

Vers des Agents Conversationnels Animés dotés d'émotions et d'attitudes sociales

Magalie Ochs, Yu Ding, Nesrine Fourati, Mathieu Chollet, Brian Ravenet,
Florian Pecune, Nadine Glas, Ken Prepin, Chloe Clavel, Catherine Pelachaud

► To cite this version:

Magalie Ochs, Yu Ding, Nesrine Fourati, Mathieu Chollet, Brian Ravenet, et al.. Vers des Agents Conversationnels Animés dotés d'émotions et d'attitudes sociales. Journal d'Interaction Personne-Système (JIPS), AFIHM, 2014, 3 (2), pp. 1-23. <<http://jips.episciences.org>>. <hal-01190028>

HAL Id: hal-01190028

<https://hal.archives-ouvertes.fr/hal-01190028>

Submitted on 1 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vers des Agents Conversationnels Animés dotés d'émotions et d'attitudes sociales

MAGALIE OCHS, YU DING, NESRINE FOURATI, MATHIEU CHOLLET, BRIAN RAVENET, FLORIAN PECUNE, NADINE GLAS, KEN PREPIN, CHLOE CLAVEL ET CATHERINE PELACHAUD

CNRS-LTCI, Télécom Paris Tech, Paris, France

Résumé : Dans cet article, nous proposons une architecture d'un Agent Conversationnel Animé (ACA) socio-affectif. Les différents modèles computationnels sous-jacents à cette architecture, permettant de donner la capacité à un ACA d'exprimer des émotions et des attitudes sociales durant son interaction avec l'utilisateur, sont présentés. A partir de corpus d'individus exprimant des émotions, des modèles permettant de calculer l'expression faciale émotionnelle d'un ACA ainsi que les caractéristiques de ses mouvements du corps ont été définis. Fondés sur une approche centrée sur la perception de l'utilisateur, des modèles permettant de calculer comment un ACA doit adapter son comportement non-verbal suivant l'attitude sociale qu'il souhaite exprimer et suivant le comportement de son interlocuteur ont été construits. Le calcul des émotions et des attitudes sociales à exprimer est réalisé par des modèles cognitifs présentés dans cet article.

Mots clés : Agent Conversationnel Animé ; émotions ; relations sociales, attitudes sociales.

Abstract: In this article, we propose an architecture of a socio-affective Embodied Conversational Agent (ECA). The different computational models of the architecture enable an ECA to express emotions and social attitudes during an interaction with a user. Based on corpora of actors expressing emotions, models have been defined to compute the emotional facial expressions of an ECA and the characteristics of its corporal movements. A user-perceptive approach has been used to design models to define how an ECA should adapt its non-verbal behavior according to the social attitude the ECA wants to display and the behavior of its interlocutor. The emotions and the social attitudes to express are computed by cognitive models presented in this article.

Key words: Embodied Conversational Agent, emotions, social relations, social attitudes.

Adresse des auteurs : Equipe Greta, (greta@telecom-paristech.fr), CNRS LTCI, Télécom ParisTech, 38 rue Dareau, 75014 Paris, France.

Les articles de JIPS sont publiés sous licence Creative Commons Paternité 2.0 Générique.

1. INTRODUCTION

Aujourd'hui, l'essor des nouvelles technologies place les systèmes interactifs dans des rôles de plus en plus variés, typiquement incarnés par des humains, tels que celui de tuteur, de compagnon de jeu, ou encore de conseiller. Lorsque ces systèmes interactifs sont dotés de caractéristiques de la communication humaine, les utilisateurs tendent à interagir naturellement et socialement avec ces derniers comme s'ils étaient humains [Kramer, 2008]. Ce rapprochement de la communication humaine dans les interfaces virtuelles peut être atteint en utilisant des artefacts humanoïdes capables de simuler la richesse de la conversation humaine. Les récents progrès dans les technologies informatiques ont rendu possible la création de ces artefacts, appelés *Agents Conversationnels Animés* (ACAs). Un *agent conversationnel animé* est un personnage virtuel créé par l'ordinateur qui peut converser avec les utilisateurs d'une façon naturelle à travers un langage verbal et non-verbal [Cassell, 2010]. Dans cet article, nous considérons les personnages virtuels humanoïdes mais les ACAs peuvent s'incarner sous d'autres formes plus ou moins réalistes.

Pour concevoir des ACAs capables d'incarner différents rôles efficacement lors d'une interaction avec l'utilisateur, à l'image de l'humain, ces derniers doivent être dotés d'une certaine forme d'*intelligence socio-émotionnelle* [Kihlstorm and Cantor, 2000 ; Salovey *et al.*, 2000]. Cette forme d'intelligence doit leur permettre de gérer la dimension *intrinsèquement* sociale et émotionnelle de l'interaction humain-machine en adoptant un comportement émotionnel et une attitude sociale adaptés au contexte de l'interaction.

Un certain nombre de recherches montrent en effet que les ACAs dotés de capacités socio-émotionnelles permettent d'améliorer significativement l'interaction [Beale and Creed, 2009]. Pour créer des agents avec lesquels l'utilisateur s'engage dans une interaction à plus ou moins long-terme, ces derniers doivent intégrer la subtilité expressive de l'humain. Par exemple, pour que leurs comportements émotionnels paraissent crédibles, les ACAs devraient être capables d'exprimer des émotions aussi bien lorsqu'ils écoutent que *en action*, *i.e.* lorsqu'ils parlent ou lorsqu'ils réalisent une action particulière [Clavel *et al.*, 2009]. Pour ce faire, dans la section « Expressions d'émotions d'un ACA » (Section 3), des modèles d'expressivité émotionnelle faciale (Section 3.1) et corporelle (Section 3.2) que nous avons construit à partir de corpus d'individus exprimant des émotions, sont présentés. De plus, les ACAs devraient pouvoir être capable d'exprimer leur attitude vis-à-vis de l'utilisateur que ce soit de l'appréciation ou de l'hostilité, non pas seulement à travers des mots mais aussi à travers leur comportement non-verbal. Pour répondre à cette problématique, dans la section « Expressions intra-personnelles des attitudes sociales d'un ACA » (Section 4.1), nous présentons un modèle computationnel, que nous avons construit à partir d'une approche centrée sur la perception de l'utilisateur, pour déterminer le comportement non-verbal d'un ACA suivant son attitude sociale. En dehors de l'aspect expressif, les ACAs devraient être capables de raisonner sur le contexte de l'interaction pour déterminer quel comportement socio-émotionnel adopter et montrer dans quelle situation. Pour calculer les émotions que l'ACA devrait exprimer, nous utilisons un modèle existant présenté dans la section « Modèle computationnel d'émotions » (Section 3.3). Concernant les attitudes sociales, un premier modèle conceptuel pour la représentation et le raisonnement est proposé dans la section « Modèle computationnel des attitudes

sociales » (Section 4.3). Impliqué dans une interaction, ce comportement de l'agent devrait considérer le comportement socio-émotionnel exprimé par l'utilisateur et la relation qui se construit entre l'ACA et l'utilisateur. Pour ce faire, le raisonnement de l'ACA intègre un réseau de neurones que nous avons développé, considérant les signaux non-verbaux exprimés par l'interlocuteur (Section 4.2 : « Expressions inter-personnelles des attitudes sociales »).

Dans cet article, nous proposons une architecture intégrant ces différentes capacités cognitives, expressives et interactives d'un ACA socio-affectif ainsi que les différents modèles computationnels sous-jacents permettant de doter un ACA de telles capacités. Plusieurs ACAs intégrant une dimension sociale ou émotionnelle ont d'ores et déjà été proposés (e.g. [Bickmore et Picard, 2005 ; DeRosis *et al.*, 2003, Dias et Paiva, 2005, Krämer, 2008]). L'enjeu des travaux de recherche présentés dans cet article est l'articulation de différents modèles au sein d'une même architecture pour doter l'ACA à la fois d'émotions (Section 3) et d'attitudes sociales (Section 4). Les pistes de recherche envisagées au sein de cette architecture sont ensuite présentées Section 5.

