WACAI 2012, Workshop Affect, Compagnon Artificiel, Interaction

Alexandre Pauchet, Dominique Duhaut, Carole Adam, Sylvie Pesty

WACAI 2012

Workshop **A**ffect, **C**ompagnon **A**rtificiel, **I**nteraction

Alexandre Pauchet Dominique Duhaut Carole Adam Sylvie Pesty

Grenoble, 15-16 Novembre 2012

Contents

Une interaction la plus riche possible et à moindre coût

Céline Jost
Lab-STICC
UBS Vannes

Brigitte Le Pévédic
Lab-STICC
UBS Vannes

Dominique Duhaut
Lab-STICC
UBS Vannes

ABSTRACT
Les jeux de stimulation cognitive ont une importance capitale pour ralentir le déclin des personnes atteintes de troubles cognitifs. Une partie de ces jeux appartient au domaine des GUI et montre des limites notamment depuis l'émergence des NUI (passivité de l'utilisateur, interaction moins riche…). Il existe également des jeux dans le domaine des NUI, mais souvent ils utilisent de la technologie onéreuse ou sont spécialisés pour un problème précis. Souvent, il n'est pas possible de modifier les exercices proposés. Cet article propose une solution pour développer des jeux de stimulation dans le domaine des NUI, à bas prix. Le principe est d'utiliser les dispositifs numériques qui sont présents pour faire un jeu attractif et réutilisable dans d'autres domaines. Cet article présente StimCards, un jeu de cartes interactif. Les utilisateurs peuvent créer leur propre carte et posséder une base de questions illimitée. Ce jeu s'adapte donc à tous les domaines, à toutes les applications. Une expérimentation a montré que StimCards est stimulant et accepté par les utilisateurs.

Author Keywords
Stimulation cognitive; serious game; interaction homme-machine; NUI

INTRODUCTION

L'interaction homme-machine a beaucoup évolué depuis ses débuts. Depuis que les premières interfaces graphiques sont apparues, les principaux outils de l'utilisateur sont : le clavier et la souris. Mais l'émergence de nouvelles technologies tend à faire disparaître ces outils de bases donnant lieu à une nouvelle génération d'interfaces : les NUI (*Natural User Interface*) [8]. L'homme est amené à interagir plus naturellement avec l'ordinateur, c'est-à-dire, à interagir avec son corps. Ce nouveau domaine pousse les chercheurs à imaginer ce que pourraient être les interfaces du futur, comme Jain et al [12] qui indiquent que l'interaction entre l'homme et le reste du monde est multimodale. Ainsi, pour que l'interaction soit une expérience naturelle, il faut préserver cette richesse de modalité et rendre la technologie transparente. En ce sens, ils citent un ensemble de travaux concernant le discours, la gestuelle, le toucher et la manipulation 3D au sein de la réalité augmentée. Mêmes certains outils vieux de 40 ans comme les télécommandes des télévisions sont revisitées [22]. Un objet a rendu possible une évolution très rapide des NUI depuis 2010 : la Kinect de Microsoft. D'après Goth [8], c'est le début de la fin de la souris. Cet objet a tout pour plaire dans le monde de la recherche puisque qu'il est bon marché et qu'il propose un SDK. L'homme entier devient le clavier et la souris. A cette échelle, les interfaces deviennent imposantes par leur taille, comme par exemple Tweetrix [7], un jeu de Tétris où l'homme joue avec son corps ou encore *the WILD Room* [2], une pièce contenant un mur composé d'écrans d'ordinateur pouvant être contrôlés par une table interactive, un système de détection de mouvement ou encore des PDAs, des tablettes tactiles... Ce système réussit le pari des NUI en permettant à l'humain qui le contrôle des mouvements naturels. Mais en dehors du monde professionnel, ce type de système est inutilisable, non seulement à cause de son prix élevé, mais encore à cause de la place qu'il prend. Les futures interfaces sont dans toutes les imaginations. Mais quelles sont-elles aujourd'hui dans notre quotidien ? La réalité économique freine l'intégration des nouvelles interfaces. Ainsi, des chercheurs préfèrent améliorer le niveau d'interaction avec des objets déjà existants, par exemple les téléphones portables. Scoditti et al [18] utilisent la surface tactile et l'accéléromètre pour améliorer la sélection d'information tandis que Francone et al [6] exploitent le suivi du visage pour créer une nouvelle technique d'interaction avec les téléphones. Cet article s'inscrit dans ce type de recherches : **comment utiliser les outils qui entourent l'humain pour créer une interaction la plus riche possible et à moindre coût ?** Le travail présenté ici propose une réponse à cette question. A travers un système informatique permettant de gérer facilement les dispositifs numériques entourant une personne, un jeu de stimulation cognitive a été créé : StimCards. Cet article présente une expérimentation qui vise à étudier l'acceptabilité de ce nouveau type de jeu.

