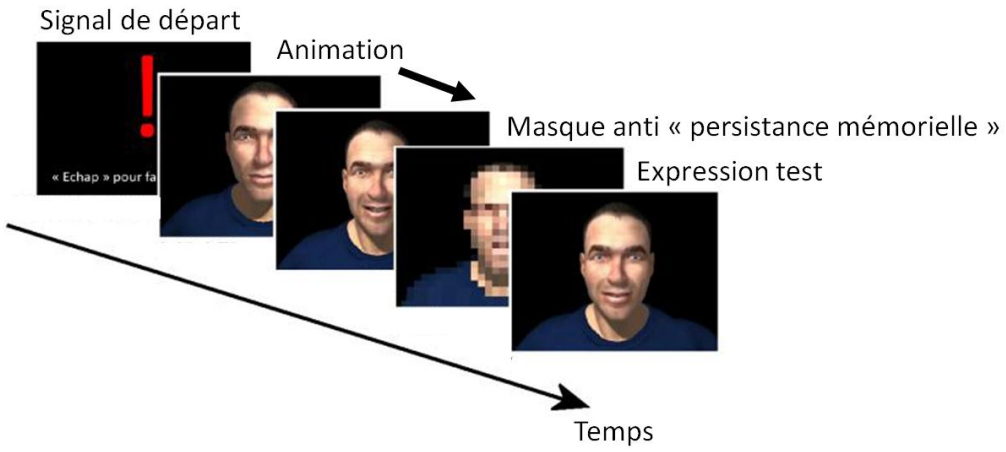


Figure 67 - Déroulement d'une vidéo stimulus pour l'étude de la perception de la dynamique des expressions.

Une image pixelisée du visage est ensuite affichée durant 250ms pour éviter l'effet de persistance mémorielle. Finalement une expression « test » statique de l'émotion cible est affichée, mais a une intensité différente (cible-30%, cible-15%, cible, cible+15%, cible+30%). L'utilisateur doit alors déterminer si l'expression « test » est plus ou moins intense que l'expression « cible ».

3.7.3 Conclusions de l'expérimentation

Les résultats de cette expérimentation suggèrent l'existence d'un *moment émotionnel* dont l'effet serait le suivant. Lorsqu'on présente à un sujet une expression faciale dont la dynamique augmente progressivement, puis que le visage est brusquement masqué, le sujet anticipe l'évolution de l'intensité. Si l'intensité au moment du masquage était forte, le sujet anticipe un déclin de l'intensité expressive. Si l'intensité au moment du masquage était plus faible, le sujet anticipe une augmentation de l'intensité expressive.

Nos mesures montrent que cet effet est observable sur toutes les catégories d'émotions. Ces résultats sont cohérents avec les résultats de Thornton (1998): avec des intensités plus élevées, les sujets anticipent un retour à une expression neutre. On peut supposer que les limites physiques de déformations faciales pourraient expliquer ce phénomène. Le sujet perçoit le stimulus de grande intensité comme la déformation maximale et anticipe un retour à l'expression neutre

Cependant, en analysant les catégories émotionnelles indépendamment, nous observons que l'effet est plus ou moins intense en fonction des catégories. Par exemple, l'effet du moment émotionnel semble plus important avec

la Colère qu'avec la Joie. L'une des explications possibles est que cela est lié aux expressions faciales utilisées: dans l'expression de Joie, la largeur de la bouche fournit un indice visuel fiable qui n'est pas disponible dans l'expression de la Colère. Ainsi, si nous postulons que les sujets se concentrent sur les traits du visage, nous pouvons également supposer que nous n'avons pas seulement mesuré le moment émotionnel, mais également un moment représentationnel des mouvements du visage.

3.7.4 Seconde étude

Pour tester cette hypothèse, nous avons procédé à une seconde expérimentation. Au vu des précédents résultats, une comparaison entre la perception de Colère et de Joie semblait pertinente. En effet, si l'effet d'anticipation mesuré est similaire pour toutes les catégories, il est nettement plus intense pour la colère que pour la Joie. Nous avons donc répété le protocole précédant, mais en utilisant cette fois quatre expressions différentes : deux de Joie, et deux de Colère, conçue de manière à ce que la quantité de mouvement sur le visage soit identique pour les deux expressions de Joie, ainsi que pour les deux expressions de Colère (Figure 68).

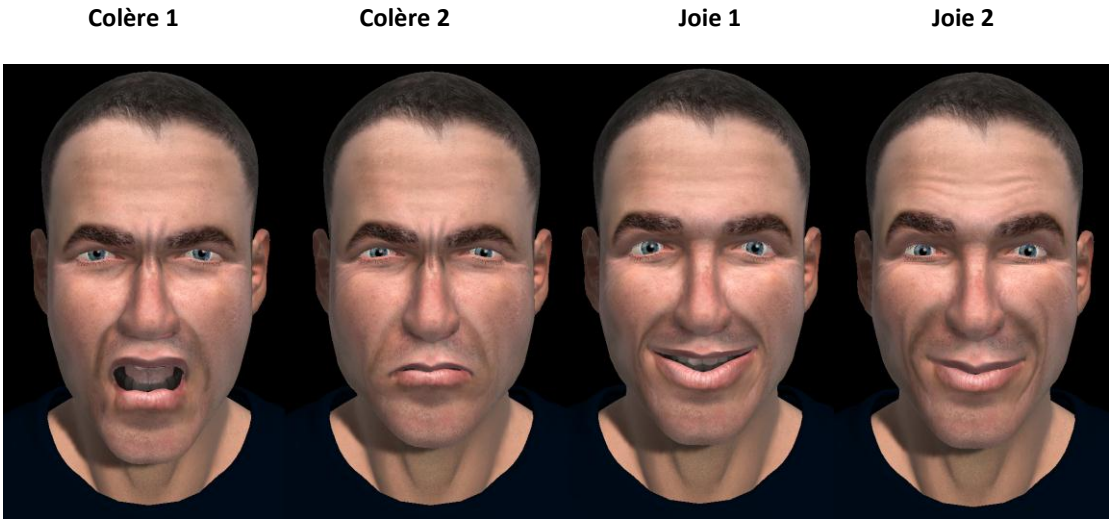


Figure 68 - Deux expressions de Colère et deux expressions de Joie inspirées de Ekman et Friesen (1975)

La quantité de déplacement est calculée en sommant le déplacement de chaque vertex du visage en 2D sur l'écran. Ainsi, seul le déplacement visible est considéré.

La comparaison entre les résultats obtenus lors de la première expérience pour les émotions Joie et Colère et les résultats obtenus à cette deuxième expérience pour les expressions Joie et Colère ont permis de révéler un élément notable : les deux expressions de Joie sont perçues différemment. En effet, même si ces expressions sont similaires, l'une des deux expressions de Joie entraîne un effet d'anticipation plus important que l'expression de Joie de la première étude.

L'analyse des résultats révèle que la présence d'indices visuels spatiaux évidents (ex : ouverture de la bouche) améliorait significativement la précision des réponses, et donc diminue l'effet d'anticipation. L'expression de Joie qui entraine un effet d'anticipation plus important ne comporte pas d'ouverture de la bouche, contrairement aux autres expressions de Joie utilisées. Ainsi, les sujets ont moins d'indices visuels, ce qui réduit la précision de leur perception, et provoque une anticipation plus importante de la dynamique expressive. En d'autres termes, la variation de l'intensité du *moment émotionnel* observé dans nos études ne dépend pas de la catégorie d'émotion mais du mouvement qu'entraîne cette émotion ou l'expression sur le visage de MARC.

Le *moment émotionnel* reflète donc la tendance naturelle des utilisateurs à anticiper la dynamique émotionnelle comme un épisode relativement court, durant laquelle l'intensité maximale est rapidement suivie d'une atténuation de l'intensité expressive. Nos résultats suggèrent donc que l'intensité maximale est suivie d'un déclin. Si on compare les différents modèles dynamiques utilisés en animation faciale catégorielle, nos résultats

semblent donc être en faveur du modèle *attack-decay-sustain-release*, plutôt que du modèle *onset-apex-offset*. En effet, le modèle *attack-decay-sustain-release* présente un pic d'intensité qui semble être inconsciemment attendu par les sujets de nos études.

3.8 Module interactif de contrôle des expressions faciales

La première application interactive conçue durant cette thèse est une application de clonage d'émotion par l'agent virtuel. On distinguera dans cette section le clonage d'expressions du clonage d'émotion. Le clonage d'expressions (Noh et Neumann, 2001) consiste à détecter et mesurer les expressions de l'utilisateur, puis à les reproduire sans y affecter de sens émotionnel. Le clonage d'émotion au contraire, n'utilise pas directement l'expression détectée, mais l'émotion qui lui est associée. Il ne s'agit donc plus de reproduire la forme de l'expression faciale de l'utilisateur, mais de reproduire son contenu émotionnel. Néanmoins, les domaines d'application de ces deux approches sont similaires.

Les applications de clonage expressives et émotionnelles sont importantes pour les systèmes de conférences basées sur des environnements virtuels (André del Valle et al., 2000). La norme MPEG4 pour l'animation faciale a été conçue en partie pour ce type d'application (André del Valle et al. 2001). Ces systèmes présentent l'avantage de réduire la quantité de données transférées entre les différents intervenants. En effet, l'échange de flux vidéo filmés requiert une large bande passante réseau, alors que les paramètres nécessaires à l'animation faciale sont relativement réduits. De plus les environnements virtuels permettent un niveau d'interactivité plus élevé que les systèmes de vidéo conférence, permettant par exemple à l'utilisateur de changer de point de vue dans la scène virtuelle, ou de modifier son apparence rapidement. Le clonage d'expression faciale ou d'émotion semble donc être un enjeu important pour les logiciels d'environnement virtuels sociaux, tels que *Second Life*, ou pour les jeux sociaux développés sur des plateformes telles que *Facebook*.

Le clonage d'expressions/d'émotion repose sur deux problématiques techniques distinctes. D'une part, comment détecter les expressions/émotion de l'utilisateur ? D'autre part, comment les répliquer ? Notre application effectue du clonage d'émotion et non d'expression. Notre objectif était de démontrer les capacités de réaction temps réel de MARC. Pour la détection des émotions de l'utilisateur, nous avons utilisé un logiciel de détection d'émotion à partir des expressions faciales nommé FaceReader. FaceReader produit dynamiquement un fichier journal qui contient le taux de reconnaissance pour six états affectifs : les six émotions de base définies par Ekman (1975). Ce fichier journal est alors analysé dynamiquement par MARC. Les taux de reconnaissance des différentes expressions sont utilisés comme intensité et les expressions sont combinées pour afficher une expression correspondant à un mélange des émotions reconnues par FaceReader. Comme nous l'avons vu, cette approche est du clonage d'émotion, non d'expressions. Ainsi, l'expression de l'utilisateur n'est pas répliquée à l'identique, mais son contenu émotionnel est comparable (Figure 69).

Cependant, notre système est dépendant de la qualité de détection du logiciel FaceReader qui impose un certain nombre de limites. D'une part, l'utilisateur doit exprimer ses émotions de manière exagérée, ce qui n'est pas le cas dans une communication naturelle. D'autre part, le logiciel peut, dans certaines circonstances, détecter trois ou quatre émotions à une probabilité élevée. Le mélange expressif résultant est donc problématique, car il cumule plusieurs expressions faciales, et le résultat obtenu est parfois trop déformé et très peu réaliste.

Néanmoins, cette première application interactive nous permet de démontrer les capacités d'interaction temps réel du moteur d'animation de MARC. En effet, si aucune évaluation formelle de ce système n'a été menée, les retours des différents utilisateurs suggèrent que la réactivité de MARC est très bien perçue. La latence réduite (de l'ordre d'une à deux secondes) semble donner aux utilisateurs le sentiment de pouvoir correctement contrôler l'agent. Certains utilisateurs ont cependant fait remarquer que l'utilisation prolongée de notre système provoque une fatigue des muscles faciaux. Nous avons attribué cette fatigue à l'exagération expressive requise par FaceReader pour reconnaitre les expressions faciales de l'utilisateur.

Pour finir, le clonage d'expression ne permet pas d'explorer différents modèles émotionnel pour la génération d'expressions faciale. Nous n'avons donc pas poursuivit nos recherches dans cette direction.

Figure 69 - Noldus FaceReader reconnait l'émotion de l'utilisateur, et MARC reproduit cette émotion avec ses propres expressions. FaceReader reconnait ici un mélange de Colère et de Dégout.

3.9 Résumé et limites de l'approche catégorielle

Dans cette première version de MARC, nous avons centré nos travaux sur le modèle catégoriel des émotions. Ce modèle est largement utilisé par les agents virtuels expressifs car il permet de représenter un nombre fini d'état émotionnel et de les faire exprimer par un agent virtuel.

Nos recherches nous ont cependant permis de mettre en lumière un certain nombre de résultats concernant la perception humaine des expressions faciales de personnages virtuels. Par exemple, nous avons montré que les sujets anticipent la dynamique d'une expression faciale selon une courbe constituée d'un pic d'intensité expressive suivit d'un retour à un état moins intense. Le phénomène que nous avons appelé le *moment émotionnel* semble donc être en faveur de l'utilisation du modèle dynamique *attack-decay-sustain-release*, plutôt que de l'utilisation du modèle dynamique *onset-apex-offset*. Nous avons également mené une étude sur la perception des expressions faciales depuis différents angles de vue, et différentes distances. Nous avons observé que l'utilisation d'un visage en gros plan et de face n'est pas nécessaire pour obtenir une bonne reconnaissance catégorielle des émotions. En effet, même de profil et/ou de loin, les catégories émotionnelles ont été correctement reconnues par nos sujets. Néanmoins, notre étude révèle que les sujets ont moins confiance en leur perception lorsque le visage est affiché de loin et de profil, malgré l'absence de différence significative dans les résultats de reconnaissance catégorielle.

Ce chapitre pose également les bases de l'animation faciale réaliste et temps réel de MARC, en utilisant des méthodes récentes basé sur le GPU. Ces travaux nous ont permis d'obtenir un réalisme visuel supérieur à celui des agents virtuels interactifs existants pour lesquels le réalisme visuel n'est pas un objectif principal (par exemple, Greta ou Max). Nous avons ainsi pu évaluer l'impact de certains indices visuels subtils sur la perception humaine des expressions faciales de personnages virtuels. Ainsi, nous avons mis en lumière que l'utilisation de rides d'expression (réalistes ou non) n'augmente pas la reconnaissance des catégories d'émotion. Cependant, nous avons montré que l'utilisation de rides d'expression réalistes augmente la préférence des utilisateurs (comparé à l'absence de ride ou à la présence de rides non réalistes) ainsi que l'intensité expressive perçue par les sujets humains.

Ce résultat nous a donc amené à développer un système de rides 3D, également basé sur l'utilisation du GPU. L'utilisation de rides 3D nous a permis d'obtenir des rides d'expressions visibles de différents angles de vue. Pour cela, nous avons proposé un système de points-clés, inspiré des points MPEG4, mais dédié aux rides d'expressions. Ainsi, notre technique de rides pourrait être ajoutée à d'autres systèmes d'animation faciale basés sur MPEG-4, sans nécessiter de modifications majeures du système d'animation.

Pour finir, nous avons utilisé MARC dans un démonstrateur de clonage d'expressions faciales. Ce démonstrateur nous a permis de démontrer l'utilisabilité de MARC dans un contexte d'interaction temps réel. Cependant, notre objectif est d'explorer plusieurs modèles émotionnels. Ainsi, si ces travaux nous ont permis de mettre en place la base de notre système d'agents virtuels expressifs, nous nous somme ensuite tournés vers d'autres approches des émotions.

Chapitre 4. Approche dimensionnelle des émotions pour l'animation faciale : le modèle P.A.D.

Sommaire du chapitre

Publications et présentations associées

- M. Courgeon, C. Jacquemin, J-C. Martin, (2008) *User's Gestural Exploration Of Diffirent Virtual Agents' Expressive Profiles*, in: Proceedings of the Internationnal 7th International Conference on Autonomous Agents and Multiagent Systems, vol 3, pp 1237-1240, Estoril, Portugal, 12-16 mai 2008
- M. Courgeon, J-C. Martin, C. Jacquemin, (2008) *Virtual Humans: Expressivity, Interactivity and Realism*, in: Digiteo Annual Forum 2008, Orsay, France, 2 octobre 2008
- M. Courgeon, J-C. Martin, C. Jacquemin, (2008) *User's Gestural Exploration of Different Virtual Agents' Expressive Profiles*, in: Proceedings of the Speech and Face to Face Communication Workshop in memory of Christian Benoît, Grenoble, France, 27-29 octobre 2008

4.1 Intérêts de l'approche dimensionnelle

Contrairement à l'approche catégorielle qui considère les émotions comme un ensemble discontinu d'états indépendants, en nombre fini, l'approche dimensionnelle considère que les émotions sont un continuum, généralement à plusieurs dimensions (Scherer, 2010). De plus, ce continuum permet de mettre en relation les émotions entre elles. Par exemple, une émotion peut être qualifiée de plus dominante qu'une autre, ou bien il est possible de comparer les distances entre plusieurs émotions.

Pour finir, contrairement à l'approche catégorielle, l'approche dimensionnelle ne soutient pas l'existence d'un processus mental spécifique à chaque émotion. Les modèles informatiques inspirés de l'approche dimensionnelle ne requièrent donc pas de spécifier les conditions de déclenchement de chaque émotion considérées, mais peuvent spécifier un traitement générique pour l'ensemble des émotions.

4.2 Architecture de MARC v2 : approche dimensionnelle

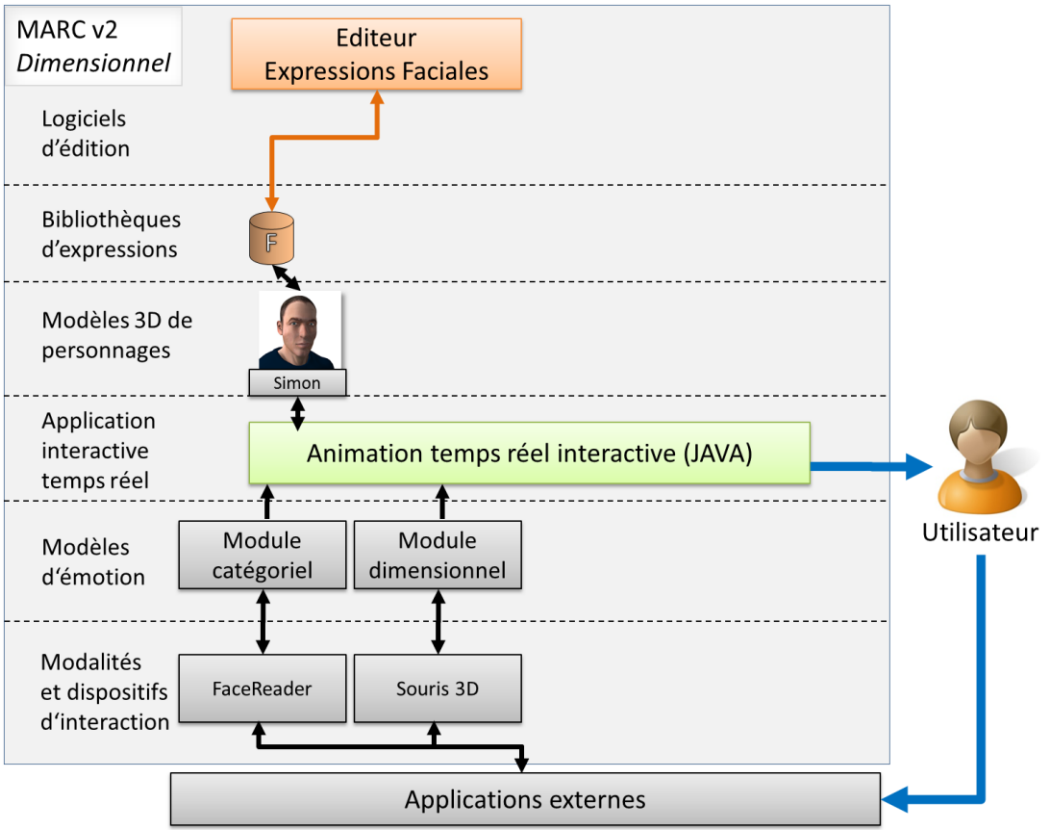


Figure 70 - Architecture de MARC v2 : Modèle Dimensionnel

4.2.1 Objectifs

L'objectif de nos travaux sur l'approche dimensionnelle des émotions est d'étudier ce qu'apporte l'utilisation d'un espace continu pour représenter les émotions. Dans notre deuxième version de MARC (Figure 70), deux axes peuvent bénéficier de l'approche dimensionnelle : l'animation faciale et l'interaction affective.

Comme nous l'avons vu, l'approche dimensionnelle permet de structurer l'espace des émotions, et permet ainsi certains raisonnements sur les émotions que l'approche catégorielle ne permet pas. Par exemple, la transition entre deux émotions ne peut s'effectuer qu'en suivant une trajectoire continue à travers l'espace dimensionnel qui les sépare. Cette approche apporte donc des règles implicites de continuité entre les émotions. Le modèle émotionnel ayant une influence directe sur l'animation faciale associée, l'animation résultante tiendra compte

des états émotionnels intermédiaires présents sur la trajectoire continue reliant une émotion à une autre. Ainsi l'animation faciale est contrainte par continuité au même titre que l'état émotionnel.

De plus, l'utilisation d'un espace à plusieurs dimensions nous permet d'appliquer des contraintes spéciales. Par exemple, il est possible de rendre certaines zones de l'espace émotionnel plus « denses » que d'autres, et ainsi, faciliter ou inhiber les états émotionnels contenus dans certaines régions de l'espace dimensionnel. Ainsi, il devient possible d'ajouter une modulation expressive dans l'espace émotionnel, et ainsi, de créer différents profils expressifs applicables à l'agent virtuel.

4.2.2 Module émotionnel dimensionnel

Pour représenter l'espace des émotions, nous avons dû choisir un système adapté à nos besoins. Nous souhaitions un espace linéaire, borné, et facile à visualiser pour l'utilisateur de notre système. Nous avons retenu l'espace P.A.D. (Russell et Mehrabian, 1977). La finalité recherchée étant de pouvoir manipuler intuitivement l'agent, l'espace P.A.D. présente l'intérêt d'être à la fois simple à visualiser, c'est un cube, et d'offrir une très bonne représentation continue de l'espace des émotions, bien adaptée à l'interaction avec des expressions fines d'émotion et des transitions progressives entre plusieurs émotions.

L'utilisation de trois dimensions permet en outre une plus grande richesse du modèle que les modèles à deux dimensions tels que le modèle Valence/Activation (Cowie et al., 2000), dans lequel la Peur et la Colère sont très proche dans l'espace. Le modèle P.A.D. a précédemment été utilisé dans plusieurs travaux sur les agents animés expressifs (Becker-Asano & Wachsmuth, 2008).

En utilisant le modèle P.A.D., nous donnons à l'utilisateur contrôle la position d'un curseur à l'intérieur du cube P.A.D. Pour simplifier la représentation mentale de l'utilisateur, nous avons choisi de disposer les émotions dans les huit coins du cube P.A.D (Figure 71). D'après Russell et Mehrabian (1977) ces émotions sont localisées à diverses positions à l'intérieur de l'espace P.A.D. Nous avons choisi de les projeter sur le coin du cube le plus proche de chaque émotion, en utilisant une méthode similaire à celle proposée par Heudin (2004), mais en utilisant un ensemble d'émotions différent. Afin de réaliser l'animation faciale associée à la dynamique émotionnelle contrôlée par l'utilisateur, il est alors nécessaire de calculer l'activation de chaque émotion en fonction de la position du « curseur » de l'utilisateur.

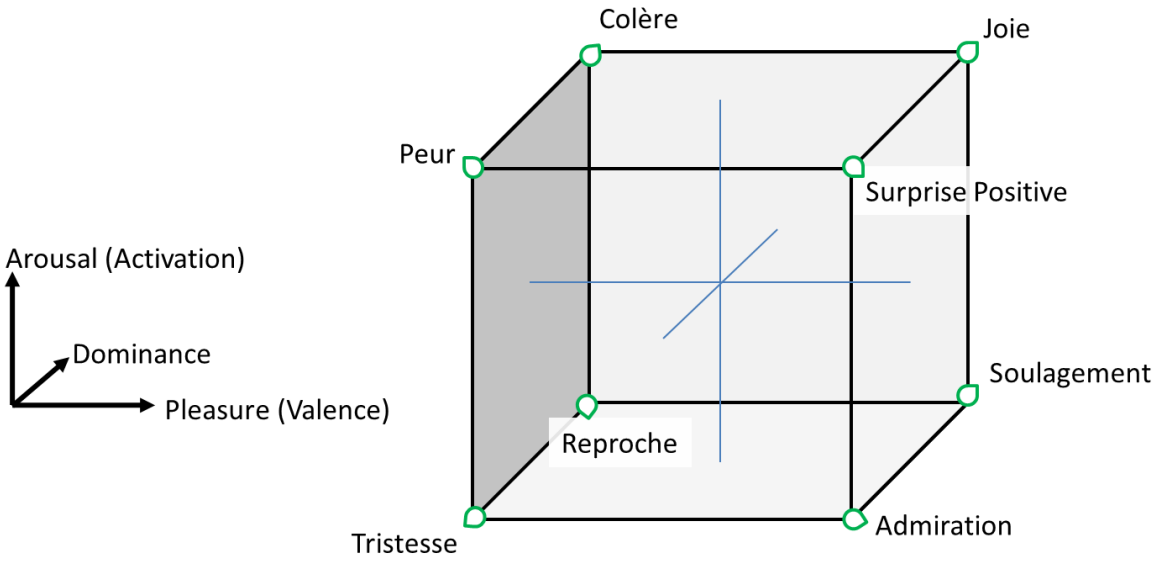


Figure 71 - Positionnement des émotions dans les coins du cube P.A.D. (Russell et Mehrabian, 1977)

Le calcul de l'activation de chaque émotion se fait par un calcul du volume du sous cube formé par le curseur (Figure 72 : cercle vert) et le coin opposé è l'émotion dans le cube. Ce volume est égal au volume total si le curseur et l'émotion sont confondus, et décroît en fonction de l'éloignement du curseur. La Figure 72 montre l'application de cette technique en projection 2D, pour simplifier la visualisation.

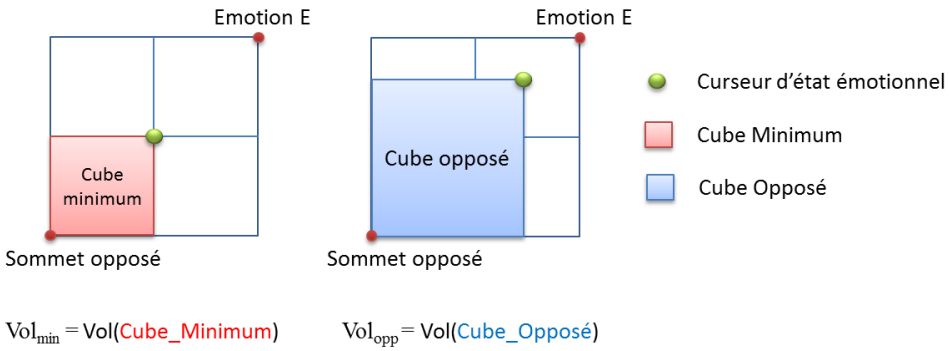


Figure 72 - Schéma explicatif de la technique du cube opposé (Projection 2D)

Le calcul de l'activation d'une émotion « E » s'écrit donc mathématiquement :

Volume du cube maximum (Curseur dans le meme coin que l'émotion E) :

$$Vol_{max} = 2.0 \times 2.0 \times 2.0 = 8.0$$

Volume du cube minimum (Curseur au centre du cube P.A.D.) :

$$Vol_{min} = 1.0 \times 1.0 \times 1.0 = 1.0$$

Volume du cube opposé a l'émotion E :

$$Vol_{opp} = (P_{cur} - P_{opp}) \times (A_{cur} - A_{opp}) \times (D_{cur} - D_{opp})$$

Où $(P_{cur}, A_{cur}, D_{cur})$ sont les coordonnées du curseur dans l'espace P.A.D.
Et $(P_{opp}, A_{opp}, D_{opp})$ sont les coordonnées du sommet opposé à l'émotions E dans l'espace P.A.D.

Activation finale de l'émotion E :

$$Act_{EmotionE} = \begin{cases} (Vol_{cub} - Vol_{Min}) / (Vol_{Max} - Vol_{Min}) & \text{Si } Vol_{cub} >= Vol_{Min} \\ 0 & \text{Sinon} \end{cases}$$

Ainsi, l'activation de l'émotion E est toujours comprise entre 0 et 1.

4.2.3 Migration de MARC vers une implémentation JAVA/OpenGL

Avec l'augmentation des calculs requis pour effectuer l'animation faciale et les pré-calculs associés, l'utilisation du Virtual Choreographer est devenue problématique. En effet, si le Virtual Choreographer est un logiciel efficace pour la gestion de scènes 3D simples, il n'a pas été conçu pour l'animation faciale complexe. Ses performances sont en effets limitées, et la programmation de scripts dans le langage XML associé devient vite très lourde. Nous avons donc exploré d'autres possibilités, et nous avons finalement opté pour l'implémentation JAVA/OpenGL. En effet, cette combinaison nous permettait de conserver l'aspect multi plateforme de MARC (Windows/Linux).

Si JAVA n'est pas généralement réputé pour sa rapidité d'exécution, l'utilisation intensive du GPU permet d'obtenir de très bonnes performances. Nous avons choisi d'utiliser l'API OpenGL 3.2 pour accéder au GPU, et le langage GLSL pour la programmation des *shaders*. OpenGL est fourni sous forme de bibliothèques natives (fichiers .dll, .so) à travers la bibliothèque open-source *Light Weight Java Game Library* (JWGL[6]).

Si cette migration a représenté une charge conséquente de programmation, elle nous a ouvert la voie pour l'implémentation de modèles plus complexes en animation et en modélisation informatiques des émotions. Nous avons donc ainsi obtenu une plateforme plus performante et plus homogène, puisque l'éditeur d'expression faciale et les modules émotionnels de MARC étaient déjà implémentés avec cette technologie JAVA/OpenGL.

4.3 Dispositif de contrôle continu des expressions

4.3.1 Module P.A.D. et manipulation directe

Comme nous l'avons vu en état de l'art, plusieurs types d'applications peuvent bénéficier de dispositif de contrôle des expressions faciales de l'agent virtuel (Jeux vidéo, mondes virtuels sociaux, etc.). Le modèle P.A.D. permet une représentation continue de l'espace des émotions dans un espace 3D qui nous a semblé adapté aux dispositifs d'interaction 3D actuels (joystick, souris 3D…) Le dispositif d'entrée doit donc permettre la manipulation d'un curseur dans un espace 3D. Nous avons expérimenté plusieurs dispositifs. Le premier dispositif que nous avons testé est un joystick. Un joystick classique contient en effet 3 axes de rotations. Le curseur se déplaçant selon trois translations, la manipulation au joystick est très peu intuitive. Le second dispositif testé se nomme SpaceNavigator (Figure 73). C'est une souris 3D possédant six degrés de liberté (3 rotations / 3 translations). Ce dispositif permet d'avoir une correspondance directe « 3 translations / 3 translations ». Nous appellerons ce type de relation un « mapping » direct entre deux espaces tridimensionnels. En effet, l'utilisation des trois degrés de translation du dispositif contrôlent directement la translation du curseur émotionnel. Les degrés de liberté en rotations disponibles sur le dispositif sont ignorés par notre système, afin d'en simplifier l'utilisation.

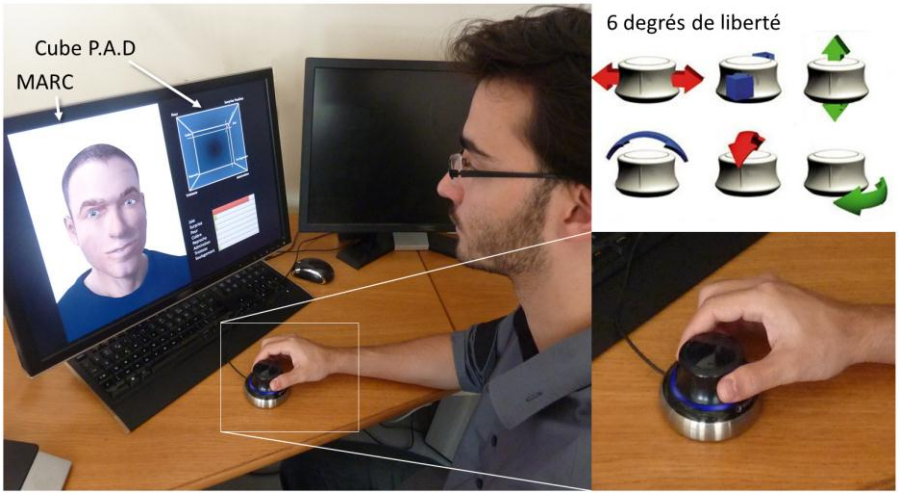


Figure 73 – Utilisateur contrôlant MARC en utilisant le SpaceNavigator (Souris 3D) de 3D Connexion

Ainsi, contrairement aux approches utilisées par Bee et al. (2009) et Jacquemin (2007), notre approche ne permet donc pas un contrôle direct de l'expressivité, mais fait une indirection en utilisant un modèle émotionnel. Ainsi, l'utilisateur ne se concentre pas sur les paramètres expressifs, mais sur l'intention communicative émotionnelle.

[6] http://lwjgl.org

4.3.2 Système de « profils expressifs » individuels

Pour pouvoir moduler l'expressivité contrôlée par l'utilisateur via le dispositif d'interaction, nous avons développé un système de profils expressifs, permettant de donner une individualité à l'agent virtuel. Les réactions expressives de l'agent sont modulées en fonction des caractéristiques de son profil expressif. Les profils expressifs définissent ainsi la réponse de l'agent aux commandes directes de l'utilisateur. Ces modulations sont effectuées de manière indépendante pour chaque émotion. Un profil expressif est défini d'une part par une courbe de modulation d'intensité, et d'autre part par deux valeurs de modulation de la dynamique.

La modulation d'intensité définit la réponse de l'agent à une émotion à l'aide d'une courbe de réponse (fonction de [0-1] dans [0-1]) (Figure 74). La courbe est définie par une fonction de Bézier à quatre points. Les points 1 et 4 dont contraints sur l'axe horizontal (respectivement à 0 et à 1), le reste des points peut être édité librement. La courbe est utilisée pour moduler l'activation de l'utilisateur. Par exemple, sur la Figure 74, la courbe module l'activation de façon à ce que si l'utilisateur demande une activation de 0.5 (50%), l'activation réelle transmise au module d'animation est d'environ 0.35 (35%) (Intersection de la courbe rouge avec la ligne grise médiane verticale. La courbe présentée en exemple aura pour effet d'atténuer l'activation de l'émotion lorsqu'elle est activée à basse intensité.

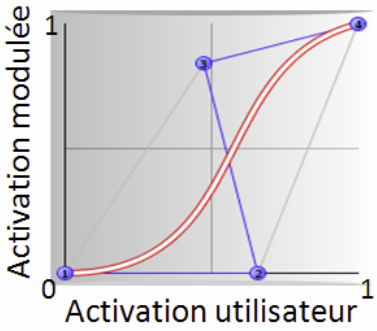


Figure 74 - Courbe de modulation de l'activation d'une émotion. Huit courbes permettent de définir un profil expressif, une courbe pour chaque émotion aux coins du cube P.A.D.