2. ARCHITECTURE D'UN ACA SOCIO-AFFECTIF

L'architecture d'un ACA, illustrée Figure 1, correspond à une extension de l'architecture SAIBA. L'architecture SAIBA [Kopp *et al.*, 2006] est une architecture de référence dans le domaine des ACAs. Cette architecture unifiée est indépendante du domaine d'application, du système d'animation et de l'environnement graphique. Elle permet la génération de comportements multimodaux d'ACA. Dans SAIBA, le *planificateur d'intentions* génère les intentions communicatives de l'ACA (par exemple, l'intention d'exprimer de la joie ou son accord). Le *planificateur de comportements* transforme ces intentions communicatives en un ensemble de signaux (e.g. voix, gestes, expressions faciales). Par exemple, ce module fournit en sortie les signaux que devraient exprimer l'ACA suivant ses émotions, son attitude et son intention (Section 3.2 et 4.1). La liste des signaux pouvant être utilisés pour exprimer chaque intention communicative est définie dans le *lexique de comportements*. Les caractéristiques morphologiques et dynamiques d'un signal pour animer le visage (par exemple un sourire) sont décrites dans le module d'*expressions faciales*. Les expressions multimodales (expressions et synchronisation des signaux sur différentes modalités) sont gérées par le *planificateur de comportements*. Enfin, le module *réalisation de comportements* fournit en sortie les paramètres d'animation pour chacun des signaux. Par exemple, ce module produit en sortie les paramètres d'animation des sourcils et des mouvements de tête suivant l'émotion de l'ACA (Section 3.1).

L'architecture SAIBA est étendue avec différents modules pour ajouter certains raisonnements cognitifs de l'agent et la prise en compte de la dimension interactive. Le module *Etat mental* contient une représentation des paramètres socio-affectifs de l'agent tels que ses émotions, ses relations sociales, ses buts et ses croyances. Il contient de plus des modèles de raisonnement permettant de calculer la dynamique de ses paramètres au cours de l'interaction. Les émotions et les relations sociales sont calculées par des modèles intégrées dans l'*Etat mental* (Section 3.3 et 4.3).

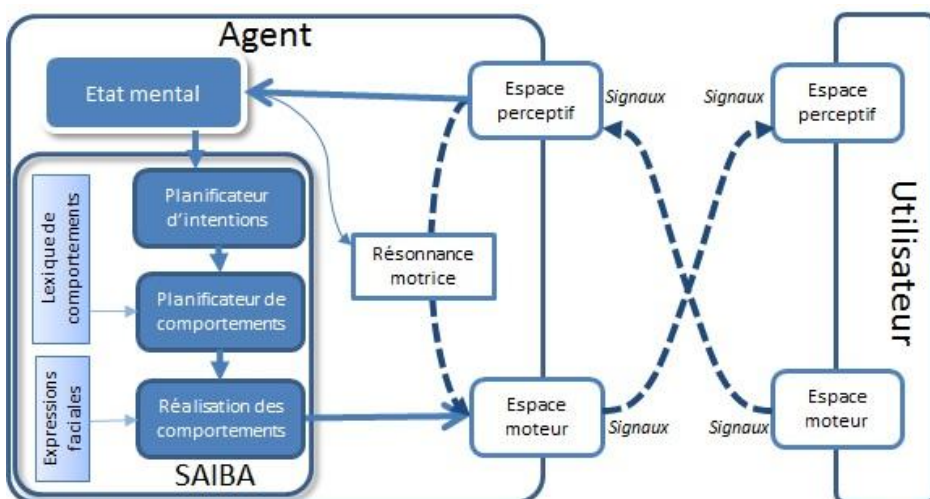


Figure 1 : Architecture d'un ACA socio-affectif

Les modules *Espace perceptif*, *Résonnance motrice* et *Espace moteur* ont été ajoutés pour permettre de simuler l'influence directe (*i.e.* par résonnance motrice, influence illustrée par les flèches en pointillées Figure 1) du comportement socio-affectif de l'utilisateur sur le comportement de l'agent. Ces modules permettent en particulier à l'ACA d'*aligner* son comportement sur celui de son interlocuteur, par exemple par mimétisme, pour traduire une attitude sociale spécifique. Ces modules sont présentés plus en détails dans la section 4.2.

A ce jour, cette architecture a été partiellement implémentée et évaluée. Certains modules sont en cours d'intégration. Dans la suite de l'article, nous décrivons plus précisément les différents modèles sous-jacents à cette architecture.

3. L'EXPRESSION D'ÉMOTIONS D'UN ACA

L'interaction humain-humain est, d'une manière générale, réalisée à travers différentes modalités telles que la parole, les gestes, les sourcils, la tête, le regard, etc. Les variations dans ces modalités permettent d'augmenter la compréhension de la parole. De plus, ces comportements non-verbaux traduisent des sentiments subtils tels que les émotions. Pour donner la capacité à un ACA d'exprimer des émotions, il est donc essentiel que ces derniers soient capables de générer des comportements non-verbaux naturels et crédibles lors d'une conversation avec un humain.

3.1. Expressions faciales affectives

Dans les relations interpersonnelles, les émotions sont principalement communiquées par l'expression faciale [Ekman and Friesen 1975]. Une expression faciale est un changement dans le visage, perceptible visuellement suite à l'activation (volontaire ou involontaire) des muscles composant le visage. Aujourd'hui, de nombreux ACA sont capables d'exprimer des émotions à travers leurs expressions faciales (par exemple, [De Rosis *et al.*, 2013]. Cependant, très peu d'entre eux sont capables d'exprimer une

émotion lorsqu'*ils parlent*. Pour ce faire, les ACAs doivent pouvoir synchroniser la parole et l'expression faciale affective.

Pour calculer automatiquement le comportement non-verbal d'un ACA synchronisé avec la parole, différentes méthodes ont été proposées. Une méthode consiste à utiliser la capture de mouvement. Cependant, cette méthode ne peut être utilisée que pour un ensemble fini de phrases (celles dont les mouvements sont capturés). Une autre méthode repose sur des règles prédéfinies. Par exemple, dans [Cassell *et al.*, 1994], les comportements sont générés selon les relations identifiées entre la prosodie de la parole et les comportements non-verbaux. Cependant, ces règles ne considèrent pas l'influence des émotions sur ces relations. Depuis une dizaine d'années, des modèles statistiques d'apprentissage ont été proposés pour synthétiser automatiquement divers comportements à partir de la parole synchronisée. Par exemple, dans [Busso *et al.*, 2007], les auteurs ont proposé d'apprendre séparément quatre modèles utilisés pour synthétiser les expressions de tristesse, de satisfaction ou de colère de manière synchronisée avec la parole. Le nombre d'émotions reste cependant très limité. De plus, les modèles proposés ne permettent généralement pas de considérer l'expression de plusieurs émotions simultanément [Ekman and Friesen, 1975].

Afin de donner la capacité à un ACA d'exprimer plusieurs émotions, nous avons développé un modèle statistique par apprentissage pour synthétiser automatiquement les mouvements des sourcils et de la tête à partir du signal de la parole. Ce modèle repose sur un modèle de Markov caché entièrement paramétré (*fully parameterized* HMM - FPHMM). Le modèle est décrit plus en détails dans [Ding *et al.*, 2013]. Ce modèle a été construit par apprentissage à partir des données d'un corpus audiovisuel contenant des vidéos d'acteurs équipés de microphones et de capteurs de mouvements du visage [Fanelli *et al.*, 2010]. Les acteurs avaient comme instructions de prononcer différentes phrases avec diverses émotions. Trois angles de rotation de tête et huit points d'animation faciale (Facial animation Points – FAPs, [Pandzic and Forcheimer, 2002]) ont été extraits comme caractéristiques du mouvement avec une fréquence de 25Hz. Les caractéristiques de la parole, extraites avec les mêmes fréquences d'échantillonnage de 25Hz sont la fréquence fondamentale et l'énergie du signal de parole. Pour la synthèse des expressions faciales affectives d'un ACA, la séquence de signaux de parole est utilisée en entrée du modèle FPHMM qui produit en sortie les mouvements de sourcils et de tête associés.

Pour évaluer ce modèle, des tests perceptifs ont été réalisés. Nous avons comparé la perception des expressions d'émotions d'un ACA résultant du modèle FPHMM produit et celles résultantes directement de la capture de mouvement. Douze émotions ont été considérées (colère, tristesse, peur, mépris, nervosité, dégoût, frustration, stress, excitation, confiance, surprise et joie)¹ et exprimées durant la prononciation de différentes phrases. Quarante utilisateurs ont indiqué leur perception des émotions ainsi exprimées. Les analyses statistiques montrent que le modèle proposé permet d'améliorer significativement la perception des émotions de peur, mépris, nervosité, stress, excitation surprise et joie en comparaison avec les expressions d'ACA générées à partir

¹ Ces émotions correspondent à celles considérées dans le corpus [Pandzic and Forcheimer 2002].

des données de capture de mouvements directement. Les résultats de l'évaluation sont détaillés dans [Ding *et al.*, 2013].

Ce modèle s'intègre dans le module de *réalisation de comportements* de l'architecture d'un ACA (Figure 1). Il permet de calculer les paramètres d'animation pour les sourcils et les mouvements de tête à partir du signal de parole de l'ACA issu de la sortie de la synthèse vocale.

