

Des agents virtuels sociaux et émotionnels pour l'interaction humain-machine

*Nicolas Ech Chafai, Magalie Ochs, Christopher Peters, Maurizio Mancini,
Elisabetta Bevacqua, Catherine Pelachaud*

{n.chafai, m.ochs, c.peters, m.mancini, e.bevacqua, pelachaud, @iut.univ-paris8.fr}
140, rue de la Nouvelle France
93100 Montreuil

RESUME

L'engagement de l'utilisateur dans une interaction humain-machine médiatisée par un agent virtuel dépend des capacités de communication et d'interaction de cet agent. De plus en plus, le contexte de l'interaction est à la fois social et émotionnel. Les agents virtuels devraient donc être capables de communiquer socialement et émotionnellement avec un utilisateur. De nombreux codes sociaux sous-tendent les partenaires d'interaction, notamment lorsque ceux-ci tentent d'initier une conversation avec un autre agent ou utilisateur, ou bien lorsque l'agent veut montrer l'intérêt qu'il porte à l'échange conversationnel. Ces nécessités impliquent d'intégrer des codes sociaux à nos agents, mais également de les doter de la capacité de percevoir leur environnement. Pour doter un agent de l'ensemble de ces capacités, un modèle de perception, d'attention, d'émotion, de rétroaction, et de contrôle de l'expressivité comportementale de l'agent sont proposés dans cet article.

MOTS CLES : Interactions sociales et émotionnelles, agents conversationnels expressifs.

1. INTRODUCTION

Les agents virtuels (2D, 3D ou simplement vocaux ou textuels) sont de plus en plus utilisés dans des interfaces des systèmes informatiques. Ils sont aujourd'hui utilisés dans divers contextes tels que les applications web, les systèmes interactifs ou les jeux vidéo. Dans les interactions humain-machine, l'utilisation de ces agents virtuels permet d'améliorer à la fois les échanges entre un utilisateur et l'environnement d'une application et le sentiment de présence offert par cette application.

Une des difficultés majeure pour la conception de ces agents réside dans le contrôle de l'animation de ces derniers : les animations pré-codées dans lesquelles le com-

portement de l'agent est prédéfini ne conviennent pas à la réalisation de systèmes véritablement interactifs. Les agents virtuels doivent être *autonomes*. Ils doivent être capables de décider quelle action entreprendre et quel comportement exhiber.

De plus, pour être *crédibles* lors de leur interaction avec un utilisateur, les agents virtuels doivent être capables de *percevoir leur environnement*. Plus précisément, ils doivent être capables d'appréhender tant leur environnement physique (les objets, leurs dispositions, etc.) que social (les autres agents à proximité, la direction du regard de l'utilisateur ou simplement sa présence.). Sans cette faculté de perception, les agents se comportent comme s'ils se mouvaient dans un environnement vide : ils n'ont pas *conscience* de leur entourage, ne prêtent pas attention aux agents ou individus à qui ils sont en train de parler, ne tentent pas de s'accaparer l'attention d'un autre agent ou de l'utilisateur.

De notre point de vue, l'élément essentiel dans une interaction humain-machine n'est pas qu'un agent virtuel se comporte exactement comme un humain, mais qu'il soit capable d'exhiber un comportement communicatif social et émotionnel, tout en restant *cohérent* avec le comportement humain réel. Ce comportement est par essence multimodal : les modalités tant verbales que non verbales jouent un rôle primordial dans les interactions conversationnelles. Le comportement non verbal est par exemple utilisé pour transmettre une information sur nos états mentaux ou émotionnels, envoyer des signaux particuliers à l'attention de nos interlocuteurs, se référer à des éléments de l'environnement, etc. Pour être capable d'adopter des comportements cohérents avec ceux d'un humain dans une interaction conversationnelle, un agent virtuel doit pouvoir (1) percevoir son environnement, le comportement et le regard des autres individus/agents présents, (2) générer des comportements communicatifs multimodaux, (3) exprimer des émotions et (4) se montrer réceptif lorsque l'utilisateur s'adresse à lui.

Dans cet article, nous présentons différents modèles permettant de doter un agent de ces capacités. Tout d'abord, un modèle de perception de l'environnement est introduit (section 2). Celui-ci est utilisé pour construire un modèle d'engagement dans une interaction (section 3). Les fonctions communicatives permettant à un agent

de dialoguer avec l'utilisateur sont décrites dans la section suivante (section 4). Un modèle comportemental définissant les capacités d'expression multimodale est présenté (section 5). Au fil de son discours, l'agent doit attirer et maintenir l'attention de son interlocuteur. Pour ce faire, des règles automatiques, introduites dans la section 6, peuvent être appliquées au comportement de l'agent. Dans la section 7, un modèle permettant à un agent d'identifier la signification émotionnelle d'une situation d'interaction est présenté. Enfin, dans la section 8, nous proposons un modèle de rétroaction de l'agent lorsque l'utilisateur est locuteur. Enfin, nous présentons un scénario d'interaction conversationnelle entre deux agents qui illustre différents modèles introduits dans cet article (section 9).