De plus, l'activation des émotions suit à nouveau une dynamique de type *onset-apex-offset*. Nous avons donc ajouté la possibilité de définir la pente maximale de ces deux phases. Deux valeurs peuvent donc être ajustées : la pente d'*onset* (vitesse maximale de transition positive vers l'émotion) et la pente d'o*ffset* (vitesse maximale de décroissance de l'émotion).

Les vitesses d'*onset* et d'*offset* permettent de créer une inertie plus ou moins importante d'une émotion. Par exemple, pour un profil expressif « positif », l'émotion Colère a un *onset* faible et un *offset* rapide. Ainsi, l'émotion doit être activée pendant une durée importante pour être activée à son intensité maximale, et une fois relâchée, son intensité va très rapidement décroitre. Au contraire, pour ce profil positif, la Joie aura un *onset* élevé et un *offset* faible. Ainsi la Joie sera très rapidement activée à son niveau maximal, et mettra un certain temps à revenir à une position neutre.

4.4 Evaluation exploratoire des profils expressifs

4.4.1 Objectifs

L'objectif de l'étude présentée dans cette section est de valider perceptivement notre système de profils expressifs individuels. En effet, ce modèle computationnel repose sur un certain nombre de choix d'implémentation et de conception. Il est donc difficile de le comparer à d'autres modèles. Ainsi, notre unique option est de mener une étude perceptive en proposant à nos participants d'interagir avec notre système.

4.4.2 Hypothèse

Nous posons pour hypothèse que les différents paramètres expressifs seront perçus par les sujets. Afin de créer une interaction cohérente, nous avons choisi de fixer les paramètres pour créer différents profils expressifs. Notre hypothèse générale est donc :

H1 : Les sujets seront en mesure de différencier le profil expressif associé à l'agent.

4.4.3 Protocole

Nous avons défini 6 profils expressifs. Ces profils ont été choisis par paires symétriques selon les trois critères suivants : Expressivité, Rapidité, et Valence. De plus, ces profils peuvent être appliqués directement et facilement sur notre modèle dimensionnel inspiré du modèle P.A.D.

1) Un profil « Très expressif » pour lequel les courbes de modulation d'activation amplifient l'activation donnée par l'utilisateur.
2) Un profil « Peu expressif » pour lequel les courbes de modulation d'activation atténuent l'activation donnée par l'utilisateur.
3) Un profil « Rapide » pour lequel les vitesses de déclin et d'attaque sont très élevées.
4) Un profil « Lent » pour lequel les vitesses de déclin et d'attaque sont très faibles.
5) Un profil « Négatif » pour lequel les émotions négatives sont amplifiées, et les émotions positives atténuées.
6) Un profil « Positif » pour lequel les émotions positives sont amplifiées, et les émotions négatives atténuées.

Avec ce système, nous avons mené une étude exploratoire visant à évaluer si les utilisateurs, confrontés de manière aléatoire à ces différents profils expressifs, étaient capables de les percevoir en interagissant avec le système. Les sujets ont donc pu manipuler l'expression émotionnelle de l'agent en utilisant la souris 3D. Pour chaque profil expressif (activés dans un ordre aléatoire) les sujets devait manipuler l'agent quelques minutes, puis évaluer différentes caractéristiques. Pour chaque caractéristique, l'utilisateur avait à sa disposition un slider continu variant entre 0% (« pas du tout ») et 100% (« tout à fait »). Ainsi, l'utilisateur pouvait moduler sa réponse avec précision pour chaque caractéristique considérée. Le niveau de hasard est donc 50%.

Les traits de personnalité à évaluer sont inspirés du modèle OCEAN. Nous avons choisi les caractéristiques suivantes : Cordial, Confiant, Curieux, Maître de Lui, Emotif, Expressif, Honnête, Positif, Social, Travailleur.

Ces caractéristiques ont été choisies pour représenter tous les facteurs du modèle OCEAN, avec deux items par facteurs. Ouverture (Curieux, Social), Conscience (Maître de soi, Travailleur), Extraversion (Cordial, Positif), Agréabilité (Confiant, Honnête), et Névrosisme (Emotifs, Expressif).

4.4.4 Résultats

Les résultats de cette étude, menée sur 17 sujets, montrent que plusieurs de ces profils expressifs sont reconnus, et en particulier les profils “Peu Expressif”, “Très Expressif”, et “Négatif”. Notre hypothèse H1 est donc validée. La Figure 75 montre les résultats de l’évaluation des profils Peu et Très expressifs. On observe une différence notable dans le jugement de l’expressivité et de l’émotivité entre les deux profils expressifs. L’agent Peu Expressif est également évalué comme plus « maître de lui-même » que l’agent Très Expressif.

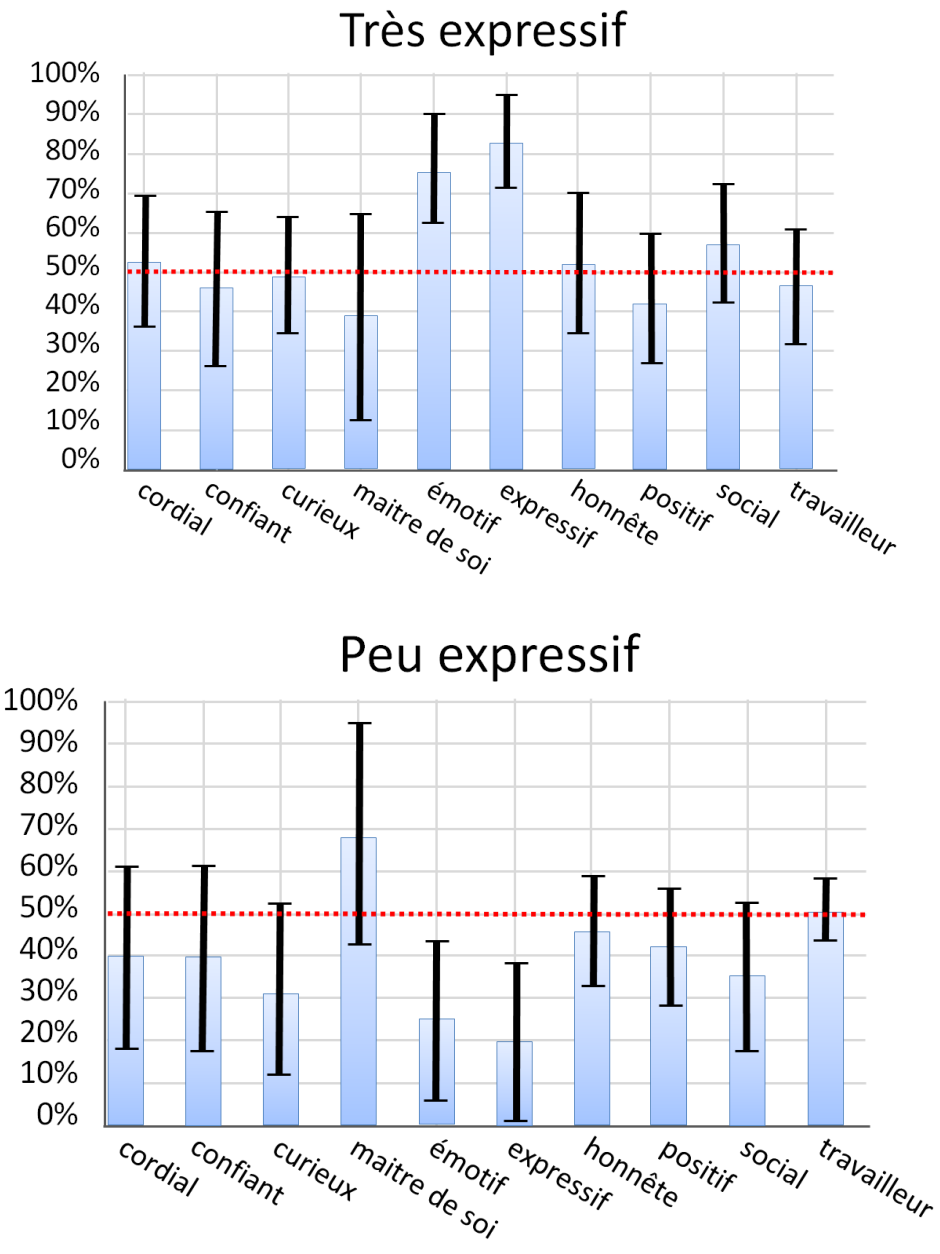


Figure 75 - Résultats de l'étude sur les profils expressifs pour les agents « Très Expressif », et « Peu Expressif ». La ligne pointillée représente le niveau de hasard (50%).

Nous pouvons également faire des comparaisons entre les différents profils expressifs. Par exemple, si nous considérons la caractéristique « Maître de lui » (Figure 76), nous observons que les profils expressifs « Lent », « Positif » et « Peu expressifs » sont évalués comme ayant un meilleur contrôle d’eux-mêmes que les profils « Négatif » et « Très Expressif ». Ces résultats semblent donc cohérents.

Nous remarquons que certaines caractéristiques utilisées dans cette expérimentation semblent ne pas être pertinentes pour l'application considérée. En effet, les caractéristiques "Social" "Travailleur" "honnête", et "Curieux" sont globalement toujours situés autour du seuil de hasard : 50%, avec un large écart type.

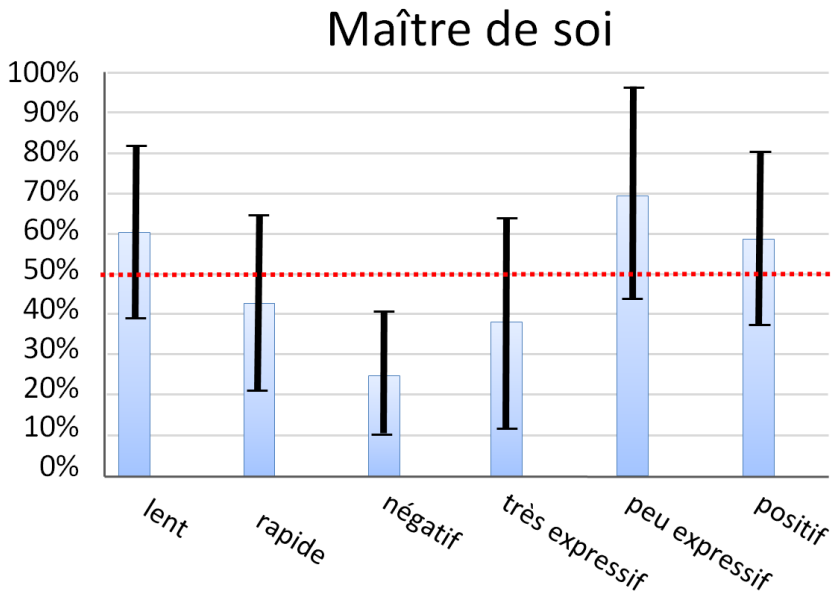


Figure 76 - Comparaison de la caractéristique "Maitre de Soi" entre les différents profils expressifs. La ligne pointillée représente le niveau de hasard (50%).

Pendant la partie entretien, les sujets ont également indiqué qu'ils avaient été perturbés par le manque d'expressivité des yeux de l'agent virtuel. Il semble que le regard fixe de notre agent est perçu comme un signe de tristesse, ce qui pourrait expliquer pourquoi le profil Positif n'est pas aussi bien perçu ainsi que le profil Négatif. L'absence de mouvements du torse et la tête semble également diminuer l'expressivité de l'agent. Ainsi, dans les versions suivantes de MARC, nous intégrerons des clignements d'yeux, ainsi que des changements de directions du regard.

4.5 Résumé et limites de l'approche dimensionnelle

Dans ce chapitre, nous avons exploré l'utilisation d'un modèle dimensionnel des émotions pour contrôler l'expressivité de notre personnage virtuel. Nous avons donc conçu un nouveau module émotionnel, fonctionnant de manière complémentaire avec le module catégoriel conçu précédemment. Ce nouveau module est basé sur l'utilisation de l'espace émotionnel P.A.D (Russell et Mehrabian, 1977). Afin de relier le module à l'animation faciale de MARC, nous avons donc proposé une méthode de calcul de l'activation émotionnelle en fonction de la position d'un curseur dans un espace dimensionnel 3D.

Une fois les activations émotionnelles calculées, nous avons pu créer une méthode de modulation expressive pour créer un ensemble de profils expressifs individuels. Ces profils nous ont permis d'ajouter à l'agent un comportement expressif modulable. Notre étude montre que les utilisateurs sont capables de distinguer certains des profils expressifs durant l'interaction avec notre système.

Cette approche permet d'imposer certaines contraintes implicites sur l'espace expressif de l'agent. Par exemple, plusieurs émotions ne peuvent être activées de manière intense simultanément. De plus, certaines transitions imposent un retour à l'état neutre. Par exemple, la transition entre Joie et Tristesse impose de passer par le centre du cube (état neutre), de plus, ces deux émotions ne peuvent donc être mélangées.

Notre méthode de calcul impose de positionner les émotions aux coins du cube, de façon similaire à l'approche utilisée par Heudin (Heudin, 2004). Cette méthode est néanmoins une simplification de l'espace P.A.D. tel que défini par Russel et Mehrabian (Russell & Mehrabian, 1977).

Si l'approche dimensionnelle apporte des contraintes qui permettent d'éviter des situations expressives critiques, elle apporte également des limitations. En effet, il devient difficile d'utiliser un grand nombre d'émotions dans une application. Notre méthode de calcul limite le nombre d'émotions à huit (une à chaque coin du cube). Cependant, l'espace émotionnel est lui-même restreint. L'utilisation d'un grand nombre d'émotions dans cet espace poserait divers problèmes. D'une part, dans un contexte de contrôle de l'agent similaire à notre application, la sélection d'une émotion particulière poserait un problème de précision spatiale. D'autre part, la transition entre deux émotions risquerait d'activer un grand nombre d'états émotionnels se trouvant dans l'espace les séparant. Ainsi, l'animation faciale correspondante serait probablement trop complexe, et composée d'un trop grand nombre d'expressions intermédiaires distinctes, ce qui ne serait pas un comportement crédible de l'agent.

Nous avons donc décidé d'étudier une autre approche des émotions. Il nous a alors semblé cohérent d'étudier les approches cognitives. Modéliser les émotions revient alors à modéliser le processus cognitif qui en est responsable. Ainsi, nous ne considérons plus l'émotion comme un état fini, mais comme un processus dynamique adapté à l'interaction homme-machine dans le cadre d'application utilisant des agents expressifs autonomes capables de prendre en compte le contexte d'interaction.

PARTIE II

De l'approche Cognitive à l'approche Cognitive et Sociale

Chapitre 5. Approche cognitive des émotions pour l'animation faciale : Le modèle CPM

Sommaire du chapitre

Publications associées

- M. Courgeon, C. Clavel, J-C. Martin, (2009) *Appraising Emotional Events during a Real-time Interactive Game*, in: Proceedings of the ICMI 2009 Workshop on Affective Computing (AFFINE), Cambrigde, U.S.A., 1-6 novembre 2009
- M. Courgeon, C. Clavel, J-C. Martin, (2011) *Real-time Interaction with a Virtual Character that Displays Facial Signs of Emotions and Appraisal: Impact on Users' Performance and Perception during a Game.* (Soumis)

5.1 Intérêts des approches cognitives

Les approches cognitives tentent de modéliser le processus cognitif à l'origine des émotions. Ainsi, une émotion est un processus dynamique d'évaluation de la situation (Scherer 2010). Pour nous, cela présente l'intérêt de considérer les émotions comme une conséquence du contexte dans lequel évolue d'agent. La prise en compte du contexte nous semble une approche importante pour la conception d'application interactive.

Comme nous l'avons vu lors de notre état de l'art, les émotions sont considérées comme des résultats du processus d'évaluation cognitive d'un événement. Dans nos travaux, nous avons sélectionné le modèle psychologique de Scherer (1984, 2001), le *Componential Process Model* (CPM) car ses descriptions détaillées des différentes étapes du processus émotionnel et des expressions faciales correspondantes en font un modèle particulièrement adapté pour une modélisation informatique.

Le modèle CPM décrit une liste de critères d'évaluation cognitive. A l'instar de Scherer, nous appellerons ces critères d'évaluation par le terme *check*. Modéliser ces checks par un modèle informatique nous permettrait donc de concevoir un système informatique capable de réagir dynamiquement aux situations et aux événements survenant lors de l'interaction avec l'utilisateur. Les réactions émotionnelles devront donc être générées dynamiquement par le modèle computationnel simulant le processus cognitif, ainsi que l'animation faciale associée.

5.2 Animation faciale basée sur la théorie des appraisals

En plus des différents checks, le modèle CPM propose une description des différentes manifestations physiologiques et motrices de ces checks. Nos travaux se baseront sur les différentes descriptions des réactions faciales associées à certains des checks (Scherer 2001).

Comme nous l'avons mentionné dans l'état de l'art, plusieurs systèmes d'animation faciale ont été inspirés par le modèle CPM (Paleari et Lisetti. 2006, Malatesta et al., 2007). Ces travaux ont mis en relief quelques problématiques liées à ce type d'animation. Par exemple, selon le modèle CPM, à certains checks correspondent certaines composantes expressives faciales. L'expression faciale émotionnelle résulterait de la combinaison des expressions des différents checks. Or, *«dans le cas de la colère, selon les paramètres des tables de prédiction d'expression faciale de la théorie du CPM, l'expression de l'évaluation « événement nouveau » inclut un haussement des sourcils. Mais, l'évaluation suivante (rapport aux buts) induit le froncement des sourcils, ce qui entre en conflit avec le haussement de sourcils précédent. Ce conflit rend problématique l'animation. Le résultat de l'animation séquentielle cumulative correspondante est source de confusion »* (Malatesta et al., 2006).

Aucun de ces travaux n'a en revanche abordé la modélisation du processus émotionnel permettant l'évaluation d'événements de manière dynamique. Ainsi, les systèmes d'animations réalisés dans les travaux cités ne sont pas des applications interactives. Leur « interaction » se limite à une présentation d'animations pré-calculées à des sujets qui doivent rapporter comment ils perçoivent ces animations, et ceci en dehors de tout contexte de tâche interactive. La modélisation informatique des processus cognitifs décrit par le modèle CPM reste donc un challenge.

5.3 Objectifs de MARC v3 : approche cognitive

Pour la conception de cette troisième version de MARC (Figure 77), notre objectif est donc double. Tout d'abord, nous souhaitons concevoir un système autonome, basé sur le modèle de Scherer. Pour cela, nous devons proposer un modèle computationnel inspiré du modèle théorique du CPM. Cette modélisation nous impose un certain nombre de choix et de simplification que nous détaillerons au fur et à mesure.

En second lieu, nous avons pour objectif de proposer un module d'animation faciale basé sur le résultat de l'évaluation effectuée par le module computationnel. Ce module devra donc proposer des solutions aux problématiques soulevées par Malatesta et al. (2006). Notre objectif est ensuite d'évaluer l'apport de ce modèle

CPM sur l'animation, par rapport au modèle catégoriel des émotions. Nous n'avons cependant pas l'ambition de créer un module cognitif capable de s'adapter à toutes les situations, et capable d'évaluer n'importe quel type d'évènement. Ce type de module ferait probablement appel à des méthodes d'*intelligence artificielle* trop éloignées de notre objectif de recherche. Nous avons donc choisi de limiter le champ d'action de notre module émotionnel en restreignant à l'analyse des événements survenant lors d'un jeu de plateau.

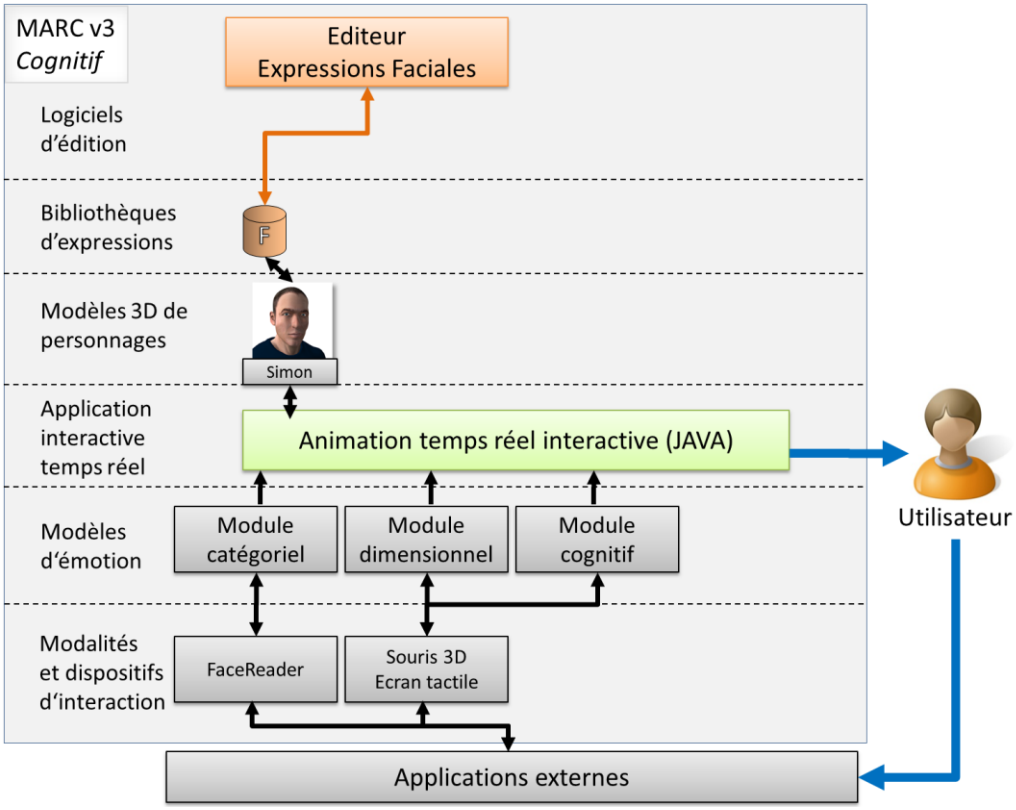


Figure 77 - Architecture de MARC v3 : Modèle Cognitif

5.4 Modèle informatique inspiré du modèle CPM

Afin de créer une situation d'interaction affective, nous avons donc opté pour l'implémentation d'un jeu de plateau. Dans le cadre de cette application ludique, l'agent virtuel joue le rôle d'adversaire de l'utilisateur. La problématique inhérente était de trouver un jeu facile à apprendre, afin de permettre au plus grand nombre de sujets de jouer, et donc, de participer à notre expérimentation.

Nous avons opté pour le jeu Reversi. Plus simple que les échecs ou les dames, ce jeu comporte peu de règles, et est donc rapide à apprendre. De plus, ce jeu est très dynamique, une partie peut basculer très rapidement en faveur d'un des joueurs, puis en faveur de l'autre. Nous pensons donc que ce jeu est adapté à notre application puisqu'il peut induire des situations émotionnelles riches et suffisamment variées.

5.4.1 Le jeu de plateau Reversi et son implémentation informatique

Nous avons donc mis en œuvre un jeu appelé Reversi. Ce jeu est joué par deux joueurs sur un plateau de huit lignes et huit colonnes. Chaque joueur possède un ensemble de pièces dont une face est blanche et l'autre noire. Chaque joueur possède l'une des deux couleurs. Le but de chaque joueur est d'avoir une majorité de pièces de sa couleur visibles sur le plateau à la fin du jeu. Le but est donc de retourner le plus grand nombre possible des

pièces de l'autre joueur. Comme nous l'avons mentionné, ce jeu est facile à apprendre et donc approprié pour mener des études expérimentales.

Dans notre étude, l'utilisateur joue contre un joueur artificiel, représenté graphiquement par notre personnage virtuel (Figure 78).

Figure 78 - Une utilisatrice jouant au reversi contre notre agent virtuel. (L'écran inférieur est tactile). Le cadre en haut à gauche montre l'image de l'utilisatrice filmée par la caméra située à gauche de l'agent virtuel.

5.4.2 MARC : Ajout du module cognitif inspiré du modèle CPM

Pour étudier les théories de l'évaluation cognitive, nous avons étendu MARC avec deux modules (Figure 79): 1) le module d'évaluation cognitive, qui évalue les événements qui se produisent pendant le jeu, et 2) le module d'animation séquentiel, qui génère les paramètres d'animation du visage correspondant aux évaluations cognitives.

- Le module d'évaluation cognitive

Ce cadre ludique est un cadre expérimental pertinent pour notre recherche car il nous permet de nous concentrer sur un ensemble restreint de situations émotionnelles avec des profils d'évaluation différents. Trois types de situations déclenchent des événements émotionnels dans le système MARC : 1) l'utilisateur joue, 2) l'agent virtuel joue, et 3) le jeu est terminé.

Nous avons adapté une sous-partie du modèle CPM présenté en état de l'art qui se rapporte à ces trois événements émotionnels. Notre système fonctionne donc suivant sept checks de l'évaluation cognitive: Nouveauté, Agrément intrinsèque, Rapport aux buts, Causalité externe, Potentiel de maitrise, Standards externes et Standards Internes. Nous avons choisi ces sept checks parce qu'ils sont pertinents dans le cadre d'un jeu. De plus, plusieurs émotions sont caractérisées sur ces checks dans les études psychologiques, ce qui facilite leur mise en œuvre dans un modèle computationnel (Scherer, 2001).

Les contextes d'interaction que nous avons considérés nécessitent un historique des événements de jeu et une anticipation à court terme des prochaines actions possibles de l'utilisateur. Anticiper le jeu permet de calculer les valeurs des checks « Nouveauté » et « Potentiel de maitrise ». Le système d'anticipation est cependant limité à deux coups d'avance. Limiter ainsi les capacités du système permet de donner au joueur une chance de gagner.

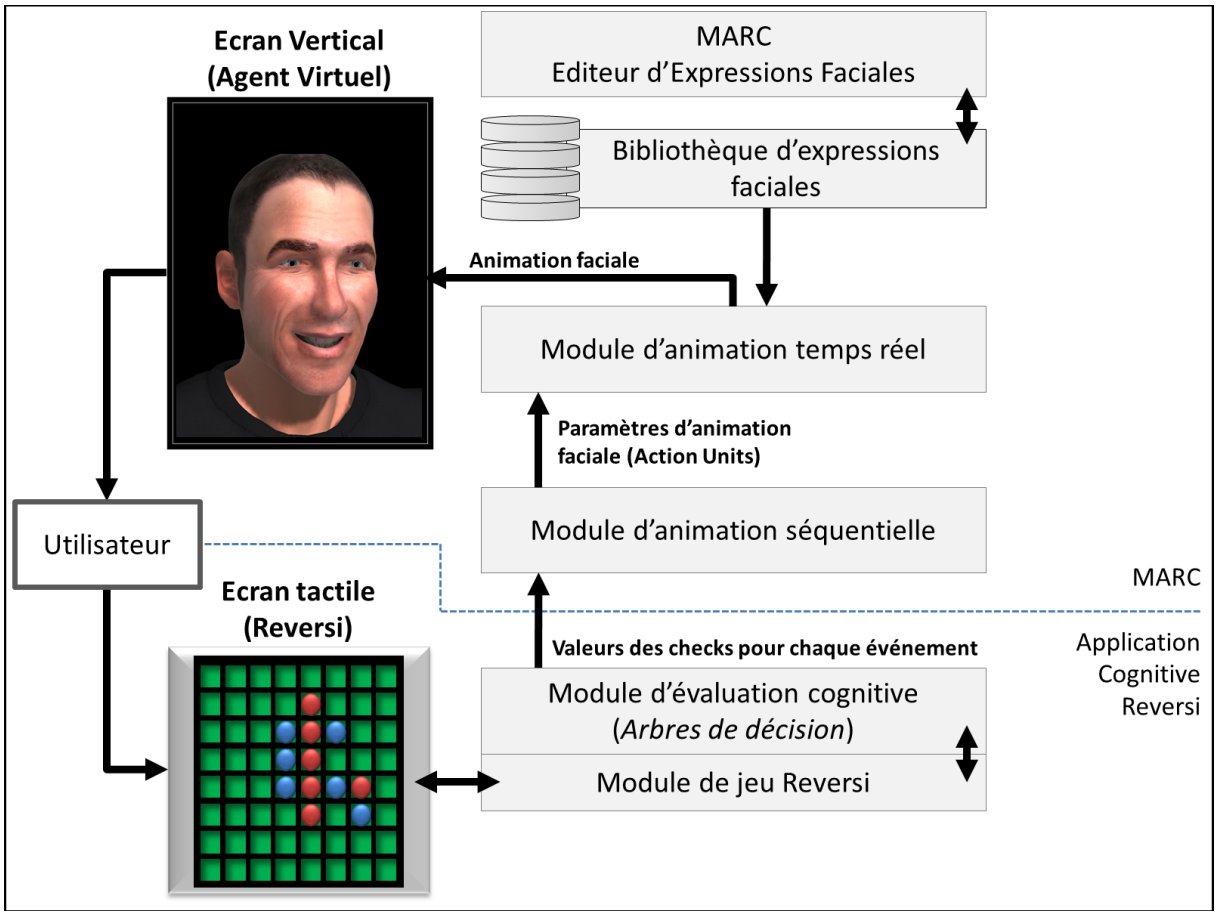


Figure 79 - Architecture logicielle de MARC version Cognitif et du jeu Reversi

Pour chaque événement, l'application d'évaluation cognitive utilise un arbre de décision pour calculer une valeur pour chacun des sept checks. La Figure 80 décrit l'un des arbres de décision utilisés pour déterminer quels événements émotionnels sont générés à la suite d'un coup de l'utilisateur. La valeur de certains checks a été définie avant l'exécution selon un ensemble de règles logiques. La valeur de certains autres checks ne peut pas être définie avant l'exécution. Certains checks sont ainsi laissés "ouverts" dans le modèle CPM. Par exemple, la valeur du check "nouveauté" pour l'émotion Joie est laissée ouverte (c'est à dire non spécifiée avant l'exécution). Ces valeurs de ces checks « ouvertes » sont ensuite calculées dynamiquement à l'exécution en tenant compte du contexte d'interaction (comme décrit dans la section suivante). Ainsi, on introduit une variabilité contextuelle dans les expressions faciales issues des mêmes évaluations dans l'arbre de décision. Les feuilles des arbres représentent l'évaluation de l'événement émotionnel. Les arbres de décision sont nécessaires pour générer un ensemble de valeurs cohérentes pour le groupe des checks que nous considérons.

Par exemple, l'utilisateur place un pion. L'état du jeu est modifié. Le système commence alors par comparer le nombre de pièces restantes de chaque couleur. L'ordinateur possède alors autant de pions de sa couleur que l'utilisateur. La réponse à la question « le système a-t-il encore deux fois plus de pièces que l'utilisateur ? » est « non ». On avance donc dans l'arbre de décision. La réponse à la question « l'utilisateur a-t-il encore deux fois plus de pièces que le système ? » est « non ». On avance donc dans l'arbre de décision. On compare alors le coup de l'utilisateur aux différents coups possibles anticipés. La réponse à la question « l'utilisateur a-t-il joué le meilleur coup possible pour lui ?» est « oui ». On avance donc dans l'arbre de décision. On compare ensuite les coups disponibles pour le système, et la réponse à la question « Le prochain coup du système est-il meilleur que le coup de l'utilisateur » est « oui ». On avance dans l'arbre de décision, et on atteint une feuille (Figure 80, en bas à droite). On obtient ainsi le résultat de l'évaluation de cette situation.

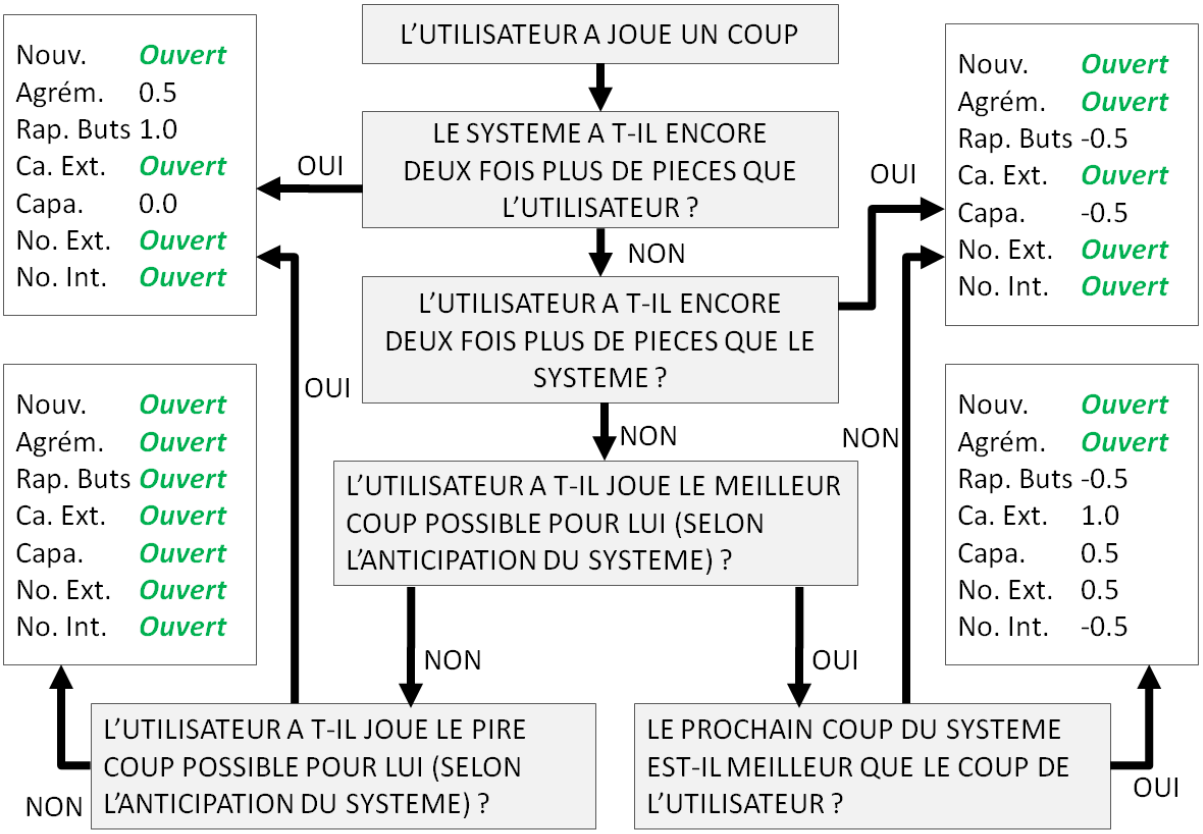


Figure 80 - L'arbre de décision utilisé suite au coup de l'utilisateur. Un arbre similaire est utilisé lorsque l'agent joue. (Nouv. = Nouveauté, Agrém. = Agrément intrinsèque, Rap. But. = Rapport aux buts, Ca. Ext. =Causes Externes, Capa. = Capacités de Maitrise, No. Ext= Normes Externes, No. Int= Normes Internes). Les valeurs Ouvertes sont calculées dynamiquement durant l'exécution.

Dans le module d'évaluation cognitive, seuls les arbres de décision sont mis à jour pour adapter notre système à une autre application.

A l'exécution, une fois que toutes les valeurs des checks de l'événement émotionnel ont été calculées, ils ont été envoyés au module d'animation séquentiel dans un message au format EmotionML tel que défini par SAIBA (Vilhjalmsson et al. 2007, Bevacqua et al, 2008)

5.4.3 Animation faciale à partir des valeurs des checks du modèle CPM

Le module d'animation séquentielle génère des expressions faciales à partir des valeurs des checks créés dynamiquement. La séquence d'expressions faciales résultante donne une animation faciale dynamique. Une fois ces signes temporaires de l'évaluation cognitive affichés sur le visage virtuel, les valeurs des checks sont utilisées pour prédire l'émotion résultant de l'évaluation cognitive. L'expression faciale de l'émotion est alors affichée en fin de séquence dynamique. Nous décrivons en détail chacune de ces étapes dans la suite.