3.2 Expressions corporelles affectives

Les émotions sont exprimées non seulement à travers l'expression faciale mais aussi à travers l'ensemble du corps. Dans un ACA, l'expression des émotions à travers ces différentes modalités permet d'améliorer la perception des émotions par l'utilisateur ainsi que la crédibilité de l'agent [Clavel *et al.*, 2009].

La synthèse des comportements émotionnels chez un ACA implique deux principales problématiques : la définition d'un système de codage des mouvements corporels expressifs (*i.e.* comment caractériser les mouvements corporels émotionnels) et la correspondance entre ce système de codage et les états affectifs (*i.e.* comment l'expression de l'émotion est décrite à travers ces caractéristiques de mouvements). Certains travaux utilisent un système de codage corporel basé sur le type de mouvement [Coulson, 2004]. Les modèles de comportement émotionnel sont donc définis à travers un ensemble discret de postures et/ou d'actions, comme le croisement des bras ou la flexion de la tête pour décrire l'expression de l'émotion. Dael *et al.* [Dael *et al.*, 2012] ont récemment proposé un nouveau système de codage des mouvements corporels basé sur le type de mouvement. Ce système décrit le mouvement du corps en termes d'unités d'action et de posture en tenant compte des différents niveaux des dimensions anatomique et directionnel. Les modèles de comportements émotionnels corporels basés sur le type de mouvement sont donc basés sur les travaux précédents portant sur l'expression d'émotion à travers les mouvements du corps [Wallbott, 1998]. Une autre méthode consiste à décrire l'expression d'émotion par la manière avec laquelle le mouvement est effectué. Ces méthodes se basent généralement sur une base de données de capture de mouvements émotionnels naturels ou actés dans le but d'en extraire des informations utiles pour la reproduction du comportement expressif et émotionnel chez un ACA. Dans cette approche, les mouvements corporels sont décrits par des paramètres définissant la qualité du mouvement, *i.e.* la manière avec laquelle le mouvement a été effectué. Dans les travaux existants, les paramètres qui contrôlent l'expressivité du mouvement se limitent généralement à un ensemble de paramètres d'expressivité calculés à partir de données de capture de mouvement pour un type de geste particulier [Roether *et al.*, 2009]. Cependant, ces paramètres sont souvent associés à un nombre limité de types de geste et ne peuvent pas être étendus à l'ensemble des gestes possibles d'un ACA. D'autres travaux proposent de contrôler l'expressivité des mouvements de l'ACA (par exemple leur amplitude) à travers des fonctions mathématiques reposant sur quelques lois invariantes du mouvement humain [Huang and Pelachaud 2012]. Ce type d'approche permet d'appliquer les mêmes paramètres pour plusieurs types de gestes. Cependant, l'ensemble des paramètres reste limité agissant principalement sur les gestes des mains.

Nous proposons un modèle de comportement émotionnel reposant sur l'analyse d'un corpus de capture de mouvements. Inspiré du système de codage défini dans [Dael *et al.*, 2012], la définition des paramètres que nous avons adoptée pour contrôler le comportement émotionnel corporel d'un ACA repose sur plusieurs niveaux de description, à la fois pour décrire un mouvement corporel et pour décrire les caractéristiques de ce mouvement. Trois principales dimensions sont considérées : une dimension *anatomique* définissant les articulations en questions (*e.g.* la tête, les bras), une dimension *directionnelle* définissant la direction du mouvement (*e.g.* latérale, verticale) et une dimension *statique/dynamique* décrivant l'aspect postural ou dynamique du geste (*e.g.* la posture statique, le changement de posture au cours du temps). Une première étude que nous avons réalisée [Fourati and Pelachaud, 2012], utilisant une partie de l'ensemble des paramètres pour analyser des marches expressives, a montré qu'on peut discriminer différents styles de marche à travers quelques caractéristiques décrivant la posture et le changement de posture au cours du temps. La marche de tristesse est par exemple caractérisée principalement par le fléchissement des coudes, l'absence d'une variation significative de la distance entre les mains, le fléchissement de la tête et l'effondrement du torse.

Nous poursuivons actuellement nos travaux à partir d'une base de données de captures de mouvements de comportements émotionnels actés que nous avons collectés. Onze acteurs ont exprimé 3 fois 8 états affectifs différents (tristesse, colère, joie, honte, fierté, anxiété, peur panique, et neutralité) en réalisant 7 actions différentes. L'objectif est d'étendre l'étude décrite ci-dessus sur la marche à d'autres mouvements tels que frapper à la porte ou déplacer des objets sur une table. Afin de vérifier que les comportements actés traduisent en effet les émotions visées et ainsi valider la base de données construite d'un point de vue perceptif, nous avons réalisé une étude perceptive, en utilisant la plateforme de crowdsourcing mechanical turk, dans laquelle les participants indiquent leur perception de l'émotion exprimée par l'acteur. Les vidéos utilisées pour cette étude perceptive représentent le mouvement corporel effectué par un acteur et reproduit sur un acteur virtuel. Les résultats de plus de 1000 participants sont actuellement en cours d'analyse.

Les comportements émotionnels corporels modélisés à partir de la base de données seront intégrés dans le module *Planificateur de comportements* de l'architecture de l'ACA illustrée Figure 1. Ainsi, les paramètres des mouvements de l'ACA pour réaliser une action pourront être calculés suivant l'émotion de ce dernier.

3.3 Modèle computationnel d'émotions

Outre la dimension expressive, un ACA doit être capable de déterminer quelles émotions exprimer durant une interaction avec l'utilisateur. Les recherches ont en effet montré qu'une même émotion exprimée par un ACA dans une situation inappropriée pouvait avoir un effet néfaste sur l'interaction [Beale and Creed, 2009]. Fondés pour la plupart sur les théories de l'évaluation cognitive des émotions [Scherer, 2005], un certain nombre de modèles computationnels a d'ores et déjà été proposé pour calculer et mettre à jour automatiquement les émotions de l'ACA suivant le déroulement de l'interaction. Dans l'architecture de l'ACA socio-affectif présentée Figure 1, le modèle computationnel FATiMA [Dias and Paiva, 2005] a été intégré dans le module *Etat mental*. Le modèle FATiMA calcule les émotions de l'ACA suivant ses croyances sur

l'état de ses buts et ses besoins. Ce modèle présente de plus l'avantage d'offrir une architecture modulaire permettant de simuler des phénomènes tels que la culture ou l'empathie dans le processus de déclenchement des émotions de l'agent.

En dehors de la dimension émotionnelle, un ACA étant engagé dans une interaction sociale avec l'utilisateur, doit être capable d'exprimer différentes attitudes vis-à-vis de ce dernier. Dans la section suivante, nous présentons des modèles permettant de doter l'ACA de telles capacités.

4. LES ATTITUDES SOCIALES D'UN ACA

Durant une interaction, suivant la relation entre deux personnes et le contexte (*e.g.* leurs rôles et leurs intentions), celles-ci ne vont pas exprimer la même attitude sociale. Scherer [Scherer, 2005] définit l'attitude sociale comme « *an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons* ». Les attitudes sociales sont souvent représentées sur deux dimensions, la *dominance* et l'*appréciation*. Cette représentation, proposée par Argyle [Argyle, 1988], est généralement illustrée sous forme d'un disque appelé le circomplexe interpersonnel (*Interpersonal Circumplex*) où une attitude sociale correspond à un point dans cet espace à deux dimensions.

Lors d'une interaction, de nombreux éléments du comportement non-verbal reflètent l'attitude sociale d'un individu [Burgoon *et al.*, 1984]. Lorsqu'un message verbal est transmis, les caractéristiques des signaux non-verbaux qui l'accompagnent peuvent traduire l'attitude sociale du locuteur. Par exemple, une attitude dominante se traduit par des gestes amples [Carney *et al.*, 2005]. De plus, la dynamique *intra-personnelle*, *i.e.* la temporalité des signaux exprimés par un même individu sur un tour de parole ou sur l'ensemble de l'interaction, caractérise une certaine attitude. Par exemple, de fréquents sourires peuvent correspondre à une attitude d'appréciation. Enfin, dans une interaction, l'attitude est exprimée à travers la dynamique *inter-personnelle*, *i.e.* la temporalité des signaux d'un individu *par rapport* à celui de son interlocuteur. Par exemple, un sourire exprimé en réponse à un sourire de l'interlocuteur peut exprimer de l'affiliation [Louwerse *et al.*, 2012].

Dans cette section, nous présentons tout d'abord des modèles permettant de déterminer comment un ACA peut exprimer une attitude sociale à travers le comportement non-verbal qui accompagne un message verbal et à travers la dynamique intra- et inter-personnelle de son comportement non-verbal. Nous présentons ensuite un modèle permettant de déterminer quelle attitude sociale l'ACA devrait exprimer durant une interaction.