2. LA PERCEPTION DE L'AGENT DE SON ENVIRONNEMENT

Si l'on conçoit des agents autonomes dans des environnements complexes, ces agents virtuels doivent être capables de percevoir leur environnement afin d'adopter un comportement approprié à la situation d'interaction. Nous proposons des agents dotés d'une vision de synthèse, de capacités d'attention et de mémorisation en temps réel [24]. Ces capacités permettent à l'agent de collecter des informations sur son environnement et ainsi de pouvoir s'y orienter. La vision de l'agent est monoculaire et dotée de multiples résolutions.

Le module de vision réalise des captures fréquentes de la région de la scène visible par l'agent. Ces informations sont ensuite transférées au module d'attention. Le but de ce module est de sélectionner un ensemble d'objets et de sites visibles dans la scène à partir d'un modèle cognitif temps réel [15]. Cet ensemble est utilisé pour calculer la *carte des espaces saillants* : un classement en niveau de gris des espaces suivant le degré avec lequel ils attirent le regard.

La carte des espaces saillants est calculée de la façon suivante : ses différentes caractéristiques sont calculées à partir du module de vision de manière concentrique, simulant les propriétés biologiques de la vision, plus sensibles aux contrastes spatiaux locaux qu'à l'amplitude d'une caractéristique particulière. L'image résultante est décomposée en quatre canaux : un lié à l'intensité, un à l'orientation, et deux autres aux couleurs. Chaque canal est utilisé comme premier niveau dans la construction d'une pyramide des images dyadiques, qui définit l'ensemble des images dans lequel chacune des images est une version filtrée de leur prédécesseur. Pour les canaux correspondant à l'intensité et à l'orientation, un filtre gaussien est appliqué. Pour l'orientation, nous utilisons un filtre Gabor ayant pour angle : 0, 45, 90, et 135 degrés. Les caractéristiques de la carte liées au champ concentrique de la vision sont obtenues à partir des images filtrées. Ce champ concentrique correspond à une mesure relative : une partie de l'image attire potentiellement l'attention en fonction des régions de l'image qui

lui sont adjacentes, de la même façon qu'un point blanc au milieu d'un tableau noir, ou qu'une ligne diagonale au milieu d'un champ de lignes verticales. Les caractéristiques de la carte pour l'intensité, la couleur et l'orientation sont ensuite combinées en trois cartes de visibilité. Chaque carte de visibilité apporte une mesure des parties de la scène qui surgissent. De la combinaison de ces cartes de visibilité nous obtenons une mesure unifiée des parties saillantes de la scène. La carte de saillance obtenue (16x16) nous donne finalement la sortie principale de notre modèle d'attention.

En fonction du parcours visuel déjà effectué par l'agent, un *système de mémoire* [25] assure que l'attention focale parcourt effectivement la scène. Associé à la carte des espaces saillants, ce système de mémoire permet d'obtenir la carte d'attention finale qui définit les espaces saillants non encore fixés par l'agent, appelés les régions d'attention. Ces régions d'attention sont utilisées pour créer le parcours visuel de l'agent.

Ces capacités permettent de créer des agents autonomes dotés de comportements réalistes : l'agent est capable de déterminer automatiquement quelles sont les régions d'attention d'une scène selon son point de vue et d'adapter son attention focale en fonction des éléments d'intérêt de la scène. Elles peuvent de plus être utilisées pour déterminer à quel moment l'agent peut réaliser des mouvements subtils des yeux tels que des clignements ou des tressautements, qui sont en partie liés aux changements de visée et au degré d'attention. Enfin, ces capacités sont utiles pour l'implémentation de mécanismes cognitifs et comportementaux de haut niveau. Dans la section suivante, nous présentons par exemple comment elles peuvent être utilisées, conjointement avec la théorie de l'esprit, pour contrôler automatiquement l'agent quand celui-ci doit initier une interaction.

3. L'INITIATION D'UNE INTERACTION CONVERSATIONNELLE

Lorsqu'un utilisateur démarre une interaction avec un agent virtuel, l'agent partage avec l'utilisateur un ensemble de comportements à forte codification sociale. L'orientation du regard, les salutations, la dénomination, sont des exemples de tels comportements. Ces derniers, acquis au fil du temps, nous semblent naturels. Pour augmenter le réalisme de l'interaction, l'agent virtuel doit également être doté de ces comportements [14]. Dans les interactions interpersonnelles, un ensemble de comportements précède généralement l'interaction en question. Les individus utilisent leurs sens pour évaluer les contacts possibles avec les personnes présentes dans leur entourage. Ils doivent gagner l'attention de la personne avec laquelle ils souhaitent interagir, signaler leur intérêt verbalement et non verbalement, chercher une coopération et évaluer la volonté d'un autre individu de participer à la conversation.