Le chapitre 2 de cet article présente le contexte de notre recherche qui a amené à la conception de StimCards. Le chapitre 3 présente StimCards en détail. C'est un jeu de la famille des *serious game*, sans souris, ni clavier, paramétrable à volonté par l'utilisateur (apparence de l'interface, questions posées...). Le chapitre 4 décrit l'expérimentation qui a été mis en place pour tester ce nouveau jeu et les résultats. Le chapitre 5 propose une discussion et ouvre de nouvelles perspectives dans la conclusion.

CONTEXTE : LE PROJET ROBADOM

Ce travail est né dans le cadre du projet Robadom [24] qui vise à étudier l'impact d'un robot d'aide à domicile auprès de personnes âgées atteintes de troubles cognitifs. Le robot qui sera fabriqué à l'issue du projet doit avoir un impact positif sur l'état psycho-affectif des personnes et améliorer leur capacité cognitive. Cela pose le problème majeur de l'acceptabilité. On sait déjà qu'il n'y a pas de corrélation entre les capacités sociales et l'acceptation des technologies [9]. Il est très difficile de savoir ce dont les personnes âgées

ont besoin et ce qu'elles veulent. D'après Vanden et al [21], les personnes âgées rejètent les ordinateurs car ils ne peuvent pas remplacer les personnes réelles. Et pourtant, si ces préjugés sont dépassés, l'ordinateur n'est pas un obstacle. Une étude sur des personnes ayant Alzheimer a indiqué que les patients pouvaient apprendre à interagir avec un ordinateur [3]. Ils en tirent même un bénéfice lorsque cet ordinateur leur permet, non seulement d'effectuer des exercices de stimulation cognitive, mais encore de communiquer avec leur famille et médecins. C'est nécessaire d'après Vanden et al [21] qui explique que les personnes âgées ont besoin de se sentir utile, de se cultiver et de rester connectées à la société. La solitude est la pire des situations pour eux. Il est donc nécessaire de créer un système qui respecte ces trois conditions de bien-être. L'idée des jeux de stimulation cognitive semble être une bonne piste. En effet, les jeux apportent des bénéfices cognitifs pendant que les utilisateurs se divertissent [4] et sont souvent utilisés dans l'éducation car ils représentent un processus d'apprentissage intéressant [23]. Les jeux qui allient exercices d'entrainement et technologies génèrent plus de plaisir que des jeux sur papiers [11] et donc favorisent l'amélioration de l'état cognitif des personnes. Mais de tels jeux existent-ils ?

La littérature fournit un ensemble riche de jeux utilisant des technologies variées et appliqués à des domaines variés. La majorité des jeux utilisent les ordinateurs et sont loin des interfaces du futur promis par le domaine des NUI. Par exemple, on trouve une application de jeux de stimulation cognitive nommée SAVION [3], une application de gestion de budgets [14] et une application web pour les enfants [10]. Tous les trois sont utilisés pour aider les personnes ayant des troubles cognitifs. Il existe également un jeu, du domaine des serious game, pour stimuler les personnes atteintes d'Alzheimer en leur présentant des situations de la vie quotidienne [11]. La personne reste dans un environnement connu, ce qui est important pour ce type de trouble cognitif. Ces jeux ont été testés sur les personnes cibles ou par simulation et obtiennent de bons résultats. Mais, ce sont des applications classiques qui risquent de trouver rapidement leur limite. Pour essayer d'intéresser d'avantage les utilisateurs, certains jeux introduisent un personnage virtuel, dans un monde 3D, se rapprochant des jeux vidéos. Certains sont très généralistes et dédiés à toutes les personnes en difficultés [1]. L'avantage est de laisser le patient évoluer seul pendant que le thérapeute le surveille à distance. D'autres applications, comme Jecripe [4] sont spécialisées pour les enfants ayant un syndrome de Down (trisomie 21). Même si Jecripe se distingue en ne laissant pas l'utilisateur passif devant l'ordinateur, en l'invitant à imiter le personnage, aucune interaction réelle n'est faite avec l'utilisateur et le jeu ne contrôle pas les actions faites par l'utilisateur et ne lui donne donc aucun retour. Au contraire, certains jeux sont trop invasifs pour les personnes en les obligeant à porter des capteurs sur elles pour pouvoir surveiller leur constantes comme ZPLAY [15]. L'acceptabilité de ce type de systèmes est discutable.