4.1. Expressions intra-personnelles des attitudes sociales d'un ACA

Différents travaux de recherche ont été menés afin de doter les ACAs de la capacité d'exprimer des attitudes sociales à travers leurs comportements non-verbaux. Par exemple, l'ACA *Laura* [Bickmore et Picard, 2005] montre son appréciation vis-à-vis de l'utilisateur par une modification de certains paramètres comme l'augmentation du nombre de hochements de tête et de gestes. Dans le projet *Demeanour* [Gillies *et al.*, 2004], les agents sont capables de changer certains paramètres de leur comportement non-verbal (la posture, les regards, et l'activation des gestes) suivant leurs attitudes représentées à l'aide de circomplexe interpersonnel.

Afin de créer un ACA capable de montrer différentes attitudes sociales à travers différentes modalités, nous avons utilisé une méthode fondée sur la perception de l'utilisateur. Nous avons collecté un corpus d'ACAs (féminin et masculin) exprimant différentes attitudes sociales (dominant, soumis, amical et hostile) en demandant à l'utilisateur de directement créer le comportement non-verbal associé. Nous avons, pour ce faire, développé et mis en ligne une application qui permet à un utilisateur d'aisément configurer les paramètres d'animation du comportement non-verbal d'un ACA pour une attitude et une intention communicative données (Figure 2).

Nous avons ainsi collecté 925 descriptions de comportements non-verbaux (170 participants). Chaque comportement du corpus est décrit par une attitude (dominant, soumis, amical et hostile), une intention communicative (informer, demander, accepter, refuser, approuver, désapprouver, que nous utilisons pour illustrer la situation), et un ensemble de valeurs pour les paramètres suivant du comportement non-verbal : la présence d'évitement de regard, la modalité utilisée (mouvement de tête, des bras, les deux ou aucun mouvement), l'expression faciale (positive, négative ou neutre), l'amplitude et la force des gestes (faible, moyenne ou forte) et l'orientation de la tête (droite, vers le bas, vers le haut ou sur le côté). Ces paramètres ont été choisis au regard de la littérature en Sciences Humaines et Sociales montrant une influence des attitudes sur ces modalités. Chaque participant avait comme tâche de configurer le comportement de l'ACA dans 6 situations différentes (une situation correspondant à une intention communicative et une attitude sociale).

A partir des données collectées, pour pouvoir calculer automatiquement les paramètres du comportement d'un ACA suivant l'attitude sociale qu'il souhaite exprimer, nous avons développé un modèle fondé sur un réseau Bayésien. Dans ce modèle, la structure des arcs entre les variables d'entrée (l'attitude et l'intention) et les variables de sorties (les paramètres du comportement non-verbal), est construite suivant les corrélations entre celles-ci, calculées statistiquement à partir des données du corpus (les résultats statistiques sont décrits plus précisément dans [Ravenet *et al.*, 2013]). Les données du corpus sont de plus utilisées pour calculer les paramètres du modèle, *i.e.* les probabilités conditionnelles des valeurs des variables.

Ce modèle permet à un ACA qui doit exprimer une intention communicative avec une certaine attitude sociale, de choisir le comportement non-verbal correspondant [Ravenet *et al.*, 2013]. Il s'intègre dans l'architecture de l'ACA (Figure 1) dans le module *Planificateur de comportements* pour la sélection du choix des signaux.

Ce modèle a été récemment été couplé avec un moteur de dialogue intégré dans le module *planificateur d'intentions* (Figure 1). Ce moteur choisit les phrases que doit dire l'agent suivant l'attitude sociale que l'agent souhaite exprimer [Callejas *et al.*, 20014]. Ce couplage a été illustré sur une application de formation aux entretiens d'embauche dans laquelle l'agent joue le rôle d'un recruteur virtuel. Le comportement verbal (phrases) et non-verbal (regard, mouvements de tête, gestes et expressions faciales) varient suivant l'attitude sociale du recruteur virtuel qui peut être définie soit comme amicale soit comme hostile. Une évaluation perceptive auprès d'utilisateurs nous a permis de valider que les attitudes étaient correctement perçues à travers le

comportement multimodal généré par les modèles (pour plus de détails sur l'évaluation, voir [Callejas *et al.*, 2014]).

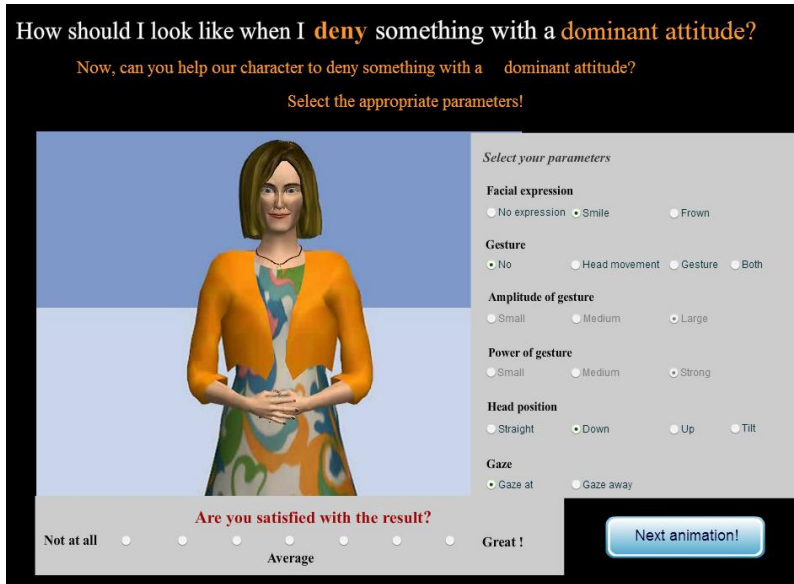


Figure 2 : Copie d'écran de l'interface de collecte de comportements non-verbaux d'ACA

L'attitude sociale se traduit non seulement à travers l'expression d'une intention communicative à un instant t mais aussi à travers la dynamique des signaux dans un tour de dialogue ou tout au long de l'interaction, *i.e.* la *dynamique intra-personnelle du comportement non-verbal*. Par exemple, de fréquents sourires suivis par une absence totale de sourire peuvent traduire une attitude d'hostilité.

Afin de modéliser la dynamique *intra-personnelle* des comportements non-verbaux qu'un ACA devrait adopter pour exprimer une attitude sociale particulière, nous proposons d'analyser des corpus d'interactions interpersonnelles. Cette méthode d'analyse de corpus a d'ores et déjà été explorée mais principalement avec le but de développer des méthodes de détection automatique des attitudes sociales des individus. Une revue de ces méthodes existantes de détection automatique d'informations sociales dans des interactions est disponible dans [Gatica-Perez, 2009]. La plupart des approches existantes se sont concentrées sur l'utilisation de mesures audio-visuelles sur l'ensemble d'une interaction pour classifier la dominance des personnes dans un groupe. Si ces approches se révèlent efficaces dans leur tâche de classification de la dominance d'une personne dans la globalité d'une interaction, elles ne permettent pas de détecter les variations locales d'attitude par rapport à la dynamique des comportements non-verbaux. De plus, il n'y a pas eu, à notre connaissance, de tentative d'intégration de ces résultats dans un modèle de comportement d'ACA.

Afin d'identifier la dynamique des comportements non-verbaux qu'un ACA devrait adopter pour exprimer une attitude durant une interaction, nous avons analysé un corpus multimodal humain-humain composé de 3 vidéos d'entretiens d'embauches, durant en moyenne 20 minutes. Ces vidéos ont été annotées à deux niveaux : (1) au niveau du comportement non-verbal des interactants en s'appuyant sur le schéma de codage MUMIN [Allwood *et al.*, 2007] et (2) au niveau de l'attitude sociale du recruteur. L'attitude sociale a été annotée par 12 personnes en utilisant l'outil *GTrace* [Cowie *et al.*, 2011] sur les deux dimensions de l'attitude sociale décrites dans la section précédente : la *dominance* et l'*appréciation*. Pour identifier des séquences de comportements non-verbaux menant à des variations locales d'attitudes, nous avons utilisé une méthode de *fouille de données séquentielles*. Une première étape de préparation des données consiste à détecter les instants où les annotateurs ont perçu une variation d'attitude. Ces variations sont regroupées en plusieurs classes en fonction de leur amplitude par le biais d'une méthode de partitionnement (*K-means clustering*). Nous obtenons ainsi des ensembles de séquences de comportements non-verbaux caractérisant variation d'attitude (augmentation ou diminution, forte ou faible d'appréciation ou de dominance). Nous appliquons à ces ensembles l'algorithme *GSP* (*Generalized Sequential Patterns*) [Srikant et Agrawal, 1996] et obtenons ainsi les séquences de comportements non-verbaux les plus fréquentes avant chaque classe de variation d'attitude. Chaque séquence est caractérisée par son *support* (*i.e.* la fréquence à laquelle cette séquence est présente pour une variation d'attitude) et sa *confiance* (*i.e.* le ratio d'occurrences de cette séquence pour cette variation d'attitude par rapport aux occurrences dans l'ensemble du corpus). Les résultats sont présentés avec plus de détails dans [Chollet *et al.*, 2014]. Nous travaillons actuellement sur un modèle graphique, construit à partir de ces séquences, permettant de calculer automatiquement le comportement non-verbal d'un ACA suivant la variation d'attitude sociale visée. Ce modèle sera intégré dans le module de *Réalisation de comportements* (Figure 1) pour modifier le comportement de l'agent suivant les attitudes sociales de l'ACA.