Un système a été développé afin de reproduire ces comportements au sein d'une interaction. Plus précisé-

ment, ces comportements ont été intégrés dans un système automatique pour permettre à l'agent virtuel d'initier une interaction avec un utilisateur virtuel. Ce système permet de doter un agent d'une perception sociale de son environnement [26] par l'intégration d'une théorie de l'esprit fondée sur la psychologie évolutionniste et soutenu par des travaux en neurophysiologie [23]. Dans notre modèle, un mécanisme de *détection de l'intentionnalité* sélectionne les utilisateurs virtuels qui ont l'intention d'entrer en interaction avec l'agent. Ces utilisateurs font alors partie du parcours visuel de l'agent. En d'autres termes, s'il y a intention d'un utilisateur d'entrer en interaction avec l'agent, alors cet utilisateur fait partie du parcours visuel de l'agent. Le corps virtuel de l'utilisateur (les yeux, la tête, et les parties inférieures du corps) est segmenté. Un *détecteur de direction de l'attention* calcule la direction des yeux, de la tête, et du reste du corps des utilisateurs virtuels par rapport à l'agent lui-même. A partir de ces données, un *niveau d'attention* est calculé. Lorsqu'il existe une attention mutuelle entre l'agent et un utilisateur (calculé par un *mécanisme d'attention mutuelle*), le niveau d'attention augmente. Toutes ces données sont stockées dans une mémoire à court terme ; elles sont ensuite utilisées pour déterminer si un utilisateur a vu l'agent et si d'autres utilisateurs virtuels ont vu l'agent le regarder. Enfin, elles permettent d'évaluer l'intérêt qu'un autre agent semble lui porter.

Ces différents modules de détection de l'attention, de l'intentionnalité et le mécanisme d'attention mutuelle déterminent une machine à état fini hiérarchique entre différents états de l'interaction tels que le contrôle de l'environnement, la capture de l'attention d'un autre et l'évaluation de la réaction d'un autre.

Ainsi, contrairement à d'autres systèmes, le comportement de l'agent n'est pas seulement déterminé par ses buts, mais aussi par les intentions des utilisateurs virtuels. Ceci permet d'éviter des conflits sociaux ; par exemple éviter que l'agent engage une conversation avec un utilisateur qui ne le souhaite pas [11].

Une évaluation des comportements d'initiation d'une interaction par l'agent avec l'utilisateur a été réalisée. L'évaluation consiste en deux études: une étude d'évaluation statique (EES), et une étude d'évaluation dynamique (EED). Nous avons demandé à 21 participants de regarder le comportement d'agents virtuels et de reporter leurs impressions vis-à-vis de leur attitude d'attention. Pour l'EES, une séquence de 25 images statiques d'un agent virtuel dans une posture droite a été montrée aux participants. Chaque image montre l'agent orienté de différentes façons en fonction du participant : dans certaines images, seuls quelques segments corporels étaient visibles. Après avoir visionné chacune de ces images, les participants devaient évaluer le degré d'attention qu'ils pensaient que l'agent leur portait ("L'attention que l'on me porte"), sur une échelle allant de AUCUNE à GRANDE. Pour l'EED, 10 animations dans lesquelles un agent se déplaçait dans son environnement virtuel, et

produisant différents comportements étaient montrées aux participants. Les comportements pouvaient aller de : ignorer le participant, le regarder, aller vers lui, à le saluer. A la fin de chaque animation, les participants devaient évaluer la façon dont ils interprétaient les actions de l'agent notamment en terme d'intérêt qu'il leur portait, ou si l'agent les avait simplement remarqués (voir Table 1 pour la liste de ces actions). Avant cette évaluation EES, une animation-test était visionnée par les participants afin de démontrer les capacités d'un agent, et ce qu'ils étaient en mesure d'en attendre. Deux cas étaient distingués dans cette animation : un cas dans lequel l'agent était perçu comme "voulant discuter avec le participant", un autre cas dans lequel l'agent se montrait à la disposition du participant si ce dernier voulait engager l'interaction. Le premier cas suggère un agent proactif et cherchant à démarrer l'interaction, tandis que le second cas suggère un agent voulant interagir, mais d'une manière plus passive.

Cas	Texte (L'agent ...)	Type / Echelle
(a)	m'a regardé	Oui / Non
(b)	sait que je suis présent	Oui / Non
(c)	est intéressé par rapport à moi	Pas du tout -> Entièrement
(d)	aimerait me parler	Pas du tout -> Entièrement
(e)	répond à ma volonté de lui parler	Pas du tout -> Entièrement
(f)	a une intention envers moi qui est	Mauvaise / Nuisible -> Bonne / Bienfaisante

Cette étude révèle un nombre intéressant de résultats (que nous détaillons dans [26]). Tout d'abord, que les agents virtuels ont la capacité de suggérer à un utilisateur qu'ils leur portent attention au travers de leur regard, de l'orientation de leur corps, de leurs gestes et de leurs comportements lorsqu'ils se déplacent. Ensuite, que la direction du regard, de la tête, du corps, de déplacement de l'agent, ainsi que son comportement dynamique lorsque l'agent se déplace permet à l'utilisateur d'évaluer l'intérêt, la recherche d'interaction et l'ouverture de cet agent. L'étude nous révèle également l'importance du regard de l'agent, ainsi que du contraste entre l'orientation des segments corporels lors de la modélisation de tout type de perception de l'attention.

Finalement, cette étude montre la pertinence de l'intégration de ces comportements sociaux non verbaux chez des agents virtuels.

4. LES FONCTIONS COMMUNICATIVES D'UN AGENT

Lors d'une conversation interpersonnelle, et par extension dans une interaction humain-machine, la communication est véhiculée par de multiples canaux d'expression [9, 16, 18, 32], comme par exemple la prosodie, les expressions faciales, la posture, le regard, les gestes. Ces comportements non verbaux sont synchronisés avec la parole, ponctuant les accentuations phonémiques et les pauses dans le discours. Plusieurs fonctions communicatives sont en jeu dans de tels comportements [7, 30] : gérer le flux discursif, mettre de l'emphase sur un terme du discours ou fournir de l'information sur nos émotions ou sur nos états mentaux.