Premièrement, le module d'animation reçoit un message encodé en EmotionML émis par le module d'évaluation cognitive contenant les valeurs des checks, et décrivant un événement du jeu venant de se produire (Figure 79). Selon le modèle des processus composants, un événement est considéré comme étant pertinent si au moins un de ses trois premiers checks (Nouveauté, Agrément intrinsèque, Rapport aux buts) n'est pas neutre. Si une (ou plusieurs) de ces trois valeurs suggère que l'événement est pertinent, alors le module d'animation génère une animation faciale. Par exemple, une action prévisible de l'utilisateur qui ne change pas la répartition des pièces de manière significative pourrait être évaluée comme n'étant pas pertinente, et ne changerait donc pas l'état émotionnel de l'agent. Aucune expression faciale ne serait alors générée.

Si l'événement est évalué comme étant pertinent, la première fonction du module d'animation est d'afficher temporairement les expressions du visage qui reflète le processus d'évaluation interne. Ces expressions ont été spécifiées en utilisant les descriptions textuelles des effets de l'évaluation suggérés par le modèle CPM (Scherer, 2001), utilisant le modèle FACS.

Le modèle CPM prédit la valeur des checks d'évaluation pour plusieurs émotions. La valeur attribuée au check «Cause Externe» peut être soit «ouvert», «externe», ou «interne». Tous les autres checks peuvent avoir l'une des valeurs suivantes: «ouverte», «très faible», «faible», «moyen», «élevé» et «très élevé». Lorsque la valeur «ouvert» est attribué à un check pour une émotion donnée, cela signifie que ce contrôle n'est pas discriminant pour cette émotion particulière. Par exemple, on peut être joyeux indépendamment de la nouveauté de l'événement qui a provoqué la joie. Par conséquent, la valeur de "nouveauté" de la joie est «ouverte».

Pour permettre d'appliquer des calculs flottants sur ces labels, nous avons fait correspondre une valeur numérique à chacun de ces adjectifs.

-1 (« très faible » et « interne »), -0,5 (« faible »), 0 (« medium »), 0.5 (« élevé »), et +1 («très élevé », et « externe »).

Lorsqu'un événement se produit, le module d'évaluation évalue la pertinence de chaque émotion possible. Pour chaque émotion, nous avons comparé son profil numérique, obtenu en utilisant des tables de Scherer (2001), avec le profil (l'ensemble des checks) de l'événement considéré. Les valeurs des checks des événements se produisant pendant le jeu en temps réel peuvent être des valeurs numériques comprises entre -1,0 et 1,0. Ainsi, plusieurs émotions peuvent être pertinentes en même temps, ce qui permet que des mélanges d'émotions se produisent.

Lorsque la séquence d'évaluation est terminée (avec l'expression temporaire des checks), un message précisant le mélange d'émotion résultant de l'analyse de l'événement est envoyé au module de rendu. Cette dernière expression est obtenue en mélangeant l'expression prototypique des émotions pertinentes à l'égard de leur intensité respective. Par exemple, si un événement a une pertinence de 0,7 pour la Joie et de 0,3 pour la Fierté, l'expression résultante sera un mélange de 70% de l'expression prototypique de Joie et de 30% de l'expression prototypique de Fierté.

Les checks ouverts ont été calculés en utilisant une échelle linéaire. Par exemple, la "nouveauté" a été calculée en utilisant la probabilité anticipée de l'action effectuée par l'utilisateur : P_{move}

$$\text{Nouveauté} = 1 - (P_{move} \times 2)$$

En utilisant cette formule, une action avec une probabilité de 1,0 aura une valeur du check nouveauté de -1,0, alors qu'un mouvement avec une probabilité de 0,05 aura une valeur du check nouveauté de 0,9.

Ces valeurs non-prototypiques affectées aux checks permettent des expressions faciales plus subtiles, ce qui concorde avec l'observation psychologique que les expressions faciales complètes et à intensité maximale sont rarement observées dans la réalité (Scherer et Ellgring, 2007).

La deuxième fonction du module d'animation a été de calculer l'état final émotionnel résultant de la séquence d'évaluation cognitive. Le module d'animation a été initialisé avec un ensemble d'émotions possibles. Chaque émotion a été décrite avec sept valeurs correspondantes aux sept checks du modèle CPM. Ces descriptions nous donnent un profil émotionnel pour chaque émotion possible. Nous avons sélectionné la liste d'émotions suivante: Joie, Peur, Colère, Tristesse et de Fierté. Cet ensemble a été choisi parce que la littérature fournit la description de leurs profils d'évaluation cognitive (les valeurs des check) (Scherer, 2001), et que ces émotions sont souvent observées lors d'une partie de jeu (Marsella et al., 2009).

Les expressions faciales de Joie, de Peur, de Colère et de Tristesse ont été définies en utilisant les travaux d'Ekman (Ekman et Friesen, 1975).

L'expression faciale de la Fierté a été inspirée par la base de données MindReading (Golan et al, 2006). A partir des six vidéos de ce corpus exprimant la fierté, nous avons extrait les caractéristiques expressives apparaissant dans au moins la moitié des vidéos (par exemple, les mouvements de sourcils).

5.4.4 Le jeu de reversi affectif présenté à la fête de la science

Figure 81 - Le jeu de Reversi affectif, présenté a la Fête de la science 2009.

L'installation interactive présentée dans ce chapitre a été exposée au grand public lors de la *fête de la science 2009*, organisée à Orsay. L'installation a donc fonctionné durant trois jours, dix heures par jour. Cela a été pour nous l'occasion de tester la robustesse de notre système. De plus, à l'instar de Pfeiffer et al. (2011) nous avons pu observer informellement que les enfants montrent moins de retenu avec l'agent virtuel que les adultes, y compris en groupe (Figure 81). L'évaluation de ce type d'installation dans un contexte grand public ouvre des perspectives de recherche pour définir des protocoles adaptés.

5.5 Etude perceptive

Ce système de jeu interactif avec l'agent doté d'un modèle cognitif des émotions nous a permis de mettre en place une étude pour comparer l'animation faciale de type catégorielle, telle que nous l'avions conçue et décrite dans la section 3.2, à l'animation faciale issue de notre simulation du processus cognitif inspirée du modèle CPM. Cette étude a été menée en collaboration avec Céline Clavel.

5.5.1 Objectifs

Afin d'évaluer l'impact des expressions faciales générées par le modèle cognitif sur la perception des utilisateurs lors d'une partie de Reversi, nous avons comparé trois conditions expressives (Figure 82):

• une condition « Pas d'émotion », le personnage virtuel ne présentait aucune expression faciale. Ce mode est une condition contrôle.

• une condition « Catégoriel », dans laquelle le personnage virtuel affiche des expressions prototypiques d'émotion catégorique.

• une condition « Cognitif », dans laquelle le personnage virtuel affiche séquentiellement les expressions faciales des checks de l'évaluation cognitive puis l'émotion catégorielle correspondante.

Figure 82 – Images issues des animations des expressions du personnage virtuel dans les trois conditions. Dans chaque condition, l'image de gauche est l'expression neutre et l'image de droite est l'expression finale (voir le texte pour les explications sur les images intermédiaires).

La Figure 82 montre, de haut en bas : la condition de Sans-émotion, la condition catégoriel (Tristesse), la condition d'évaluation cognitive (une séquence menant à la tristesse avec les valeurs suivantes des checks de l'évaluation: inattendue, désagréable, entrave des buts, faible adaptation possible). Les valeurs des trois autres checks (Causes externes, Normes internes et externes) ne sont pas exprimées, mais sont pris en compte lors du calcul final des probabilités des émotions résultantes. Les animations du mode catégoriel et du mode cognitif ont la même durée.

5.5.2 Hypothèses

Pour évaluer l'influence du mode expressif sur la perception des utilisateurs, nous proposons les hypothèses suivantes:

H1: L'expressivité émotionnelle de l'agent virtuel perçue par l'utilisateur augmente avec le nombre d'expressions faciales utilisées lors de l'animation.

Plus le nombre d'expressions affichées par le personnage virtuel est élevé, plus il devrait être perçu comme expressif. La condition sans-émotion utilise une seule expression (neutre). Chacune des animations affichées dans l'état catégoriel utilise deux expressions: une expression neutre et une expression prototypique pour chaque émotion (et les expressions intermédiaires correspondent à une interpolation entre l'expression neutre et l'expression cible). Enfin, les animations affichées dans la condition Cognitif utilisent une séquence de quatre expressions différentes (pour les checks) suivie par l'expression faciale de l'émotion qui en résulte. Cette différence en termes de nombre d'expressions utilisées par le personnage virtuel dans les trois conditions devrait influencer la perception de l'expressivité par les utilisateurs.

En revanche, seul le mode Sans-émotion devrait être perçu comme non expressif, car les deux autres modes génèrent des expressions faciales d'émotions.

H1A: Les utilisateurs perçoivent moins d'émotions dans les expressions du personnage virtuel dans la condition sans-émotion que dans les conditions catégorielles et évaluation cognitive.

Perception des expressions de l'émotion:

Sans émotion < catégoriel = cognitif

Nous nous attendions à ce que les utilisateurs perçoivent aussi la différence de dynamique entre les animations utilisées dans les trois conditions.

H1B: Les utilisateurs perçoivent une dynamique plus élevée de l'expression émotionnelle quand le nombre d'expressions du visage est plus élevé.

Perception de la dynamique de l'expression émotionnelle:

Sans émotion < catégoriel < cognitif

H2: L'attribution (par les sujets) d'états émotionnels et cognitifs au personnage virtuel dépend de la condition expérimentale.

Parce que l'agent virtuel affiche des expressions émotionnelles dans le mode catégoriel et dans le mode cognitif, nous nous attendons à ce que les utilisateurs attribuent des états mentaux internes au personnage virtuel (par exemple, le personnage virtuel est fier quand il a gagné la partie).

H2A: Les utilisateurs attribuent moins d'états émotionnels au personnage virtuel dans la condition sans-émotions que dans les deux autres modes.

Attribution d'états mentaux émotionnels:

Sans-émotion < catégoriel = cognitif

La condition Cognitif présente des expressions liées à l'évaluation cognitive de la situation, ce qui n'est pas le cas des modes Catégoriel et Sans-émotions. Ainsi, nous nous attendons à ce que les utilisateurs attribuent davantage d'états mentaux non-émotionnels (par exemple, « Réfléchir ») pour le personnage virtuel dans le mode Cognitif que dans les autres modes.

H2B: Les utilisateurs attribuent plus d'états mentaux non-émotionnels à l'agent virtuel dans le mode Cognitif que dans les modes Sans-émotions et Catégoriel.

Attribution d'états mentaux non-émotionnels:

Sans-émotion = catégoriel < cognitif

H3: L'utilisateur gagne plus souvent et passe plus de temps à jouer lorsque le nombre d'expressions faciales est plus élevé.

Nous nous attendons à ce que les utilisateurs gagnent plus souvent dans le mode Cognitif et le mode catégoriel que dans la condition Sans-émotion parce qu'ils peuvent utiliser plus de réactions pour comprendre et prédire le comportement du personnage virtuel.

H3A: Les utilisateurs gagnent plus souvent lorsque le nombre d'expressions faciales est plus élevé.

L'utilisateur gagne:

Sans-émotion < Catégoriel < Cognitif

Enfin, nous nous attendons à ce que les utilisateurs passent plus de temps à préparer et à effectuer leurs actions dans la condition Cognitif que dans la condition Catégoriel parce qu'ils ont à interpréter un plus grand nombre et une plus grande variété d'expressions faciales pour prédire le comportement du personnage virtuel. De même, nous nous attendons à ce que les utilisateurs passent plus de temps à jouer dans le mode Catégoriel que dans la condition Sans-émotion.

H3B: Les utilisateurs passent plus de temps à jouer lorsque le nombre d'expressions faciales est plus élevé.

Durée de jeu:

Sans-émotion < Catégoriel < Cognitif

5.5.3 Protocole

Design

L'expérience est conçue avec une seule variable dépendante (Sans-Emotion, Catégorique et Cognitif). Nous avons utilisé le même personnage virtuel et le même module d'évaluation cognitive pour les trois conditions (y compris les mêmes arbres de décision). La seule différence entre les trois conditions a été le mode d'affichage des expressions faciales.

Utilisateurs

Soixante utilisateurs ont participé à l'étude, tous de langue maternelle française. Les sujets se composent de 17 femmes et 43 hommes. L'âge moyen est de 26 ans. Tous les utilisateurs ont un diplôme d'études secondaires ou un diplôme universitaire. Les utilisateurs étaient filmés pour évaluer la manière dont ils regardaient le personnage virtuel et réagissaient à ses expressions.

Procédure

Les utilisateurs ont été informés qu'ils allaient jouer à Reversi contre un personnage virtuel. Un consentement éclairé écrit a été obtenu de chaque participant avant l'expérience, en particulier du fait qu'ils acceptaient d'être filmés.

Les utilisateurs ont été répartis au hasard dans trois groupes de vingt utilisateurs chacun. Le premier groupe d'utilisateurs a joué dans la condition Sans-émotion. Le second groupe d'utilisateurs a joué face au mode Catégoriel. Le troisième groupe d'utilisateurs a joué dans la condition Cognitive.

Dans les trois conditions, l'agent était animé de mouvements de tête périodiques en direction du plateau de jeu, environs toute les 10 secondes. Le reste du temps, il regardait en direction de l'utilisateur. De plus, ainsi que suggéré par l'étude présentée dans la section 4.4, l'agent était animé de clignements réguliers et aléatoires de paupières avec une période semi-aléatoire de 6 à 10 secondes, afin de lui donner un aspect moins robotique. Ces animations étaient identiques pour les trois conditions expérimentales et visaient à faciliter l'interaction grâce à un comportement moins statique et à encourager l'utilisateur à regarder le personnage virtuel.

Déroulement

Les utilisateurs reçoivent des instructions décrivant la tâche. Ils sont isolés avec le système de jeu dans une pièce calme. Les utilisateurs sont ensuite informés qu'à la fin du jeu, ils auront à répondre à un questionnaire sur leurs perceptions des états mentaux, des émotions et des comportements du personnage virtuel. Le visage des utilisateurs a été enregistré en vidéo pendant la partie entière.

Une fois qu'il est prêt, l'utilisateur joue en premier. Le plateau de jeu est ensuite assombri de manière à attirer l'attention de l'utilisateur vers les expressions faciales de l'agent virtuel. Le personnage virtuel exprime ensuite l'émotion résultant de l'évaluation du coup joué par l'utilisateur. Puis, l'agent virtuel joue, et son visage exprime l'émotion qui reflète l'évaluation de la nouvelle situation. Le plateau est ensuite à nouveau affiché normalement afin que l'utilisateur joue son prochain coup.

Mesures

L'annotation manuelle a posteriori de la vidéo de la direction du regard de chaque utilisateur dans les vidéos a permis de mettre en relief que plusieurs utilisateurs n'ont pas suffisamment regardé l'agent virtuel, mais se sont à la place concentrés sur le plateau de jeu. L'étude se rapportant à la perception des expressions de l'agent virtuel par les utilisateurs, seules les données des utilisateurs qui ont regardé l'agent virtuel au moins 30% de la durée du jeu ont été gardées pour analyse. L'échantillon d'utilisateurs ainsi sélectionné est composé de 26 utilisateurs: 6 utilisateurs dans la condition de non-émotion, 9 utilisateurs dans l'état catégorique, et 11 utilisateurs dans la condition d'évaluation.

Mesures subjectives

Le questionnaire comprend deux parties. La partie 1 vise à analyser la perception des utilisateurs des expressions faciales affichées par le personnage virtuel. La partie 2 vise à étudier les états mentaux internes émotionnels et non-émotionnels que les utilisateurs ont attribués au personnage virtuel.

Dans chacune de ces deux parties, une liste de questions a été proposée. Pour chacune de ces questions, les utilisateurs devaient déclarer leur niveau d'accord selon une échelle de Likert à cinq points. Les questions ont été inspirées par des questionnaires de mesure de l'intelligence émotionnelle (Mayer et al, 2003).

La fiabilité de notre questionnaire a été vérifiée à l'aide de l'alpha de Cronbach. Cette mesure calcule la corrélation moyenne des réponses de plusieurs questions destinées à mesurer une seule dimension. L'American Psychological Association (APA) estime qu'un questionnaire est acceptable lorsque le coefficient alpha est supérieur à 0,70. Nous avons calculé l'alpha de Cronbach pour les dimensions liées à la perception de l'expression faciale, l'attribution d'états mentaux émotionnels et l'attribution d'états mentaux non-émotionnels cognitifs. Nos questions sont valides selon les critères de l'APA (Alpha de Cronbach entre 0,70 et 0,94).

<u>Partie I: la perception des utilisateurs des expressions faciales affichées par le personnage virtuel</u>

L'objectif de cette première section était de tester notre première hypothèse (H1): différentes conditions expressives devraient conduire à des différences dans la perception des expressions faciales de l'agent virtuel. Cette partie du questionnaire comptait trois sections:

• Perception des expressions émotionnelles. Cinq éléments ont été utilisés (coefficient alpha de Cronbach: 0,85), par exemple, «Marc a exprimé qu'il appréciait jouer à ce jeu".

• La perception d'une absence d'expression émotionnelle. Cinq éléments ont été utilisés (coefficient alpha de Cronbach: 0,94), par exemple, «Marc est resté imperturbable durant le jeu".

• Perception de la dynamique émotionnelle des expressions. Quatre rubriques ont permis d'estimer si les utilisateurs avaient vu une expression unique ou plusieurs expressions au cours de l'expression d'une émotion (alpha de Cronbach: 0,75), par exemple, «Quand Marc a joué, son visage a exprimé les différentes phases de sa pensée».

Partie II: Attribution des états mentaux émotionnels et non-émotionnels cognitifs par les utilisateurs au personnage virtuel

La deuxième partie du questionnaire concernait l'attribution d'états mentaux internes par les utilisateurs au personnage virtuel. Il se rapporte à notre deuxième hypothèse (H2): l'attribution d'états mentaux émotionnels et cognitifs au personnage virtuel dépend de la condition expressive. Cette partie du questionnaire comportait deux sections:

• L'attribution d'états émotionnels. Sept éléments ont été utilisés (alpha de Cronbach: 0,79), par exemple, «Marc était fier de certains de ses mouvements".

• Attribution d'états mentaux cognitifs non-émotionnels. Sept éléments ont été utilisés (alpha de Cronbach: 0,70), par exemple, «Marc n'a pas l'air de réfléchir à la situation ».

Mesures Objectives

Des mesures objectives ont également été réalisées pour tester la troisième hypothèse (H3). Les résultats de chaque session de jeu ont été recueillis comme une variable binaire (1 si l'utilisateur a gagné la partie, 0 si le personnage virtuel a gagné la partie). Nous avons enregistré le temps de chaque action de l'agent et des utilisateurs. Ces données nous ont permis de calculer le temps global de jeu, et le temps de préparation de chaque coup de l'utilisateur (incluant le temps de réflexion).

5.5.4 Résultats

Les résultats présentés dans cette section sont statistiquement significatifs ($p < 0,05$). Les résultats sont explicitement mentionnés comme «tendanciels» si *p* est compris entre 0,05 et 0,1. Les données recueillies pour les mesures subjectives ont été analysées en utilisant le test T-Student.

Mesures subjectives

H1: L'expressivité émotionnelle perçue de l'agent virtuel augmente avec le nombre d'expressions faciales utilisées dans le mode d'expression.

H1A: Les utilisateurs perçoivent moins d'émotions dans les expressions du personnage virtuel dans la condition Sans-émotion que dans les conditions Catégorielles et Cognitives.

La quantité d'expressions émotionnelles rapportée était plus élevée dans la condition cognitive que dans la condition de Sans-émotion ($t(15) = -3,93$, $p < 0,001$).

Les utilisateurs du mode Catégoriel ont perçu plus d'expressions émotionnelles que les utilisateurs de la condition de non-émotion ($t(13) = -2,12$, $p < 0,05$). L'agent de la condition de Sans-émotion a été jugée globalement moins expressif que l'agent du mode Catégoriel ($t(13) = -3,88$, $p < 0,002$).

Ces résultats confirment en partie notre hypothèse H1A. L'agent en mode Cognitif a été perçu comme exprimant plus d'émotions plus que l'agent dans les conditions Catégorique et Sans-émotion (Figure 83). Le score a été obtenu en sommant les scores des cinq questions se rapportant à l'hypothèse. Chaque question étant sur une échelle de Likert en 5 points (score de 0 à 4), le score maximal est de 20.

H1B: Les utilisateurs perçoivent une dynamique expressive plus élevée quand le nombre d'expressions du visage est plus élevé.

La perception de la dynamique émotionnelle est plus élevée dans la condition Cognitive que dans la condition de Sans-émotion ($t(15) = -2,06$, $p < 0,05$).

Les conditions Sans-émotion et Catégorielles sont équivalentes en termes de perception de la dynamique expressive (t (13) = -0,83, p <0,42 NS).

Enfin, la comparaison entre les modes Catégoriel et Cognitif ne révèle qu'un effet tendanciel. Le mode Cognitif semble être perçu aussi comme plus dynamique que le mode Catégoriel (t (18) =- 1,87, p <0,08). La Figure 84 illustre ces résultats. Les scores sont également obtenus en sommant les scores des échelles de Likert et varient également entre 0 et 20.

Ces résultats confirment en partie notre hypothèse H1B.

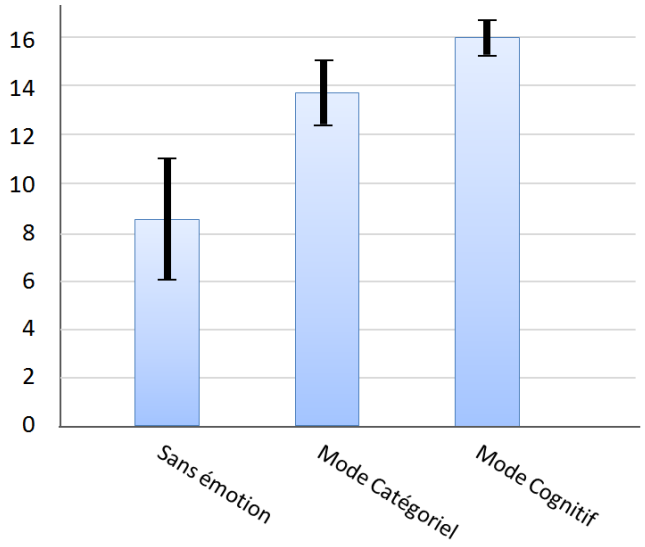


Figure 83 - Perception rapportée des expressions des émotions de l'agent virtuel en fonction des conditions expérimentales (Sans émotion, Catégoriel et Cognitif). Les scores peuvent varier entre 0 et 20.

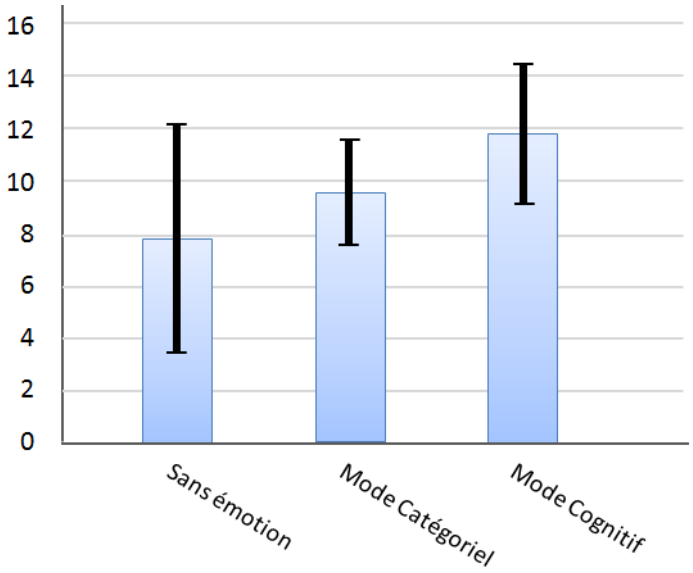


Figure 84 – Perception rapportée de la dynamique expressive en fonction de la condition expérimentale (Sans émotion, Catégoriel, Cognitif) Les scores varient entre 0 et 20.

En résumé, nous observons les relations suivantes, qui confirment partiellement notre hypothèse H1 :

- Perception de l'expressivité émotionnelle:

  Sans émotion < Catégoriel < Cognitif

- Perception de la dynamique expressive :

  Sans émotion = Catégoriel < Cognitif

H2: l'attribution par l'utilisateur d'états émotionnels et cognitifs au personnage virtuel dépend de la condition expressive :

H2A: Les utilisateurs attribuent moins d'états émotionnels au personnage virtuel en condition Sans-émotion que dans les conditions Catégorielles et Cognitive.

Les utilisateurs attribuent un plus grand nombre d'états émotionnels à MARC dans les conditions Catégorielles et Cognitive que les utilisateurs de la condition de Sans-émotion ($t(13) = -2,40$, $p < 0,03$; $t(15) = -4,06$, $p < 0,001$).

Nous n'avons pas observé de différences significatives entre les conditions Catégoriel et Cognitif en termes d'attribution d'états mentaux émotionnels. La Figure 85 illustre ces résultats. Les scores sont obtenus par somme des scores des échelles de Likert et varient entre 0 et 28 (Sept questions). Cependant, ces résultats valident notre hypothèse H2A. Les sujets attribuent en effet un plus grand nombre d'états mentaux émotionnels lorsque l'agent exprime des émotions (mode Catégoriel ou Cognitif) que lorsqu'il n'en exprime pas (mode Sans émotion).

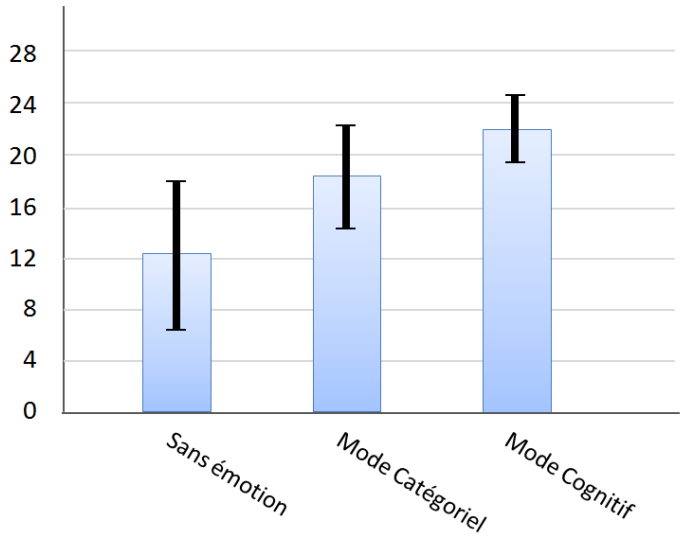


Figure 85 – Attribution d'états mentaux par l'utilisateur à l'agent en fonction de la condition expressive. Les scores varient entre 0 et 28.

H2B: Les utilisateurs attribuent plus d'états mentaux cognitifs non-émotionnels à l'agent virtuel dans la condition Cognitive que dans les conditions Sans émotion et Catégorielle.

L'attribution d'états mentaux cognitifs non-émotionnels est plus élevée dans la condition Cognitive que dans la condition Sans émotion ($t(15) = -2,87$, $p < 0,01$) (Figure 86). Les conditions Catégorielle et Sans-émotion sont équivalentes.

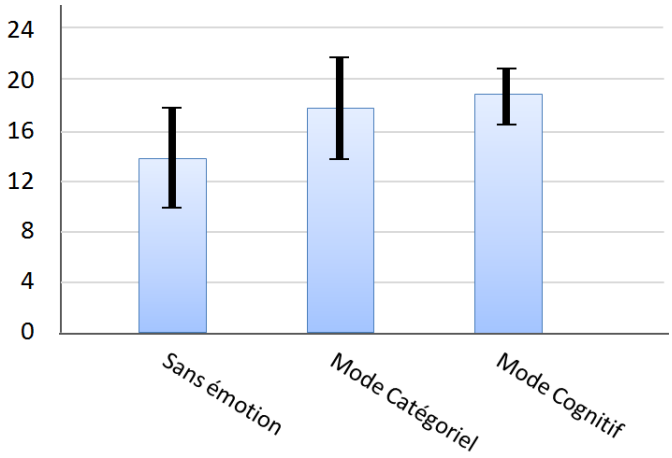


Figure 86 - Attribution d'états mentaux non-émotionnels par l'utilisateur à l'agent en fonction de la condition expressive. Les scores varient entre 0 et 28

Pourtant, aucune différence significative n'est observée entre les conditions Catégoriel et Cognitive.

La Figure 86 illustre ces résultats. Les scores sont obtenus par somme des scores des échelles de Likert et varient entre 0 et 28 (Sept questions).

Ces résultats ne valident pas tout à fait notre hypothèse H2B. Les utilisateurs n'ont pas perçu l'agent Cognitif comme ayant plus d'états mentaux non-émotionnels cognitifs que l'agent Catégoriel.

Pour résumer, nous avons observé les relations suivantes:

• L'attribution d'états mentaux émotionnels :

Sans émotion < Catégoriel = Cognitif

• Attribution d'états mentaux cognitifs non-émotionnels :

Sans émotion < Cognitif **et** Sans émotion = Catégoriel **et** Catégoriel = Cognitif

Mesures objectives

H3: Les utilisateurs gagnent plus souvent et passent plus de temps à jouer lorsque le nombre d'expressions faciales est plus élevé

Les résultats des parties de jeu ont été analysés avec la méthode *chi–squared* (χ^2). Nous avons observé que l'issue du jeu dépend de la condition expressive (χ^2 (2) = 6, 39, p <0,04). Les utilisateurs de la condition Sans émotion ont perdu la partie plus souvent les utilisateurs des conditions Catégorielle et Cognitive. Il semble que l'interaction avec un agent virtuel qui exprime des émotions améliore les performances des utilisateurs. Les utilisateurs ont donc perçu ces expressions d'émotions et semblent les utiliser pour être plus performants.

Le temps de réflexion des utilisateurs a été analysé par analyse de variance en utilisant deux variables. 1) le mode expressif et 2) l'émotion exprimée par le personnage virtuel avant chaque coup. Nous avons également utilisé le LSD de Fisher pour les comparaisons par paires post-hoc. Nos résultats révèlent que la durée totale des actions d'un utilisateur ne dépend pas de la condition expressive (F (2, 755) = 0,39, p = 0,68 NS). La durée moyenne d'un jeu est de 9 minutes 50 secondes. Toutefois, la condition Cognitive montre des différences de durée de jeu en fonction de l'émotion exprimée par l'agent avant le coup de l'utilisateur. Nous avons en effet observé un effet tendanciel en fonction de l'expression émotionnelle de l'agent avant l'action d'un utilisateur (F

(2, 317) = 2,41, p = 0,09). Par exemple : les utilisateurs ont joué plus lentement après que l'agent ai exprimé de la tristesse que lorsque l'agent virtuel est resté neutre avant l'action de l'utilisateur. Un résultat significatif a également été observé pour l'émotion exprimée par le personnage virtuel suite à sa propre action (F (2, 317) = 2,9265, p = 0,05) : les utilisateurs ont joué plus lentement, quand le personnage a exprimé de la fierté après avoir joué, que quand il a exprimé de la joie.

Ces résultats confirment en partie notre hypothèse H3. Les joueurs ont été plus efficaces et ont gagné plus souvent quand ils jouaient contre un agent virtuel expressif et encore plus souvent lorsque l'agent a montré des signes de son évaluation séquentielle de la situation (condition Cognitive). Par ailleurs, nous avons observé que l'émotion exprimée par le personnage avait une certaine incidence sur le comportement des utilisateurs dans le mode Cognitif.

5.5.5 Discussion

Cette étude a exploré la façon dont les utilisateurs perçoivent les expressions faciales émotionnelles affichées par un personnage virtuel au cours d'un jeu interactif. Nous avons comparé trois conditions différentes: une condition sans-émotion, une condition catégorielle, et une condition cognitive. Les résultats que nous avons observés soulèvent plusieurs questions.

Notre expérience suggère que l'affichage dynamique des expressions émotionnelles inspiré par les théories d'évaluation cognitive peut présenter un intérêt pour l'animation de personnages virtuels. En condition Cognitive (lorsque l'agent virtuel affiche des expressions faciales séquentielles des checks d'évaluation), les utilisateurs attribuent plus d'états mentaux non-émotionnels à l'agent virtuel que dans la condition Sans-émotion. Ce résultat est en accord avec Smith et Scott. (1997), qui font valoir que l'expression faciale de chaque check a un sens émotionnel particulier. Notre étude confirme que ces expressions de checks transmettent un message concernant le processus d'évaluation de l'agent; l'expression faciale d'un check peut exprimer à la fois une composante cognitive et une composante émotionnelle. Ainsi, présenter des signes faciaux de l'évaluation cognitive pourrait augmenter la perception que les utilisateurs ont des capacités cognitives de l'agent virtuel. En outre, aucune différence n'a été observée entre le mode Catégoriel et le mode Sans-émotions en termes d'attribution d'états mentaux cognitifs non-émotionnels.

Cependant, nous n'avons pas observé de différence entre les modes Catégoriel et Cognitif en termes d'attribution d'états mentaux non-émotionnels. Une des explications possibles est que l'expression de l'émotion affichée à la fin de chaque séquence d'évaluation du mode Cognitif a été exactement la même que celle affichée dans le mode Catégoriel. Cette caractéristique commune à ces deux conditions pourrait expliquer pourquoi les utilisateurs attribuent autant d'états mentaux non-émotionnels à ces deux conditions. Ce résultat est conforme à Grandjean et Scherer (2008), qui ont observé que des images statiques d'expressions faciales des émotions de base permettent de déduire à la fois les catégories émotionnelles et l'évaluation cognitive associée. Une autre explication pourrait être la nature rapide de la séquence de signes d'évaluation affichés avant l'image finale. Une étude plus approfondie est nécessaire pour explorer l'impact de ces aspects temporels et comment ils sont perçus par les sujets.

L'expression émotionnelle du personnage virtuel a eu un impact sur le comportement des utilisateurs pendant la partie. Les utilisateurs semblent avoir utilisé les émotions exprimées par l'agent pour guider leurs actions et leur réflexion. Les sujets jouant face à un personnage virtuel exprimant ses émotions ont utilisé ces émotions pour influencer leur stratégie de jeu. En conséquence, ils ont gagné plus souvent que lorsque l'agent n'a pas exprimé d'émotion. Ce résultat aurait peut-être été différent si l'agent n'avait pas toujours été « honnête » et exprimé sincèrement son évaluation de la situation.

Nous avons également observé que les utilisateurs de la condition Cognitive ont pris les indices subtils des checks en compte dans leur stratégie, et que cela a influencé leur comportement. Nous avons observé des différences dans le temps de jeu de ces utilisateurs, et que ce temps de jeu dépend de l'émotion affichée par l'agent juste avant leur coup. Les utilisateurs ont pris plus de temps pour préparer un coup après que le

personnage virtuel ait affiché une expression de fierté. Cet effet n'a pas été observé dans la condition Catégoriel. Cette observation pourrait être expliquée par les significations sous-jacentes de la fierté en termes d'interactions sociales et de jeu de stratégie.