Dans la section suivante, nous présentons un modèle permettant de modéliser la dynamique *inter-personnelle* du comportement non-verbal d'un ACA pour exprimer différentes attitudes sociales durant une interaction.

4.2. L'expressions inter-personnelles des attitudes sociales de l'ACA

Durant une interaction, comme le soulignent Chindamo *et al.* [Chindamo *et al.*, 2012] "*stances are constructed across turns rather than being the product of a single turn*". En effet, lorsque les attitudes de chaque partenaire de l'interaction sont mises en présence, une attitude de la dyade émerge de l'alignement diachronique entre les interactants. Par exemple, l'alignement du comportement non-verbal d'un locuteur et de son interlocuteur (*e.g.* l'expression de sourires synchronisés) peut traduire une attitude d'intérêt et de plaisir mutuel [Louwerse *et al.*, 2012].

Plusieurs agents virtuels capables d'aligner leur comportement non-verbal sur celui de l'utilisateur ont été développés. Par exemple, dans [Bevacqua *et al.*, 2010], un modèle est proposé pour permettre à un ACA qui écoute l'utilisateur d'automatiquement exprimer des *backchannels*, *i.e.* des signaux non-verbaux, comme des hochements de tête ou des sourires, alignés sur ceux de l'utilisateur. Ces travaux montrent que ces

signaux de *backchannels* facilitent en effet l'allocation de l'utilisateur et améliorent sa perception de l'agent. Ces types de modèle proposent une adaptation du comportement non-verbal de l'agent *en réaction* à celui de l'utilisateur, et non *en interaction* avec ce dernier. Le comportement de l'agent s'aligne en effet sur celui de l'utilisateur par une expression de signaux de mêmes types que ceux détectés de l'utilisateur. Afin de permettre à un ACA d'aligner dynamiquement son comportement non-verbal *en interaction* avec celui de l'utilisateur, nous avons développé un modèle permettant de modifier en temps réel les caractéristiques morphologiques et dynamiques des signaux non-verbaux de l'agent suivant les réactions de son interlocuteur. Nous illustrons ce modèle sur le signal du *sourire*. Lors d'une interaction, si l'ACA sourit à l'utilisateur, et que ce dernier lui sourit en réponse, l'ACA devrait *renforcer* son sourire pour lui montrer un intérêt et un plaisir partagés [Louwerse *et al.*, 2012]. En d'autres termes, l'ACA devrait être capable de moduler dynamiquement son sourire suivant sa perception directe du sourire de son interlocuteur. Pour modéliser le sourire d'un ACA, nous avons fondé nos travaux sur les caractéristiques morphologiques et dynamiques des sourires d'un ACA identifiées dans des travaux précédents [Ochs *et al.*, 2012]. Un sourire est ainsi défini par sa taille (*i.e.* l'amplitude des muscles zygomatiques), le soulèvement des pommettes, l'ouverture de la bouche et la tension des lèvres. Ainsi, un ACA peut renforcer son sourire en amplifiant la durée et l'amplitude de ces caractéristiques.

Pour permettre à un ACA de modifier dynamiquement et en temps réel son sourire durant une interaction, nous avons développé un modèle computationnel fondé sur un réseau de neurones. Le réseau de neurones permet de définir une architecture neurone par neurone contrôlée dynamiquement. Comme illustré Figure 3, les neurones représentent les caractéristiques du sourire. La couche perceptive (*i.e.* *Espace perceptif*) correspond aux caractéristiques perçues alors que la couche générative (*i.e.* *Espace moteur*) correspond aux caractéristiques exprimées par l'agent. Les liens entre ces deux couches représentent l'influence des caractéristiques perçues sur celles exprimées (ce modèle est décrit plus en détails dans [Prepin *et al.*, 2013]).

L'alignement dynamique des sourires de l'ACA, *i.e.* les liens entre la couche perceptive et générative (le module de *Résonnance motrice*), vont dépendre de l'*attitude sociale* que veut exprimer l'agent. Cette influence est représentée dans l'architecture de l'ACA (Figure 1) par le lien entre le module *Etat mental*, et le module *Résonnance motrice*.

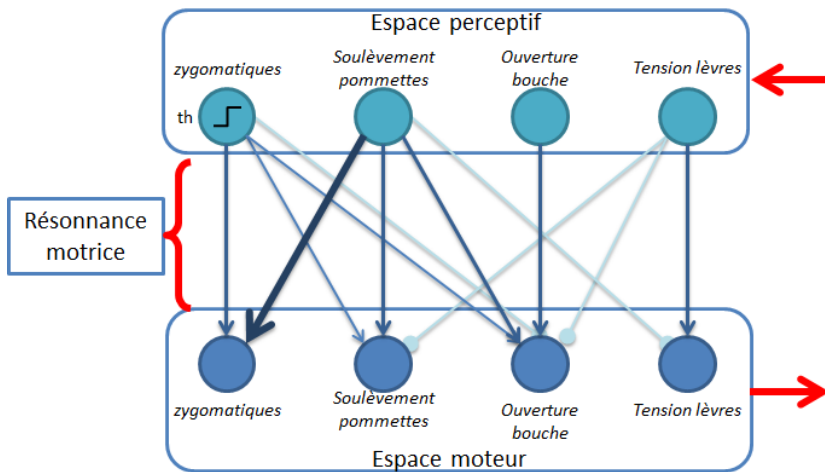


Figure 3 : Modèle de résonance motrice
 th représentant le seuil d'activation du neurone (threshold)

Ainsi, si l'agent décide de se montrer amical, l'influence sera positive : les sourires de l'ACA seront renforcés par la perception des sourires de l'utilisateur. En d'autres termes, lorsque l'ACA perçoit un sourire de l'utilisateur, suivant les caractéristiques du sourire perçu, l'amplitude et la durée des caractéristiques de son sourire seront augmentées. Dans le cas contraire, les sourires de l'ACA ne seront pas influencés par ceux de l'utilisateur.

Cette modélisation du lien entre les attitudes sociales de l'ACA et l'alignement de son comportement non-verbal est fondée sur les résultats d'une récente étude perceptive que nous avons réalisée [Ochs *et al.*, 2013]. Dans cette étude, nous avons simulé une interaction entre deux ACAs (dotés du modèle de résonance motrice décrit ci-dessus) et comparé la perception que les utilisateurs ont de l'un des deux ACA lorsque celui renforce ou non son sourire en réponse à celui de son interlocuteur ACA. Les résultats confirment que l'alignement de ce comportement non-verbal a un effet significatif sur la perception de son attitude, le renforcement du sourire améliorant la perception de l'attitude. Nous travaillons actuellement sur l'intégration du système permettant la détection automatique temps réel des caractéristiques des signaux exprimées par l'utilisateur (*e.g.* les caractéristiques du sourire). La prochaine étape vise à valider, à travers des tests perceptifs, la capacité d'un ACA à exprimer différentes attitudes par l'alignement de son comportement non-verbal (incluant le sourire mais aussi d'autres signaux tels que les hochements de tête), lorsque celui-ci est impliqué dans une interaction directe avec l'utilisateur.

4.3. Modèle computationnel des attitudes sociales

Afin de déterminer l'attitude sociale que devrait exprimer un ACA durant une interaction avec un utilisateur, un modèle computationnel de la dynamique de ces attitudes doit être défini. Comme l'indique Scherer [Scherer 2005], les attitudes sociales sont influencées par différents paramètres qui peuvent être de nature contextuelle, stratégique ou interpersonnelle. Pour qu'un ACA exprime des attitudes sociales

appropriées à la situation d'interaction, les mécanismes inhérents aux relations sociales et leur évolution au cours des interactions doivent être modélisés.