Les domaines des NUI tentent de résoudre les problèmes présentés par les jeux sus-cités. On trouve des jeux de stimulation cognitive dans le domaine de la réalité augmentée [25], dans le domaine des jeux vidéos avec la console WII, d'une part [16] ou avec un vélo d'appartement relié à un écran, d'autre part [5]. On trouve des jeux dépendant des vecteurs de communication : par exemple, un jeu urbain de construction d'histoire, basé sur une infrastructure ubiquitaire permettant de faire jouer des centaines de joueurs en même temps [17] ou encore des traitements pour des maladies d'yeux sur iPod Touch [20] ou bien encore un robot à domicile dont l'étude montre que les participants préfèrent largement le robot à la version sur ordinateur [19]. Ces jeux sont tous spécialisés et inadaptables à un autre contexte, ou alors difficilement.

Nous proposons un jeu de stimulation cognitive tout public et entièrement configurable pour pallier à ce manque.

STIMCARDS, LA STIMULATION PAR LE JEU DE CARTES

Le jeu que nous proposons est un jeu de cartes interactif. Le principe de base est illustré sur la Figure 1. Il consiste à présenter une carte de jeu à la caméra. Celle-ci scanne le code barre situé au verso de la carte et peut alors accéder à ces informations : question, proposition de réponses, indices... La Figure 1 montre un exemple où l'environnement de la carte est chargé dans une interface graphique Stim'env et deux robots permettent l'interaction avec l'utilisateur : un bioloïd (le robot humanoïde à gauche) et le nabaztag (le robot lapin à droite).

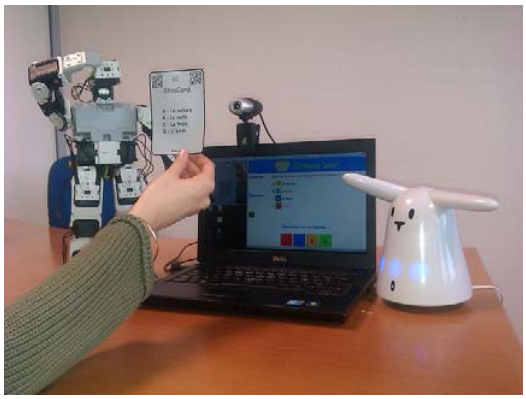


Figure 1. StimCards : le jeu de cartes interactif

Chaque carte de jeu est associée à un fichier XML qui contient le libellé de la question et une photo associée, le type de la question (question à choix multiple, question ouverte), une catégorie de carte (divertissement, sciences,

maths, …), la couleur de fond de l'interface graphique, la couleur de la police, un ensemble d'indices qui peuvent être donnés à l'utilisateur, l'ensemble des réponses proposés à l'utilisateur et la vraie réponse. Ces informations sont utilisées par l'interface graphique Stim'env qui affiche le contenu de la carte. La Figure 2 montre un exemple de carte chargée.

L'utilisateur dispose d'une tablette tactile pour répondre à la question en sélectionnant directement la bonne réponse. Il est possible d'accompagner ce jeu de tout type de compagnons : robot, avatar virtuel ou un logiciel informatique communicant avec l'utilisateur. Ce compagnon devient le référent du jeu pour l'utilisateur et l'aide à comprendre et utiliser StimCards. Nous utilisons la synthèse vocale de Windows : Microsoft Speech API. Chacun des compagnons utilisent la même voix de synthèse : Virginie de Scansoft.

Le déroulement du jeu n'est pas fixe. Nous utilisons MICE [13] qui est un environnement permettant de faire communiquer ensembles des dispositifs numériques de façon simple et rapide. Il est possible de créer des scénarios d'interaction grâce à une interface de programmation visuelle simplifiée. Ainsi, StimCards est doublement configurable : il est possible de créer des nouvelles cartes et il est possible de définir le déroulement du jeu.