Dans notre modèle, nous nous sommes fondés sur la taxonomie des fonctions communicatives proposée dans [30]. Quatre catégories principales sont distinguées selon que le comportement fournit de l'information sur les croyances du locuteur, ses intentions, son état affectif ou sur des informations méta-cognitives liées aux états mentaux du locuteur. Une information peut être exprimée par des signaux verbaux et/ou non-verbaux. Il peut de plus exister plusieurs façons d'exprimer une même information. Par exemple, pour indiquer un objet dans la scène, la direction du regard et/ou un geste de pointage peuvent être utilisés. Nous avons défini un ensemble de signaux (comme par exemple la direction du regard) auxquels sont associés une ou plusieurs significations (caractère *polysémique* d'un comportement non verbal). Une signification peut également être attachée à plusieurs signaux (caractère *polysigne*).

Afin de contrôler le comportement de l'agent, un langage de représentation appelé "Affective Presentation Markup Language" (APML) est utilisé. Ce langage permet de décrire le contenu textuel prononcé par l'agent, ainsi que l'ensemble des étiquettes caractérisant les fonctions communicatives définies précédemment. A travers ce langage, il est donc possible de décrire à quel moment l'agent doit marquer l'emphase sur un mot ou exprimer une émotion particulière lorsqu'il prononce une phrase [9].

5. LE MODELE COMPORTEMENTAL D'UN AGENT

Le comportement d'un agent doit être adapté à sa personnalité, son humeur, son état émotionnel, et à ses caractéristiques expressives personnelles [33]. Ceci permet de créer des agents virtuels capables d'exhiber des comportements distinctifs. Ainsi, même si les intentions communicatives et/ou les états émotionnels de deux agents sont semblables, il se peut qu'ils agissent de façon différente.

Pour ce faire, un ensemble de paramètres permettant de décrire des profils de comportement a été défini. Ces profils, définis en fonction des intentions communicatives et des états émotionnels de l'agent, sont décrits en termes de *signaux préférentiels* pour la communication et en termes d'*expressivité* de ces signaux. Ces deux notions sont décrites plus précisément ci-dessous.

5.1. La multi-modalité

Un agent virtuel doit décider quelle modalité utiliser, parmi un ensemble de possibles, pour exhiber son comportement. Son discours, ses gestes, son regard, ses expressions faciales, ses mouvements corporels ou sa posture sont autant de modalités capables de véhiculer de l'information [22, 33]. Un mécanisme de décision a été mis en place afin de permettre à l'agent de sélectionner la modalité à utiliser. Ce mécanisme est basé sur les préférences d'usage des modalités parmi celles disponibles.

5.2. L'expressivité des comportements

Un même geste peut être réalisé de diverses manières. L'*expressivité d'un comportement* permet de mettre en évidence la manière donc est exécuté un comportement [36]. Nous décrivons l'expressivité d'un comportement par les paramètres suivants : la vitesse du geste/de la tête, son accélération, l'amplitude spatiale occupé par le geste, l'énergie et la fluidité du geste/de la tête et le nombre de répétitions du geste/du mouvement de la tête [13]. Les comportements (mouvement de tête, mouvement des bras et des mains, expressions du visage) sont définis par des positions clés qui simulent leurs caractéristiques temporelles telles que définies par Ekman [9] et Kendon [16]. Ces paramètres agissent directement sur la spécification des comportements pour chacune des positions clés.

5.3. Les caractéristiques expressives personnelles

Les *caractéristiques expressives personnelles* décrivent les tendances d'un agent à utiliser une modalité particulière pour s'exprimer. Dans notre modèle, les *tendances globales* sont distinguées des *tendances locales* (tendances propres à des situations particulières). Imaginons un individu qui produit généralement des gestes larges et des mouvements rapides dans une conversation. Cet individu aura recours plus fréquemment à ses gestes pour transmettre une information qu'à d'autres modalités non verbales (expressions faciales, regard, posture, mouvements de tête). Cette tendance générale est ce que nous appelons un indicateur *global* du comportement. D'un autre côté, certaines situations peuvent amener cet individu à bousculer ces tendances et à adopter un comportement différent. Par exemple, une personne qui n'a pas l'habitude de produire des gestes ou des mouvements de tête lorsqu'elle parle, pourra avoir tendance à en produire lorsqu'elle est en colère. Nous appelons cette tendance un indicateur *local* du comportement. Dans notre modèle, l'indicateur local dépend des intentions communicatives et des états mentaux et émotionnels de l'agent. Les indicateurs *local* et *global* du comportement déterminent les signaux préférentiels de l'agent et leur expressivité. L'indicateur *local* modifie l'indicateur *global* localement. C'est-à-dire les caractéristiques expressives comportementales de l'agents telles que spécifiées par l'indicateur *global* (degré de préférence d'usage et spécification de l'expressivité pour chaque modalité) sont modifiées localement.

6. L'ATTRACTION DE L'ATTENTION

A travers la conversation, un agent virtuel fournit à l'utilisateur des informations sur une application donnée, maintient son intérêt sur un aspect précis d'une tâche ou converse simplement avec lui comme compagnon virtuel. Si l'on conçoit ces agents comme des compagnons à présence continue, il faut les doter de capacités de communication suffisantes pour maintenir l'intérêt de l'utilisateur et focaliser son attention à des instants clés de l'interaction (comme par exemple lorsqu'une information importante doit être délivrée à l'utilisateur).