Dans le domaine des ACAs, la plupart des modèles computationnels se focalisent sur les relations sociales. Même si ces dernières sont essentielles pour déterminer l'attitude sociale d'un ACA, d'autres éléments du contexte de l'interaction peuvent amener l'ACA à exprimer des attitudes différentes de ces relations « ressenties ». De plus, les modèles existants des relations sociales utilisent principalement les émotions pour calculer la dynamique des relations sociales, par exemple [Prendinger and Descamps and Ishizuka 2002]. Comme proposé dans [Dias and Paiva 2013], la dynamique peut être aussi modélisée à partir de théories de psychologie cognitive telle que la théorie de la congruité [Osgood and Tannenbaum 1955] mettant en évidence comment les relations sociales entre les individus s'articulent.

Dans notre modèle, pour calculer l'attitude sociale d'un ACA, nous avons proposé une représentation des différentes relations sociales de l'agent et de l'utilisateur. Comme les attitudes sociales, les relations sociales peuvent être définies selon les deux axes de dominance et d'appréciation, via le circomplexe interpersonnel (décrit précédemment). Dans notre modèle, nous utilisons deux représentations du circomplexe, définissant deux types de relations bien distinctes : pour les relations de l'agent envers l'utilisateur, nous représentons la relation par un vecteur démarant à l'origine du circomplexe [Gurtman 2009]. Les relations de l'utilisateur envers l'agent - qui correspondent aux croyances de l'agent - sont modélisées par un sous-ensemble du circomplexe prenant ainsi en compte la part d'incertitude relative à ces croyances.

A partir de la définition de Scherer présentée ci-dessus, nous proposons un modèle computationnel dans lequel nous représentons les **relations de l'agent envers l'utilisateur** par trois vecteurs différents sur ce même circomplexe interpersonnel : (1) la *relation sociale ressentie* ; (2) la *relation sociale idéale* et (3) l'*attitude sociale*.

La *relation ressentie* représente le sentiment de l'ACA envers l'utilisateur (e.g. l'agent n'aime pas l'utilisateur). Afin de calculer cette relation sociale ressentie, nous avons défini un modèle formel de la dominance, de l'appréciation et de la familiarité. Ces trois éléments sont formellement représentés par des combinaisons d'attitudes mentales particulières de croyances et de buts de l'agent. Par exemple, la dominance de l'agent dépend de ses croyances sur sa capacité à réaliser ou menacer des buts ou des préconditions des buts de l'utilisateur. Ainsi, suivant l'état mental défini de l'agent, sa relation sociale ressentie peut être calculée. Le modèle est décrit plus en détails dans [Pecune *et al.*, 2013]. La *relation idéale* indique la relation que l'agent souhaiterait idéalement exprimer dans une situation particulière (e.g. un enseignant peut vouloir montrer de la dominance envers ses élèves). La taille et la direction de ce vecteur dépendent principalement du rôle de l'agent et de la manière dont l'agent souhaite exprimer ce rôle. Enfin, l'*attitude sociale* correspond à la relation exprimée par l'agent. Cette attitude est calculée en fonction de la *relation ressentie* et de la *relation idéale*. De plus, l'humeur de l'agent influence la valeur d'appréciation (e.g. un agent de mauvaise humeur aura tendance à se montrer plus hostile) [Isen *et al.*, 1992]. La valeur de dominance est aussi influencée par l'état des ressources de l'agent et de l'utilisateur [Raven, 2008].

La dynamique de l'attitude sociale de l'agent, *i.e.* la relation sociale que l'agent va exprimer durant l'interaction, dépend fortement des croyances de l'ACA sur les relations sociales de l'utilisateur envers l'agent. Le comportement de l'agent va en effet être différent suivant que l'utilisateur montre par exemple de l'appréciation ou de l'hostilité envers lui. Dans notre modèle, les **relations de l'utilisateur envers l'agent** sont représentées par les 2 relations suivantes : (1) la *relation désirée* est la relation que l'agent désire voir exprimée chez l'utilisateur [Schutz, 1958]. Cette relation est calculée à partir de la *relation idéale*, en accord avec la théorie de la complémentarité interpersonnelle [Sadler and Ethier and Woody 2011]. D'après cette théorie, la *relation désirée* est une symétrie de la *relation idéale* selon l'axe d'appréciation : si l'agent veut montrer de la dominance et de l'appréciation, il s'attend à ce que l'utilisateur exprime de la soumission et de l'appréciation ; (2) la *relation perçue* décrit la relation que l'agent pense que l'utilisateur ressent envers lui (*e.g.* l'agent pense que l'utilisateur l'apprécie).

Afin de modéliser la dynamique de ces relations et ainsi calculer l'attitude sociale que devrait exprimer l'ACA, notre modèle s'inspire de la *Théorie de la congruité* [Osgood and Tannenbaum 1955]. Nous nous basons donc sur un schéma triangulaire entre un agent A, un utilisateur U et une situation S. Les arêtes du triangle décrivent la relation de A envers l'utilisateur U (*relation ressentie*), la relation que A souhaite exprimer dans la situation S (*relation idéale*) et l'estimation de A sur la relation que U devrait exprimer dans la situation S (*relation désirée*). Par exemple, l'agent apprécie l'utilisateur, veut montrer de la dominance et s'attend donc à ce que l'utilisateur adopte une attitude soumise. L'interaction est dite *congruente* quand la *relation perçue* de U dans la situation S est en adéquation avec la *relation désirée* (Figure 3). Si l'on reprend l'exemple précédent, il y a congruence si l'utilisateur montre de la soumission. Dans ce cas, la *relation ressentie* de l'ACA va tendre vers la *relation idéale*. En revanche, l'interaction est dite *non-congruente* si l'agent perçoit de la dominance chez l'utilisateur. L'agent peut alors adopter trois différentes stratégies : (a) il peut changer sa *relation ressentie* envers l'utilisateur, (b) il peut changer sa propre *relation idéale* et (c) il peut essayer de changer la *relation exprimée* par l'utilisateur. Finalement, l'attitude sociale exprimée par l'ACA va dépendre des stratégies adoptées par l'agent.

Notre prochaine étape vise à modéliser la prise de décision quant à la stratégie à adopter par l'agent dans le cas d'une non-congruence. De plus, un modèle d'actions pour mettre en place la stratégie (c) décrite ci-dessus doit être mis en place pour déterminer comment l'ACA peut changer la relation exprimée par l'utilisateur (et donc la *relation perçue* par l'agent).

Ce modèle fait partie intégrante du module *Etat mental* de l'architecture de l'ACA (Figure 1). Il permet de fournir en sortie l'attitude sociale et les intentions communicatives que l'ACA devrait exprimer étant donnée la situation d'interaction.

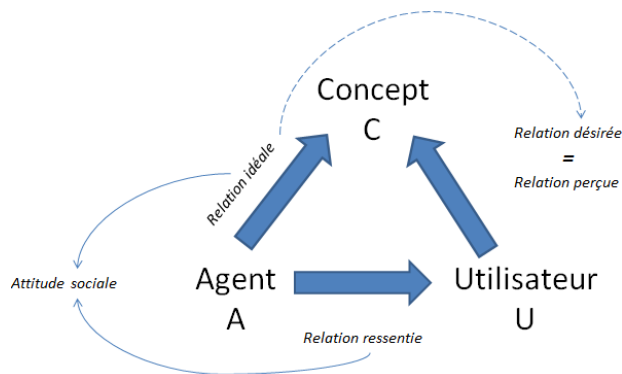


Figure 3 : Exemple d'interaction congruente

5. PERSPECTIVES DE RECHERCHE

Dans cet article, nous avons présenté une architecture incluant les différents modèles permettant de donner la capacité à un ACA d'exprimer des émotions et des attitudes sociales à travers son comportement. La prochaine étape de nos travaux se concentrera sur l'évaluation d'un tel ACA socio-affectif. Tout d'abord, chaque modèle devra être évalué séparément pour vérifier que le comportement exprimé par l'ACA traduit en effet les émotions et attitudes sociales visées. Comme présenté dans les sections précédentes, ce type d'évaluation peut être réalisé à travers des tests perceptifs auprès d'utilisateurs [Ding *et al.*, 2013, Fourati and Pelachaud, 2012, Ochs *et al.*, 2012]. Enfin, l'architecture dans son ensemble, *i.e.* avec l'ensemble des modules présentés, doit être évaluée *en interaction*, *i.e.* lorsque l'utilisateur est impliqué dans une interaction directe avec l'ACA [McRorie *et al.*, 2012]. Il s'agira alors de vérifier non seulement la perception par l'utilisateur des comportements socio-affectifs de l'ACA [Ochs *et al.*, 2012, 39] mais aussi l'effet de ces comportements sur l'engagement de l'utilisateur, sa satisfaction et ses performances dans la réalisation d'une tâche.