Figure 2. Exemple d'affichage d'une carte dans l'interface Stim'env

Pour étudier l'acceptabilité de StimCards, nous avons créé un scénario simple. Le compagnon se présente et explique les règles du jeu. C'est un jeu de calcul mental à 5 niveaux, du plus simple au plus complexe : jaune clair, jaune foncé, vert clair, vert foncé et bleu. L'utilisateur doit présenter une carte à la caméra. S'il ne fait rien pendant une minute, le compagnon lui demande à nouveau de présenter une carte à la caméra. S'il prend une mauvaise carte, le compagnon lui rappelle le niveau de la carte qu'il doit présenter. La question s'affiche dans Stim'env, par exemple : « Quel est le résultat de 28+31 ? » L'utilisateur saisit sa réponse sur la tablette tactile à l'aide de l'interface illustrée sur la Figure 3. Si l'utilisateur donne la bonne réponse, il monte d'un niveau. S'il donne une mauvaise réponse ou s'il ne répond pas, il est invité à prendre une nouvelle carte du même niveau.

Il dispose en effet de 5 cartes par niveau. L'utilisateur est invité à refaire la même opération 9 fois afin de vider le paquet de cartes de plus haut niveau s'il ne se trompe jamais.

Le scénario du jeu est montré sur la Figure 4. Les phrases dites par le compagnon sont les suivantes :

A : Bonjour je suis le compagnon. Tu te souviens ? Je me suis présenté la dernière fois. Comment t'appelles-tu ? (A' = Quel est ton nom ?)

B : Nous allons jouer à un jeu ensemble. Pour cela tu as besoin du jeu de cartes sur la table devant toi, de la caméra qui se trouve dans le boitier noir, du clavier tactile devant toi et de la poubelle à gauche du clavier tactile. Est-ce que tu vois tout cela ?

C : Prends un instant pour repérer les touches sur le clavier tactile, il y a des touches jaunes avec des chiffres de 0 à 9 pour taper ta réponse. Ta réponse s'affichera dans la case blanche. Il y a aussi une touche rouge en haut à droite pour effacer ta réponse si tu te trompes. Il y a aussi une touche verte en bas à droite pour valider ta réponse. Est-ce que tu vois tout cela ?

Figure 3. Interface proposée sur la tablette tactile

D : Les cartes de jeu sont réparties dans des paquets de couleurs différentes : jaune clair, jaune foncé, vert clair, vert foncé et bleu. Tu vois ces paquets ?

E : Je vais t'expliquer les règles du jeu. Prends une carte et retourne là. Tu vois il y a un code barre derrière la carte. Quand je te dirais, "montre le code barre à la caméra", tu glisseras la carte dans le boitier noir en faisant attention de mettre le code barre vers la caméra ensuite je te lirais la question. Tu as compris ?

F : Pour répondre à la question tu tapes la réponse sur le clavier tactile puis tu valides ta réponse. Tu as compris ?

G : Tu as une minute pour répondre. Quand tu as répondu a la question tu peux poser la carte dans la poubelle. Tu as tout compris ?

H : Maintenant nous allons jouer et je vais te guider pendant le jeu.

I : Prends une carte dans le paquet jaune clair et glisse la carte dans le boîtier.

I’ : Je t'explique à nouveau. Prends une carte dans le paquet jaune clair et glisse la carte dans le boîtier.

I’’ : Attention, tu t'es trompé de carte. Prends une carte dans le paquet jaune clair et glisse la carte dans le boîtier.

J : La question est la suivante : Quel est le résultat de […] ?

K : Bravo c'est une bonne réponse.

K’ : Le temps est écoulé, mais ce n'est pas grave, nous allons essayer une nouvelle carte.

K’’ : Ce n'est pas la bonne réponse mais ce n'est pas grave nous allons essayer une nouvelle carte

L : Le jeu est terminé. Merci beaucoup d'avoir joué avec moi. A bientôt !

Les phrases de B à G possèdent une phrase alternative, de B' à G', dans le cas où l'enfant n'a pas compris la consigne. Les phrases I, I' et I" évoluent en fonction du niveau de la question : jaune clair, jaune foncé, vert clair, vert foncé et bleu.