Nos résultats suggèrent une différence entre la perception instantanée inconsciente des signes de l'évaluation cognitive par les utilisateurs (qui influe sur la manière dont ils jouent dans la condition Cognitive) et le rapport post-hoc de la perception des états mentaux cognitifs non-émotionnels (pas de différence entre les conditions Catégorielle et Cognitive). Cette différence montre une fois de plus la complexité bien connue de mesurer la perception émotionnelle après, mais aussi durant l'interaction.

Enfin, le plateau de jeu a été affiché sur un écran horizontal alors que le personnage virtuel était affiché sur un écran vertical. Plusieurs utilisateurs ont été tellement engagés dans le jeu qu'ils se concentraient uniquement sur le plateau et non sur le personnage virtuel. Certains utilisateurs finissent par ignorer les réactions de l'agent virtuel, surtout quand il n'a exprimé aucune émotion. MARC fournit alors peu d'informations, ou même aucune information, quand il n'est pas expressif. Certains utilisateurs se focalisent alors uniquement sur le jeu de Reversi, car leur objectif principal est de gagner la partie.

5.6 Limitations du modèle informatique proposé

Dans la section 5.4.2, nous avons expliqué comment nous avons étendu notre système d'agent virtuel MARC avec la capacité d'évaluer la situation dynamiquement au cours d'un jeu interactif. Le personnage virtuel affiche dynamiquement une séquence d'expressions faciales correspondant aux évaluations cognitives séquentielles. Nous avons effectué une étude expérimentale dans laquelle nous avons comparé trois modes expressifs (Sans émotion, Catégoriel, et Cognitif). Certains de nos résultats suggèrent que les systèmes d'animation faciale pourraient bénéficier de l'utilisation d'animation séquentielle plutôt que d'utiliser uniquement des expressions faciales prototypiques.

Cette étude peut être étendue dans plusieurs directions. Le contexte d'interaction de notre application de jeu a été volontairement limité à certains événements qui se produisent pendant le jeu pour mieux maîtriser nos conditions d'expérimentation. Nous avons seulement examiné sept checks du modèle CPM. Cette étude peut être étendue pour inclure d'autres checks et d'autres types événements. L'utilisation d'un contexte plus complet et interactif pourrait conduire à un ensemble de plus d'émotions (y compris des émotions complexes) et des comportements peut-être plus réalistes pourraient être affichés par l'agent virtuel. De même, certaines théories de l'évaluation cognitive conceptualisent le processus cognitif comme un processus récursif : c'est un effort constant de l'individu pour affiner les résultats de son évaluation et les mettre en conformité avec la réalité (Scherer, 2010). Le résultat est un changement constant de l'état affectif et de l'intensité de l'émotion. La mise en œuvre d'un tel processus de réévaluation chez notre agent virtuel pourrait aussi contribuer à donner la perception d'une plus grande «intelligence émotionnelle» de l'agent.

Nos résultats indiquent également que les utilisateurs distinguent l'expressivité du personnage virtuel et l'émotion qu'il exprime. Afficher plus d'expressions d'émotions augmente l'expressivité perçue et l'évaluation de la dynamique sous-jacente, mais pas le nombre des états émotionnels internes attribués à l'agent. D'autres recherches sont nécessaires pour évaluer l'impact de notre modèle émotionnel sur la crédibilité perçue des comportements de l'agent. Certains utilisateurs ont mentionné qu'ils n'avaient pas regardé MARC parce qu'ils ne savaient pas s'ils pouvaient lui faire confiance. En effet, la perception des expressions doit être distinguée de l'attribution des états mentaux internes à autrui. Par exemple, on peut exprimer intentionnellement de la colère pour effrayer quelqu'un d'autre sans être en colère. Symétriquement, il est possible de percevoir la colère, sans attribuer l'état mental associé à l'individu exprimant la colère (Fridlund, 1994). Cette observation ouvre des questions de recherche concernant les stratégies qui devraient être mises en œuvre dans l'agent virtuel (par exemple, le masquage des émotions) et leur impact sur les performances de l'utilisateur et sur sa perception.

Dans le système présenté dans cette expérimentation, seules les actions de l'utilisateur sur le plateau de jeu sont considérées comme entrées du système. En dehors de ces événements, aucune information n'a été considérée. Nous pourrions avoir besoin d'étudier les moments où l'utilisateur est distrait de l'application, ou quand il

exprime des états affectifs. En effet, prendre en compte l'état affectif de l'utilisateur est essentiel pour créer une boucle d'interaction affective complète et pour permettre l'élaboration de stratégies affectives plus sophistiquées chez le personnage virtuel.

5.7 Résumé et limites de l'approche cognitive

L'approche cognitive constitue une perspective de recherche riche et prometteuse. Elle est en revanche confrontée à un certain nombre de verrous théoriques et technologiques. En complément aux modèles informatiques cognitifs existants, l'étude présentée dans ce chapitre ouvre des perspectives sur la modélisation d'approches cognitives dans le cadre de l'interaction avec des agents virtuels expressifs dans le cadre d'une tâche.

D'une part, l'approche cognitive permet de simuler une dynamique affective plus riche et plus complexe que les approches catégorielles classiques. D'autre part, les modèles tels que CPM donnent les briques de base pour construire le comportement expressif associé à la simulation affective. Ainsi, il est possible de générer un comportement expressif dynamique et subtil, et d'explorer de nouvelles interactions entre l'agent virtuel et l'utilisateur.

Notre modèle expressif dynamique basé sur la théorie du CPM suggère que ces expressions dynamiques sont bien perçues par les utilisateurs, et qu'elles modifient leur perception.

Notre étude suggère également que l'utilisation de modèles cognitifs modifie le comportement de l'utilisateur. D'autres études sont nécessaires pour évaluer le potentiel des théories cognitives dans cette direction.

Le modèle implémenté dans cette première version est cependant limité par un certain nombre de choix d'implémentation. Notre modèle implique de définir à l'avance les émotions possibles, avec leur profil en termes de checks d'évaluation cognitive. De plus, la théorie complète considère un nombre plus important de checks, qui permettent donc une plus grande finesse du processus cognitif simulé.

Nous avions également limité les possibilités de situation par un protocole strict, nous permettant de bien contrôler le déroulement de l'étude. Nous avions donc décidé d'interdire des situations telles que la triche et le « bluff », avec lesquelles nous aurions pu faire varier les checks relatifs aux normes sociales, et introduire plus de variété dans les réactions de l'agent.

Pour finir, plusieurs études mettent en relief l'importance du contexte social dans l'évaluation cognitive. Dans notre étude, il aurait été préférable de pouvoir mesurer l'état émotionnel de l'utilisateur pour le prendre en compte dans l'évaluation cognitive de l'agent. L'unique entrée utilisateur (à savoir le plateau de jeu) constitue une entrée trop incomplète pour permettre une vraie boucle d'interaction affective.

Néanmoins, les premiers résultats de cette étude sont encourageants, et mettent en lumière les bénéfices potentiels des modèles cognitifs en situation d'interaction affective avec un agent virtuel.

Pour étendre ce travail, nous pouvons nous orienter vers des interactions sociales dans lesquelles le personnage virtuel est la seule interface visible. Les utilisateurs ne seront pas distraits par un plateau de jeu et pourront alors se concentrer sur les expressions faciales du personnage virtuel. L'autre aspect serait de créer une situation collaborative (avec objectifs communs) pourrait faciliter l'interaction affective entre l'utilisateur et le personnage virtuel. Nous émettons l'hypothèse que l'utilisation de MARC comme agent collaboratif pourrait aider les utilisateurs à faire confiance au personnage virtuel. Les utilisateurs serraient alors peut-être plus engagés dans l'interaction et fourniraient des commentaires plus précis sur la façon dont ils perçoivent le personnage virtuel lors de l'interaction.

Chapitre 6. Approche sociale des émotions pour l'animation faciale : le cas du social appraisal

Sommaire du chapitre

6.1 Intérêt de l'approche cognitive et sociale pour les interactions entre personnages virtuels

Peu d'études considèrent comment l'expressivité d'un agent virtuel influence l'expressivité d'autres personnages virtuels ou l'expérience de l'utilisateur. Pourtant, plusieurs théories de la psychologie traitent du caractère social de l'émotion (Averill, 1985). Comme nous l'avons vu dans l'état de l'art, le phénomène du *social appraisal*, ou évaluation cognitive sociale, fait référence à l'influence de l'expression d'autrui sur l'évaluation cognitive d'un individu, ainsi que sur sa réaction émotionnelle.

Plusieurs problématiques émergent. Comment adapter les théories cognitives et sociales pour animer plusieurs personnages virtuels ? L'utilisateur est-il capable de percevoir l'influence sociale d'un agent virtuel sur un autre ? Peut-on utiliser le principe de social appraisal pour influencer l'évaluation émotionnelle d'un utilisateur ?

L'utilisation de scènes virtuelles dans lesquelles évoluent plusieurs personnages virtuels impose de gérer à la fois les interactions avec l'utilisateur et les interactions entre les personnages virtuels (André et Rist, 2000). Concevoir un modèle où les émotions d'un personnage sont automatiquement prises en compte dans les évaluations et les réactions expressives des autres personnages permettrait donc de ne pas avoir à scripter ces comportements.

Pour étudier comment de tels mécanismes sont perçus par l'utilisateur, nous avons choisi de nous intéresser à un phénomène particulier : le *social appraisal*, dont plusieurs études ont montré les effets sur la perception humaine (Mumenthaler et Sander, 2009).

6.2 Architecture de MARC v4 : approche sociale

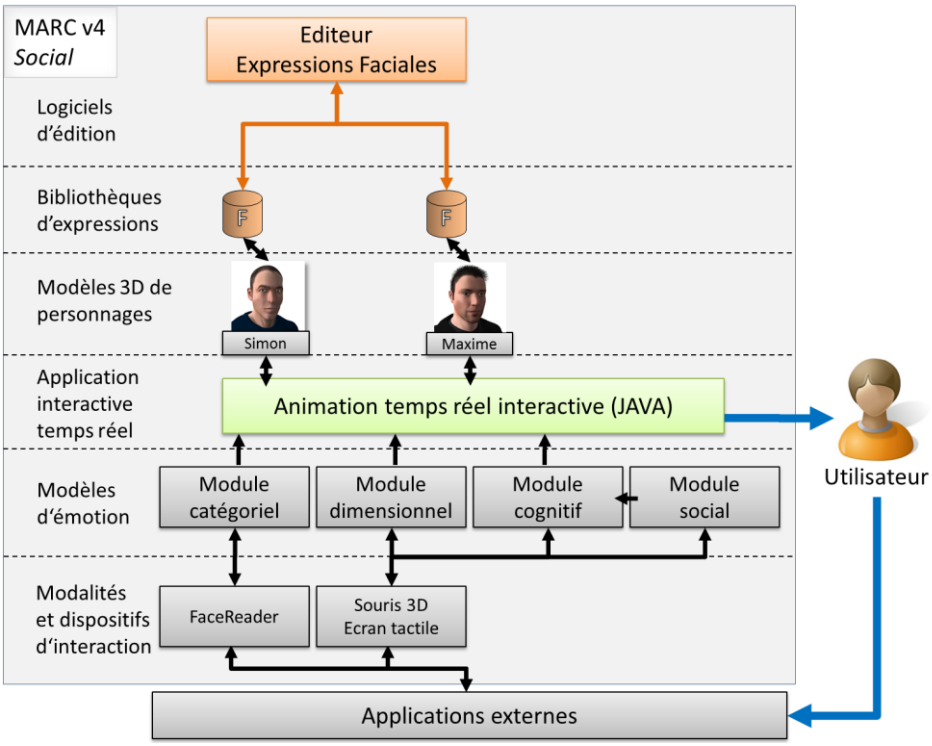


Figure 87 - Architecture de MARC v4: Modèle Social

6.2.1 MARC : Ajout du module de social ré-appraisal

En collaboration avec Christian Mumenthaler et David Sander de l'université de Genève, nous avons donc ajouté un module d'évaluation cognitive sociale à MARC. Contrairement aux autres modules émotionnels de MARC, ce module requiert deux agents dans la scène virtuelle. Le premier des personnages, appelé par la suite *"Observateur"*, effectue une réévaluation cognitive sociale sur la base de l'observation de l'expression de l'autre personnage, nommé dans la suite *"Observé"*.

Nous avons donc ajouté un deuxième modèle de personnage 3D. Entre les deux personnages, la relation est la suivante : l'un des personnages, appelé par la suite *"Observateur"*, effectue une réévaluation cognitive sociale sur la base de l'observation de l'expression de l'autre personnage, nommé dans la suite *"Observé"*.

Pour modéliser le phénomène du social appraisal, nous avons proposé une architecture logicielle, basée sur le module cognitif existant dans MARC (Figure 88) et décrit dans la section 5.4 de ce document.

Lors de la première phase, les deux agents effectuent une évaluation de la situation, qui peut être différente entre les deux agents. Le visage de chaque agent reflète alors son évaluation en utilisant le module cognitif de MARC. L'agent *observant* regarde ensuite automatiquement l'agent *observé*. Il peut alors percevoir l'expression de l'agent *observé*, et déduire quelle évaluation a été faite par celui-ci. En effet, Bänzinger et al. (2009), ont montré que l'expression faciale d'un humain permet de déduire quelle est l'évaluation qu'il fait d'une situation. En revanche, cette déduction peut introduire des erreurs. Notre architecture prend en compte cette marge d'erreur en incluant un « filtre perceptif ». Les valeurs des checks que l'agent *observant* va attribuer à l'agent *observé* peuvent donc être altérées par rapport aux checks exprimés par l'agent *observé*.

Une fois que l'agent *observant* a pris en compte l'expression de l'agent *observé*, le module de réévaluation cognitive sociale va définir, en fonction des deux premières réactions, la nouvelle évaluation cognitive de l'agent *observant*. Cette nouvelle évaluation va donner lieu à une nouvelle séquence d'expression faciale.

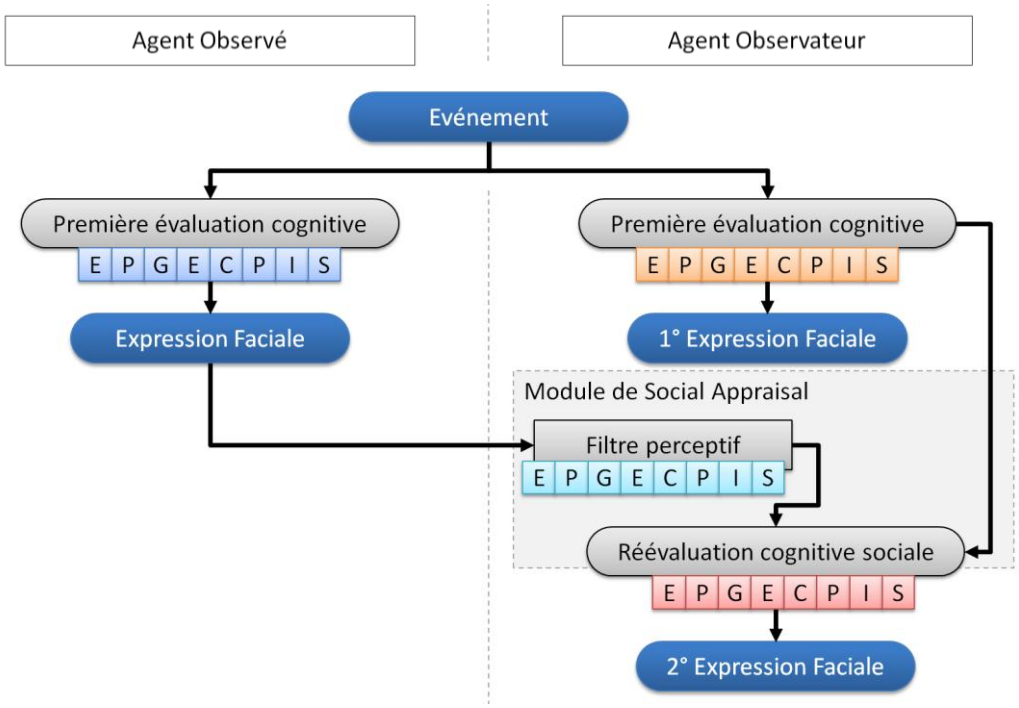


Figure 88 - Architecture du module de Social Appraisal

6.2.2 Les différentes approches proposées

Très peu de données sont disponibles sur le rôle de l'évaluation cognitive sociale dans l'évaluation faite par l'agent *observant* et sur l'émotion qu'il exprime. Nous avons décidé de tester plusieurs variantes du module de réévaluation cognitive sociale. Nous avons donc défini deux approches. La première consiste à copier certains des checks perçus chez l'agent *observé*, pour « mélanger » les deux évaluations. La seconde approche consiste à définir, de manière empirique, un certain nombre de règles logiques qui seront utilisées pour définir la seconde évaluation.

6.2.3 Règles logiques du module d'évaluation cognitive sociale

La description des règles est faite check par check, et d'un point de vue interne à la personne qui fait la réévaluation cognitive. Par exemple, « Si j'ai évalué la situation de telle manière, et que j'observe que l'autre agent a fait telle évaluation, alors mon évaluation change de telle façon. ».

Les règles ont été écrites dans un contexte où les agents sont en ***collaboration***, et non en compétition. Ainsi, ils partagent par exemple les mêmes buts. Chaque règle est décrite en français puis sous forme de règles logiques.

Pour les règles logiques, nous utiliserons les symboles suivants :

- $Prem_{Ch}$ correspond à la valeur du check « Ch » de la première évaluation de *l'observant*.

- $Second_{Ch}$ correspond à la valeur du check « Ch » de la réévaluation sociale de *l'observant*.

- $Percu_{Ch}$ correspond à la valeur perçue par l'*observant* du check « Ch » de l'évaluation exprimée par l'*observé*.

Unexpectedness:

Si j'ai évalué l'événement comme attendu et que j'observe que l'autre agent est surpris par cet événement, alors je suis moi-même surpris (par la réaction de l'autre agent).

$$(Prem_{Unexpec}<=0) \ \wedge \ (Percu_{Unexpec}>0) \ \rightarrow Second_{Unexpec} = Percu_{Unexpec}$$

Par contre, si j'ai évalué l'événement comme attendu et que j'observe la même évaluation chez l'agent observé, alors je ne change pas mon évaluation initiale (je ne suis pas surpris par cet événement).

$$(Prem_{Unexpec}<=0) \ \wedge \ (Percu_{Unexpec}<=0) \ \rightarrow Second_{Unexpec} = Prem_{Unexpec}$$

Si j'ai évalué l'événement comme inattendu, alors cette surprise disparait lors de ma seconde évaluation. (Car ce n'est plus un événement nouveau), quelle que soit l'expression de l'agent observé.

$$(Prem_{Unexpec}>0) \ \rightarrow Second_{Unexpec} = 0$$

Unpleasantness

Si j'ai évalué l'événement comme plaisant, mais que j'observe que l'autre agent l'a évalué comme obstructif, et qu'il pense ne pas avoir le contrôle sur cet événement, mon évaluation de l'aspect plaisant l'événement sera réduite d'autant que l'événement est obstructif agent observé.

$$(Prem_{Pleasant}>=0) \ \wedge \ (Percu_{Obstruct}>=0) \ \wedge \ (Percu_{Control}<0) \ \rightarrow Second_{Pleasant} = Prem_{Pleasant} - Percu_{Obstruct}$$

Cela donne, par exemple, que si l'événement est joyeux pour moi, mais qu'il déclenche de la tristesse chez l'agent observé, alors ma joie est diminuée.

Sinon, mon évaluation de l'événement ne change pas selon ce critère.

$(Prem_{Pleasant}<0)$ v $(Percu_{Obstruct}<0)$ v $(Percu_{Control}>=0)$ → $Second_{Pleasant} = Prem_{Pleasant}$

Goal Hindrance

Si j'ai évalué l'événement comme obstructif pour moi, et que j'évalue que mon niveau de contrôle sur cet événement est négatif, et que l'autre agent l'a perçu comme plaisant et non obstructif, alors l'événement me parait encore plus obstructif.

$(Prem_{Obstruct}>=0) \wedge (Prem_{Control}<0) \wedge (Percu_{Pleasant}>=0) \wedge (Percu_{Obstruct}<0)$ → $Second_{Obstruct} = Prem_{Obstruct}$ x 2

Par exemple, si l'événement est source de tristesse pour moi, et que l'autre agent exprime de la joie, alors je peux penser que l'autre agent n'a pas correctement analysé la situation, ce qui est à l'encontre de mes buts, puisque l'autre agent doit comprendre le problème pour contribuer remédier au problème.

Sinon, mon évaluation de l'événement ne change pas selon ce critère.

$(Prem_{Obstruct}<0)$ v $(Prem_{Control}>=0)$ v $(Percu_{Pleasant}<0)$ v $(Percu_{Obstruct}>=0)$ → $Second_{Obstruct} = Prem_{Obstruct}$

Coping Control

Si j'ai évalué l'événement comme obstructif, que mon contrôle est négatif, et que l'autre agent l'a perçu comme plaisant et non obstructif, alors mon sentiment de contrôle est encore plus réduit.

$(Prem_{Obstruct}>=0) \wedge (Prem_{Control}<0) \wedge (Percu_{Pleasant}>=0) \wedge (Percu_{Obstruct}<0)$ → $Second_{Control} = Prem_{Control}$ x 2

Par exemple, si l'événement est source de tristesse pour moi, et que l'autre agent exprime de la joie, alors je peux penser que l'autre agent n'a pas correctement analysé la situation, ce qui diminue mes possibilités contrôler l'événement, puisque l'autre agent doit comprendre le problème pour m'aider.

Sinon, j'adopte le sentiment de contrôle maximum entre le mien, et celui perçu chez l'autre agent.

$(Prem_{Obstruct}<0)$ v $(Prem_{Control}>=0)$ v $(Percu_{Pleasant}<0)$ v $(Percu_{Obstruct}>=0)$ → $Second_{Control} = \max(Prem_{Control}, Percu_{Control})$

Dans le cas où mon évaluation est neutre, et que l'autre agent exprime de la joie, alors je peux penser que l'autre agent envisage des manières de contrôler les conséquence de l'événement que je n'ai pas imaginé, et qu'il peut donc m'aider. J'adopte donc son point de vue.

Coping Power

Si j'ai évalué l'événement comme obstructif, que mon contrôle est négatif, et que l'autre agent l'a perçu comme plaisant et non obstructif, alors mon power augmente vers le positif, car je me repose sur l'autre agent qui semble juger l'événement de manière optimiste.

(A) $(Prem_{Obstruct}>=0) \wedge (Prem_{Control}<0) \wedge (Percu_{Pleasant}>=0) \wedge (Percu_{Obstruct}<0)$ → $Second_{power} = Prem_{power} - Percu_{Obstruct}$

Par exemple, si l'événement est source de tristesse pour moi, et que l'autre agent exprime de la joie, alors je peux penser que l'autre agent envisage des ressources possibles que je n'ai pas envisagées moi-même et qu'il peut donc m'aider. Mon niveau de power est donc revu à la hausse.

Si l'autre agent et moi avons tous les deux évalué l'événement comme obstructif, jugé notre niveau de contrôle élevé, mais que nous avons évalué le power différemment, alors son évaluation m'influence, et mon niveau de power tend vers le sien.

(B) $(Prem_{Obstruct}>=0) \wedge (Percu_{Obstruct}>=0) \wedge (Prem_{Power} \times Percu_{power} < 0) \rightarrow Second_{Power} = (Prem_{power} + Percu_{Power}) / 2$

Par exemple, si l'événement est source de Colère pour moi, et que l'autre agent exprime de la Peur, alors je peux penser que l'autre agent n'envisage pas autant de ressources possibles que moi. Mon niveau de power est donc revu à la baisse.

Dans les autres cas, j'adopte le power maximum entre le mien, et celui perçu chez l'autre agent.

$\neg(A) \wedge \neg(B) \rightarrow Second_{Power} = \max (Prem_{power}, Percu_{Power})$

Dans le cas ou mon évaluation est neutre, et que l'autre agent exprime de la Colère, alors je peux penser que l'autre agent envisage des ressources possibles que je ne vois pas. J'adopte donc son point de vue.

6.3 Evaluation du module d'évaluation cognitive sociale

6.3.1 Matériel expérimental

Afin d'évaluer l'impact du module de social appraisal sur la perception des utilisateurs. Nous avons mené une étude préliminaire pour comparer plusieurs conditions impliquant deux agents.

L'agent observé a un comportement identique dans toutes les conditions expérimentales. L'agent observant est lui dans l'une des catégories de conditions suivantes :

• L'agent observateur exprime son émotion initiale, regarde l'agent observé et n'affiche pas de seconde expression faciale (condition contrôle)

• L'agent observateur exprime son émotion initiale, regarde l'agent observé et affiche alors son expression issue de la réévaluation cognitive sociale en mode « Copie ».

• L'agent observateur exprime son émotion initiale, regarde l'agent observé et affiche alors une nouvelle expression issue du mode « Règles logiques ».

La catégorie « Copie » a été séparée en sous catégories. Nous avons choisi de ne copier qu'un check à la fois, en nous limitant aux checks qui ont une influence sur le visage. Quatre conditions « Copie » ont donc été mises en place, et donc, six conditions ont été utilisées pour l'agent observateur dans cette étude :

- « règles logiques »
- « copie du check Expectedness »
- « copie du check Pleasantness »
- « copie du check Goal-obstruction »
- « copie du check Coping-Ability »
- Condition contrôle « Pas de réévaluation »

6.3.2 Hypothèses

H1: Hypothèse concernant l'expérience émotionnelle

H1. L'agent Observant devrait être perçu comme exprimant plusieurs émotions uniquement lorsqu'il effectue une réévaluation sociale. De plus les règles logiques étant plus complexes, elles donnent lieu à des émotions plus variées qui seront mieux perçues. La relation attendue est donc :

Contrôle < Copies < Règles

H2: Hypothèses concernant l'influence entre agent

Seul l'agent Observant devrait être perçu comme influencé par l'agent Observé et ce sentiment d'influence devrait être d'autant plus fort que le système de réévaluation sociale sous-jacent est complexe.

H2A. Perception influence Observé sur Observant

Contrôle < Copies < Règles

H2B. Perception influence Observant sur Observé

Contrôle ≈ Copies ≈ Règles

H3: Hypothèses concernant la pertinence de l'expression issue de la réévaluation sociale

H3. La pertinence de la seconde réaction perçue devrait suivre la relation suivante :

Contrôle < Copies < Règles

6.3.3 Protocole

Le questionnaire est présenté sous forme de page web. Une vidéo de 700 par 360 pixels est présentée en haut de la page. Nous avons utilisé pour cette étude deux modèles 3D différents. Nous avons contrebalancé l'effet du personnage, en alternant leur apparence et leur rôle (observé / observateur). L'observateur est toujours présenté à droite, et l'observé à gauche (Figure 89). Au début d'une animation les deux personnages regardent dans la direction de l'utilisateur. Au milieu de l'animation, l'agent observateur tourne la tête et regarde l'agent observé. Il revient ensuite face à l'utilisateur.

Le formulaire web se compose d'un ensemble de questions auxquelles l'utilisateur répond par une échelle de Likert à 5 points. Chaque question est posée deux fois, une fois pour chaque agent. La liste des questions a été établie pour mesurer les effets cités dans nos hypothèses, par exemple : « *Maxime est influencé par l'expression émotionnelle de Simon* ». De plus nous avons effectué certains contrôles, tels que « *Au cours de cette scène, Maxime a regardé Simon* ».

6.3.4 Sujets

24 sujets ont terminé le questionnaire, 11 femmes et 13 hommes, (âge moyen 27,81 ans), et en un temps moyen de 39 minutes.

Figure 89 - Image extraite de l'une des vidéos stimuli de l'étude du social appraisal. A gauche, Maxime joue le rôle de l'observé. A droite, Simon joue le rôle de l'observant. En haut, les deux personnages sont vus de face. En bas, Simon observe la réaction de Maxime.

6.3.5 Résultats

Mumenthaler et Sander (2009) ont montré l'importance de la perception du jeu de regard pour la perception du *social appraisal*. Pour cette raison, nous avons mesuré que les sujets percevaient correctement le jeu de regard. Les résultats indiquent (Figure 90) que l'agent observé a bien été perçu comme ne regardant pas l'agent observant. Au contraire, l'agent observant a bien été perçu comme regardant l'agent observé. Ces données valident donc que le jeu de regard entre les deux agents a bien été perçu.

La perception de l'expressivité émotionnelle a été mesurée en demandant aux sujets s'ils avaient perçu aucune, une, ou plusieurs émotions. Les résultats obtenus aux questions axées sur l'expression sont identiques à ceux obtenus aux questions axées sur le ressenti des émotions (Figure 91).

La condition « Règles-logiques » semble donc permettre d'un peu mieux exprimer le changement d'état affectif que les autres conditions, bien que cette différence ne soit significative ($p<0.05$) qu'avec « Copie Expectedness » et tendancielle ($p<0.1$) avec « Copie Goal-obstruction » et « Copie Coping-ability ». La condition « Règles-logiques » ne présente en revanche pas de différence significative avec la condition « Copie Pleasantness ».

Bien que ces effets soient en partie tendanciels, notre hypothèse H1 est confirmée. Nous observons en effet la relation suivante sur le nombre d'émotions perçue chez l'agent observant :

Contrôle < Copies < Règles

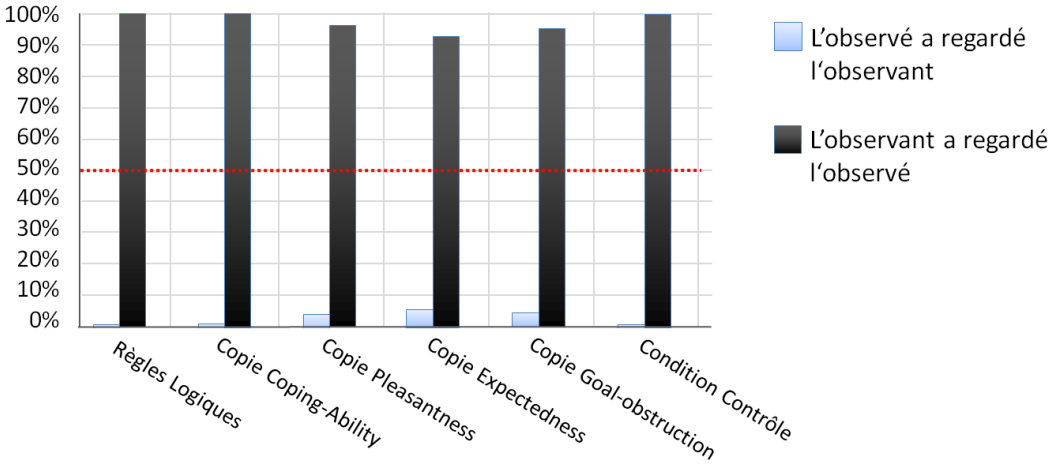


Figure 90 - Perception du jeu de regard entre l'agent observé et l'agent observant en fonction du mode de réévaluation sociale. La ligne pointillée représente le seuil de hasard (50%)

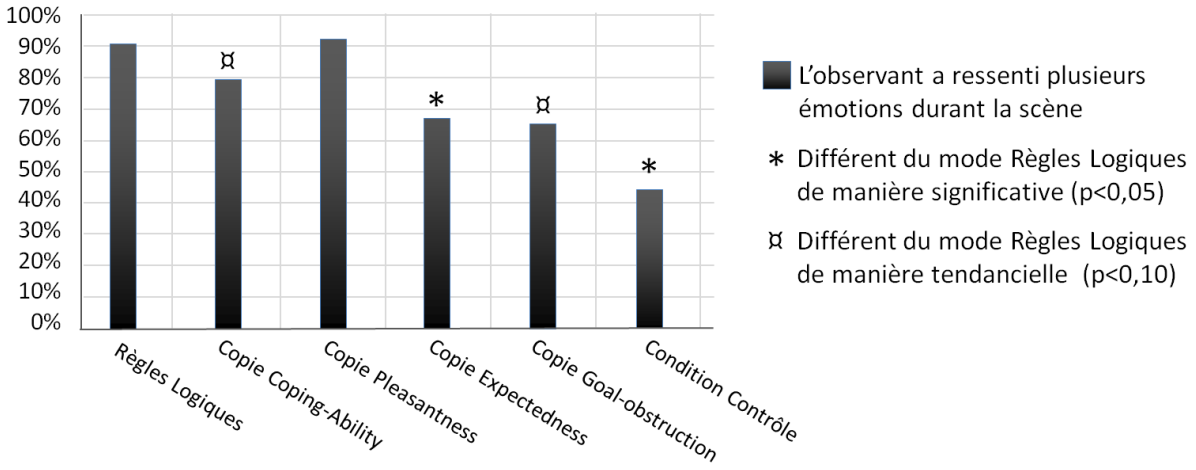


Figure 91 - Perception de l'expression de l'agent observant. "L'observant a ressenti plusieurs émotions durant la scène". 100% correspond à "tout à fait d'accord", 0% correspond à "pas du tout d'accord".

En ce qui concerne la perception de l'influence (Figure 92), nous avons mesuré que l'agent observant est généralement perçu comme étant influencé par l'agent observé, sauf dans la condition contrôle. A l'inverse, l'agent observé est généralement perçu comme n'étant pas influencé par l'agent observant.

On observe également que dans la condition « contrôle », l'agent observant et l'agent observé sont perçus de manière similaire comme non influencé par l'autre. Ainsi, la seule présence d'un jeu de regard ne suffit pas à créer la perception d'une influence émotionnelle.

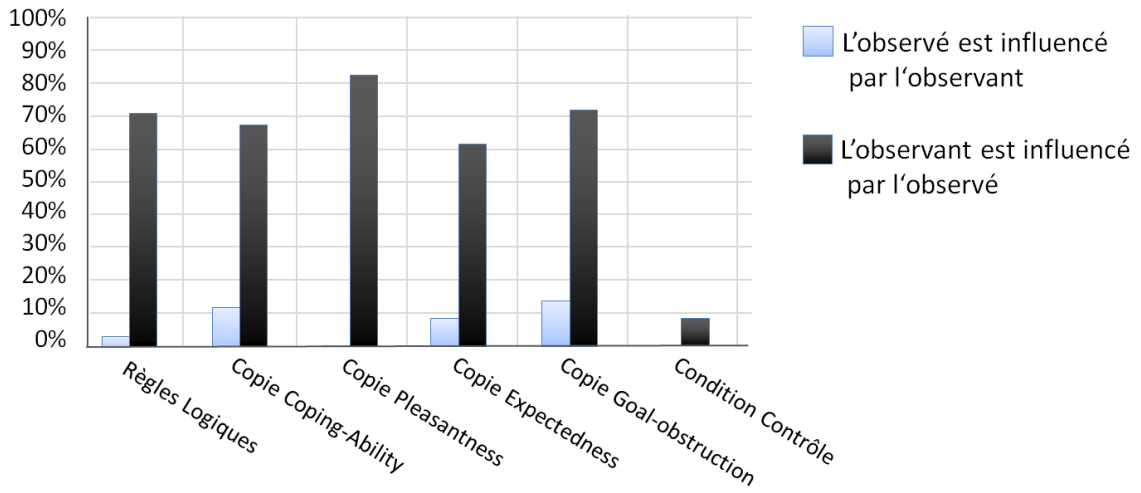


Figure 92 - Comparaison des influences entre l'agent observé et l'agent observant

Lorsqu'on compare la perception des utilisateurs en fonction des conditions, on observe également que la perception de l'auto influence des deux agents est similaire dans les conditions Copies et dans la condition Règles-Logiques.