Les modèles présentés dans cet article ont été construits dans un contexte culturel précis (principalement européen). La perception du comportement non-verbal et les codes reliés étant dépendant du contexte culturel, d'autres évaluations auprès de population de cultures différentes devront être réalisées pour valider ces travaux dans d'autres contextes culturels. Les méthodes présentées pourront pour ce faire être réutilisées.

Les travaux présentés ci-dessus se concentrent principalement sur la modélisation du comportement socio-affectif *non-verbal* de l'ACA. Outre les perspectives présentées dans les sections précédentes concernant l'amélioration de chacun des modèles du comportement non-verbal, nous souhaitons intégrer dans l'ACA des modèles permettant d'étendre cette modélisation au comportement verbal à travers, d'une part, la modélisation des intentions communicatives de l'agent et d'autre part la réalisation de ces intentions, au niveau verbal. Le premier défi sera de modéliser l'influence du contexte socio-affectif sur le choix des intentions communicatives et nous travaillons actuellement sur une extension du module de dialogue DISCO [Rich and Sidner 2012] aujourd'hui intégré dans le module *planificateur d'intentions* de l'architecture présentée

Figure 1. Le deuxième défi sera de déterminer comment l'ACA va exprimer verbalement ses intentions communicatives à travers le contenu linguistique (lexique, syntaxe) et paralinguistique (*e.g.* disfluences, prosodie, pause, respiration) suivant ses émotions et ses attitudes sociales et comment ces phénomènes langagiers se synchronisent avec les comportements non-verbaux présentés dans cet article.

Enfin, nous souhaitons travailler sur la compréhension du contexte de l'interaction en entrée du module *Etat mental* et de son modèle computationnel des attitudes sociales présenté précédemment. L'enjeu sera ici de construire un modèle socio-affectif correspondant à l'utilisateur (à partir des informations obtenues en sortie des différents modules d'analyse des signaux vidéo et audio) qui soit pertinent pour la définition de la stratégie à adopter par l'agent.

REMERCIEMENTS

Les travaux de recherche présentés dans cet article sont financés par les projets européens SSPNet, VERVE, REVERIE et TARDIS et par les projets ANR Anipev, MoCa, A1:1, et IMMOMO. Les auteurs de l'article tiennent à remercier *André-Marie Pez* et *Pierre Philippe* pour leur aide précieuse pour la partie développement de l'ACA.

BIBLIOGRAPHIE

- ALLWOOD, J. AND KOPP, S. AND GRAMMER, S. AND AHLSEN, E. AND OBERZAUCHER, E. AND KOPPENSTEINER, M. 2007. The Analysis of Embodied Communicative Feedback in Multimodal Corpora: a Prerequisite for Behavior Simulation. *Language Resources and Evaluation* 41, 3-4 (Dec. 2007), 252–272.
- ARGYLE, M. 1988. Bodily Communication. University paperbacks. Methuen.
- BEALE, R. AND CREED, C. 2009. Affective interaction: How emotional agents affect users. *Int. J. Hum.-Comput. Stud.* 67(9): 755-776.
- BEVACQUA, E. HYNIEWSKA, S. and PELACHAUD, C. 2010. Positive influence of smile backchannels in ECAs, in *Proc. of Workshop on Interacting with ECAs, AAMAS Conference*, p. 13-20.
- BICKMORE, T.W. AND PICARD, R.W. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum.Interact.* 12, 2, 2005, 293-327.
- BURGOON, J.K., BULLER, D.B., HALE, J.L., de TURCK, M.A. 1984. Relational Messages Associated with Nonverbal Behaviors. *Human Communication Research* 10(3), 352-378.
- BUSO, C. AND DENG, Z. AND GRIMM, M. AND NEUMANN, U. AND NARAYANAN, S. 2007. Rigid head motion in expression speech animation: Analysis and synthesis. *IEEE Trans. on Audio, Speech & Language Processing* 15(3), 1075-1086.

- CALLEJAS, Z. AND RAVENET, B. AND OCHS, M. AND PELACHAUD, C. 2014. A Computational model of Social Attitudes for a Virtual Recruiter. *Proceedings of Autonomous Agent and Multiagent Systems (AAMAS)*. 93-100.
- CARNEY, D., HALL, J., LEBEAU, L. 2005. Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior* 29(2), 105-123.
- CASSELL, J. 2000. More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM* 43, 70-78.
- CASSELL, J. AND PELACHAUD, C. AND BADLER, N. AND STEEDMAN, M. AND ACHORN, B. AND BECHET, T. AND DOUVILLE, B. AND PREVOST, S. AND STONE, M. 1994. Animated conversation: Ruled-based generation of facial expression gesture and spoken intonation for multiple conversational agents. In: *Computer Graphics*, 413-420.
- CHINDAMO, M. ALLWOOD, J. and AHLSEN, E. 2012. Some suggestions for the study of stance in communication, in *Proc. SocialCom Conference*, 617-622.
- CHOLLET, M. AND OCHS, M. AND PELACHAUD, C. 2014. Mining a multimodal corpus for non-verbal behavior sequences conveying attitudes. In *Proceedings of the 9th. Language Resources and Evaluation Conference*. 3417-3424
- CLAVEL, C., PLESSIER, J., MARTIN, J. C., ACH, L., and MOREL, B. 2009. Combining facial and postural expressions of emotions in a virtual character. In *Proc. IVA Conference*, 287-300.
- COULSON, M. 2004. Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior* 28 (2), 117-139.
- COWIE, R. AND COX, C. AND MARTIN, J.-C. AND HEYLEN, D. AND KARPOUZIS, K.. 2011. Issues in Data Labelling. In Petta, P. and Pelachaud, P. and Cowie, R. (Eds.) *Emotion-Oriented Systems: The Humaine Handbook*, Springer-Verlag Berling Heidelberg, 215-244.
- DAEL, N., MORTILLARO, M., and SCHERER, K. R. 2012. The Body Action and Posture Coding System (BAP): Development and Reliability. *Journal of Nonverbal Behavior* 36, 97-121.
- DE ROSIS, F. AND PELACHAUD, C. AND POGGI, I. AND CAROFIGLIO, V. AND DE CAROLIS, B. 2003. From Greta's mind to her face: modeling the dynamics of affective states in a conversational embodied agent. *Int. J. Hum.-Comput. Stud.* 59(1-2), 81-118.
- DIAS, J. AND PAIVA, A. 2013. I Want to Be Your Friend: Establishing Relations with Emotionally Intelligent Agents. *Proc. of AAMAS*, 777-784.
- DIAS, J. and PAIVA, A. 2005. Feeling and reasoning: A computational model for emotional characters. In *Proc. of EPIA*, 127-140.

- DING, Y. AND PELACHAUD, C. AND ARTIÈRE, T. 2013. Modeling Multimodal Behaviors From Speech Prosody. In *Proc. of IVA*. 217-228.
- EKMAN, P. AND FRIESEN, W. V. 1975. Unmasking the Face. A guide to recognizing emotions from facial clues. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- FANELLI, G. AND GALL, J. AND ROMSDORFER, H. AND WEISE, T. AND VAN GOOL, L. 2010. A 3-D Audio-Visual Corpus of Affective Communication, *IEEE Transactions on Multimedia* 12(6). 519-598.
- FOURATI, N. and PELACHAUD, C. 2012. Toward new expressive movement characterizations, In *Proc. Motion in Games*
- GATICA-PEREZ, D. 2009. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing* 27, 12, 1775-1787.
- GILLIES, M. AND CRABTREE, I. AND BALLIN, D. 2004. Customisation and context for expressive behaviour in the broadband world. 2004. *BT Technology Journal* 22, 2, 16-17.
- GURTMAN, M. B. 2009. Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass* 3, 4, 601-619.
- HUANG, J. and PELACHAUD, C.. 2012. Expressive body animation pipeline for virtual agent In *Proc. IVA Conference*, 355-362.
- ISEN, A. M.; NIEDENTHAL, P. M. AND CANTOR, N. 1992. An influence of positive affect on social categorization. *Motivation and Emotion*, 16, 1, 65-78.
- KIHLSTORM, J. and CANTOR, N. 2000. Social intelligence. In *Handbook of Intelligence*, Cambridge University Press, 359-379.
- KOPP, S. KRENN, B. MARSELLA, S. MARSHALL, A. N. PELACHAUD, C. PIRKER, H. THRISSON, K.R. and VILHJLMSSON. H. 2006. Towards a common framework for multimodal generation: The behavior markup language, In *IVA Conference*, 21-23.
- KRAMER, N. C. 2008. Social effects of virtual assistants. A review of empirical results with regard to communication. In *Proc. IVA Conference*, 507-508
- LOUWERSE, M. DALE, R. BARD, E. and JEUNIAUX, P. 2012. Behavior matching in multimodal communication is synchronized, *Cognitive Science* 36 (8), 1404-1426.
- MCRORIE, M.; SNEDDON, I.; MCKEOWN, G.; BEVACQUA, E.; DESEVIN, E.; PELACHAUD, C. 2012. Evaluation of Four Designed Virtual Agent Personalities, *IEEE Transactions on Affective Computing* 3 (3), 311-322