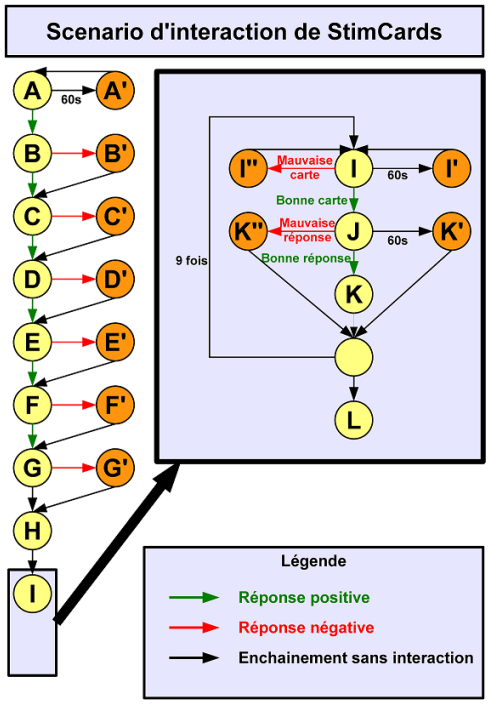


Figure 4. Scénario d’interaction de StimCards

ETUDE DE L’ACCEPTABILITE DE STIMCARDS

L'objectif de cette expérimentation était double. Premièrement, elle nous a permis de vérifier l'acceptabilité et l'utilisabilité de StimCards. Deuxièmement, elle nous a permis de tester différents compagnons pour déterminer le compagnon préféré des participants. Nous avons proposé quatre compagnons : l'ordinateur, le personnage (personnage virtuel projeté sur un écran), le robot et l'animal (le robot déguisé avec un habit de peluche). L'expérimentation s'est décomposée en quatre sessions de 10 minutes. A chaque session, un compagnon différent était proposé. Les participants ont donc interagit avec chacun des compagnons.

Participants

Les participants étaient des élèves de CM2 (moyenne d'âge : 10,27 ans). Nous avons choisi de tester StimCards sur de jeunes enfants avant de le soumettre aux personnes âgés ciblées par le projet Robadom. Nous voulions d'abord tester si le jeu était suffisamment simple et facile à utiliser avant de proposer des jeux de stimulation cognitive. Nous avons choisi des enfants de CM2 car ils n'ont pas de préjugés sur la technologie puisqu'ils ont toujours connus les outils informatiques et parce qu'ils ne sont pas experts ce qui représente un large échantillon de la population. Cette expérimentation a été effectuée par 52 élèves de deux écoles françaises (28 filles et 24 garçons).

Méthode

Equipement

L'expérimentation se passait dans deux écoles différentes. Il n'était donc pas possible, *a priori*, d'obtenir le même cadre expérimental pour tous les élèves puisque les pièces fournies par les directeurs étaient différentes. Nous avons donc construit une structure, placée à l'intérieur de chaque pièce. La Figure 5 montre l'organisation de la structure. C'était une cabine cubique d'environ 1m60, fermée par des rideaux verts. Le vert avait été choisi non seulement pour augmenter la luminosité dans la cabine mais encore pour sa gaieté et sa nature apaisante.

Deux miroirs sans tain se trouvaient sur les murs latéraux de la cabine, afin de contrôler le bon déroulement de l'expérimentation, puisque l'enfant était isolé à l'intérieur. Un bureau et une chaise se trouvaient au fond de la cabine, face à l'entrée. Deux caméras positionnées sur la gauche filmaient l'interaction. Une première caméra filmait le visage de l'enfant, tandis qu'une deuxième caméra filmait l'enfant en entier, sous un autre angle, pour apercevoir les mouvements des jambes et des mains et la posture générale. Un projecteur illuminait l'intérieur de la cabine. Le matériel sur le bureau était à disposition des enfants pendant la durée de l'expérimentation. Un ordinateur affichait l'interface Stim'env. Les cartes de jeu étaient posées devant l'ordinateur. Elles étaient séparées en cinq paquets : jaune clair, jaune foncé, vert clair, vert foncé et bleu. Chaque couleur représentait un niveau de difficulté, du plus facile au plus difficile. Une tablette tactile était posée devant l'enfant et affichait l'interface graphique de la Figure 3. Une

caméra se trouvait dans un boîtier noir, à gauche de l'ordinateur. Les enfants devaient positionner une carte dans la fente, afin que l'ordinateur puisse lire le code barre et traiter la question. Une boîte « poubelle »permettait de déposer chaque carte, une fois qu'elle avait été utilisée. La zone A était l'emplacement réservé au compagnon. Grâce à cette mise en place, le cadre expérimental était le même pour chacun des participants.