Pour ce faire, nous avons étudié la valeur pragmatique des gestes. Au cours d'une interaction conversationnelle, les gestes aident non seulement le locuteur à formuler et à illustrer ses pensées, mais expriment en même temps son effort de communication (reflétant généralement l'importance pour le locuteur que son intention communicative soit réalisée). Cet effort est révélé (entre autres) par l'expressivité des gestes (la force, la fluidité, l'expansion spatiale ou la répétition) (voir section 5.2). Cette expressivité représente un indice pour l'utilisateur lui permettant d'identifier l'effort communicatif de l'agent conversationnel.

Grâce notamment à des techniques de suivi du regard, de précédents travaux ont permis d'établir quelles sont les caractéristiques des gestes qui déclenche l'attention focale de l'interlocuteur [3, 12]. Ces travaux ont permis de mettre en évidence le fait que le regard est plus attiré par des gestes déictiques (qui pointent vers un objet concret ou abstrait – *e.g.* lorsque le locuteur prend à parti une personne absente, mais qu'il localise dans son espace gestuel) ou à forte expansion spatiale. Dans nos travaux, nous avons collecté des données subjectives à partir de vidéos d'animation traditionnelle 2D.

A partir de ces vidéos, nous avons étudié le lien entre les variations de l'expressivité des gestes du locuteur et son intention de communiquer une information particulière. Au cours d'un discours ou d'une interaction conversationnelle, le locuteur n'emploie pas continuellement la même expressivité dans ses gestes. Les variations d'expressivité peuvent correspondre à des variations dans ses intentions communicatives. Nous distinguons deux types de variations : les *irrégularités* qui correspondent à une variation très localisée de l'expressivité dans le temps et les *discontinuités* qui correspondent à une brusque variation mais sur une période indéfinie. Nous avons pu observer deux fonctions jouées par ces variations : une relation rhétorique de similarité d'une part et une relation rhétorique de contraste d'autre part, entre des éléments verbaux ou non verbaux de l'énoncé [8]. Cette relation de similarité accentue la segmentation de l'énoncé (*e.g.* lorsque le locuteur ponctue son discours de gestes répétitifs), tandis que la relation de contraste reflète l'effort du locuteur de communiquer une intention particulière.

L'utilisation de ces deux types de variations dans un agent virtuel peut permettre d'améliorer l'interaction, en

agissant sur le maintien de l'intérêt de l'utilisateur par l'accentuation de la segmentation de l'énoncé (par l'utilisation d'*irrégularités*) et en agissant sur son attention en signalant une intention particulière du locuteur (par l'utilisation de *discontinuité*).

7. LA MODELISATION DES EMOTIONS

Ces dernières années, un intérêt grandissant est apparu pour la création d'agents virtuels capables d'exprimer des émotions. Les expressions d'émotions permettent en effet de créer une illusion de vie et ainsi d'augmenter leur réalisme (traduction du terme anglais *believability* [4]). De plus, de récentes recherches ont montré que l'expression d'émotions (particulièrement celle d'*empathie*) d'un agent virtuel semble permettre d'améliorer la perception par l'utilisateur de l'agent, d'induire des émotions positives et d'augmenter les performances et l'engagement de l'utilisateur dans la réalisation d'une tâche durant l'interaction [5, 17, 21]. Pour être capable d'exprimer des émotions, un agent doit savoir *comment* les exprimer afin de les rendre perceptibles à l'utilisateur mais aussi savoir *dans quelles circonstances quelles émotions* sont déclenchées afin de déterminer lesquelles exprimer et lesquelles sont potentiellement ressenties par ses agents voisins. Ceci est essentiel pour un agent empathique qui, par définition, exprime les émotions qu'il pense que son interlocuteur ressent [29].

La théorie de l'évaluation cognitive (*appraisal*) [35] – qui vise à expliquer ce qui conditionne l'émergence d'une émotion particulière pour un individu donné – peut être utilisée pour donner la capacité à un agent d'identifier la signification émotionnelle d'une situation. Suivant cette théorie, une émotion est issue de l'évaluation subjective d'un événement. Lorsqu'un événement se déroule dans l'environnement, l'individu évalue ce dernier à partir de ses croyances et ses buts.

Notre objectif est de créer un modèle de déclenchement d'émotions indépendant du domaine et permettant à un agent d'identifier ses propres émotions et celles potentiellement ressenties par ses interlocuteurs. Pour ce faire, nous proposons une représentation des conditions de déclenchement des émotions fondée sur une approche BDI [31]. Les agents de type BDI (*Belief Desire Intention*) sont dotés d'*états mentaux* composés d'attitudes mentales, telles que les croyances et les buts, qui vont leur permettre de raisonner et d'agir sur leur environnement. L'état mental d'un agent correspond à une représentation cognitive du monde à un instant donné. Il inclut une représentation des événements perçus dans l'environnement. Par conséquent, un événement déclencheur d'émotion est aussi représenté par des attitudes mentales. Ces événements peuvent donc être représentés par un état mental particulier. Fondé sur les théories en psychologie cognitive [35, 20], nous avons décrit en termes de croyance et de buts les conditions de déclenchement d'une émotion. Par exemple, une émotion de *joie* est générée par la croyance qu'un événement qui vient de se dérouler a permis à un agent de réaliser un de ses buts.