- OCHS, M. NIEWIADOMSKI, R. BRUNET, P. and PELACHAUD C. 2012. Smiling Virtual Agent in Social Context, *Cognitive Processing, Special Issue on "Social Agents. From Theory to Applications"* 13 (22), 519-53.
- OCHS, M. PREPIN, P. and PELACHAUD, C. 2013. From emotions to interpersonal stances: Multi-levels analysis of smiling virtual characters, in *Proc. of ACII*, 258-263.
- OSGOOD, C. AND TANNENBAUM, P. 1955. The Principle of Congruity in the Prediction of Attitude Change. *Psychological Review* 62, 1, 42-55.
- PANDZIC, I. AND FORCHEIMER, R. 2002. MPEG4 Facial Animation - The standard, implementations and applications. John Wiley & Sons.
- PECUNE, F.; OCHS, M.; PELACHAUD, C. 2013. A formal model of social relations for artificial companions. To appear in *Proc. of European Workshop on Multi-Agent Systems (EUMAS)*.
- PRENDINGER, H.; DESCAMPS, S., AND ISHIZUKA, M. 2002. Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence* 16 (7-8), 519-553.
- PREPIN, K. OCHS, M. and PELACHAUD, C. 2013. Beyond backchannels: co-construction of dyadic stance by reciprocal reinforcement of smiles between virtual agents, *International Conference CogSci*.
- RAVEN, B.H. 2008. The Bases of Power and the Power/Interaction Model of Interpersonal Influence. *Analyses of Social Issues and Public Policy* 8, 1, 1-22
- RAVENET, B. AND OCHS, M. AND PELACHAUD, C. 2013. From a User-Created Corpus of Virtual Agent's Non-Verbal Behavior to a Computational Model of Interpersonal Attitudes, *Proc. of IVA*, 263-274.
- RICH, C. AND SIDNER, C. 2012. Using collaborative discourse theory to partially automate dialogue tree authoring. *Intelligent Virtual Agents*. Springer Berlin Heidelberg.
- ROATHER, C. L., OMLOR, L., CHRISTENSEN, A., and GIEASE, M. A. 2009. Critical features for the perception of emotion from gait. *Journal of Vision* 9 (6), 1-32.
- SADLER, P.; ETHIER, N. AND WOODY, E. 2011. Interpersonal Complementarity. Dans *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*, New York: Horowitz, L.M. ; Strack, S, 123-142.

- SALOVEY P., BEDELL B., DETWEILER J., MAYER J. 2000. Current Directions in Emotional Intelligence Research, Lewis M., Haviland-Jones J., Eds., *Handbook of Emotions*, The Guilford Press, New York, 504-520
- SCHERER, K. 2005. What are emotions? And how can they be measured? *Social Science Information* 4(4), 2005, 685-729.
- SCHUTZ, W. C. 1958. FIRO: A three-dimensional theory of interpersonal behavior. 267 pages
- SRIKANT, R., AGRAWAL, R. 1996. Mining sequential patterns: Generalizations and performance improvements. *Advances in Database Technology* 1057, p. 1-17
- WALLBOTT, H. G. 1998. Bodily expression of emotion. *European Journal of Social Psychology*, 28(6), 879–896.



Magalie Ochs est chargée de recherche au CNRS-LTCI à Télécom ParisTech. Après l'obtention de son master en intelligence artificielle à l'Université de Montréal, Magalie Ochs a réalisé son doctorat en informatique à l'Université Paris 8 sur la modélisation et la formalisation de l'empathie dans les agents virtuels. Elle a mené des recherches dans le domaine de l'informatique affective à Orange Labs (France Télécom R&D), au Laboratoire d'informatique de l'Université Paris 6 et au National Institute of Informatics à Tokyo. Elle mène aujourd'hui des recherches au CNRS sur l'intégration d'une dimension sociale et émotionnelle dans les agents conversationnels animés pour améliorer l'interaction humain-machine.



Yu Ding est doctorant au Laboratoire Traitement et Communications de l'Information de Telecom ParisTech depuis 2011. Ses activités de recherche portent sur la synthèse d'animation de visage à partir des signaux de parole. Il est diplômé d'un Master en 2010 à l'Université Pierre et Marie Curie en France et d'une Licence en 2006 à l'Université Xia Men en Chine.



Nesrine Fourati est doctorante au Laboratoire CNRS Traitement et Communications de l'Information de Telecom ParisTech depuis 2011 et titulaire d'un Master « Réseaux Image Parole » de l'Université Paris Descartes. Ses recherches actuelles portent la caractérisation des comportements émotionnels corporels à partir des données de capture de mouvement 3D.



Mathieu Chollet est doctorant au Laboratoire Traitement et Communications de l'Information de Telecom ParisTech depuis 2011, et titulaire d'un Master Recherche en Informatique de l'Université de Rennes 1 et d'un diplôme d'Ingénieur de Telecom Bretagne. Ses activités de recherche portent sur la modélisation des dynamiques des comportements non-verbaux pour l'expression des attitudes interpersonnelles d'Agents Conversationnels Animés.



Brian Ravenet est doctorant au Laboratoire Traitement et Communications de l'Information de Telecom ParisTech depuis 2012. Son travail de thèse porte sur la modélisation de l'influence des attitudes sociales sur le comportement non-verbal des agents conversationnels animés. Il est titulaire d'un Master en architecture logicielle de l'Université de Nantes en 2010 après avoir effectué un stage de recherche au sein du projet OpenSim à Tokyo et a ensuite travaillé comme ingénieur d'études et développement en technologie .Net en 2011.



Florian Pecune est doctorant au Laboratoire Traitement et Communications de l'Information de Telecom ParisTech depuis 2012, et titulaire d'un Master Ingénierie et Cognition de l'Université Paris 8. Ses activités de recherche portent sur la modélisation de la prise de décision des compagnons artificiels dans différents contextes d'interaction sociale.



Nadine Glas est doctorante au Laboratoire Traitement et Communications de l'Information de Telecom ParisTech depuis 2013 et titulaire d'un Master "Erasmus Mundus Language and Communication Technologies" de l'université de Groningen et de l'Université de Lorraine. Ses recherches portent sur l'engagement dans l'interaction multimodale humain-agent virtuel.



Ken Prepin a reçu un Master en Mathématiques et un Master en Sciences Cognitives de l'Université Paris 6. Il a effectué son doctorat au sein du laboratoire de Psychologie du Développement sous la responsabilité de Jacqueline Nadel à La Salpêtrière et du laboratoire ETIS en neuro-cybernétique sous la direction de Philippe Gaussier. Sa thèse interdisciplinaire a porté sur la modélisation des interactions sociales via l'imitation entre des humains et des robots. Il est actuellement chargé de recherche au CNRS LTCI à Telecom-ParisTech. Ses intérêts de recherche incluent les agents conversationnels animés, l'émergence de synchronisation non-verbale durant des situations de dialogue et le couplage de systèmes dynamiques.



Chloé Clavel est enseignant-chercheur à Télécom-ParisTech. Ses activités se focalisent sur deux thématiques connexes appartenant au domaine de l'informatique affective : la reconnaissance acoustique des manifestations émotionnelles, et l'analyse des opinions et des sentiments dans les textes. Après son doctorat, elle a mené ses recherches en milieu R&D au sein de deux centres de recherche : Thales Research and Technology et EDF R&D. A Telecom-ParisTech, elle mène ces recherches dans le contexte de l'interaction humain-agent.



Catherine Pelachaud est Directrice de Recherche au CNRS. Ses intérêts de recherche portent sur les agents conversationnels animés, la modélisation de la communication non-verbale et des comportements expressifs et les interfaces multimodales. Elle a participé et participe à plusieurs projets Européens portant sur la communication multimodale, les agents conversationnels animés crédibles, les émotions et les comportements sociaux. Elle est éditeur associé de plusieurs journaux dont IEEE Transactions on Affective Computing et Journal on Multimodal User Interfaces. Elle a co-édité plusieurs livres sur les agents virtuels et les systèmes d'informatique émotionnelle.