Déroulement de l'expérimentation

L'expérimentation a duré 4 jours dans chacune des écoles. Chaque jour, les enfants passaient les uns après les autres pour une série d'exercices. La durée était d'environ 10 minutes, comprenant les explications et le jeu. La différence entre les 4 jours était le contenu de la zone A : l'ordinateur, un personnage animé projeté sur un écran, un robot et un animal (le robot déguisé en habit de peluche). La Figure 6 montre chacun des compagnons proposés. L'interlocuteur du jeu était soit l'ordinateur, le personnage, le robot ou l'animal et énonçait les consignes. Pour l'école 1, l'ordre de passage était le suivant : ordinateur, robot, personnage et animal. Pour l'école 2, l'autre de passage était le suivant : animal, personnage, robot et ordinateur.

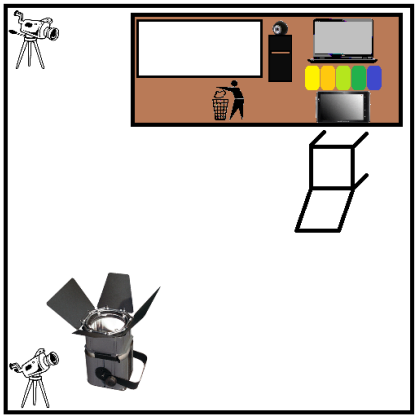


Figure 5. Cadre de l'expérimentation

Collecte de données et analyses

Pour chaque compagnon, les participants devaient répondre aux questions suivantes :

1. As-tu trouvé les exercices faciles?
2. Penses-tu que la présence du compagnon t'ait aidé pour répondre aux questions ?
3. As-tu aimé jouer avec le compagnon ?
4. T'es-tu senti(e) encouragé(e) par le compagnon pendant le jeu ?
5. Trouves-tu le compagnon sympa ?
6. Trouves-tu le compagnon énervant ?
7. Aimerais-tu avoir le compagnon pour t'aider à faire tes devoirs à la maison ?
8. Penses-tu que le compagnon sait faire des calculs mentaux ?
9. Penses-tu que le compagnon ait compris tes réponses ?
10. Penses-tu que le compagnon puisse te voir ?
11. Penses-tu que le compagnon puisse t'entendre ?
12. Penses-tu que le compagnon t'aime bien ?
13. Penses-tu que le compagnon soit content de jouer avec toi ?

A la fin des sessions d'expérimentation, les participants devaient répondre à un questionnaire général sur le jeu de cartes et sur les compagnons :

14. As-tu aimé jouer à ce jeu ?
15. Aimerais-tu avoir ce jeu à la maison ?
16. Est-ce que les règles du jeu étaient faciles ?

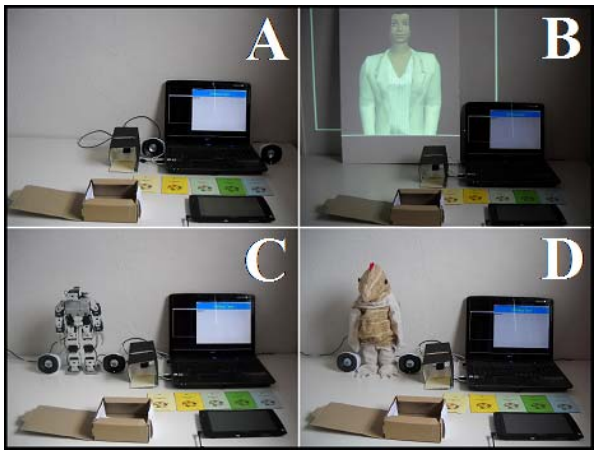


Figure 6. Les quatre compagnons de l'expérimentation

17. Est-ce que tes interlocuteurs sont intervenus au bon moment ?
18. Est-ce que tu voudrais pouvoir faire intervenir ton interlocuteur quand tu veux et comme tu veux ?
19. Est-ce que tu aimerais que ton interlocuteur soit personnel (spécialement adapté pour toi) ?
20. Quel est ton ordre de préférence des compagnons ?
21. Si tu pouvais choisir un compagnon pour t'aider dans la vie quotidienne, lequel choisirais-tu ?