La relation observée pour l'influence est donc :

Pour l'agent observé : Contrôle ≈ Copies ≈ Règles

Pour l'agent observant : Contrôle < Copies ≈ Règles

Ainsi, l'hypothèse H2 n'est que partiellement validée, car nous n'observons pas la relation attendue dans la perception de l'agent observant en condition « règles-logiques » et dans les conditions « copies »

Notre dernière hypothèse concerne la pertinence de la seconde réaction de l'agent observant. Les sujets devaient donc indiquer leur accord/désaccord avec l'affirmation : « La réaction de Maxime s'est bien adaptée à l'expression de Simon ».

Globalement, nous observons (Figure 93) que l'agent observant est perçu comme s'adaptant mieux que l'agent observé, sauf dans la condition « Contrôle ». Cependant, aucune différence significative n'apparait entre la condition « Règles-logiques » et les conditions « Copie ». Ces résultats suggèrent donc que la présence de la réévaluation permet de donner l'impression que l'agent observant « s'adapte » à la réaction de l'agent observé.

Notre hypothèse H3 n'est donc pas validée, car nous obtenons la relation suivante :

Contrôle < Copies ≈ Règles

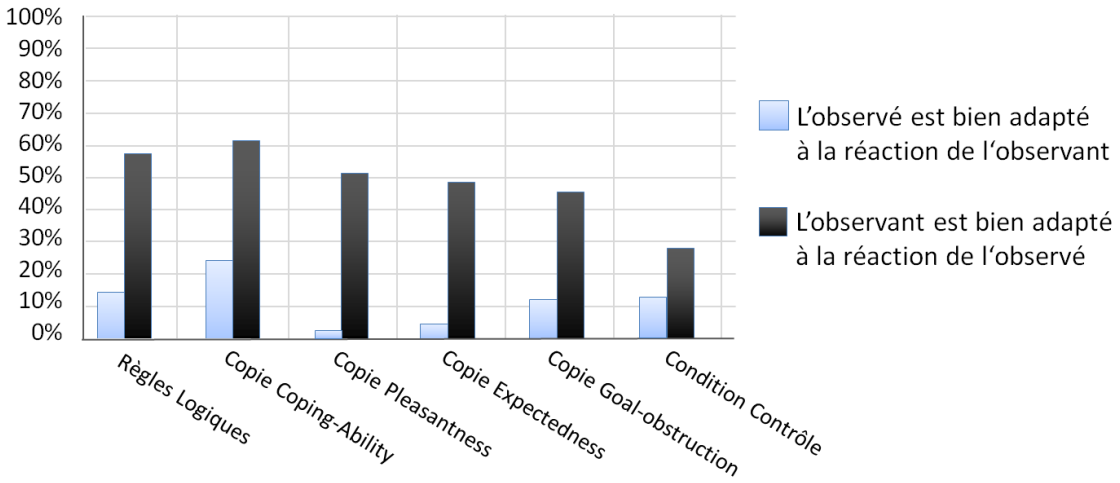


Figure 93 - Perception de l'adaptation d'un agent en fonction de la réaction de l'autre.

6.3.6 Conclusion

Notre première étude exploratoire sur les modèles computationnels de réévaluation sociale suggère donc que l'utilisation d'un modèle computationnel social permet d'améliorer la perception de l'expressivité de l'agent observant. Les sujets perçoivent en effet plus d'émotions exprimées par l'agent observant lorsqu'il réévalue la situation. Ce résultat est important, car il montre l'intérêt de l'utilisation de modèles sociaux.

Cette étude ne nous a pas permis de différencier clairement nos deux approches pour la modélisation de la réévaluation sociale. En effet les modes « Copies », et en particulier le mode « Copie Pleasantness », montrent des résultats similaires au mode « Règles-logique ». D'autres approches devront donc être explorées pour la modélisation de la réévaluation sociale. Par exemple, l'intégration d'un contexte plus large, intégrant les relations sociales entre les deux personnages virtuels, pourrait permettre d'améliorer la pertinence des réactions sociales générées notamment avec le mode « Règles-logiques ».

Nos travaux mettent également en relief la problématique de mener des études perceptives sur des comportements sociaux. Le manque de résultats significatifs pourrait provenir de nos stimuli. L'utilisation de vidéo de courte durée peut en effet être problématique pour la perception de relation sociale à plus long terme. De plus, l'absence de contexte situationnel pourrait nuire à l'évaluation par le sujet de la pertinence de la réaction de l'agent observant. En effet, les sujets semblent s'imaginer une situation, et donc la réaction sociale peut ne pas être cohérente avec la situation imaginée par le sujet.

Ces résultats sont cependant encourageants, car ils montrent que les utilisateurs perçoivent les réactions sociales des personnages virtuels. Ainsi, ces travaux ouvrent des perspectives en termes de modélisation des relations sociales entre l'agent et l'utilisateur, ainsi que pour la modélisation de comportements sociaux dans un groupe d'agents virtuels expressifs.

6.4 Résumé et limites de l'approche cognitive sociale

L'approche sociale des émotions a été moins explorée que les approches catégorielles ou cognitives. Pourtant, les émotions ont un rôle social important (Averill, 1985, Parkinson, 1996, Rimé et al. 1992). Peu de modèles computationnels ont été proposés en s'inspirant de l'approche sociale des émotions. Souvent, ces modèles se limitent à la simulation de l'empathie (McQuiggan et Lester, 2006, Ochs et al., 2008, Leite et al., 2010).

Les modèles empathiques proposés sont généralement uniquement capables d'exprimer de l'empathie, et donc, ils ne permettent pas une interaction générale avec l'utilisateur (McQuiggan et Lester, 2006, Leite et al., 2010). Pour rendre ces modèles plus génériques, il semble nécessaire de les combiner avec d'autres approches des émotions. Par exemple, Ochs et al. (2008) ont combiné leur modèle empathique avec l'approche cognitive via le modèle OCC (Ortony, Clore et Collins, 1988).

Nos travaux combinent également l'approche cognitive (inspirée du modèle CPM) avec l'approche sociale des émotions. Cependant, nous ne nous sommes pas limités à l'empathie, mais nous avons exploré plus largement les réactions sociales à une situation. Le modèle que nous avons proposé pourrait permettre, via un ensemble de règles logiques dédiées, d'explorer le principe d'empathie.

Les premières règles logiques que nous avons proposées empiriquement se limitent à un nombre de situations restreint. Nos premiers résultats semblent pourtant suggérer que notre approche est pertinente. En effet, les sujets de notre étude ont bien perçu que l'agent observateur s'adaptait à la réaction émotionnelle de l'autre agent. Ainsi, le développement de règles logiques plus riches pourrait permettre de mieux modéliser un comportement social crédible, et permettre un des situations plus variées.

Nos règles logiques semblent pouvoir être modulées par des aspects de personnalité de l'agent qui effectue la réévaluation sociale, ainsi que par des informations sur la relation sociale entre les personnages. Ainsi, notre approche semble avoir un grand potentiel pour modéliser toute la complexité du comportement social émotionnel entre deux agents virtuels.

PARTIE III

Cas d'utilisation de la plateforme MARC

Chapitre 7. Cas d'application de la plateforme MARC

Sommaire du chapitre

Publications associées

- Toni Vanhala, Veikko Surakka, M. Courgeon, and J.C. Martin, (2012) *Voluntary Facial Activations Regulate Physiological Arousal and Subjective Experiences During Virtual Social Stimulation*, In: ACM Transactions on Applied Perception, (à paraître)
- M. Courgeon, C. Clavel, N. Tan, J.C. Martin, (2011) *Front View vs. Side View of Facial and Postural Expressions of Emotions in a Virtual Character*, In: LNCS Transactions on Edutainment, 6, pp 132-143
- N. Tan, G. Pruvost, M. Courgeon, C. Clavel, Y. Bellik, J.C. Martin (2011) *A Location-Aware Virtual Character in a Smart Room: Effects on Performance, Presence and Adaptivity*, in: Proceedings of the International Conference of Intelligent User Interface (IUI 2011), Palo Alto, U.S.A, 13-16 février 2011
- O. Grynszpan, J. Nadel, J. Constant, F. Le Barillier, N. Carbonell, J. Simonin, J-C. Martin, M. Courgeon, (2011), *A new virtual environment paradigm for high functioning autism intended to help attentional disengagement in a social contex*t, in: Journal of physical therapy education, 25(1), pp 42-47
- M. Courgeon, M. Rebillat, B. Katz, C. Clavel, J-C. Martin, (2010) *Life-Sized Audiovisual Spatial Social Scenes with Multiple Characters: MARC & SMART-I2*, in: Proceedings of the national conference of AFRV, 6-8 decembre 2010
- N. Tan, C. Clavel, M. Courgeon, J-C. Martin (2010), *Postural Expression of Action Tendency*, in: Proceedings of the ACM Multitimedia 2010's Workshop on Social Signal Processing, 8 septembre 2010
- Clay, M. Courgeon, N. Couture, E. Delord, C. Clavel, J-C. Martin (2009) *Expressive Virtual Modalities for Augmenting the Perception of Affective Movements*, in: Proceedings of the ICMI09 International workshop on Affective Computing (AFFINE), Cambridge, USA, 1-6 novembre, 2009

Ce chapitre présente brièvement les autres outils et utilisations de la plateforme MARC. En effet, en dehors du cadre de cette thèse, centrée sur l'animation faciale en situation d'interaction, d'autres fonctionnalités complémentaires ont été ajoutées à MARC, et d'autres projets de recherche et applications artistiques ont su tirer parti de la plateforme développée. Certaines de ces applications relèvent de la collaboration scientifique, tandis que d'autres ne sont qu'une utilisation des outils proposés par notre plateforme MARC.

7.1 MARC : Outils complémentaires

L'un des aspects nécessaires au réalisme d'un agent virtuel est l'utilisation de la multi modalité. Les postures corporelles et la dynamique corporelle sont des modalités très importantes pour la conception d'un agent expressif complet. De même, la modalité vocale est importante dans un contexte d'interaction, où l'agent doit pouvoir verbaliser certaines informations. Avant de présenter les applications de MARC, nous allons donc présenter l'intégration de ces différentes modalités dans la plateforme MARC.

7.1.1 Animation corporelle expérimentale

7.1.1.1 Objectif

L'expressivité d'un personnage virtuel est multi-modale. Même si cette thèse est focalisée sur les expressions faciales, la plateforme MARC que nous avons développée pourrait bénéficier d'autres modalités expressives. Dans ce contexte, l'animation du corps du personnage nous a semblé pouvoir apporter une modalité expressive supplémentaire importante (Wallbot, 1998). Ainsi, pour permettre l'animation du corps des personnages de MARC, nous avons ajouté un éditeur dédié, associé à un système d'animation temps réel. Cet éditeur présente des fonctions similaires à certains outils d'animation de personnages virtuels (par exemple Poser). En développer un qui soit dédié à notre plateforme d'animation nous permet de coordonner finement les animations faciales et posturales et d'initier des recherches sur les expressions posturales d'émotion.

7.1.1.2 MARC : Ajout des animations posturales

Le système d'animation postural de MARC est basé sur le principe du « *skeleton rigging* ». Un squelette virtuel est ajouté sous la structure géométrique (maillage), et chaque os influence une partie du maillage. Le système d'influence est similaire à celui utilisé pour les points-clés de l'animation faciale. En revanche, contrairement au point clé, dont la translation est reportée sur le maillage, c'est la rotation des os qui est transposée sur le maillage du corps.

Comme pour l'animation faciale, l'éditeur interactif permet de définir la position des os à l'intérieur du maillage, ainsi que l'influence des os sur le maillage. L'influence est calculée de manière semi-automatique. Une première estimation de l'influence du squelette est générée automatiquement par MARC, puis elle peut être retouchée manuellement pour chaque os à l'aide d'un système de brosse de pondération, identique à celui de l'éditeur d'expression faciale.

L'édition se fait par manipulation directe, en utilisant les mêmes contrôles que l'éditeur d'expression faciale. Une des différences entre les deux éditeurs est que l'édition d'animation corporelle permet l'édition de séquence de posture sur une ligne temporelle. L'interpolation de type SLERP (*spherical linear interpolation*) (Shoemake, 1985) est faite automatiquement sur les quaternions associés aux articulations du squelette en respectant la temporalité définie par le concepteur de l'animation. Ainsi, le concepteur n'édite pas une « pose » immobile, comme dans l'éditeur d'expression faciale, mais une dynamique entière. La Figure 94 montre l'interface de l'éditeur de posture de MARC. La partie inférieure est la ligne temporelle, sur laquelle les blocs de couleur représentent chacun une posture élémentaire.

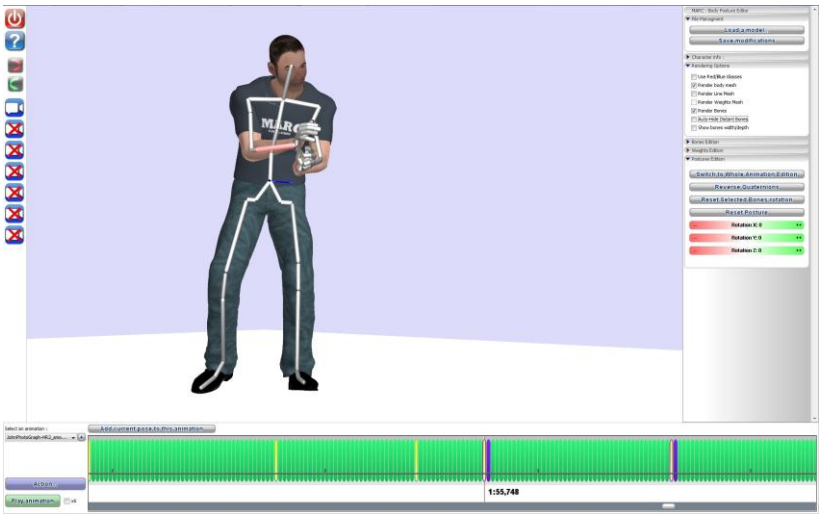


Figure 94 – Interface de l'éditeur de postures de MARC

Afin d'améliorer la qualité de la dynamique, nous avons conçu un système de lissage temporel. Cet outil effectue un lissage de type gaussien à une dimension (temps) sur les rotations des articulations du squelette (Figure 95). L'avantage de cette approche est de réduire l'effet « robot » sur les animations éditées. En revanche, elle absorbe les changements brusques, ce qui peut parfois atténuer l'expressivité de l'animation. Par exemple, dans une animation où le personnage tape dans ses mains, le lissage crée un amortissement de la dynamique au moment du contact des mains. Si on lisse plusieurs fois, les mains peuvent finir par ne même plus entrer en contact. Pour éviter cela, le concepteur doit choisir avec précision la zone temporelle qu'il souhaite lisser.

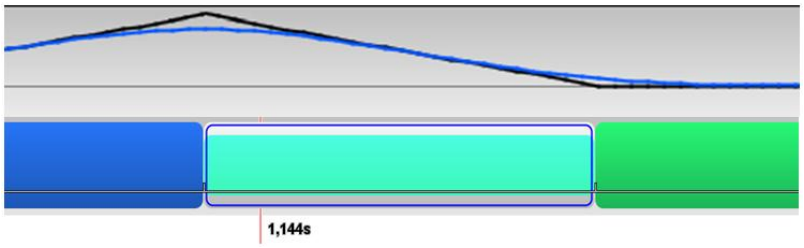


Figure 95 – Exemple de lissage de la dynamique angulaire d'une articulation. L'axe vertical est l'angle de rotation (selon l'axe Y), l'axe horizontal est le temps. En noir, la courbe d'origine avec une dynamique linéaire (robotique). En bleu, la courbe lissée correspondante.

L'édition manuelle d'animation du corps de l'agent est rapidement limitée. Le processus de conception est souvent long, et il est très difficile de créer une animation réaliste en termes d'équilibre, de dynamique, etc.

Les équipes AMI et AA du LIMSI-CNRS s'étant équipées d'un système de capture de mouvements[7], nous avons intégré à MARC la possibilité d'importer des fichiers de mouvements capturés au format standard BVH[8]. Ces fichiers sont ensuite traités, et convertis dans le format d'animation de MARC. Les animations sont alors visibles sur la ligne temporelle, les postures peuvent être modifiées manuellement, et le lissage peut être appliqué. En effet, la captation optique introduit un certain bruit, qui une fois dans MARC se traduit par une légère vibration du personnage virtuel. L'utilisation de l'algorithme de lissage permet d'absorber ces vibrations sans dégrader la dynamique corporelle capturée.

Chaque animation est enregistrée par MARC avec un nom unique, modifiable, qui l'identifie dans la bibliothèque d'animations posturales. Chaque personnage de MARC possède sa propre bibliothèque. Durant l'animation temps réel, ces mouvements peuvent être rejoués et enchainés, en les déclenchant via des scripts BML.

[7] Optitrack® Arena
[8] (Biovision Hierarchy).

7.1.2 Parole et synchronisation labiale

Lors d'expérimentations impliquant l'interaction entre un utilisateur et un personnage virtuel, il est fréquent de devoir « faire parler » le personnage. Bien que n'ayant pas des objectifs de recherche en synthèse audiovisuelle de la parole, MARC a été utilisé dans différentes expérimentations au LIMSI ou par des partenaires extérieurs qui nécessitaient de pouvoir faire parler les personnages virtuels. Pour cela, deux choix sont possibles. Utiliser des fichiers audio de parole enregistrée, ou bien utiliser un synthétiseur vocal, générant à la fois le son et les paramètres d'animation des lèvres du personnage virtuel.

Les développements décrits dans cette section ont été soutenus en partie par l'Action Sur Programme « Tête Parlante » au LIMSI.

7.1.2.1 Détection automatique des phonèmes par analyse fréquentielle

Lorsqu'on utilise un fichier audio enregistré, il faut détecter les différents sons afin d'animer correctement les lèvres du personnage virtuel. MARC possède pour cela une procédure d'analyse simple basée sur l'analyse fréquentielle par transformée de Fourier.

L'analyse implémentée fonctionne en trois étapes. Pour commencer, on effectue la décomposition fréquentielle du son par transformée de Fourier. A partir de cette décomposition, on détecte les changements de fréquences, qui séparent les différents phonèmes contenus dans le son. Une fois que ces phonèmes sont isolés, on analyse les fréquences contenues pour déterminer de quel phonème il s'agit, et donc du visème associé. Pour cela, on considère trois types de phonèmes. Les fricatives, dont les fréquences élevées sont activées (par exemple, les sons « S » ou « F »). Les voyelles, dont les fréquences basses sont activées (par exemple, les sons « A » ou « O »), et les autres, pour lesquelles la détection a échoué.

L'algorithme implémenté dans MARC souffre en effet de limitations importantes. D'une part, la détection des transitions entre phonèmes manque de précision. D'autre part, la détection du type de phonème est très simple, et ne différencie pas les voyelles entre elles. Ainsi, la chaîne de visèmes générée est complétée de manière aléatoire selon les types de phonèmes détectés. L'animation résultante est donc une approximation. Une implémentation plus précise utilisant une transformée de Fourier plus fine permettrait d'obtenir un résultat de meilleure qualité. Cependant, notre algorithme permet de générer en temps réel une animation acceptable dans un contexte limité ou l'articulation labiale de l'agent virtuel n'est pas l'objet d'une attention particulière.

Lorsqu'au contraire, le nombre d'animation à réaliser est important ou que le contenu linguistique doit être choisi dynamiquement, il devient alors nécessaire d'utiliser un système de synthèse de la parole. Nous avons intégré deux synthétiseurs dans MARC. Ces deux systèmes étant conçus pour être intégrés dans des architectures logicielles modulaires, leur intégration n'a posé aucune difficulté technique.

7.1.2.2 Intégration du système de synthèse vocale du LIMSI

Le projet *Tête Parlante* est un projet fédérateur mené au LIMSI-CNRS sous la direction de Jean-Claude Martin entre 2005 et 2007. Ce projet est une collaboration entre trois groupes du LIMSI-CNRS : le groupe *Architectures et Modèles pour l'Iinteraction (AMI),* dans lequel sont notamment intervenus Christian Jacquemin, Jean-Claude Martin, Jean-Paul Sansonnet, et Sébastien Morel. Le groupe *Audio Acoustique (AA),* dans lequel sont intervenus Brian Katz et Christophe d'Alessandro. Et pour finir, le groupe *Langues, Informations et Représentation (LIR),* dans lequel est intervenu Aurélien Max.

Les objectifs scientifiques de ce projet consistent à initier des recherches dans cette direction en intégrant et fédérant plusieurs expertises complémentaires développées ou émergentes dans le département CHM : la synthèse de la parole à partir de texte et la spatialisation du son (PS), la communication multimodale via des agents conversationnels animés et la synthèse de scènes appliquée à la synthèse de visages parlants (AMI), la génération d'énoncés en langage naturel (LIR).

Ce projet a abouti à la mise en place d'un visage virtuel simple et capable d'énoncer des phrases en synchronisant l'articulation des lèvres du visage virtuel avec la voix de synthèse générée par le TTS développé AA et spécifique à la langue française. Le visage virtuel développé par Christian Jacquemin dans le cadre de ce projet a d'ailleurs servi de base pour la conception de la première version de MARC (v1, basée sur le logiciel Virtual Choreographer).

La version actuelle de MARC prend donc toujours en charge les protocoles mis en place dans le projet Tête Parlante, ce qui lui permet de pouvoir utiliser le système de synthèse vocale du LIMSI. Cependant, si l'animation des lèvres est de meilleure qualité grâce aux informations générées par le système Tête Parlante (en comparaison de l'approche par analyse des fichiers sons), la synthèse vocale génère une voix robotisée. L'animation audio-visuelle résultante n'est donc pas très réaliste, et peut être mise en rapport avec l'*uncanny valley* selon les critères de Mitchell et al., (2011) sur les décalages de réalisme entre la voix et le rendu visuel.

7.1.2.3 Intégration du système de synthèse vocale « Open Mary »

Le second système de TTS disponible est le système OpenMary, développé au DFKI (Schroder et al. 2010). OpenMary ne gère en revanche pas le français. Il peut être utilisé pour l'anglais, l'allemand, et d'autres langues. Les informations générées sont similaires à celles du TTS « tête parlante », mais la qualité audio est supérieure. OpenMary permet de sélectionner plusieurs types de voix (masculine/féminine) et d'accents (britannique/américain), et d'y appliquer des effets (écho, modulation de fréquence, etc.).

Ces deux synthétiseurs vocaux sont donc complémentaires, ce qui justifie leur intégration logicielle dans MARC.

7.1.3 Ajout de scènes virtuelles 3D

Afin de pouvoir intégrer les personnages virtuels de MARC dans un environnement virtuel, nous avons ajouté à MARC la gestion d'objets et de l'environnement 3D. Nous avons donc mis en place un système modulable qui permet de sélectionner l'environnement parmi une liste, ou de le supprimer.

La liste des environnements disponibles est définie par un ensemble de fichiers XML de configuration. Ainsi, chaque environnement est décrit par un fichier XML décrivant la position des objets dans la scène, ainsi que leurs propriétés graphiques (géométrie sous forme de fichier X3D, textures, etc…) et sonores (fichier .WAV)

Une fois la scène chargée, les objets composant la scène peuvent être manipulés à travers une extension du langage BML (Vilhjalmsson et al., 2007). En effet, le langage BML permet d'être étendu en utilisant des *namespaces*. Dans notre cas, nous avons utilisé le *namespace* « marc » pour tous les ajouts non standards à l'interpréteur BML de MARC, et ajouté la balise "environment" pour toutes les commandes relatives à l'environnement des agents virtuels.

Ainsi un objet peut être déplacé durant l'interaction temps réel avec l'agent par le script BML suivant :

```
<bml xmlns:marc="http://marc.limsi.fr/schema">

      <marc:environment>

            <object_move name="chair" x="5" y="0" z="3.2" />

      </marc:environment>

</bml>
```

Il est ainsi possible de situer l'interaction avec les agents dans un contexte audio-visuel particulier (Figure 96).

Figure 96 - Plusieurs personnages de MARC dans un environnement 3D.

7.2 Collaborations scientifiques

7.2.1 Etude sur la perception des combinaisons d'expressions faciales et posturales

L'intégration de l'animation posturale dans la plateforme MARC nous permet d'évaluer l'impact de la posture sur l'expressivité globale du personnage virtuel. Plusieurs études comparent la perception des émotions posturales et faciales, en évaluant chaque modalité, séparément, de manière congruente et incongruente (Clavel et al., 2009).

L'animation posturale de MARC a été utilisée dans une étude publiée dans le journal international *Springer Transactions on Edutainment*. Cette section présente un récapitulatif des objectifs et des conclusions de cette étude.

L'expérience présentée dans cette section vise à évaluer l'impact du point de vue et des modalités expressives sur la perception émotionnelle. Nous avons fait varier d'une part l'angle de vue (personnage affiché de Face versus de Profil) et d'autre part les modalités expressives, pour lesquelles nous considérons 1) l'expression monomodale (posture seule et visage seul) versus 2) combinaisons congruentes (visage et posture exprimant une émotion unique) versus 3) combinaisons incongruentes (visage et corps exprimant différentes émotions).

Nous avons sélectionné quatre émotions pour cette étude: Joie, Colère, Tristesse, et Relaxation. Ces émotions ont été choisies car elles sont équilibrées en termes de valence et d'activation (Russell et Mehrabian, 1977). La joie est une émotion positive avec activation élevée. La relaxation est une émotion positive avec activation faible. La colère est une émotion négative avec activation élevée. La tristesse est une émotion négative avec activation faible. En outre, des combinaisons incongrues de la Joie, colère et la tristesse ont été utilisées précédemment dans une étude similaire en utilisant des vues de face (Clavel et al., 2009). Sélectionner ces trois émotions de base nous permet de comparer les résultats avec ceux de l'étude citée précédemment.

Cette étude a fait l'objet d'une collaboration au LIMSI avec Céline Clavel et Ning Tan.

7.2.1.1 Hypothèses

H1: Les combinaisons congruentes devrait être mieux reconnues que les images de visage seul et posture seule.

H2: Le point de vue devrait avoir un impact sur la reconnaissance des émotions.

H3: Dans les combinaisons incongruentes, les sujets doivent percevoir la catégorie de l'émotion exprimée dans le visage, mais le niveau d'activation exprimé dans le corps.

H4: La confiance des sujets devrait être plus faible pour des vues de profil que pour les vues de face; plus faible pour les combinaisons incongruentes que pour les combinaisons congruentes.

7.2.1.2 Résultats et perspectives

Nos résultats présentent des taux de reconnaissance élevés pour des images congruentes et monomodales de l'émotion quel que soit le point de vue et la distance. Cependant, comme dans notre étude sur la perception des expressions faciales de différents points de vue, nous observons que le point de vue a un impact sur la perception, et la confiance rapportée. Selon nous résultats, les hypothèses H1, H3 et H4 est validée par nos résultats. En revanche, l'hypothèse H2 n'est pas validée.

Dans cette expérience, nous n'avons utilisé que deux angles (face et profil). L'utilisation des angles intermédiaires pourrait fournir des résultats intéressants sur la limite à laquelle l'angle de vue devient problématique pour la perception et conduit à une diminution de la confiance rapportée.

En ce qui concerne les mélanges d'émotion, nos résultats confirment les résultats antérieurs (Clavel et al., 2009) Dans les combinaisons incongruentes du visage et de la posture, les sujets rapportent la catégorie émotionnelle affichée sur le visage, alors qu'ils rapportent le niveau d'activation de l'émotion qui est véhiculée par la posture.

Plusieurs études soulignent l'impact de la dynamique par rapport à des stimuli statiques sur la perception des expressions des émotions (Bänziger et Scherer, 2007). Il serait donc intéressant d'étendre cette étude à des animations du personnage virtuel. Pour finir, il serait intéressant d'explorer les combinaisons avec d'autres modalités telles que le contexte exprimé par l'environnement 3D, la parole et la situation sociale.

7.2.2 Intégration de MARC dans un système de réalité virtuelle : SMART-I²

7.2.2.1 Objectifs

L'intégration de MARC dans un système de réalité virtuelle ouvre la porte à des applications immersives qui ne peuvent être réalisées sur un ordinateur de bureau. En outre, la stéréoscopie visuelle (vision binoculaire en relief) offre une perception différente des agents virtuels, qui sont alors affichés en taille réelle. Grâce à la stéréovision, ils peuvent également bénéficier d'une présence physique plus importante et ainsi mieux engager l'utilisateur dans la tâche (Traum et Rickel, 2002). Plusieurs applications (ex : entrainement militaire) ont déjà été conçues en utilisant des humains virtuels en réalité virtuelle immersive, mais peu d'applications étudient le lien avec les modèles affectifs et de l'expressivité des agents.

C'est dans ce contexte que MARC a été intégré au système SMART-I2[9]. Si aucune recherche n'a pour le moment été menée avec cette installation, une application de démonstration a été réalisée pour les journées de l'Association Française de Réalité Virtuelle (AFRV) organisées en 2010.

7.2.2.2 Le système SMART-I²

SMART-I² est un système de rendu audio-visuel spatialisé. D'un point de vue acoustique, il utilise la technique de synthèse de champ sonore pour simuler les ondes sonores émises par les objets virtuels. Cette technique utilise le principe de Huygens-Fresnel (1816), selon lequel: « *Tout champ de pression créé par une source primaire peut être reproduit en sommant les contributions de sources secondaires.* ». Plus précisément, le système est composé d'un ensemble de haut-parleurs indépendants. Pour simuler une source sonore localisée, le système SMART-I² calcule les contributions sonores individuelles de chaque haut-parleur afin de générer un champ sonore identique à celui qu'aurait généré l'objet virtuel depuis sa position. Ce phénomène est représenté sur la Figure 97. Cette technique présente l'avantage de ne pas avoir à suivre l'utilisateur dans le dispositif, car les ondes sonores générées sont indépendante de la position de celui-ci.

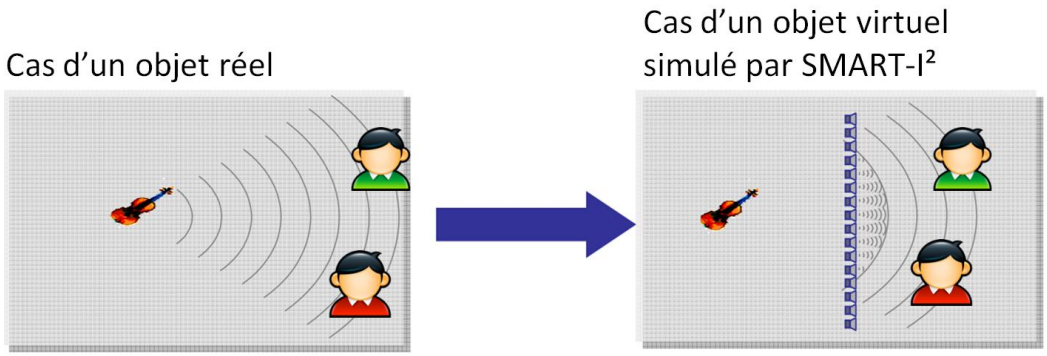


Figure 97 - Principe de la synthèse de champ sonore du système SMART-I²

Pour le rendu visuel stéréoscopique, le système SMART-I² utilise un double projecteur à lumière polarisée. L'utilisateur doit donc porter des lunettes polarisées. C'est ce qu'on appelle de la stéréoscopie passive. Contrairement au système audio, le système vidéo nécessite de localiser l'utilisateur dans la scène pour adapter la perspective visuelle des écrans. Pour cela, le système SMART-I² est équipé du système de suivi RigidBody de la société Optitrack.

Le système SMART-I² est composé de deux écrans perpendiculaires de trois mètres de large sur deux mètres soixante de haut. Chaque écran est équipé d'une ligne de hauts parleurs utilisés pour la synthèse de champ sonore, et sert de support pour deux projecteurs vidéo (soit quatre projecteurs au total).

[9] SMART-I2 : Marc Rébillat et Brian Katz (Rébillat et al. 2008)

7.2.2.3 Intégration de MARC dans le système SMART-I²

L'objectif à long-terme de l'intégration de MARC dans le système SMART-I² est d'envisager un rendu réaliste cohérent en visuel et en audio. Cela pourrait être utile par exemple pour immerger l'utilisateur dans des scènes sociales virtuelles dans lesquelles le rendu spatialisé de la parole fournirait à l'utilisateur des indices sur le ou les personnages en train de parler.

Cette intégration a posé deux problématiques techniques. D'une part, la gestion des quatre écrans de rendu, nécessaire à la stéréoscopie sur deux surfaces perpendiculaires. D'autre part, l'adaptation des paramètres de projection pour chaque œil de l'utilisateur, et en fonction de la position de l'utilisateur dans l'espace du système SMART-I² (position fournie par le système de tracking Optitrack). Ces deux aspects ont nécessité des implémentations spécifiques dans MARC que nous ne détaillerons pas ici. Cependant, cela nous a permis d'ouvrir à MARC la possibilité d'être utilisé dans un environnement immersif (Figure 98).

Nous avons ensuite conçu une application de démonstration de la plateforme MARC+SMART-I² pour la présenter à l'AFRV 2010. Pour cela, nous avons enregistré deux acteurs en leur demandant d'improviser un scénario à deux. Ces acteurs ont été équipés de marqueurs de motion capture (système Arena, Optitrack) afin de pouvoir reproduire leur gestuelle dans MARC. De plus, nous avons simultanément enregistré leurs dialogues à l'aide de microphone. La scène résultante a été intégrée dans un environnement 3D, et présentée à la conférence AFRV.

Dans cette démonstration de l'intégration technique, aucune expression faciale n'a été mise en place sur les personnages virtuels et les modèles émotionnels de MARC n'ont pas été utilisés. Pourtant, plusieurs utilisateurs nous ont rapporté avoir perçu des expressions faciales d'émotion dans la scène présentée. Nous avons donc émis l'hypothèse que l'expressivité des mouvements posturaux pourrait être à l' origine d'un biais dans la perception du visage neutre affiché par les personnages virtuels. De plus, il est possible que cet effet soit accentué par l'immersion dans la scène, due au système SMART-I².

Ces premières observations empiriques sont encourageantes. Elles semblent suggérer que les personnages virtuels expressifs peuvent bénéficier, en termes d'expressivité, de leur intégration dans des applications de réalité virtuelle. Cependant, plusieurs études seront nécessaires pour évaluer de manière formelle l'impact de ces technologies sur l'interaction avec des agents virtuel expressifs.