Les événements déclencheurs d'émotion sont ainsi représentés par un état mental particulier, *i.e.* par des combinaisons d'attitudes mentales (croyances et buts) particulières. A partir de cette formalisation, l'agent peut identifier en temps réel ses émotions. Suivant les buts de ses interlocuteurs, il peut déterminer leurs émotions potentielles. Un modèle d'expressions faciales des émotions a été construit afin que l'agent puisse exprimer ses émotions (pour plus de détails voir [19]).

Ce modèle d'émotions a été intégré dans un agent rationnel dialoguant. Fondés sur la théorie des actes de langage [2], ces agents de type BDI sont dotés d'une représentation des actes communicatifs en terme de croyances et d'intentions. Ainsi, à partir d'un énoncé, ils peuvent déduire certaines croyances et intentions de leur interlocuteur. Par exemple, si l'utilisateur demande à l'agent rationnel dialoguant l'adresse d'un restaurant, l'agent en déduit que l'utilisateur ne connaît pas actuellement l'adresse du restaurant et a l'intention de connaître cette adresse. L'agent adopte alors l'intention de l'utilisateur (pour plus de détails voir [34]). Ainsi, à partir de sa représentation des conditions de déclenchement des émotions, il exprimera par exemple une émotion négative d'empathie lorsqu'il pense qu'il n'a pas réussi à réaliser une intention de l'utilisateur.

8. DES AGENTS ATTENTIFS AUX UTILISATEURS

Durant une conversation, il y a deux participants : le locuteur et celui qui écoute. L'agent virtuel se trouve souvent dans la position de celui qui écoute lorsqu'il interagit avec un utilisateur. Il adopte alors un comportement non verbal particulier afin d'exprimer son intérêt, son accord ou désaccord, son étonnement, etc. En effet, une conversation n'est pas un flot d'information unilatéral du locuteur vers l'interlocuteur. Il correspond à un échange continu d'information entre les différents participants. Par son comportement non verbal (tel que des hochements de tête, la direction du regard, les froncements des sourcils, etc.), l'interlocuteur manifeste son intérêt et son engagement dans la conversation. Le locuteur utilise ces informations pour estimer si son interlocuteur comprend ou accorde du crédit au discours qu'il lui tient [1, 28]. Ce sont de tels comportements, correspondant à un ensemble de signaux appelés *rétroactions*, qui rendent possible et plus aisée la communication.

Deux types de signaux de rétroaction sont distingués : les *rétractions cognitives*, et les *rétroactions réactives*. Les *rétractions cognitives* correspondent aux signaux émis après une évaluation cognitive de ce que le locuteur a dit et fait. L'interlocuteur peut par exemple montrer qu'il est d'accord ou n'est pas d'accord, qu'il croît ou ne croît pas, qu'il comprend ou ne comprend pas, qu'il accepte ou n'accepte pas, ce que le locuteur lui a dit ou a fait. Les *rétractions réactives* correspondent quant à eux à des signaux émis inconsciemment par l'interlocuteur. Par exemple, le froncement des sourcils peut être

réalisé involontairement lorsque nous avons des difficultés à suivre le discours du locuteur, il s'agit dans ce cas d'une rétroaction réactive [1].

Ces deux types de rétroaction sont nécessaires pour développer un modèle réaliste du comportement d'un interlocuteur. D'un point de vue système, la prise en compte des rétroactions cognitives nécessite de représenter l'information sémantique contenue au niveau du discours ainsi que les états mentaux de l'interlocuteur (l'état de ses connaissances, ses croyances, ses choix, etc.). Pour le moment, nous travaillons plus particulièrement sur les rétroactions réactives. A partir d'une analyse de corpus [6, 10, 18], nous avons pu déterminer un ensemble de règles. Ainsi, les signaux (hochements de tête, réponses verbales) sont généralement synchrones avec les variations de la tonalité du flux verbal du locuteur. Les froncements des sourcils, les mouvements corporels et les changements de direction du regard sont produits lorsque le locuteur montre de l'incertitude dans son discours. De plus, les expressions faciales, la posture et la direction du regard reproduisent généralement ceux émis par le locuteur par un processus de mimésis.

Ces comportements non verbaux, participant largement à la réussite d'une interaction conversationnelle, doivent être intégrés dans les agents virtuels.

9. SCENARIO

Ci-dessous est présenté le scénario d'un dialogue entre deux agents virtuels. Il nous permet d'illustrer différents modèles introduits : du modèle perceptif de l'agent, de ses capacités d'initiation d'une interaction, et de déclenchement des émotions. Un agent virtuel, Lucy, veut s'entretenir avec un autre agent virtuel, Greta.

La perception de l'environnement de Greta, et l'initiation de l'interaction – sections 2 et 3

Dans le champ visuel de Greta se trouve Lucy. Le *module de vision* permet de détecter la présence de Lucy, qui porte un regard soutenu sur Greta.

Un regard soutenu constitue un code social permettant de communiquer une intention d'entrer en interaction avec un autre individu. Parmi tous les agents virtuels présent dans le champ de vision de Greta, le *module de détection de l'intentionnalité* sélectionne l'ensemble des agents voulant entrer interaction avec Greta, en l'occurrence Lucy de part son regard soutenu (un *détecteur de direction de l'attention* nous permettant de calculer la direction du regard de Lucy).