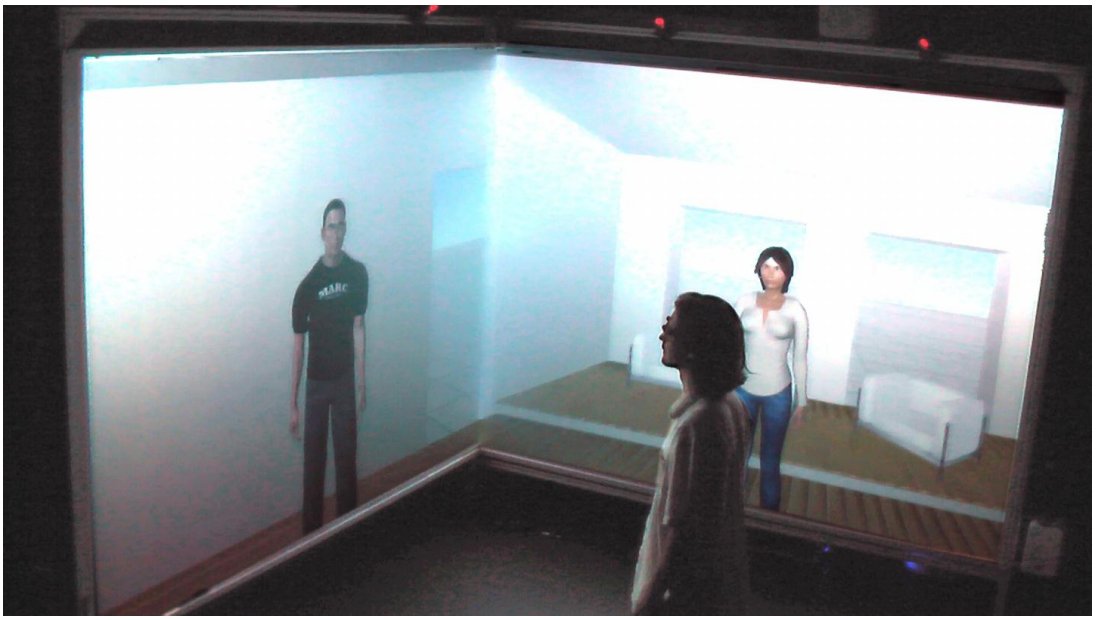


Figure 98 - Un utilisateur et les agents virtuels de MARC dans le système SMART-I² (Rébillat et al., 2008)

7.2.3 MARC et l'informatique ambiante : l'iRoom

Ces dernières années ont vu un intérêt croissant pour les systèmes ambiants, des pièces intelligentes et les personnages virtuels interactifs. Une pièce intelligente est un espace accessible et habitable dans lequel sont intégrés des services numériques et possédant les propriétés suivantes: Informatique ubiquitaire (invisible), communication ubiquitaire et interfaces utilisateur intelligente (Ducatel et al. 2001). Les technologies de suivi de l'utilisateur peuvent être utilisées dans des environnements ambiants pour fournir des informations sur la position de l'utilisateur et mouvements à l'intérieur d'une salle (Steggles and Gschwind, 2005). Les personnages virtuels interactifs combinent des modalités de communication naturelle et permettent donc d'envisager une communication face-à-face pertinente avec les utilisateurs d'une pièce intelligente (Tan et al. 2010).

C'est dans ce contexte que MARC a été intégré à la pièce intelligente iRoom du LIMSI (Bellik et al. 2009) en collaboration avec Yacine Bellik, Céline Clavel, Gaëtan Pruvost, et Ning Tan (Tan et al. en préparation).

Bellik et al (2009) développent depuis plusieurs années la pièce intelligente nommée iRoom au LIMSI-CNRS. L'objectif de l'iRoom est de mener des études sur les techniques d'interaction en environnement ambiant (Figure 99). Pour cela, les coins de la pièce sont équipés de capteurs Ubisense® qui permettent de suivre l'utilisateur et plusieurs objets simultanément dans la pièce.

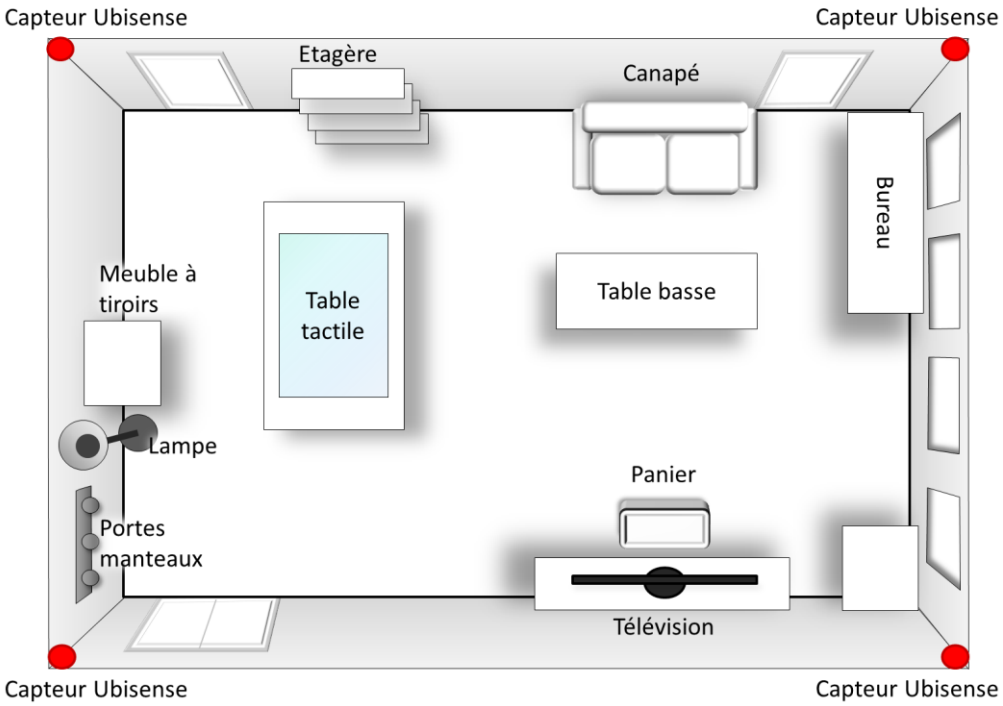


Figure 99 - Plan de l'iRoom (Bellik et al., 2009) Les points rouges représentent les capteurs Ubisense.

Nous avons intégré la plateforme MARC dans l'environnement de l'iRoom. MARC est affiché sur un écran de télévision 42 pouces, et contrôlé par le système informatique de l'iRoom, qui lui permet de réagir en fonction du comportement de l'utilisateur. Les mouvements de la tête, les gestes, l'orientation du corps et le regard du personnage virtuel peut être dirigé vers un emplacement spécifique dans la pièce (ex : l'emplacement de l'utilisateur ou l'emplacement d'un objet).

Pour évaluer l'impact de l'agent dans l'environnement ambiant, nous avons conçu une tâche de recherche d'objets cachés, assistée par l'agent virtuel (Tan et al. 2010). L'objectif de l'étude est d'évaluer l'impact de l'adaptabilité de l'agent sur les performances et la perception des utilisateurs.

Quatre actes communicatifs ont été sélectionnés comme étant pertinents pour la tâche de recherche: pointer, confirmer, infirmer et féliciter. Lors de l'étude en interaction, nous avons choisi une tâche dans laquelle

l'utilisateur doit trouver des objets dans la pièce intelligence. Nous avons choisi cette tâche car l'utilisateur doit se déplacer dans la pièce et fournit ainsi une information de localisation riche et continue au système.

Six objets cibles ont été équipés de balises Ubisense® de sorte que le système sache où se trouvent ces objets et puisse guider l'utilisateur vers chacun de ces objets (adaptateur, enveloppe, veste, DVD, boîte de médicaments et dossier). Certains objets étaient cachés (par exemple dans un tiroir), tandis que d'autres étaient visibles (par exemple sur la table à café). La position initiale des objets cibles a été choisie de manière à utiliser tout l'espace et à obliger l'utilisateur à se déplacer dans toute la pièce. Trois distracteurs (objets qui ressemblent à des objets cibles) ont également été équipés de balises Ubisense. L'agent virtuel a été affiché sur un écran de télévision 42 pouces.

Au début de la session, l'agent accueille l'utilisateur et décrit verbalement le premier objet à rechercher. Le système "sachant" où l'objet cible est, il sait quand l'utilisateur est proche de l'objet cible, et distingue quand l'utilisateur saisi l'objet cible ou un distracteur. Lorsque l'utilisateur a trouvé l'objet, l'agent demande à l'utilisateur de le déposer dans un panier devant la télé et décrit verbalement l'objet suivant à rechercher. Toute la session est chronométrée. En fin de session l'utilisateur répond à un questionnaire sur la présence perçue et l'adaptabilité de l'agent virtuel.

Les sujets ont été répartis en deux groupes. Le premier groupe était confronté à un agent « adaptatif », utilisant des gestes communicatifs et des postures expressives pour diriger le sujet (Figure 100). Le second groupe a été confronté à un agent « non adaptatif », ne s'expriment que verbalement.

Figure 100 - Utilisatrice interagissant dans l'iRoom (Bellik et al, 2009)

L'hypothèse était que les comportements non verbaux spatiaux tels que la proximité, le corps d'orientation et la posture pouvait fournir une aide intuitive aux utilisateurs dans une tâche de recherche spatiale. Les évaluations ont été réalisées en termes de performances objectives (durée de la tâche) et la perception subjective (de la présence et l'adaptabilité de l'agent). Les résultats, décrits dans la thèse de Ning Tan, confirment que les fonctions d'adaptation de l'agent virtuel augmentent la présence perçue et l'adaptabilité perçue de l'agent virtuel.

7.2.4 Assistance aux personnes agées

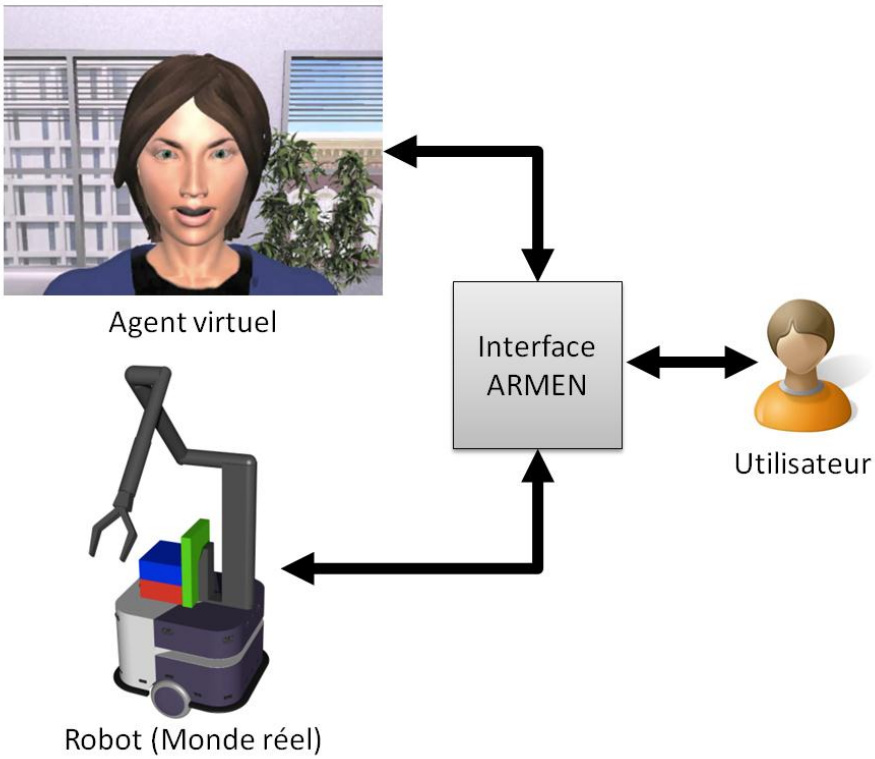


Figure 101 - Architecture de l'installation du projet ARMEN

L'objectif du projet ARMEN[10] est la conception d'un robot d'assistance simple d'utilisation mais procurant des fonctions évoluées pour aider au maintien à domicile des personnes en situation de perte d'autonomie. Le robot est secondé par un personnage virtuel ayant pour rôle de faire l'interface entre le robot assistant et l'utilisateur (Figure 101).

Le robot virtuel a pour objectif de se repérer seul dans l'environnement, et de pouvoir apporter et déplacer des objets. L'agent virtuel est utilisé comme interface visuelle et comme stimulation cognitive.

L'application ARMEN n'a pas pour objectif de remplacer le personnel d'assistance médicale, mais de compenser le manque de disponibilité et d'effectif. Ainsi, il s'agit de restructurer le temps de présence du personnel de soin et non de systématiquement diminuer leur temps de présence.

Pour MARC, ce projet permet d'investiguer l'interaction vocale et émotionnelle à long terme avec une population particulière. En effet, ces personnes auront probablement une perception de l'agent différente de celle des individus mobilisés dans nos études.

[10] ANR ARMEN: laboratoire CEA LIST, association APPROCHE, laboratoire LASMEA, industriel Robosoft, industriel Voxler, et laboratoire LIMSI.

7.2.5 MARC et interaction haptique

Les interactions affectives sont par essence multimodales. La combinaison des informations provenant de nos différents sens permet une communication plus naturelle, riche et précise. Si l'utilisation d'expressions faciales d'agent expressif constitue l'une des modalités les plus courantes de l'interaction homme-machine affective, d'autres modalités, telles que le toucher, sont moins exploitées.

Le toucher est en effet un sens qui possède la particularité d'être à la fois une action (toucher quelqu'un) et une perception (être touché). De plus, de récentes études ont d'ailleurs montré que ce sens, est très efficace pour l'interaction affective et relationnelle (Bickmore et al., 2010). L'utilisation combinée d'expressions haptiques et d'expressions faciales permettrait ainsi une amélioration de la sensation de présence et de l'intensité des expériences émotionnelles perçue par l'utilisateur. Pour finir les dispositifs haptiques pourraient représenter un dispositif d'entrée important pour permettre à l'utilisateur d'exprimer ses émotions, et ainsi d'augmenter la boucle affective d'une nouvelle source d'informations sur l'état émotionnel de l'utilisateur.

Dans ce contexte, nous avons collaboré au LIMSI avec Meddhi Ammi, Yoren Gaffary et David Bonnet pour mettre en place plusieurs études sur l'interaction haptique émotionnelle combinée à l'expressivité du visage de MARC. La première étape de cette collaboration a été de créer un corpus d'expressions d'émotion par le canal haptique. Ce corpus a été enregistré en demandant à 40 utilisateurs d'exprimer gestuellement un ensemble d'émotions en utilisant un *Phantom OMNI (SensAble)* (Figure 102), bras haptique à retour d'effort. L'ensemble d'émotion est le suivant : Joie, Exaltation, Dégout, Mépris, Inquiétude, Peur, Irritation, et Rage.

Figure 102 - Bras haptique Phantom OMNI (SensAble) pour l'interaction avec MARC

Les expressions collectées ont ensuite été validées perceptivement par un deuxième ensemble de 25 sujets, et les stimuli les mieux reconnus ont été sélectionnés pour l'étude suivante.

La deuxième partie de ces travaux consistera à comparer la perception des émotions en monomodal (Haptique seul ou Visage seul), avec la perception des émotions en multimodal (Haptique et Visage combinés). Cette étude est actuellement en cours de réalisation par Yoren Gaffary, sous la direction de Medhi Ammi et Jean-Claude Martin.

7.2.6 Le Projet Autisme

Le but du projet Autisme[11] est de concevoir des outils multimédias pour l'éducation sociocognitive des personnes autistes de haut niveau. Une première version de ces travaux avait été réalisée en utilisant des agents virtuels générés par le logiciel Poser® (non temps réel) et utilisant uniquement les expressions faciales (Buisine et al. 2010). La principale hypothèse de l'introduction des postures corporelles dans cet outil a été de soutenir les changements attentionnels afin d'améliorer le traitement des stimuli sociaux par les personnes autistes.

MARC a donc été utilisé pour créer une seconde version des outils combinant les animations posturales avec les animations faciales utilisées dans la première version des travaux. Les résultats de l'étude menée avec cette seconde version suggèrent que l'utilisation de différentes postures de « repos » (« idle mouvements » en anglais) peuvent influer les processus attentionnels et augmenter la perception des émotions (Buisine et al. Soumis). Les émotions sont mieux perçues lorsque le personnage affiche une animation posturale de repos. L'utilisation de postures au repos influence également les processus émotionnels, peut-être en raison d'un effet attentionnel: les sujets plus attentifs percevraient mieux les expressions émotionnelles. Afficher des postures au repos sur le personnage virtuel peut donc constituer un compromis pour augmenter l'expressivité des personnages.

L'utilisation de postures émotionnelles semble également impacter les processus attentionnels et la perception des émotions. De plus, les postures d'émotion congruentes avec l'expression faciale améliorent la reconnaissance des émotions. Ces résultats complètent les données fournies par la littérature (Meeren et al, 2005, Clavel et al, 2009, Courgeon et al, 2011). Dans notre expérience, les postures congruentes augmentent également l'intensité perçue des émotions, et améliorent le réalisme des animations. Le réalisme est une caractéristique qualitative qui est habituellement cherchée dans la conception de personnages virtuels émotionnels. Ce concept se réfère également au naturel, ou à la crédibilité, qui sont des buts difficiles à atteindre dans ce domaine de recherche. La crédibilité est souvent considérée comme liée au réalisme visuel de l'agent et sur sa capacité à générer des comportements verbaux et non verbaux pendant l'interaction avec l'utilisateur (Johnson et al., 2000). Dans notre étude, les postures émotionnelles permettent d'augmenter le réalisme et l'efficacité de la communication (perception et reconnaissance des émotions). Ce résultat est très positif par rapport à d'autres recherches suggérant que le réalisme n'est pas toujours corrélé à l'efficacité de la communication (Calder et al, 2000; Buisine et al, 2010). L'utilisation de postures émotionnelles semble donc être un moyen efficace de gagner à la fois en efficacité de communication émotionnelle et en réalisme.

Les perspectives de futures recherches de ces travaux porteront sur les expérimentations similaires avec des personnes autistes de haut niveau. À cet égard, le principal défi sera de confirmer si les résultats obtenus avec des personnes non autistes se confirment avec des personnes autistes. De tels résultats permettraient de compléter les considérations théoriques sur l'autisme et permettraient la conception de nouveaux outils susceptibles d'aider les personnes autistes de haut niveau.

[11] Projet Autisme Fondation de France. Coordinateur : Ouriel Grynszpan Nadel (Université Pierre et Marie Curie et Hôpital La pitié Salpêtrière), Jacqueline Nadel (Université Pierre et Marie Curie et Hôpital La pitié Salpêtrière), docteur Constant et Florence le Barillier (hôpital de Chartres), Noelle Carbonnel et Jérome Simonin (Laboratoire LORIA de Nancy), et Jean-Claude Martin (laboratoire LIMSI, Orsay). Le stage d'Aurélien Charles a été co-encadré avec Stéphanie Buisine (ENSAM).

7.3 Utilisations de MARC comme outil d'animation interactif de personnages expressifs

7.3.1 Traitement des phobies sociales

Les travaux de doctorat de Toni Vanhala (Université de Tampere, Finlande) portent sur le traitement et la remédiation des phobies sociales. Il travaille à l'aide d'agents virtuels pour stimuler l'anxiété de personnes sujettes aux phobies sociales. Dans l'une de ses études, Toni Vanhala a sollicité notre collaboration pour utiliser MARC et créer un contexte social propice à l'étude des phobies sociales. L'un des personnages de MARC (en alternance féminin ou masculin) apparait dans la scène, présentée sur un écran d'ordinateur. Au départ, l'agent virtuel est à une certaine distance de l'utilisateur. L'utilisateur doit alors faire approcher l'agent virtuel grâce à des capteurs d'activité musculaire situés sur son visage. En contractant certains muscles de son visage (les sourcils ou les zygomatiques, en fonctions des groupe expérimentaux), l'utilisateur contrôle ainsi la distance entre le personnage virtuel et lui.

Lorsque l'utilisateur estime que l'agent est suffisamment proche, il décontracte son visage et arrête ainsi la progression de l'agent. L'agent énonce alors une tâche logique (type opération mathématique) que l'utilisateur doit résoudre. Durant cette tâche, le niveau de conductivité de la peau de l'utilisateur est enregistré. Il sert de mesure objective. De plus, l'utilisateur doit remplir en fin de test un questionnaire subjectif sur les émotions qu'il a ressenti lors de la passation.

Cette étude a permis de montrer que l'activation volontaire des muscles visage pourrait offrir une méthode de régulation de l'activation physiologique et subjective lors de l'exposition à des stimuli sociaux artificiels. L'activation des sourcils semble améliorer la relaxation des sujets et a été relativement bien maîtrisée par les sujets durant l'expérimentation. De plus, dans la condition où les participants contrôlent l'agent avec l'activation de leurs zygomatiques, les évaluations subjectives de l'expérience émotionnelle suggèrent que les participants les plus socialement anxieux étaient les moins à l'aise de devoir sourire au personnage virtuel.

La détection de tels schémas de réponse et le suivi de leur évolution pourrait être utile pour évaluer le degré d'anxiété sociale de patients, et par exemple, de suivre les progrès de la thérapie du patient. De plus, les deux types d'activations des muscles faciaux (Zygomatiques et Sourcils) produisent des changements physiologiques compatibles et potentiellement bénéfiques pour les objectifs du traitement de la pathologie. En conclusion, les résultats de cette étude fournissent plusieurs pistes prometteuses pour la recherche et constituent une base solide pour poursuivre les travaux dans l'étude des activations faciales volontaires en tant que méthode de régulation des émotions assistée par ordinateur.

Bien que dans cette étude les capacités expressives de MARC n'ont pas été utilisées, nous envisageons d'explorer cette voie dans une future collaboration avec l'équipe de Finlande. De plus, ces travaux mettent en relief la possible utilisation de MARC dans la conception de logiciel de remédiation de troubles psychologiques autres que l'autisme.

7.3.2 Le projet ANR CARE sur la danse « augmentée »

L'utilisation d'agents virtuels dans des performances artistiques est encore peu explorée. L'une des principales raisons est le manque de réactivité et d'interactivité des agents existants. En effet, la plupart des agents interactifs sont extrêmement scriptés, et leur comportement est donc très prévisible. Dans un contexte d'interaction de courte durée, et souvent en laboratoire, cet aspect scripté permet un meilleur contrôle de l'agent. En revanche, dans une application artistique, c'est un frein à la créativité et à la dynamique artistique.

Notre plateforme MARC présente l'intérêt d'animer des agents pouvant être soit autonomes (en utilisant un modèle émotionnel informatique), soit entièrement pilotés en temps réel et de manière continue par une application externe, comme une « marionnette ». Le système permet donc un contrôle dynamique de l'expressivité de l'agent, ce qui fait de MARC une plateforme adaptée aux applications artistiques temps réel impliquant un personnage virtuel.

Dans ce contexte, l'objectif du projet ANR CARE[12] a visé à augmenter un spectacle de danse avec des modalités virtuelles dans un but pédagogique (par exemple faciliter la compréhension des émotions exprimées un danseur par des spectateurs novices) et créatif (le personnage virtuel étant utilisé par le danseur dans un but créatif).

MARC a ainsi été intégré dans un spectacle artistique augmenté traitant des émotions dans la danse. Dans cette performance, un danseur exprime par les mouvements de son corps un contenu émotionnel. Les gestes effectués par le danseur sont capturés en temps réel grâce à une combinaison Moven, et transmis au logiciel eMotion de l'ESTIA (Clay et al. 2009) qui classifie les mouvements reconnaître quelques catégories d'émotion. Une fois les émotions détectées, elles sont transmises à deux modules expressifs : 1) Shadoz, logiciel d'animation jouant avec l'ombre d'un personnage en fil de fer. Shadoz permet d'exprimer les émotions via la couleur (de l'ombre du personnage) et via les mouvements du personnage. 2) MARC, qui permet d'exprimer les émotions via les expressions faciales. Ces deux logiciels trouvent donc une complémentarité dans les modalités utilisées.

Notre objectif est de déterminer quel type d'augmentation ou combinaison d'augmentations aident le public à mieux reconnaître les émotions exprimées par le danseur, et favorisent ainsi le lien de communication entre l'artiste et le public.

Dans un premier temps, nous avons mené une étude perceptive en laboratoire en présentant aux sujets plusieurs séquences de danse sous différentes modalités : la vidéo du danseur filmé (sans augmentation), Shadoz (sans ombre), Shadoz (ombre colorée), le visage de MARC seul, et enfin, Shadoz (ombre colorée) + MARC. Cinq séquences de danse ont été sélectionnées, exprimant chacune une émotion différente. Pour chaque émotion, nous avons généré les stimuli vidéo des cinq modalités choisies. Nous avons donc obtenu 25 stimuli. Chacun des sujets a visionné 5 vidéos tirées aléatoirement parmi les 25 possibles.

Les résultats montrent que les émotions complexes sont plus difficiles à reconnaître que les émotions de base. Néanmoins, lorsque le public bénéficie de modalités augmentées, on constate une augmentation du taux de reconnaissance des émotions complexes. Ces premiers résultats montrent donc l'intérêt d'augmenter virtuellement la performance artistique pour la communication émotionnelle entre l'artiste et son public.

La première représentation publique du spectacle a été donnée au ballet de Biarritz et au festival Les Ethiopiques en 2010.

[12] ANR CARE : Coordination : IMMERSION, collaboration avec l'ESTIA, l'IRIT, Metapages, UJF UTT.

7.3.3 Les spectacles artistiques « Oh Peer, My Teddy » et « Beautiful Beast »

7.3.3.1 Objectifs

Le projet « Oh Peer, My Teddy !», est un projet artistique proposé par Pascale Barret, et qui traite de l'interaction entre l'homme et la virtualité. Pascale Barret s'intéresse à l'utilisation des avatars dans les performances artistiques (par exemple, les humains virtuels de *Second Life*). Grâce à ces avatars, elle questionne le rapport réel/virtuel à travers une vision artistique. MARC contribue donc à faire avancer ses travaux artistiques en raison de son focus sur le rendu des émotions.

Inversement, l'intérêt pour MARC est d'explorer des champs de communication émotionnelle non standard (avec d'autres interfaces de contrôle, avec des paramétrages excessifs...) qui ne peuvent être explorés dans les environnements de laboratoire. En effet, cette application explore des dispositifs d'interaction originaux, et permet donc d'investiguer une autre approche de l'interaction et du contrôle d'agent virtuel. Ces travaux entrent en résonance avec l'installation *Pogany* (Jacquemin, 2007) dans laquelle un dispositif tangible en forme de visage est utilisé pour contrôler l'expression de l'agent virtuel. Les travaux effectués en collaboration avec Pascale permettent donc à MARC d'aborder sous un autre angle les problématiques d'interaction entre l'humain et le personnage virtuel, dans un contexte artistique. Cette utilisation de MARC nous a également permis de vérifier sa robustesse technique lors de spectacle.

7.3.3.2 Le projet « Ho Peer, My Teddy »

Dans un premier temps, Pascale s'est associée à Rudi Giot. Rudi allie des connaissances techniques en électronique et en informatique, ainsi qu'une passion pour la musique, qui le poussent à créer de nouvelles interfaces, de nouveaux sons, de nouveaux instruments. Ensemble, Rudi et Pascale mettent en place une peluche augmentée. La peluche est dotée d'un gyromètre de positionnement en 3 dimensions, d'un microphone, de senseurs de flexion aux bras, de senseurs lumière dans les yeux, ainsi que de 5 zones de contacts électriques sur les pattes et le nez. En corps à corps avec le performeur, la peluche permet d'atteindre un contrôle sensible dans l'espace de l'installation, sur l'environnement sonore et visuel, et enfin, de communiquer avec un personnage virtuel doté d'expressions émotionnelles.

C'est dans ce cadre que MARC intervient dans l'installation de Pascale. Via la peluche, Pascale interagit avec MARC, et joue en temps réel un scénario les diverses modalités d'entrée incluses dans la peluche (Figure 103).

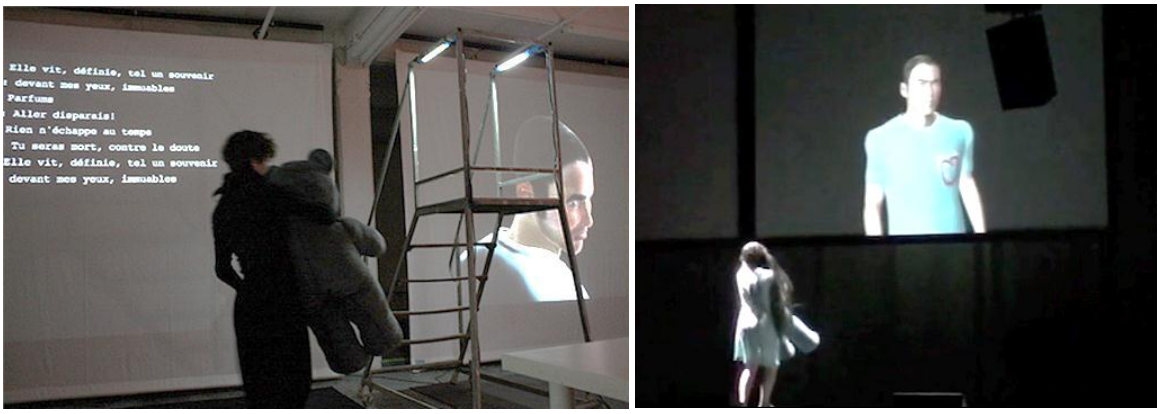


Figure 103 - Pascale Barret manipulant « Teddy », l'ours en peluche augmenté, pour contrôler MARC.

Dans la première étape de cette collaboration, Pascale s'est intéressée aux expressions faciales proposées par MARC. Ces travaux ont donné lieu à une représentation publique de la performance de Pascale. Cette collaboration nous a apporté une vision neuve sur nos travaux. Ici les émotions ne sont plus considérées d'un point de vue de la psychologie ou de l'informatique affective, mais pour leurs aspects artistiques, communicatifs et esthétiques. Le dispositif d'interaction lui-même n'est pas conçu pour être ergonomique, mais dans une

optique artistique, en se concentrant sur l'interaction elle-même, un corps à corps entre l'humain et la machine. Cette collaboration nous a également permis de mettre à l'épreuve la généricité et la robustesse de notre logiciel. En effet, l'utilisation faite par Pascale de MARC est très différente de ce pourquoi il a été conçu. Cela a d'ailleurs donné lieu à diverses implémentations spécifiques et a ainsi contribué à la robustesse et la généricité de MARC. Pour finir, cette collaboration nous a permis de diffuser nos travaux dans une communauté différente, et ainsi d'ouvrir la voie à de nouvelles collaborations art-science.

Une deuxième étape de ce travail est actuellement en cours de réalisation, intitulée *Beautiful Beasts*, et dont une première présentation a été donnée à Porto le 16 septembre 2011 dans le cadre du Kyma International Sound Symposium (KISS 2011). Pascale cherche à y créer l'ambigüité entre les trois protagonistes de la performance artistique : l'ours en peluche, le performeur (elle-même), et l'agent virtuel MARC. De plus, cette nouvelle version explore l'expressivité gestuelle de MARC. Les mouvements de Pascale ont été capturés avec le dispositif de capture de mouvements du LIMSI. Les fichiers capturés sont ensuite rejoués par le personnage virtuel MARC lors du spectacle en les combinant avec des expressions faciales.

PARTIE IV

Conclusions et futures directions de recherche

Chapitre 8. Conclusions et perspectives

Sommaire du chapitre

8.1 Rappel des objectifs de recherche

8.1.1 Introduction

L'objectif principal de cette thèse était de modéliser, implémenter, et évaluer différentes approches des émotions et de l'animation faciale temps-réel afin de contribuer à l'amélioration de l'interaction temps-réel entre les agents expressifs et l'utilisateur. Pour cela, nous avons exploré deux axes. D'une part, nous avons proposé différents modèles émotionnels informatiques, inspirés de différentes approches des émotions issues de la psychologie. D'autre part, nous avons cherché à évaluer l'apport de plusieurs techniques de rendu graphique réalistes et temps réel sur la perception de l'agent et de son expressivité par des utilisateurs.

Pour chaque modèle émotionnel informatique que nous avons proposé, nous avons également exploré ses liens avec l'animation de visages virtuels. Afin d'évaluer ces travaux, nous avons mis en place plusieurs expérimentations perceptives et différentes applications interactives.

8.1.2 Besoins et contraintes

L'interaction avec un agent virtuel nécessite l'utilisation d'un modèle computationnel des émotions. Dans la littérature en psychologie des émotions, diverses approches se confrontent. Proposer un modèle computationnel des émotions nécessite donc une sélection parmi les différentes approches de la psychologie, en identifiant celles qui permettront d'obtenir la qualité d'animation faciale et le niveau d'interactivité requis.

Comme nous l'avons vu, nos travaux considèrent également un certain nombre de contraintes techniques. Ces contraintes sont principalement liées à l'aspect interactif temps réel des applications que nous avons ciblées. D'une part, les modèles computationnels proposés doivent être exécutables en temps réel, afin de fournir un retour expressif immédiat et pertinent avec la situation d'interaction. D'autre part, les techniques d'animations utilisées doivent être suffisamment efficaces pour permettre une animation fluide et dynamiquement contrôlable.

8.2 Approche méthodologique

Durant cette thèse, nous avons opté pour un processus itératif nous permettant d'explorer différentes approches des émotions. En parallèle, nous avons développé un système d'animation basé sur des techniques récentes d'animation GPU et exploré différentes techniques de rendu de rides d'expression. Ces deux directions de recherche que nous avons sélectionnées sont complémentaires. Les modèles des émotions permettent de représenter des états internes émotionnels complexes, que notre système d'animation faciale nous permet d'exprimer en prenant en compte leur subtilité.

Nos travaux ont été validés par des études perceptives. En effet, comme Wallraven *et al.* (2005) l'ont montré, les agents virtuels peuvent être utilisés pour étudier la perception humaine, et réciproquement, la perception humaine peut être utilisée pour évaluer, étudier et améliorer les modèles et les rendus d'agents virtuels. Ainsi, nous avons conçu différents protocoles d'évaluation pour chacun des modèles que nous avons explorés. Les résultats de ces études ont ainsi contribué à l'élaboration des modèles successifs que nous avons proposés.

8.3 Résumé de la thèse

Les travaux itératifs présentés durant cette thèse ont mené à la conception de la plateforme MARC (Multimodal Affective and Reactive Characters). MARC est issu de notre démarche itérative de conception des modèles et des outils dédiés à l'interaction avec un visage virtuel expressif.

Dans la version 8.7.0 disponible à la fin de cette thèse (Figure 104), MARC comporte 5 personnages virtuels distincts, capables d'interagir avec l'utilisateur via quatre modèles computationnels des émotions, issus des quatre approches des émotions sélectionnées. Ces différents modèles computationnels ont donné lieu à quatre

modules logiciels indépendants. Ces personnages peuvent être animés indépendamment. Ils peuvent s'exprimer via leurs expressions faciales ou via leurs postures, dont ils possèdent chacun une bibliothèque d'expressions. Ces personnages peuvent également être intégrés dans des environnements virtuels.

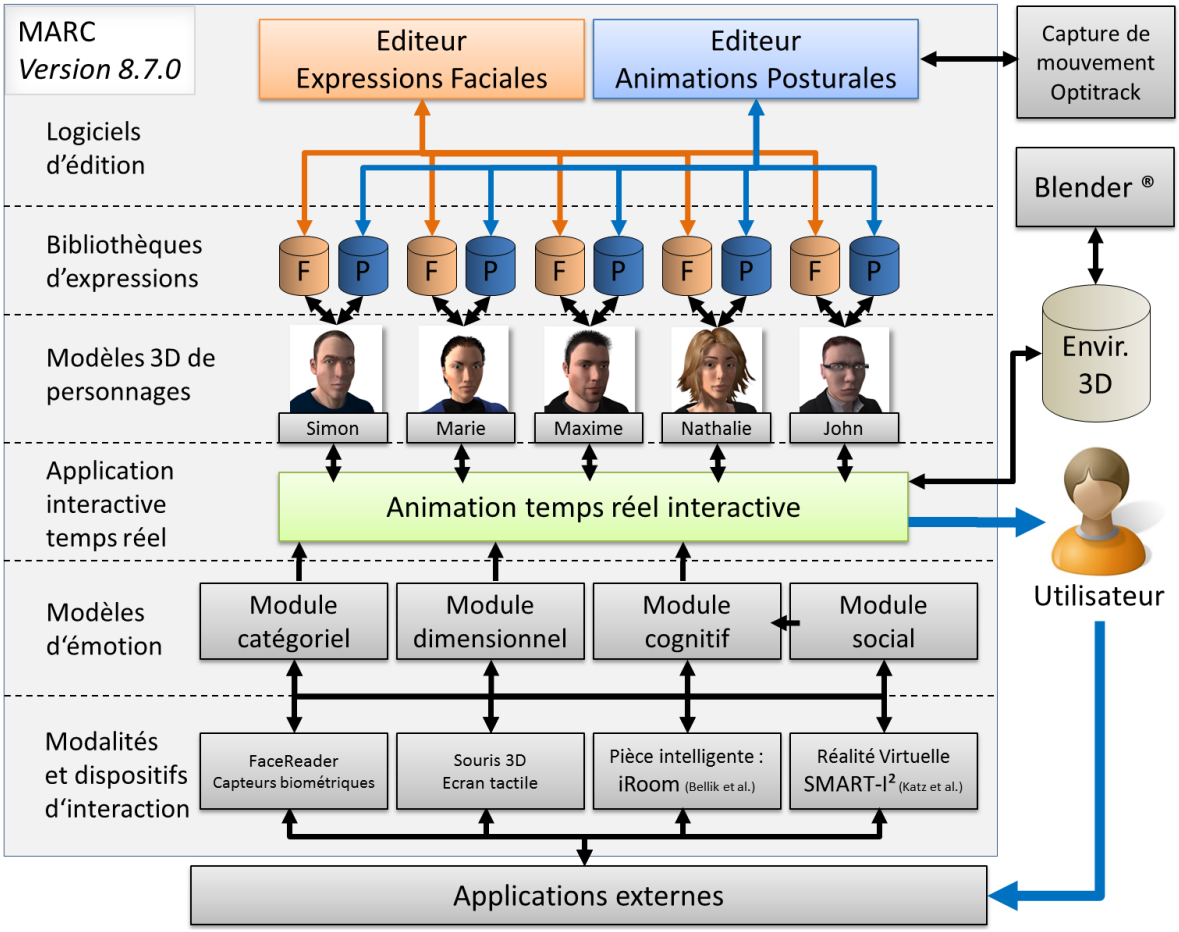


Figure 104 - Architecture de MARC à la fin de la thèse

Dans sa version actuelle, MARC se compose d'environs 300 classes Java, pour environs 100.000 lignes de code (Classes JAVA et Shaders GLSL inclus), et sans prendre en compte le code de la bibliothèque 3D, indépendante de MARC, mais développée en majeure partie pour répondre aux besoins de notre plateforme d'agents virtuels.

En plus du logiciel d'animation temps réel, MARC est composé de deux logiciels d'édition (expressions faciales et postures) qui permettent la mise en place de bibliothèques expressives utilisables par le module interactif temps réel. Chaque modèle de personnage possède ses propres bibliothèques d'expressions.

MARC a été connecté avec succès à un système de capture de mouvement (Optitrack Arena), un système de réalité virtuelle (SMART-I², Rébillat et Katz, 2010), un environnement ambiant (iRoom, Bellik et al. 2009), et divers capteurs physiologiques (Performances de Pascale Barret, Vanhala et al, 2010). D'un point de vue logiciel, MARC est en mesure de communiquer avec des logiciels de séquençage haut niveau (PureData, Max/MSP, etc.), ou plus généralement avec toute application capable d'utiliser des communications UDP/TCP. De plus, MARC est multiplateforme (Windows et Linux).

Comme nous l'avons vu, nous avons exploré quatre approches des émotions. Pour chaque approche, nous avons proposé un modèle informatique qui a donné lieu au développement d'un module logiciel spécifique. Chacun des modèles computationnels proposés nous a permis de mener un certain nombre d'expérimentations perceptives.

8.3.1 Approche catégorielle

La première approche que nous avons explorée est l'approche catégorielle des émotions. Après avoir conçu un système d'animation faciale permettant de contrôler l'expressivité de l'agent par des labels émotionnels, nous avons mené plusieurs études perceptives.

Nous avons commencé par évaluer la reconnaissance catégorielle des expressions des émotions de base mise en place avec notre système d'animation. Nous avons montré que ces expressions étaient correctement reconnues par les participants, avec des taux similaires à ceux de la littérature en perception des expressions humaines, issus de la liste des études établie par Russell (1994).

Nous avons ensuite évalué l'apport des rides d'expressions et de leur réalisme sur la reconnaissance d'expressions faciales d'émotions de base, et d'émotions complexes. Dans une première étude, nous avons pu montrer que l'utilisation de rides d'expression n'augmente pas la reconnaissance catégorielle des émotions. En revanche, nous avons montré que les rides d'expressions augmentent l'intensité expressive perçue, et augmente la préférence des utilisateurs. Ces résultats sont en accord avec l'étude de De Melo et al. (2009) effectuée sur la combinaison de plusieurs signes faciaux d'émotion. Cependant, nous avons ensuite montré que l'utilisation de rides géométriques 3D permet d'augmenter la reconnaissance des catégories d'émotion sur des visages vus de côté. Ainsi, nos résultats montrent divers intérêts à l'utilisation de rides d'expressions 3D et réalistes sur les visages d'agents virtuels expressifs.

Nous avons ensuite mené une étude sur la perception de la dynamique des expressions faciales. Dans cette étude, nous avons montré que les sujets anticipent la dynamique expressive du visage selon une courbe composée d'un pic expressif suivi d'un déclin d'intensité. Avec cette étude, nous avons observé le phénomène d'anticipation suggéré par Thornton et al. (1998) et que nous avons appelé le *moment émotionnel*.

Pour finir, nous avons mis en place une application de clonage émotionnel basée sur le logiciel FaceReader en entrée, et sur notre plateforme MARC en sortie. Cette première application interactive nous a permis de démontrer les capacités temps réel et interactives de MARC, ainsi que sa capacité à se connecter avec d'autres applications dédiées à l'informatique affective.

Ces différentes études nous ont permis de nous confronter à un certain nombre de limitations de l'approche catégorielle. Par exemple, le modèle catégoriel impose de spécifier chaque état émotionnel, ainsi que la ou les expressions faciales associées. De plus, il n'établit pas de relation entre les émotions. Les émotions sont considérées comme des états indépendants et possédant leurs propres mécanismes neuronaux. Cette approche limite donc les raisonnements de plus haut niveau sur les émotions, puisque chaque émotion doit avoir ses propres conditions d'apparition, indépendantes des autres émotions.

8.3.2 Approche dimensionnelle

Notre démarche itérative nous a ensuite conduits à aborder l'approche dimensionnelle des émotions. Dans le modèle informatique à trois dimensions (inspiré du modèle P.A.D., Russell et Mehrabian, 1977) que nous avons proposé, nous obtenons un espace continu qui nous permet de créer des relations entre les émotions. Ces relations imposent un certain nombre de contraintes implicites. Par exemple, l'état affectif de l'agent doit conserver une trajectoire continue dans l'espace, et ne peut passer d'une position à l'autre de manière discontinue.

En utilisant notre modèle informatique, nous avons mis en place un système de profils expressifs individuels permettant de moduler l'expressivité du personnage en fonction de paramètres liées aux dimensions P.A.D. Ainsi, la manipulation par l'utilisateur de l'expression faciale du personnage était modulée par des paramètres expressifs liés au modèle émotionnel dimensionnel. L'étude que nous avons menée nous a permis de montrer d'une part que l'utilisation d'un espace dimensionnel des émotions permet la manipulation de l'expression d'un

agent virtuel, et d'autre part que les profils expressifs modulant l'expression sont bien perçus par les sujets manipulant l'agent virtuel.

Cependant, l'approche dimensionnelle ne nous a pas semblé suffisamment complexe pour permettre à l'agent de manifester un comportement autonome, car elle ne modélise pas le traitement de l'information, et ne permet donc pas de déterminer la réaction émotionnelle dans un contexte dynamique.

8.3.3 Approche cognitive

Pour aborder l'approche cognitive des émotions, nous avons choisi de nous inspirer du modèle CPM et de proposer un modèle dynamique de l'émotion. Ce modèle nous a semblé pertinent pour créer une interaction avec un agent virtuel autonome car il simule une évaluation cognitive de la situation. Ainsi, le système est capable de gérer de manière continue l'état émotionnel et l'expressivité de l'agent virtuel durant une interaction avec l'utilisateur. Pour spécifier notre module, nous avons effectué un certain nombre de choix d'implémentations. De plus, dans notre étude, nous avons limité le contexte d'interaction à un jeu de plateau de société, afin de mieux contrôler les différentes situations possibles durant l'interaction.

Ce système nous a permis de comparer l'expressivité générée par le modèle cognitif proposé avec l'expressivité issue de l'approche catégorielle des émotions. Nous avons pu montrer que l'utilisation d'un modèle cognitif modifie la perception que les utilisateurs ont de l'agent. En effet, nos résultats montrent que les sujets attribuent plus d'états mentaux à l'agent lorsqu'il exprime son évaluation de la situation à travers ses expressions faciales. De plus, l'agent est perçu comme étant plus expressif avec le modèle cognitif qu'avec le modèle catégoriel. Pour finir, nos résultats montrent que le modèle cognitif modifie le comportement de l'utilisateur. En effet, en mode cognitif, les utilisateurs ont passé plus de temps à jouer, mais ont gagné plus souvent que dans le mode catégoriel.

8.3.4 Approche cognitive et sociale

Considérant les limites de notre modèle cognitif, nous avons cherché à prendre en compte les aspects sociaux du processus cognitif émotionnel. Pour cela, nous avons étudié un phénomène social particulier nommé le *social appraisal*. Comme nous l'avons vu, le social appraisal, ou évaluation cognitive sociale, est défini par Manstead et Fisher (2001) comme la prise en compte de l'évaluation cognitive d'autrui dans notre propre évaluation d'un événement. Nous avons donc proposé un modèle de réévaluation des stimuli cognitifs, basé sur la prise en compte de l'évaluation d'un second personnage virtuel (agent observé). Ce modèle s'appuie sur notre modèle cognitif pour réaliser l'animation faciale.

Aucune donnée n'étant disponible sur les mécanismes responsables de l'influence sociale sur le processus d'évaluation cognitive, nous avons proposé deux approches. La première consiste à copier le résultat de l'un des critères d'évaluation de l'agent observé. La seconde méthode consiste à utiliser un ensemble de règles logiques pour établir la seconde réaction de l'agent social, en fonction des premières évaluations des deux agents. Ces deux approches ont été comparées lors d'une expérimentation perceptive.

Dans cette étude, nous avons montré que l'utilisation d'un modèle de réévaluation sociale permet d'augmenter la perception de l'expressivité de l'agent dans le contexte social. Cependant, nous n'avons pas pu déterminer quelle méthode de simulation du processus social était la plus efficace. Pour cela, d'autres modèles devront être proposés et évalués. Néanmoins, les sujets semblent percevoir la communication émotionnelle entre les deux agents. L'agent doté d'une réaction sociale est en effet perçu comme s'adaptant mieux à l'autre agent que l'inverse.

Malgré ses limites, notre modèle expérimental de social appraisal semble donc permettre la prise en compte automatique de réactions sociales dans l'évaluation cognitive. Ces résultats sont encourageants, et d'autres travaux devront être effectués dans ce sens, et en reliant notre travaux à des phénomènes sociaux mieux étudiés, tels que l'empathie.

8.4 Perspectives de recherche

8.4.1 Perspectives spécifiques aux approches émotionnelles étudiées

8.4.1.1 Approche Catégorielle

Les perspectives ouvertes par nos travaux sur l'approche catégorielle concernent principalement l'animation faciale. Pour commencer, plusieurs questions restent ouvertes concernant la gestion de la dynamique des émotions et de leurs expressions faciales. En effet, si plusieurs travaux adressent ce problème (Pelachaud et al, 2006), le phénomène du *moment émotionnel* que nous avons mesuré pourrait permettre de proposer de nouvelles approches sur la gestion de la dynamique des émotions et de leurs expressions faciales. En particulier, notre étude semble montrer les limites perceptives du modèle dynamique *onset-apex-offset*.

Nos travaux n'abordent que partiellement les mélanges d'émotions. Pourtant, ces phénomènes sont régulièrement observés dans l'interaction humaine (Scherer, 1998, Abrillian et al. 2005), et des phénomènes tels que le masquage d'expressions, la politesse et plus généralement le contexte social, ont une grande influence sur nos émotions et la façon de les exprimer. L'approche catégorielle des émotions, par sa simplicité théorique, permet d'explorer ces phénomènes émotionnels et leurs expressions faciales.

8.4.1.2 Approche Dimensionnelle

Notre modèle dimensionnel inspiré de P.A.D. pourrait être étendu de plusieurs façons. D'une part, nous avons effectué un certain nombre de simplifications, comme le placement des émotions aux coins du cube. En utilisant un plus grand nombre d'émotions, localisées différemment dans l'espace dimensionnel, nous pourrions obtenir une expressivité plus fine. De plus, nos travaux sur les profils expressifs individuels montrent que l'approche dimensionnelle peut être mise en relation avec des paramètres individuels. Nos profils expressifs continus pourraient donc être étendus à des phénomènes affectifs à plus long terme, tels que les humeurs, où la personnalité.

L'interaction que nous avons sélectionnée (souris 3D) permet un *mapping* direct pour contrôler l'émotion et l'expression de l'agent. Pourtant, d'autres dispositifs moins directs pourraient être explorés. Par exemple, les interfaces proposées par Pascale Barret (Peluche augmentée) et Christian Jacquemin (*Pogany*) pourraient être adaptés pour contrôler non pas directement les paramètres expressifs, mais des paramètres émotionnels de plus haut niveau. Nos travaux soulèvent donc également des questions de recherche sur les modalités et les dispositifs d'interaction avec l'agent virtuel.

8.4.1.3 Approche Cognitive

Notre premier modèle cognitif inspiré du modèle CPM, associé à l'animation faciale des évaluations séquentielles, nous a donc permis de montrer l'apport d'un modèle cognitif sur l'animation faciale en situation d'interaction. Cependant, pour permettre une interaction plus riche et à plus long terme, le modèle doit être étendu pour prendre en compte l'historique de l'interaction. De plus, pour permettre la simulation de raisonnements cognitifs plus complexes, le modèle nécessiterait une représentation plus détaillée de la situation et de l'état interne de l'agent virtuel. Par exemple, l'intégration d'un modèle de type *Belief-Desire-Intention* permettrait de modéliser l'état interne de l'agent et d'apporter des informations importantes pour l'évaluation cognitive des événements. Certaines parties des arbres de décision utilisés dans notre modèle peuvent en effet se rapporter à des croyances et des désirs. L'agent « désire » gagner, ainsi il est en mesure d'évaluer qu'un événement est obstructif par rapport à ses buts. D'autre part, le mécanisme d'anticipation lui permet de « croire » que l'utilisateur va placer son pion à une certaine position. L'intégration d'un modèle BDI permettrait donc de remplacer les arbres de décision de notre modèle par une approche plus générique, et ainsi de rendre notre modèle cognitif plus rapidement applicable à d'autres applications.

Notre modèle cognitif pourrait également bénéficier de la prise en compte de l'état émotionnel de l'utilisateur. En effet, il aurait été préférable de pouvoir mesurer cet état émotionnel pour le prendre en compte dans l'évaluation cognitive de l'agent. Ainsi, l'ajout de systèmes de capture d'expressions faciales et de capteurs physiologiques (Knapp et al., 2011) permettant de modéliser l'état affectif de l'utilisateur pourrait améliorer notre système cognitif et ainsi de créer une boucle affective complète.

8.4.1.4 Approche Sociale

En ce qui concerne notre modèle d'évaluation cognitive sociale, nos travaux sur le *social appraisal* ne sont que préliminaires. Le modèle que nous avons proposé est empirique et l'étude effectuée nécessite d'être étendue à d'autres émotions, et d'être menée sur un plus grand nombre de sujets. Cependant, à l'instar des travaux de Mumenthaler et Sander, (2009), nos premiers résultats sont encourageants et suggèrent que notre approche est pertinente.

Pour étendre ce modèle, l'une de nos perspectives est d'utiliser l'utilisateur comme source de l'influence sociale. Ainsi, l'agent aurait un comportement social vis-à-vis de l'utilisateur. En effet, en plus de la prise en compte de l'état émotionnel de l'utilisateur dans le traitement cognitif (comme suggéré plus haut), rendre l'agent capable d'interpréter les réactions faciales de l'utilisateur pour effectuer une seconde évaluation cognitive et sociale permettrait une plus grande adaptabilité de l'agent, et ainsi une boucle d'interaction affective plus complexe.

En ce qui concerne l'utilisation de plusieurs agents, nous envisageons de permettre aux deux agents d'effectuer plusieurs évaluations sociales. Ainsi, les deux agents seraient capables de produire un comportement social, et de s'auto-influencer. Ainsi, nous créerions une boucle d'interaction affective entre les deux agents virtuels.

Cette boucle d'interaction sociale ouvre la voie à une autre problématique, celles des relations sociales et hiérarchiques. En effet, nous posons l'hypothèse que la relation de statut entre les deux personnages influence l'effet de l'évaluation cognitive sociale. Si l'un des agents est le supérieur hiérarchique de l'autre, son influence sera probablement différente que si les deux agents ont le même statut. La modélisation de ces relations permettrait donc d'obtenir une évaluation sociale plus complexe, prenant en compte des informations sociales plus larges.

Une autre extension possible et ambitieuse de nos travaux serait d'intégrer notre modèle de réévaluation sociale dans les systèmes de simulations sociales multi-agents à plus large échelle. Ainsi, il serait possible d'augmenter le nombre d'agents pris en compte dans l'évaluation sociale. Nous pourrions ainsi simuler des évaluations cognitives sociales dans un groupe d'agents virtuels. A plus grande échelle, ces travaux pourraient également contribuer à la simulation de comportements de foule, notamment sur des phénomènes de panique. Ces phénomènes font cependant intervenir des analyses complexes de la communication émotionnelle et de la dynamique des scènes sociales. Nous aborderions alors des problématiques très différentes de l'interaction entre un visage virtuel expressif et un utilisateur.

L'utilisation de dispositifs de réalité virtuelle tels que SMART-I² permettent effectivement d'interagir simultanément avec plusieurs agents virtuels. Dans ce contexte, l'utilisation de modèles sociaux semble pertinente, car elle permet de générer une interaction de groupe au lieu de générer un ensemble d'interactions deux à deux.

8.4.2 Perspectives générales

8.4.2.1 Rendu graphique et animation temps réel

Le système de rendu du visage que nous avons mis en place dans MARC utilise des techniques d'illumination récentes basées sur le GPU. L'une des manières possibles d'améliorer le rendu que nous obtenons est d'utiliser des données géométriques et lumineuses de meilleures qualités. Ces données peuvent être obtenues par un système de capture haute qualité tel que LightStage (Debevec et al., 2000), et permettrait d'augmenter le réalisme du rendu du visage. Dans le cadre du projet FUI ADN-TR[13], qui démarre en octobre 2011, notre équipe aura accès à de telles données. Il sera donc nécessaire d'évaluer perceptivement l'impact des rendus que nous obtiendrons sur la perception de l'agent, en les comparants aux rendus actuels de MARC. Nous pouvons en effet nous attendre à une augmentation du réalisme perçu, néanmoins, il n'est pas certain que cela améliore l'expressivité de l'agent, lié en grande partie à la dynamique de l'animation.

En ce qui concerne l'animation, MARC utilise une approche paramétrique, pour laquelle nous avons proposé plusieurs techniques de simulation des rides d'expressions. Nous avons tout d'abord utilisé la technique du *bump-mapping* pour créer des rides d'expressions sur le visage de notre agent virtuel. Comme nous l'avons vu, cette approche permet de modéliser des rides d'expression fines, indépendantes de la topologie du maillage. Cependant, elle ne permet pas de créer des rides de plus grande ampleur. Donc avons donc proposé une méthode à base de points-clés dédiés aux rides pour déformer le maillage et ainsi créer des rides géométriques en déformant la structure 3D du visage. Cette technique est cependant liée à la finesse du maillage, et ne permet donc pas la création de rides très fines. Nous pensons donc qu'il est nécessaire de proposer un système hybride, combinant ces deux approches complémentaires. Les rides géométriques telles que nous les avons proposées permettraient de déformer le visage et de créer du relief réel, alors que l'utilisation complémentaire du *bump mapping* serait réservée aux rides très fines.

L'utilisation d'un système d'animation paramétrique tel que celui utilisé dans MARC pose également des problématiques en termes de réalisme. En particulier, l'interpolation linéaire d'expressions n'est pas perçue comme réaliste (Cosker et al., 2010). Pour augmenter le réalisme de la dynamique des expressions faciales, plusieurs méthodes peuvent être considérées. L'utilisation de paramètres temporels différents et adaptés à chaque zone du visage, l'utilisation d'un système biomécanique, ou encore l'utilisation d'un système hybride. L'implémentation de ces différents modèles et la mise en place de protocoles d'évaluation communs représenteraient un bénéfice significatif pour comprendre les apports de ces différentes méthodes dans le cadre des agents virtuels expressifs.

8.4.2.2 La boucle d'interaction d'affective

L'un des aspects importants de l'interaction affective que nous n'avons pas abordé est l'utilisation du langage naturel pour communiquer avec l'agent virtuel. Par exemple, les guides de musée *Ada et Grâce* (Swartout et al., 2010) utilisent la modalité vocale pour communiquer avec les visiteurs. Le système facilite ainsi l'engagement de l'utilisateur dans l'interaction. Ce type d'interaction ouvre à de nombreuses problématiques scientifiques, telles que la reconnaissance et la synthèse de la parole, le traitement de l'information et la génération dynamique de réponses adaptées. La plupart des systèmes actuels simplifient ces problématiques en repérant des mots clés dans les phrases prononcées par l'utilisateur et en sélectionnant des phrases prédéfinies et les comportements expressifs associés. Si cette approche permet de faire illusion lors d'une interaction à court terme, elle mène rapidement à une répétition des mêmes séquences par l'agent, qui peut mener à un désengagement de l'utilisateur (Swartout et al., 2010). De plus, les phrases prononcées par l'utilisateur peuvent contenir un contenu émotionnel important, soit à un niveau sémantique, soit dans leur prosodie. Ainsi, pour créer une boucle d'interaction affective riche, l'analyse de la voix ne doit pas se limiter à une analyse sémantique, mais doit concerner également une analyse émotionnelle (Devillers et al., 2005).

[13] Projet AAP FUI 11 : Agence Doublure Numérique – Temps Réel - Coordinateur : Cédric Guiard
http://www.adnda.com/accueil.php

En effet, la prise en compte de l'état émotionnel de l'utilisateur est fondamentale pour créer une boucle d'interaction affective complète. Parmi les différentes applications que nous avons mises en place, nous avons investigué plusieurs facettes de l'interaction : interaction ludique, interaction artistique, informatique ambiante, réalité virtuelle, etc. Cependant, aucune de ces applications ne nous a permis d'obtenir et de considérer l'état émotionnel de l'utilisateur. Une boucle affective complète, dans lequel l'agent perçoit et considère correctement les émotions de l'utilisateur, permettrait de concevoir des applications affectives plus riches. Par exemple, dans le cadre du projet ARMEN, l'une des entrées du système porte sur la détection des émotions dans la voix de l'utilisateur.

Cependant, la boucle émotionnelle n'est pas uniquement limitée par les capacités du système à détecter les émotions de l'utilisateur, mais également par les capacités expressives de l'agent virtuel. Nos travaux se limitent en effet à l'expressivité faciale. Cependant, d'autres modalités expressives peuvent être exploitées par les agents virtuels : les gestes, les jeux de regards (Kulms et al., 2011), l'expressivité vocale (Schroder et al., 2010), les mouvements de tête, etc. En ajoutant des modèles d'interaction dédiés à ces modalités, la plateforme MARC permettrait de les explorer indépendamment et en les combinant. De plus, l'utilisation d'un système de réalité virtuelle tel que SMART-I² apporte une dimension immersive favorisant l'étude des comportements spatiaux, tels que la posture et le regard, ainsi que des comportements interpersonnels avec l'utilisateur. En effet, les personnages virtuels étant présenté en 3D stéréoscopique et en taille réelle, il devient possible d'étudier des modalités spatiales telles que la proxémie (distance interpersonnelle) entre un ou plusieurs agents virtuels et l'utilisateur. Ces aspects seront étudiés en utilisant MARC en collaboration avec Brice Isableu (Université Paris-Sud 11) et dans le cadre de la thèse de Tom Giraud qui commencera en octobre 2011.

8.4.2.3 Modèles émotionnels et autres phénomènes affectifs

La plateforme logicielle MARC que nous avons développé au cours de cette thèse permet d'explorer différentes approches des émotions. Si nos travaux se sont focalisés sur quatre d'entre elles, d'autres approches pourraient être pertinentes pour l'animation faciale interactive temps réel. Par exemple, dans la classification de Scherer (2010), on trouve les approches motivationnelles et adaptatives, que nous n'avons pas abordées.

De plus, nous n'avons pas étudié la relation entre les différentes approches des émotions afin de déterminer si elles sont ou non exclusives. En effet, le continuum proposé par Gross et Feldman-Barrett (2011) suggère que les différentes approches partagent certaines caractéristiques et qu'elles ne sont donc pas opposées. De même, la classification de Scherer montre certaines correspondances entre les différentes approches. Du point de vue des modèles informatiques, nos travaux sur l'approche sociale mélangent l'approche cognitive et l'approche sociale. De plus, certains travaux (ex : Becker-Asano et Wachsmuth, 2008) présentent également des systèmes tirant partie de plusieurs approches simultanément. Ainsi, si nos travaux présentent un certain nombre d'informations importantes pour choisir entre différentes approches des émotions, de nombreuses études sont encore nécessaires pour établir clairement les apports et les limites de chacune, ainsi que la manière de les combiner, pour se diriger vers un modèle intégratif, capable de combiner simultanément plusieurs approches des émotions.

Dans nos travaux, nous nous sommes limités à une liste restreinte d'états affectifs émotionnels. De plus, nous n'avons pas considéré les phénomènes affectifs de plus longue durée, telles que l'humeur ou la personnalité. Pourtant, ces phénomènes sont importants pour la modélisation des comportements affectifs d'agent virtuels (André et al., 2000) et des agents relationnels (Bickmore et al., 2011). Ils impactent à la fois l'évaluation cognitive à l'origine de l'émotion et la manière d'exprimer l'émotion. Nos modèles cognitifs semblent pertinents pour permettre de prendre en compte ces phénomènes. En effet, il serait ainsi possible de moduler l'expressivité de l'agent et d'influencer la simulation du processus émotionnel en fonction de paramètres à plus long terme tels que la personnalité (Clavel et Martin, 2009). Cependant, cela soulève de nombreuses problématiques, en particulier : comment modéliser et représenter la personnalité de l'agent ? Comment la personnalité de l'agent impacte son évaluation cognitive ? Néanmoins la prise en compte d'informations à plus long terme semble nécessaire pour concevoir des systèmes avec lesquels l'utilisateur doit interagir sur une période de temps plus étendue.

En effet, la prise en compte d'un contexte temporellement plus large est nécessaire à la modalisation d'un processus émotionnel complet (Scherer, 2001). Plusieurs aspects de la mémoire à long terme sont envisageables. Par exemple, la mémoire autobiographique permet de comparer les événements survenant avec des événements antérieurs (Ho et Watson, 2006), ou encore la mémoire émotionnelle, résultat de nos expériences émotionnelles antérieures (Kensinger et Corkin, 2004). Ces aspects à long terme ont une influence sur notre processus cognitif, nos décisions, nos réactions et notre expressivité. Le modèle FATIMA (Dias et Païva, 2005) utilise par exemple une combinaison du modèle cognitif OCC et d'une mémoire autobiographique pour la génération de comportements affectifs.

Le traitement des informations à long terme permettrait donc d'apporter une cohérence dans une interaction prolongée avec l'utilisateur. Ainsi, il serait possible d'éviter que l'agent ne devienne répétitif, et que l'utilisateur se désengage de l'interaction.

8.5 Conclusion générale

Toutes ces perspectives montrent que les agents virtuels expressifs sont à l'intersection de nombreux domaines de recherche en informatique et en sciences humaines et sociales. Nos travaux se sont focalisés uniquement sur l'informatique affective (et plus particulièrement la simulation des émotions), et sur l'animation faciale interactive. Pourtant, les agents virtuels sont également liés à d'autres domaines, telles que la synthèse et la reconnaissance de la parole, l'intelligence artificielle, le traitement automatique des langues, la représentation des connaissances, la psychologie, les sciences sociales, etc. Ainsi, les agents virtuels sont un carrefour interdisciplinaire à l'intersection duquel de nombreuses collaborations scientifiques sont possibles. Les travaux que nous avons présentés nous ont permis de répondre à certaines questions de recherches, mais nous sommes encore loin de savoir simuler le comportement affectif humain dans toute sa complexité. Cet objectif ne peut être atteint que par une collaboration interdisciplinaire forte.

Cependant, les agents virtuels expressifs ne se limitent pas à la recherche. Leurs applications industrielles sont de plus en plus variées. Ainsi, les agents expressifs peuvent également servir de pont entre l'industrie et la recherche, en permettant de nombreuses collaborations. Par exemple, l'industrie du jeu vidéo a rapidement compris l'intérêt d'inclure des comportements affectifs dans ses personnages virtuels pour favoriser l'immersion du joueur. Certains jeux sont même aujourd'hui axés entièrement sur cet aspect (ex : L.A. Noire, Rockstar). Les agents web représentent un autre exemple intéressant d'application industrielle. Les agents conversationnels non expressifs se répandent aujourd'hui sur le web et sur les applications mobiles car ils sont conviviaux et simples d'accès. Il est probable que les agents conversationnels de demain seront dotés de modèles émotionnels. En effet, les technologies utilisées sont de plus en plus performantes et permettent un rendu visuel de plus en plus détaillé et réaliste. Il est donc nécessaire que le comportement des agents virtuels évolue également. L'utilisation de modèles émotionnels sophistiqués semble donc une direction pertinente pour contribuer à cette évolution.

Pour finir, les agents virtuels expressifs peuvent servir d'outils pour étudier la perception humaine et la communication émotionnelle. En retour, ces études permettent de contribuer à l'amélioration des systèmes d'agents virtuels expressifs en fournissant des informations importantes et des règles de conception basées sur des validations perceptives. En appliquant ce principe d'enrichissement réciproque, nous espérons que nos travaux, ainsi que la plateforme MARC qui en résulte, contribueront à mieux comprendre et à améliorer l'interaction avec les agents virtuels expressifs réalistes et temps-réel.

Bibliographie

Abrilian, S., Devillers, L., Buisine, S., & Martin, J.-C. (2005). EmoTV: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces. *HCI Internationnal*.

Adam, C. (2007). Emotions: from psychological theories to logical formalization and implementation in a BDI agent. *Thèse de l'Université Paul Sabatier*.

Albrecht, I., Schröder, M., Haber, J., & Seidel, H. (2005). Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. *Virtual Reality, 8(4)*, 201-212.

Alexander, O., Rogers, M., Lambeth, W., Chiang, J.-Y., Ma, W.-C., Wang, C.-C., et al. (2010). The Digital Emily Project: Achieving a Photorealistic Digital Actor. *IEEE Journal on Computer Graphics and Applications*, 20-31.

André, E., & Rist, T. (2000). Presenting through performing: on the use of multiple lifelike characters in knowledge-based presentation systems. *Proceedings of the 5th international conference on Intelligent user interfaces*, 1-8.

André, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. *Affective interactions*, 150-165.

Andrés del Valle, A. C., Dugelay, J.-L., & Pelé, D. (2001). *Overview of face animation in MPEG-4 and study of the compliance level of Eurecom's face animation-teleconferencing system.* A video conference system under MPEG-4.

Andrés del Valle, A. C., Dugelay, J.-L., Garcia, E., & Valente, S. (2000). Acquisition et animation de clones réalistes pour les communications. *Compression et Représentation des Signaux Audiovisuel*, 19-20.

Arnold, M. B. (1960). *Emotion and personality*. New York: Columbia University Press.

Averill, J. (1985). The Social Construction of Emotion: With Special Reference to Love. Dans *The Social Construction of a Person* (pp. 89-109). New York, Springer.

Balci, K. (2005). Xface: Open source toolkit for creating 3d faces of an embodied conversational agent. *Smart Graphics*, 924-932.

Bänziger, T., & Scherer, K. (2007). Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus. *{Affective computing and intelligent interaction*, 476-487.

Bänziger, T., Grandjean, D., & Scherer, K. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion, 9*, 691-704.

Barkhuysen, P., Krahmer, E., & Swerts, M. (2010). Cross-modal and incremental perception of audiovisual cues to emotional speech. *Language and speech, 53(1)*, 3-30.

Baron-Cohen, S. (2007). *Mind Reading: The Interactive Guide to Emotions.* . Jessica Kingsley Publishers.

Barrett, J. J. (2011). Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View. *Emotion Review*, 8-16.

Bartneck, C. (2002). Integrating the OCC Model of Emotions in Embodied Characters. *Workshop on Virtual Conversational Characters: Applications, Methods and Reseach Challenges.*

Beaudouin-Lafon, M. (2004). Designing interaction, not interfaces. *Advanced visual interfaces* (pp. 15-22). Gallipoli, Italy: ACM.

Becker-Asano, C., & Ishiguro, H. (2011). Evaluating facial displays of emotion for the android robot Geminoid F. *Affective Computational Intelligence*, (pp. 1-8). Paris, France.

Becker-Asano, C., & Wachsmuth, I. (2008). Affect Simulation with Primary and Secondary Emotions. *Intelligent Virtual Agent*, (pp. 15-28).

Bee, N., Falk, B., & André, E. (2009). Simplified Facial Animation Control Utilizing Novel Input Devices: A comparative Study. *International Conference on Intelligent User Interfaces*, 197-206.

Bellik, Y., Rebaï, I., Machrouh, E., Barzaj, Y., Jacquet, C., Pruvost, G., et al. (2009). Multimodal interaction within ambient environments: an exploratory study. *Human-Computer Interaction*, 89-92.

Bevacqua, E., Prepin, K., de Sevin, E., Niewiadomski, R., & Pelachaud, C. (2008). Reactive behaviors in SAIBA architecture. *Autonomous Agents and Multi Agents Systems*, 9-12.

Bickmore, T., Fernando, R., Ring, L., & Schulman, D. (2010). Empathic Touch by Relational Agents. *IEEE Transactions on Affective Computing*, 60-71.

Bickmore, T., Pfeifer, L., & Schulman, D. (2011). Relational Agents Improve Engagement and Learning in Science Museum Visitors. *Intelligent Virtual Agents*, (pp. 55-67). Reyjavik, Iceland.

Borod, J., Yecker, S., Brickman, A., Moreno, C., Sliwinski, M., Foldi, N., et al. (2004). Changes in Posed Facial Expression of Emotion Across the Adult Life Span. *Experimental Aging Research: a Journal Devoted to the Scientific Study of the Aging Process, 30(4)*, 305-331.

Boukricha, H., Becker, C., & Wachsmuth, I. (2007). Simulating Empathy for the Virtual Human Max. *International Workshop on Emotion and Computing*, (pp. 22-27).

Bratman, M. (1987). Intentions, PLans, and Practial Reason. *Harvard University Press*.

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: does it exist ? *HCI human-animated character interaction*, 16-20.

Bridson, R., Marino, S., & Fedkiw, R. (2003). Simulation of Clothing with Folds and Wrinkles. *ACM SIGGRAPH Eurographics Symposium on Computer Animation*, 28-36.