Enfin, le *mécanisme d'attention mutuelle* assure que l'attention est partagée par les deux agents virtuels, ce qui permet à l'agent Greta d'initier l'interaction.

Scénario

Greta : Bonjour Lucy

Lucy : Bonjour Greta

Greta : Comment ça va ?

Lucy : Oh ... j'ai eu une journée difficile: j'ai eu une démo à faire pour une conférence. Mais maintenant c'est fini et ça c'est bien passé !

Greta : Super !

Des interlocuteurs rétroactifs – section 8

Dans ce scénario, Greta et Lucy prennent successivement les tours de parole. Elles sont donc tour à tour locutrice et interlocutrice. A partir d'un tel dialogue et d'un comportement supposé du locuteur de chaque tour de parole, certaines rétroactions réactives sont déclenchées.

L'interlocuteur reproduisant généralement les expressions faciales du locuteur, Lucy émet un sourire lorsque Greta lui souhaite le bonjour en tout début d'interaction. De la même façon, lorsque Lucy énonce avec dégoût les difficultés de sa journée, Greta émet une rétroaction réactive de dégoût.

Si Greta détecte de l'incertitude lorsque Lucy énonce [Oh ...], alors Greta a également la possibilité d'émettre une rétroaction en fronçant les sourcils, marquant qu'elle détecte une incertitude chez la locutrice Lucy.

Finalement, une dernière rétroaction est élicitée lorsque Lucy change de tonalité dans son discours, introduit par son [Mais maintenant c'est fini ...]. Cette variation de tonalité élicite chez Greta un hochement de tête, tandis qu'une expression de sourire chez Lucy pendant cette proposition élicite un sourire chez Greta.

Ce scénario simple montre ainsi combien les rétroactions sont largement présentes dans une interaction, et apportent une dynamique essentielle pour le bon fonctionnement de cette interaction.

Le déclenchement des émotions de Lucy – section 7

Pendant la démonstration de la conférence, Lucy a l'état mental suivant :

- *But* : la démonstration se passe bien
- *Croyance* : il est possible que la démonstration se passe mal.

Cet état mental, et plus particulièrement le fait que Lucy croit qu'il se peut qu'un de ses buts échoue, implique une émotion *peur* chez Lucy.

Après la démonstration, Lucy a l'état mental suivant :

- *But* : la démonstration se passe bien
- *Croyance* : la démonstration c'est bien passée

Cet état mental implique une émotion de *soulagement* chez Lucy car elle n'a plus la croyance qu'un de ses buts peut échouer. De plus, Lucy a une émotion de *joie* car elle a la croyance qu'un de ses buts vient d'être réalisé.

Le déclenchement des émotions empathiques de Greta – section 7

Pendant la conversation, Greta exprime de l'empathie à travers son expression de *joie* pendant son interaction avec Lucy. A partir du modèle d'émotion, Greta connaît les états mentaux déclencheurs d'émotions, et donc ceux générateurs de joie. De plus, au cours du dialogue, Lucy informe Greta de son but ("J'avais une démo à faire pour la conférence"). Avec ces connaissances, Greta déduit que la réalisation du but de Lucy entraîne une émotion de joie. Elle exprime alors, par empathie, de la joie pour Lucy.

Animation des agents

Suivant [36], la colère 'chaude' est associée à des mouvements brusques émis avec une grande force, tandis que la joie est montrée par des mouvements rapide et fluides. L'animation des agents suivent ces caractéristiques.

10. CONCLUSION

Nous avons présenté dans cet article plusieurs modèles permettant de doter un agent de la capacité d'interagir socialement et émotionnellement avec un utilisateur. Aujourd'hui, le comportement rétroactif et les expressions d'émotions de l'utilisateur ne sont pas pris en compte. La prochaine étape vise à coupler les modèles présentés avec des techniques permettant d'identifier les comportements non verbaux de l'utilisateur et d'ainsi prendre en considération ces comportements pour déterminer la réaction de l'agent. Cet article vise ainsi à améliorer les capacités d'interaction entre un utilisateur humain et une interface intégrant un agent virtuel, en proposant des modèles sociaux et émotionnels pour des agents aux nombreuses modalités d'expression.

BIBLIOGRAPHIE

1. Allwood, J., J. Nivre, and E. Ahlson. On the semantics and pragmatics of linguistic feedback. *Semantics*, 9(1), 1993.
2. Austin, J. L., *Quand dire c'est faire*, Editions du Seuil, Paris, 1962
3. Barrier, G., J. Caelen, and B. Meillon. La visibilité des gestes: Paramètres directionnels, intentionnalité du signe et attribution de pertinence. In *Workshop Français sur les Agents Conversationnels Animés* (Grenoble, France), 2005, pp. 113-123.
4. Bates, J. The role of emotion in believable agents. *Communications of the ACM - CACM*, 37(7), 1994, pp. 122-125.
5. Brave, S., Nass, C., and Hutchinson, K. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, (62), 2005, pp. 161-178.

6. Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, M. Stone, Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. *Proceedings of SIGGRAPH'94*, ACM Special Interest Group on Graphics, 1994, pp. 413-420.
7. Chovil, N. Social determinants of facial displays. *Journal of Nonverbal Behavior*, 15(3), Autumn 1991, pp. 141-154.
8. Ech Chafai, N., C. Pelachaud, D. Pelé, G. Breton, Gesture expressivity modulations in an ECA application. In *Intelligent Virtual Agents – IVA06*, (Marina Del Ray, CA.), August 2005, pp. 181-192.
9. Ekman, P. About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, editors, *Human ethology: Claims and limits of a new discipline: contributions to the Colloquium*, pages 169-248. Cambridge University Press, Cambridge, England; New-York, 1979.
10. Friedman, D., and M. Gillies. Teaching virtual characters to use body language. In *Intelligent Virtual Agents, Lecture Notes in Artificial Intelligence*. Springer-Verlag, 2005.
11. Goffman, E. *Behaviour in public places: notes on the social order of gatherings*. The Free Press, New York, 1963.
12. Gullberg, M., and K. Holmqvist. Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition*, (7), 1999, pp. 35-63.
13. Hartmann, B., M. Mancini, and C. Pelachaud. Towards affective agent action: Modelling expressive ECA gestures. In *Proceedings of the IUI Workshop on Affective Interaction* (San Diego, CA), January 2005.
14. Isbister, K. *Better Game Characters by Design: A Psychological Approach*. Elsevier Science and Technology Books, 2006.
15. Itti, L. *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology, 2000.
16. Kendon, A. Movement coordination in social interaction: Some examples described. In S. Weitz, editor, *Nonverbal Communication*. Oxford University Press, 1974.
17. Klein, J., Moon, Y., and Picard, R. This computer responds to user frustration. In *Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania), ACM Press, 1999, pp. 242-243
18. Maatman, R.M., J. Gratch, and S. Marsella. Natural behavior of a listening agent. In *5th International Conference on Interactive Virtual Agents* (Kos, Greece), 2005.
19. Ochs, M., R. Niewiadomski, C. Pelachaud, and D. Sadek. Intelligent expressions of emotions. In Jianhua Tao, Tieniu Tan, and Rosalind W. Picard, editors, *The 1st International Conference on Affective Computing and Intelligent Interaction - ACII05*, (Beijing, China), Springer, 2005, pp. 707-714.
20. Ortony, A., G. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge University Press, 1988.
21. Partala, T. and Surakka, V. The effects of affective interventions in human-computer interaction. *Interacting with computers*, (16), 2004, pp.295-309.
22. Pelachaud, C., V. Carofiglio, B. De Carolis, and F. de Rosis. Embodied contextual agent in information delivering application. In *First International Joint Conference on Autonomous Agents and Multi-Agent Systems – AAMAS* (Bologna, Italy), July 2002.
23. Perrett, D.I. and N.J. Emery. Understanding the intentions of others from visual signals: Neurophysiological evidence. *Current Psychology of Cognition*, (13), 1994, pp. 683-694.
24. Peters, C. *Bottom-Up Visual Attention for Autonomous Virtual Human Animation*. PhD thesis, Department of Computer Science, Trinity College Dublin, 2004.
25. Peters, C. Designing Synthetic Memory Systems for Supporting Autonomous Embodied Agent Behaviour, *Proceedings of the 15th International Symposium on Robot and Human Interactive Communication*, September 2006, pp. 14-19.
26. Peters, C. Direction of attention perception for conversation initiation in virtual environments. In *International Working Conference on Intelligent Virtual Agents* (Kos, Greece), September 2005, pp. 215-228.
27. Peters, C. Evaluating perception of interaction initiation in virtual environments using humanoid agents. In *Proceedings of the 17th European Conference on Artificial Intelligence* (Riva Del Garda, Italy), August 2006, pp. 46-50.
28. Poggi, I. Backchannel: from humans to embodied agents. In *Conversational Informatics for Supporting Social Intelligence and Interaction - Situational and Environmental Information Enforcing Involvement in Conversation workshop in AISB05* (University of Hertfordshire, Hatfield, England), 2005.
29. Poggi, I. Emotions from mind to mind. In *Proceedings of the Workshop on Empathic Agents*. AAMAS, 2004.

30. Poggi, I. Mind markers. In N. Trigo M. Rector, I. Poggi, editor, *Gestures. Meaning and use*. University Fernando Pessoa Press, Oporto, Portugal, 2003.
31. Rao, A.S. and M.P. Georgeff. Modeling rational agents within a BDI architecture. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning - KR* (San Mateo, CA), USA, 1991, pp. 473-484.
32. Reilly, S. Believable Social and Emotional Agents. *Computer science*, University of Carnegie Mellon, 1996.
33. Ruttkay, Z., and C. Pelachaud. Exercises of style for virtual humans. In *Symposium of the AISB02 Convention*, Volume Animating Expressive Characters for Social Interactions, London, 2002.
34. Sadek, D., *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*, Thèse de l'Université de Rennes 1, 1991
35. Scherer, K. Criteria for emotion-antecedent appraisal: A review. In Hamilton, V., Bower, G., and Fridja, N., editors, *Cognitive perspectives on emotion and motivation* (Dordrecht: Kluwer), 1988, pp. 89-126.
36. Wallbott, H.G., Bodily expression of emotion. *European Journal of Social Psychology*, (28) 1998, pp.879-896.