Broekens, J., & DeGroot., D. (2004). Scalable and Flexible Appraisal Models for Virtual Agents. *Computer Games: Artificial Intelligence, Design and Education.*

Broekens, J., DeGroot, D., & Kostersa, W. (2008). Formal models of appraisal: Theory, specification, and computational model. *Cognitive Systems Research, 9(3)*, 173-197.

Buisine, S., Wang, Y., & Grynszpan, O. (2010). Empirical investigation of the temporal relations between speech and facial expressions of emotion. *Journal on Multimodal User Interfaces, 3*, 263-270.

Busso. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. *Internationnal Conference of Multimodal Interfaces*, (pp. 205-211).

Calder, A., Rowland, D., Young, A., & Nimmo-Smith, I. (2000). Caricaturing Facial Expressions. *Cognition & Emotion, 76(1)*, 105-146.

Cassell, J. (2000). Embodied conversational interface agents. *Commun. ACM 43(4)*, 70-78.

Cassell, J. (2001a). Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine 22(4)*, 67-84.

Cassell, J., Vilhjàlmsson, H., & Bickmore, T. (2001b). BEAT: the behavior expression animation toolkit. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, (pp. 477--486).

Cavazza, M., Pizzi, D., Charles, F., Vogt, T., & André, E. (2009). Emotional input for character-based interactive storytelling. *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, 1*, 313-320.

Charles, F., Pizzi, D., Cavazza, M., Vogt, T., & André, E. (2009). EmoEmma: Emotional Speech Input for Interactive Storytelling. *The Eighth International Conference on Autonomous Agents and Multiagent Systems*, 1381-1382.

Choe, B., Lee, H., & Ko, H. (2001). Performance-Driven Muscle-Based Facial Animation. *The Journal of Visualization and Computer Animation, 12(2)*, 67-79.

Chuand, E., & Bregler, C. (2002). *Performance Driven Facial Animation using Blendshape Interpolation.* Standford University.

Clavel, C., & Martin, J. C. (2009). PERMUTATION: A Corpus-Based Approach for Modeling Personality and Multimodal Expression of Affects in Virtual Characters. *Digital Human Modeling*, 211-220.

Clavel, C., Plessier, J., Martin, J., Ach, L., & Morel, B. (2009). Combining facial and postural expressions of emotions in a virtual character. *Intelligent Virtual Agents*, 287-300.

Cosker, D., Krumhuber, E., & Hilton, A. (2010). Perception of linear and nonlinear motion properties using a FACS validated 3D facial model. *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, (pp. 101--108).

Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schröder, M. (2000). 'FEELTRACE': An instrument for recording perceived emotion in real time. *ISCA Workshop on Speech and Emotion*, (pp. 19-24).

Damasio, A. (1994). *Descartes' error: emotion, reason, and the human brain.* Putnam's Sons, New York.

Dariouch, B., Ech Chafai, N., Mancini, M., & Pelachaud, C. (2004). Tools to Create Individual ECAs. *Workshop HUMAINE*. Santorini.

de Melo, C., & Gratch, J. (2009). Expression of Emotions Using Wrinkles, Blushing, Sweating and Tears. *Intelligent Virtual Agent*, (pp. 188-200).

de Sevin, E., Niewiadomski, R., Bevacqua, E., Pez, A., Mancini, M., & Pelachaud, C. (2010). Greta: Une Plateforme d'Agent Conversationnel Expressif et Interactif. *TSI. Technique et science informatiques, 29(7)*, 751-776.

Debevec, P., Hawkins, T., Tchou, C., Duiker, H., Sarokin, W., & Sagar, M. (2000). Acquiring the Reflectance Field of a Human Face. *SIGGRAPH* , (pp. 145-156).

Dehn, D., & van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies, 52(1)*, 1-22.

Deng, Z., & Ma, X. (2008). Perceptually Guided Expressive Facial Animation. *SIGGRAPH Symposium on Cumputer Animation*, 67-76.

D'Eon, E., Luebke, D., & Enderton, E. (2007). Efficient Rendering of Human Skin. *Eurographics Symposium on Rendering.*

Devillers, L., Cowie, R., Martin, J.-C., Douglas-Cowie, E., Abrilian, S., & McRorie, M. (2006). Real Life Emotions in French and English TV video clips: an Integrated Annotation Protocol Combining Continuous and Discrete Approaches. *international conference on Language Resources and Evaluation*.

Devillers, L., Vidrascu, L., & Lamel, L. (2005). Emotion detection in real-life spoken dialogs recorded in call center. *Journal of Neural Networks, Emotion and Brain, 18(4)*, 407-422.

Dias, J., & Païva, A. (2005). Feeling and reasoning: a computational model for emotional agents. *EPIA*, 127-140.

Donner, C., & Jensen, H. (2005). Light Diffusion in Multi-Layered Translucent Materials. *SIGGRAPH 2005*, (pp. 1032-1039).

Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., & Burgelman, J.-C. (2001). Scenarios for Ambient Intelligence. *European Commission*.

Dutreve, L., Meyer, A., & Bouakaz, S. (2009). Real-Time Dynamic Wrinkles of Face for Animated Skinned Mesh. *International Symposium on Visual Computing*, 25-34.

Dutreve, L., Meyer, A., & Bouakaz, S. (2011). Easy Acquisition and Real-Time Animation of Facial Wrinkles. *Computer Animation and Virtual World*.

Dutreve, L., Meyery, A., & Bouakaz, S. (2008). Feature Points Based Facial Animation Retargeting. *Virtual Reality Software and Technology*, (pp. 197--200).

Ekman, P. (1994). Strong evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique. *Psychological bulletin, 115(2)*, 268-287.

Ekman, P. (1999). Basic Emotions. Dans T. Dalgleish, & M. Power, *Handbook of Cognition and Emotion*. Sussex, U.K.: John Wiley & Sons, Ltd.

Ekman, P. (2003). *Emotions revealed.* New York: Times Books.

Ekman, P., & Friesen, W. (1975). *Unmasking the Face. A guide to recognizing facial clues.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Ekman, P., & Friesen, W. (1986). A new pancultural expression of emotion. *Motivation and Emotion, 10*, 159-168.

Ekman, P., Davidson, R., & and Friesen, W. (1990). The Duchenne Smile: Emotional Expression and Brain Physiology II. *Journal of Personality and Social Psychology, 50(2)*, 342-353.

Ekman, P., Friesen, W., & Hager, J. (2002). *Facial Action Coding System - The Manual.* A Human Face.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. Dans R. J. Davidson, H. Goldsmith, & K. R. Scherer, *Handbook of Affective Sciences.* New York and Oxford: Oxford University Press.

Fresnel, A. (1816). Oeuvres. *Annales des Chimie et des Physique, 1(2)*, 89-129.

Freyd, J., & Finke, R. (1985). A Velocity Effect for Representational Mementum. *Bulletin of the Psychonomic Society, 23(6)*, 443-446.

Fridlund, A. J. (1994). Human facial expression: An evolutionary view. *San Diego, CA: Academic Press.*

Frijda, N. H. (1986). *The emotions.* London: Cambridge University Press.

Frijda, N. H. (1988). The Laws of Emotion. *American Psychologist, 43*, 349-358.

Frijda, N. H., & Swagerman, J. (1987). Can Computer Feel? Theory and Design of an Emotional System. *Cognition and Emotion, 1(3)*, 235-257.

Gamond, L., George, N., Lemaréchal, J.-D., Hugueville, L., Adam, C., & Tallon-Baudry, C. (2011). Early influence of prior experience on face perception. *Neuroimage, 54(2)*, 1415-1426.

Gerrig, R. J., Maloney, L. T., & Tversky, A. (1991). Validating the dimensional structure of psychological spaces: Applications to personality and emotions. *Frontiers of Mathematical Psychology*, 138-165.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *IEEE International Workshop on Robot and Human Interactive Communication*, 55-60.

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The cambridge mindreading (CAM) face-voice battery : Testing complex emotion recognition in adults with and without asperger syndrome : Affective/emotional experiences of people on the autism spectrum. *Journal of autism and developmental disorders, 36*, 169-183.

Graf, M., Reitzner, B., Corves, C., Casile, A., M., G., & W., P. (2007). Predicting point-light actions in real-time. *NeuroImage, 2*, 22-32.

Grandjean, D., & Scherer, K. R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion, 8(3)*, 341-351.

Gratch, J., & Marsella, S. (2004). Evaluating a general model of emotional appraisal and coping,. *AAAI Spring Symposium on Architectures for Modeling Emotion: Cross-disciplinary Foundations*.

Grosjean, F. (1996). Gating. *Language and Cognitive Processes, 11(6)*, 597-604.

Gross, J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271-299.

Gross, J., & Feldman Barrett, L. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion review, 3(1)*, 8-16.

Hadap, S., Bangerter, E., Volino, P., & Magnenat-Thalmann, N. (1999). Animating Wrinkles on Clothes. *Conference on Visualization*, 175-182.

Hess, U., Adams, R., & Kleck, R. E. (2004). Facial Appearance, gender, and Emotion Expression. *Emotion, 4(4)*, 378-388.

Heudin, J.-C. (2004). Evolutionary Virtual Agent. *International Conference on Intelligent Agent Technology*, (pp. 93-98).

Ho, W., & Watson, S. (2006). Autobiographic knowledge for believable virtual characters. *Intelligent Virtual Agents*, 383-394.

Izard, C. E. (1977). *Human emotions.* New York: Plenum Press.

Jacquemin, C. (2004). Architecture and experiments in networked 3d audio/graphic rendering with virtual choreographer. *Proceedings of Sounds and Music Computing.*

Jacquemin, C. (2007). Pogany: A tangible cephalomorphic interface for expressive facial animation. *second International Conference on Affective Computing and Intelligent Interaction*, (pp. 558-569).

James, W. (1884). What is an emotion ? *Mind (9)*, 188-205.

Jaques, P. A., Lehmann, M., & Pesty, S. (2009). Evaluating the affective tactics of an emotional pedagogical agent. *ACM symposium on Applied Computing*, 104-109.

Jarraya, M., Amorim, M.-A., & Bardy, B. (2005). Optical flow and viewpoint change modulate the perception and memorization of complex motion. *Perception & Psychophysics*, 951-961.

Jörding, T., & Wachsmuth, I. (1997). An anthropomorphic agent for the use of spatial language. *Workshop on Representation and Processing of Spatial Expressions*, 41-53.

Kahler, K., Haber, J., & Seidel, H. (2001). Geometry-based muscle modeling for facial animation. . *Graphics Interface*, 37-46.

Kensinger, E., & Corkin, S. (2004). Two Routes to Emotional Memory: Distinct Neural Processes for Valence and Arousal. *National Academy of Sciences of the United States of America*, (pp. 3310-3320).

Kline, J. (2005). Psychological Testing: A Practical Approach to Design and Evaluation. *Sage Publications, Inc*.

Kline, P. (1987). A Handbook of Test Construction: Introduction to Psychometric Design. *Methuen & Co. Ltd.*

Knapp, R., Kim, J., & André, E. (2011). Physiological Signals and Their Use in Augmenting Emotion Recognition for Human--Machine Interaction. *Emotion-Oriented Systems*, 133-159.

Kopp, S., & Jung, B. (2000). An Anthropomorphic Assistant for Virtual Assembly: MAX. *Workshop on Communicative Agents in Intelligent Virtuel Environments*.

Kopp, S., Stocksmeier, T., & Gibbon, D. (2007). Incremental multimodal feedback for conversational agents. *Intelligent Virtual Agents*, (pp. 139-146).

Kriegel, M., Aylett, R., Cuba, P., Vala, M., & Paiva, A. (2011). Robots Meet IVAs: A Mind-Body Interface for Migrating Artificial Intelligent Agents. *Intelligent Virtual Agents*, (pp. 282-295). Reykjavik, Iceland.

Kulms, P., Krämer, N. C., Gratch, J., & Kang, S. H. (2011). It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze. *Intelligent Virtual Agents*, 80–92.

Lanctôt, N., & Hess, U. (2007). The timing of appraisals. . *Emotion, 7*, 207-212.

Larboulette, C., & Cani, M.-P. (2004). Real-Time Dynamic Wrinkles. *Computer Graphics International*, 522-525.

Lazarus, R. (1968). Emotions and Adaptation: Conceptual and Empirical relations. *Nebraska Symposium on Motivation, 16(1)*, 175-270.

Lazarus, R., & Folkman, S. (1984). *Stress Appraisal and Coping*.

Lee, Y., Terzopoulos, D., & Waters, K. (1993). Constructing physics-based facial models of individuals. *Graphics Interface*, 1-8.

Leite, I., Castellano, G., Pereira, A., & P., M. (2009). Designing a Game Companion for Long-Term Social Interaction. *International Conference on Multimodal Interaction AFFINE Workshop*.

Leite, I., Martinho, C., Pereira, A., & Paiva, A. (2008). iCat: an affective game buddy based on anticipatory mechanisms. *Autonomous agents and multiagent systems*, (pp. 1229-1232).

Leite, I., Mascarenhas, S., Pereira, A., Martinho, C., Prada, R., & Paiva, A. (2010). " Why Can't We Be Friends?" An Empathic Game Companion for Long-Term Interaction. *Intelligent Virtual Agents*, (pp. 315-321).

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhoga, R. (1997). The persona effect: affective impact of animated pedagogical agents. *SIGCHI conference on Human factors in computing systems*, 359-366.

Loviscach, J. (2006). Wrinkling Coarse Meshes on the GPU. *Computer Graphics, 25(3)*, 467-476.

Ma, W., Hawkins, T., Peers, P., Chabert, C., Weiss, M., & Debevec, P. (2007). Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Sperical Gradient Illumiation. *Eurographics Symposium on rendering*, (pp. 1-11).

MacDorman, K. F., Coram, J., Ho, C., & Patel, H. (2010). Gender Differences in the Impact of Presentational Factors in Human Character Animation on Decisions in Ethical Dilemmas. *Presence: Teleoperator and Virtual Environment*.

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort: Uncanny responses to computer generated faces. *Computers in Human Behavior, 25(3)*, 695–710.

Magnenat-Thalmann, N., Primeau, E., & Thalmann, D. (. (1988). Abstract Muscle Action Procedures for Human Face Animation. *The Visual Computer, 3(5)*, 571-586.

Malatesta, C., Fiore, M., & Messina, J. (1987). Affect, Personality, and Facial Expressive Characteristics of Older People. *Psychology and Aging, 2(1)*, 64-69.

Malatesta, L., Raouzaiou, A., Karpouzis, K., & Kollias, S. (2007). Towards modeling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. *Applied Intelligence*, 58-64.

Manstead, A. S., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. Dans K. R. Scherer, A. Schorr, & T. Johnstone, *Appraisal processes in emotion: Theory, methods, research.* (pp. 221-232).

Marsella, S. (2010). Modeling Emotion and Its Expression in Virtual Humans. *Intelligent Tutoring Systems, 1(2)*, 1--2.

Marsella, S., & Gratch, J. (2006). EMA: A computational model of appraisal dynamics. *Agent Construction and Emotions.*

Marsella, S., Gratch, J., Wang, N., & Stankovic, B. (2009). Assessing the validity of a computational model of emotional coping. *International Conference on Affective Computing and Intelligent Interaction*, (pp. 1-8). Amsterdam, The Netherlands.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.

McQuiggan, S., & Lester, J. (2006). Learning empathy: A data-driven framework for modeling empathetic companion agents. *International joint Conference on Autonomous Agents and Multi-Agent Systems.*

Meeren, H., van Heijnsbergen, C., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *National Academy of Sciences of the United States of America, 102(45)*, 16518.

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception, 2(1)*, 10-12.

Mori, M. (1970). The Uncanny valley. *Energy*, 33-35.

Moundridou, M., & Virvou, M. (2002). Evaluating the Persona Effect of an Interface Agent in an Intelligent Tutoring System. *Journal of Computer Assisted Learning, 18(2)*, 253--261.

Mumenthaler, C., & Sander, D. (2010). Social Appraisal, how the Evaluation of Others Influences our Own Perception of Emotional Facial Expressions. *XVII Annual Cognitive Neuroscience Society Meeting*, (pp. 17-20).

Niewiadomski, R., & Pelachaud, C. (2010). Affect expression in ECAs: Application to politeness displays. *International Journal of Human-Computer Studies, 68*, 851-871.

Niewiadomski, R., & Pelachaud, C. (2010). Affect expression in ECAs: Application to politeness displays. *International Journal of Human-Computer Studies, 68(11)*, 851-871.

Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C. (2009). Greta: an interactive expressive ECA system. *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, (pp. 1399-1400).

Niewiadomski, R., Hyniewska, S., & Pelachaud, C. (2010). Introducing multimodal sequential emotional expressions for virtual characters. *International Conference on Kansei Engineering and Research.*

Noh, J., & Neumann, U. (2001). Expression cloning. *Computer graphics and interactive techniques*, 277-288.

Noldus. (2007-2011). FaceReader. http://www.noldus.com/human-behavior-research/products/facereader.

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2005). Intelligent Expressions of Emotions. *Affective Computing and Intelligent Interaction* , (pp. 707-714).

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2006). Expressions intelligentes des émotions. *Revue d'Intelligence Artificielle, 20(4)*, 607-620.

Ochs, M., Pelachaud, C., & Sadek, D. (2008). An Empathic Virtual Dialog Agent to Improve Human-Machine Interaction. *Autonomous Agents and Multiagent Systems*, (pp. 12-16).

Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions.* Cambridge: Cambridge.

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandanavian Journal Of Psychology, 7*, 1-30.

Paiva, A., & Machado, I. (2002). Life-long training with vincent, a web-based pedagocial agent. *Internationnal Journal of Continuiing Engineering Edution and Life-Long Learning 12(1)*, 254 - 266.

Paleari, M., & Lisetti, C. L. (2006). Psychologically Grounded Avatar Expressions. *Workshop of Emotions and Computing.*

Pandzic, I. S., & Forchheimer, R. (2003). *Mpeg-4 Facial Animation: The Standard, Implementation and Applications.* New York, NY, USA.: John Wiley & Sons, Inc.,.

Parke, F. I. (1974). A parametric model for human faces, PhD Thesis, University of Utah.

Parke, F. I. (1982). Parameterized Models for Facial Animation. *IEEE Computer Graphics and Applications*.

Parkinson, B. (1996). Emotions are Social. *British Journal of Psychology, 87*, 663-683.

Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. *ACM international conference on Multimedia*, (pp. 683--689).

Pelachaud, C., & Bilvi, M. (2003). Computational model of believable conversational agents. *Communication in Multiagent Systems*, 300-317.

Pereira, A., Martinho, C., Leite, I., & Paiva, A. (2008b). iCat, the chess player: the influence of embodiment in the enjoyment of a game. *Autonomous agents and multiagent systems*, (pp. 1253-1256).

Pereira, D., Oliveira, E., & Moreira, N. (2008a). Formal Modelling of Emotions in BDI Agents. *Computational Logic in Multi-Agent Systems*, (pp. 62-81).

Pfeiffer, T., Liguda, C., Wachsmuth, I., & Stein, S. (2011). Living with a Virtual Agent: Seven Years with an Embodied Conversational Agent at the Heinz Nixdorf MuseumsForum. *Conference on Re-Thinking Technology in Museum.*

Picard, R. (2010). Affective Computing: From Laughter to IEEE. *IEEE Transactions on Affective Computing, 1(1)*, 11-17.

Picard, R. W. (1997). *Affective Computing.* MIT Press.

Pighin, F., & Lewis, J. P. (2006). Introduction to performance driven facial animation. *ACM SIGGRAPH 2006 Courses*, (pp. 1-2).

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. Dans R. Plutchik, & K. H., *Emotion: Theory, research, and experience, 1* (pp. 3-33). New York: Academic.

Prendinger, H., Becker-Asano, C., & Ishizuka, M. (2006). A study in users' physiological response to an empathic interface agent. *International Journal of Humanoid Robotics, 3(3)*, 371-391.

Rapcsak, S., Galper, S., Comer, J., Reminger, S., Nielsen, L., Kaszniak, A., et al. (2000). Fear recognition deficits after focal brain damage. *Neurology, 54(3)*, 575-583.

Rébillat, M., Corteel, E., & Katz, B. (2008). SMART-I²: A Spatial Multi-users Audio-visual Real Time Interactive Interface. *Convention of the Audio engineering Society*.

Rehm, M., & André, E. (2005). Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. *Intelligent Virtual Agents*, (pp. 241--252).

Reisenzein, R. (1996). Emotional Action Generation. *Processes of the molar regulation of behavior*, (pp. 151-165).

Rimé, B., Philippot, P., Boca, S., & Mesquita, B. (1992). Long-lasting cognitive and social consequences of emotion: Social sharing and rumination. *European review of social psychology, 3(1)*, 225-258.

Rist, T., André, E., & Müller, J. (1997). Adding animated presentation agents to the interface. *2nd international conference on Intelligent user interfaces (IUI '97)*, (pp. 79-86).

Rivière, J., & Pesty, S. (2010). Actes des conversations multimodaux et émotions. *Workshop sur les Agent Conversationnel Animé.*

Riviere, J., Adam, C., Pesty, S., Pelachaud, C., Guiraud, N., Lorini, E., et al. (2011). Expressive Multimodal Conversational Acts for SAIBA agent. *Intelligent Virtual Agents*, (pp. 316–323).

Roseman, I. J. (1984). Cognitive Determinants of Emotion: A Structural Theory. *Review of Personality & Social Psychology,5,* 11-36.

Russell, J. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115(1),* 102-114.

Russell, J. A. (1989). Measures of emotion. *Emotion: Theory, research, and experience, 4*, 83-111.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Research on Personality 11(3),* 273-294.

Ruttkay, S., Noot, H., & Hagen, P. (2003). Emotion Disc and Emotion Squares: tools to explore the facial expression face. *Computer Graphics Forum, 22(1),* (pp. 49-53).

Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, 303-316.

Sander, D., Grandjean, D., & Scherer, K. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks, 18*, 317-352.

Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. *Approaches to emotion*, 293-317.

Scherer, K. R. (1999). Appraisal theory. Dans T. Dalgleish, & M. Power, *Handbook of cognition and emotion* (pp. 637-663). Chichester: Wiley.

Scherer, K. R. (2001). Appraisals Considered as a Process of Multilevel Sequential Process. *Emotion: Theory, Methods, Research*, 92-120.

Scherer, K. R. (2005). What are emotions? And how can they be measured ? *Social Science Information, 44(4),* 695-729.

Scherer, K. R. (2010). The component process model: a blueprint for a comprehensive computational model of emotion. Dans K. Scherer, T. Bänziger, & E. Roesch, *A Blueprint for Affective Computing: A sourcebook and manual*.

Scherer, K. R., & Ceschi, G. (1997). Lost Luggage: A Field Study of Emotion–Antecedent Appraisal. *Motication and Emotion, 21(3),* 211-235.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7(1),* 113-130.

Scherer, K. R., & Peper, M. (2001). Psychological theories of emotion and neuropsychological research. *Handbook of Neuropsychology, 5*, 17-48.

Scherer, K. R., & Sangsue, J. (2004). Le système mental en tant que composant de l'émotion. *Cognitions et emotions*, 11-36.

Schroder, M., Burkhardt, F., & Krstulovic, S. (2010). Synthesis of Emotional Speech. *Blueprint for affective computing*, 222-231.

Shaver, P. R., Wu, S., & Schwartz, J. C. (1992). Cross-cultural similarities and differences in emotion and its representation. *Emotion, Review of personality and social psychology, 13*, 172-212.

Shoemake, K. (1985). Animating rotation with quaternion curves. *SIGGRAPH '85*, 245-254.

Sifakis, E., Neveroc, I., & Fedkiw, R. (2005). Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Transaction on Graphics 24 (3)*, 417-425.

Sifakis, E., Selle, A., Robinson-Mosher, A., & Fedkiw, R. (2006). Simulating speech with a physics-based facial muscle model. *SIGGRAPH/Eurographics symposium on Computer animation*, 261--270.

Sloman, A., & Croucher, M. (1981). Why Robots will have Emotions. *International Joint Conference on Artificial Intelligence.*

Smith, A. (1759). *The Theory of Moral Sentiments.* Edinburgh.

Smith, C., & Ellsworth, P. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813-838.

Smith, C., & Scott, H. (1997). A Componential Approach to the Meaning of Facial Expressions. Dans J. Russell, & J. Fernandez Dols, *The Psychology of Facial Expression* (pp. 229-254). Cambrigde University Press.

Song, J., Choi, B., Seol, Y., Noh, J. ,., & 22:187–194. (2011). Characteristic facial retargeting. *Computer Animation and Virtual Worlds, 22*, 187–194.

Steggles, P., & Gschwind, S. (2005). The Ubisense smart space platform. *Adjunct Proceedings of the Third International Conference on Pervasive Computing*, 73-76.

Stoiber, N., Breton, G., & Seguier, R. (2010). Modeling Short-Term Dynamics and Variability for Realistic Interactive Facial Animation. *IEEE Computer Graphics and Applications, 30(4)*, 51-61.

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., & Bronnenkant, K. (2010). Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. *Intelligent Virtual Agents*, (pp. 286-300).

Tan, N., Clavel, C., Courgeon, M., & Martin, J. C. (2010). Postural Expression of Action Tendency. *Proceedings of the ACM Multitimedia 2010's Workshop on Social Signal Processing* .

Terzopoulos, D., & Waters, K. (1993). Analysis of Facial Image Sequences Using Physical and Anatomical Models. *Pattern Analysis and Machine Intelligence, 15(6)*, 569-576.

Thórisson, K. R. (1999). A Mind Model for Multimodal Communicative Creatures and Humanoids. *International Journal of Applied Artificial Intelligence, 13(4)*, 449-486.

Thornton, I. (1998). *The Perception of Dynamic Human Faces.* PhD Thesis, University of Oregon.

Tomkins, S. S. (1984). Affect theory. Dans K. Scherer, & P. Ekman, *Approaches to emotion* (pp. 163-195). Hillsdale, NJ: Erlbaum.

Traum, D., & Rickel, J. (2002). Embodied agents for multi-party dialogue in immersive virtual worlds. *Autonomous Agents and Multi Agent Systems*, 766-773.

Vanhala, T., Surakka, V., Courgeon, M., & Martin, J. C. (2011). Voluntary Facial Activations Regulate Physiological Arousal and Subjective Experiences During Virtual Social Stimulation. *ACM Transactions on Applied Perception*, à paraitre.

Vilhjalmsson et al. (2007). The behaviour markup language : recent developments and challenges. *Intelligent Virtual Agent*, (pp. 99-110).

Vinayagamoorthy, V., Steed, A., & Slater, M. (2005). Building Characters: Lessons Drawn from Virtual Environments. *Toward Social Mechanisms of Android Science*, 119–126.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*, 879-896.

Wallraven, C., Breidt, M., Cunningham, D., & Bülthoff, H. H. (2005). Psychophysical evaluation of animated facial expressions. *2nd Symposium on Applied Perception in Graphics and Visualization* (pp. 17-24). ACM Press, New York, NY.

Walters, M., Syrdal, D., Dautenhahn, K., te Boekhorst, R., & Koay, K. (2008). Avoiding the Uncanny Valley. Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion. *Journal of Autonomous Robots, 24(2)*, 159–178.

Waters, K. (1987). A muscle model for animation three-dimensional facial expression. *ACM SIGGRAPH Computer Graphics*, 11-24.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. (2000). Stydying the Dynamics of Emotional Expression via Synthesized Facial Muscle Movements. *Journal of Personality and Social Psychology, 78(1)*, 105-119.

Weyrich, T., Matusik, W., Pfister, H., Bickel, B., Donner, C., Tu, C., et al. (2006). Analysis of human faces using a measurement-based skin reflectance model. *ACM SIGGRAPH*, (pp. 1013-1024).

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 460-473.

Wu, Y., Kalra, P., & Magnenat-Thalmann, N. (1997). Physicallybased Wrinkle Simulation & Skin Rendering. *Computer Animation and Simulation*, 69-79.

Wu, Y., Magnenat-Thalmann, N., & Thalmann, D. (1995). A Dynamic Wrinkle Model in Facial Animation and Skin Aging. *Journal of Visualization and Computer Animation, 6(4)*, 195-205.

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 1-10).

Yoshikawa, S., & Sato, W. (2008). Dynamic Facial Expressions of emotion induce Emotional Momentum. *Cognitive, Affective, and Behavior*, 25-31.

Publications

En cours de soumission:

- M. **Courgeon**, S. Buisine, J-C. Martin (2011) *Expressive Wrinkles in a Virtual Character's Facial Expressions: Impact on Emotion Recognition, Perceived Expressiveness and Users' Preferences*, (Soumis)

Journaux internationaux :

- Toni Vanhala, Veikko Surakka, M. **Courgeon**, and J.C. Martin, (2012) *Voluntary Facial Activations Regulate Physiological Arousal and Subjective Experiences During Virtual Social Stimulation,* In: ACM Transactions on Applied Perception, (à paraître, janvier 2012, impact factor : 1.447)
- M. **Courgeon**, C. Clavel, N. Tan, J.C. Martin, (2011) *Front View vs. Side View of Facial and Postural Expressions of Emotions in a Virtual Character*, In: LNCS Transactions on Edutainment, 6, pp 132-143
- O. Grynszpan, J. Nadel, J. Constant, F. Le Barillier, N. Carbonell, J. Simonin, J-C. Martin, M. **Courgeon**, (2011), *A new virtual environment paradigm for high functioning autism intended to help attentional disengagement in a social context,* in: Journal of physical therapy education, 25(1), pp 42-47

Conférences internationales :

- N. Tan, G. Pruvost, M. **Courgeon**, C. Clavel, Y. Bellik, J.C. Martin (2011) *A Location-Aware Virtual Character in a Smart Room: Effects on Performance, Presence and Adaptivity*, in: Proceedings of the International Conference of Intelligent User Interface (IUI 2011), Palo Alto, U.S.A, 13-16 février 2011
- M. **Courgeon**, M-A. Amorim, C. Giroux, J-C. Martin (2010), *Do Users Anticipate Emotion Dynamics in Facial Expressions of a Virtual Character?*, in: Proceedings of the 23rd International Conference on Computer Animation and Social Agents (CASA 2010), Saint Malo, France, 31 mai - 2 juin 2010
- C. Jacquemin, W.K. Chan, M. **Courgeon**, (2010) *Bateau Ivre: An Artistic Markerless Outdoor Mobile Augmented Reality Installation on a Riverboat*, in: Proceedings of the ACM Multimedia 2010 (Interactive Art Program)
- M. **Courgeon**, S. Buisine, J-C. Martin (2009) *Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions*, In: Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA 09), pp 201-214, Amsterdam, The Netherlands, 10-12 septembre 2009
- M. **Courgeon**, C. Jacquemin, J-C. Martin, (2008) U*ser's Gestural Exploration Of Different Virtual Agents' Expressive Profiles*, in: Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 08), 3, pp 1237-1240, Estoril, Portugal, 12-16 mai 2008

Workshops:

- **M. Courgeon**, N. Tan, C. Clavel, C. Zakaria, V. Eyharabide, J. Jacquemot, C. Jacquemin, J.C. Martin, (2011) *Humains Virtuels & Humains Réels,* in: Digiteo Annual Forum, Palaiseau, France, 18 octobre 2011
- N. Tan, C. Clavel, M. **Courgeon**, J-C. Martin (2010), *Postural Expression of Action Tendency*, in: Proceedings of the ACM Multimedia 2010's Workshop on Social Signal Processing, 8 septembre 2010
- M. **Courgeon**, M. Rebillat, B. Katz, C. Clavel, J-C. Martin, (2010) *Life-Sized Audiovisual Spatial Social Scenes with Multiple Characters: MARC & SMART-I2,* in: Proceedings of 5th conf. of AFRV, Orsay, France 6-8 décembre 2010
- M. **Courgeon**, C. Clavel, J-C. Martin, (2009) *Appraising Emotional Events during a Real-time Interactive Game,* in: Proceedings of the ICMI 2009 Workshop on Affective Computing (AFFINE), Cambrigde, U.S.A., 1-6 novembre 2009
- Clay, M. **Courgeon**, N. Couture, E. Delord, C. Clavel, J-C. Martin (2009) *Expressive Virtual Modalities for Augmenting the Perception of Affective Movements,* in: Proceedings of the ICMI09 International workshop on Affective Computing (AFFINE), Cambridge, USA, 1-6 novembre, 2009
- C. Arrat, M. **Courgeon**, (2009) Particules, in : Journée d'études Simulation Technologique et Matérialisation Artistique, Béton Salon, Paris
- M. **Courgeon**, J-C. Martin, C. Jacquemin, (2008) *MARC: a Multimodal Affective and Reactive Character* In: Proceedings of the ICMI workshop on Affective Computing (AFFINE), Chania, Greece, 20-22 octobre 2008
- M. **Courgeon**, J-C. Martin, C. Jacquemin, (2008) *MARC : Un Personnage Virtuel Réactif Expressif*, In: Proceedings of the 2nd Workshop on Animated Conversational Agents, Paris, France, 25-26 novembre 2010
- M. **Courgeon**, J-C. Martin, C. Jacquemin, (2008) *Virtual Humans: Expressivity, Interactivity and Realism,* in: Digiteo Annual Forum, Orsay, France, 2 octobre 2008
- M. **Courgeon**, J-C. Martin, C. Jacquemin, (2008) *User's Gestural Exploration of Different Virtual Agents' Expressive Profiles,* in: Proceedings of the Speech and Face to Face Communication Workshop in memory of Christian Benoît, Grenoble, France, 27-29 octobre 2008
- C. Jacquemin et al. (2008) Capture and Machine Learning of Physiological Signals, in: Proceedings of the eNTERFACE 08 Workshop, Orsay, France

Présentations et Posters

- Conférence Internationale AAMAS 08 (Estoril, Portugal) Présentation d'un poster : *User's Gestural Exploration of Different Virtual Agents' Expressive Profiles*
- Forum Digiteo 2008 (Saclay, France) Présentation d'un poster : *Virtual Humans: Expressivity, Interactivity and Realism*
- Conférence Internationale ICMI-MLMI 2008 (Chania, Grèce) Présentation orale : *MARC: a Multimodal Affective and Reactive Character*
- Conférence Nationale WACA 2008 (Paris, France) Présentation orale : *MARC : Un Personnage Virtuel Réactif Expressif*
- Conférence Internationale IVA 2009 (Amsterdam, Pays-Bas) Présentation orale *: Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions*
- Conférence Internationale ICMI-MLMI 2009 (M.I.T. Cambridge, USA) Présentation orale : *Appraising Emotional Events During a Real-time Interactive Game*
- Conférence Internationale CASA 2010 (St Malo, France) Présentation Orale : *Do Users Anticipate Emotion Dynamics in Facial Expressions of a Virtual Character?*
- Séminaire LIMSI-CHM, Juin & Septembre 2010 (Orsay, France) Présentation orale : *OpenGL 3.1 (R)évolution ?*
- Conférence de l'AFRV 2010 (Orsay, France) Présentation orale *: Life-Sized Audiovisual Spatial Social Scenes with Multiple Characters: MARC & SMART-I²*
- Salon Display et Solutions Numérique 2011 : Applications de la réalité augmentée, (Porte de Versailles, Paris, France) Présentation orale : *Bateau Ivre: An Artistic Markerless Outdoor Mobile Augmented Reality Installation on a Riverboat.*
- Forum Digiteo 2011 (Palaiseau, France) Présentation d'un poster : *Humains Virtuels, Humains réels*

Table des illustrations