Pour les questions 1-19, l'échelle de Likert était utilisée pour les propositions de réponse : Pas du tout, Un peu, Moyen, Beaucoup, Énormément. Dans le reste de l'article, nous écrirons 0 pour *Pas du tout*, 1 pour *Un peu*, 2 pour *Moyen*, 3 pour *Beaucoup* et 4 pour *Énormément*.

Les analyses statistiques ont été réalisées avec le logiciel Minitab 15©. Le test du Chi Deux a permis de vérifier les réponses significatives. Le seuil de significativité (p) a été

fixé à 0,05. A chaque question de 1 à 13, les statistiques ont été calculées pour chaque compagnon (ordinateur, personnage, robot et animal) pour tous les élèves confondus, mais également en séparant les deux écoles. Les statistiques ont également été calculées pour l'ensemble des compagnons confondues et pour tous les élèves. En plus des statistiques, chaque question a reçu un score qui correspond à une somme pondérée (valeur de la réponse en fonction du nombre de votes obtenus). Ainsi, plus un score est élevé, plus la question a reçu de vote positif et inversement.

Résultats

L'ordre de passage, n'ayant pas eu d'influence sur les résultats, ceux-ci ne sont pas présentés distinctement.

Tout élève confondu, les exercices ne sont pas jugés difficiles que ce soit avec l'ordinateur (0, X^2=23,3846, dl=4, p=0,000), avec le personnage (0, X^2=9,53846, dl=4, p=0,049) ou avec le robot (0, X^2=11,2692, dl=4, p=0,024). Les résultats ne sont pas significatifs dans l'ensemble pour l'animal. La première école juge les exercices plus faciles avec l'animal (4, X^2=22, dl=4, p=0,000) et ni faciles, ni difficiles avec l'ordinateur (2, X^2=18,5217, p=0,001). Les autres résultats ne sont pas significatifs. La tendance est inversée pour la deuxième école qui juge les exercices plus faciles avec l'ordinateur (4, X^2=16,6897, dl=4, p=0,002) et ni faciles, ni difficiles avec l'animal (2, X^2=16,3448, dl=4, p=0,003). L'ordinateur reçoit le score le plus élevé et le personnage le score le moins élevé.

Les résultats ne sont pas significatifs concernant la question 2, sauf pour l'école 1 où la présence du personnage n'a pas aidé les élèves pour répondre aux questions (0, X^2=11,1305, p=0,025). Globalement, les réponses sont mitigées mais Pas du tout obtient le plus grand nombre de voix. Le score le plus élevé est obtenu par l'ordinateur et le plus faible par le personnage.

Les enfants ont aimé jouer avec les compagnons (4, X^2=117,529, p=0,000 pour l'ensemble des compagnons). Il n'y a aucune différence significative entre les différents compagnons. Le score le plus élevé est obtenu par l'animal et le score le plus faible est obtenu par le robot.

Les résultats ne sont pas significatifs pour la question 4, excepté pour l'animal où les élèves se sont sentis encouragés (les deux écoles : 3, X^2=10,8846, p=0,028, l'école 1 : NS, l'école 2 : 4, X^2=34,2759, p=0,000). Le score le plus élevé est obtenu par l'animal et le score le plus faible est obtenu par le robot.

Les élèves ont trouvé tous les compagnons sympas (Ordinateur : 4, X^2=10,1154, p=0,039, Personnage : 4, X^2=18,9615, p=0,001, Robot : 4, X^2=18,5769, p=0,001, Animal : 4, X^2=39,1538, p=0,000). Le score le plus élevé est obtenu par l'animal et le score le plus faible est obtenu par l'ordinateur.

Concernant la question 6, les élèves n'ont pas jugé les compagnons énervants (Ordinateur : 0, X^2=73,5769, p=0,000, Personnage : 0, X^2=66,8462, p=0,000, Robot : 0, X^2=68,9615, p=0,000, Animal : 0, X^2=113,385, p=0,000). Le meilleur score, qui est ici le score le plus faible est obtenu par l'animal, tandis que le moins bon score est obtenu par le robot.

De façon générale, les élèves aimeraient avoir le compagnon à la maison pour les aider à faire leurs devoirs (4, X^2=56,4541, p=0,000). Dans le détail, les résultats ne sont pas significatifs pour l'ordinateur. Le score maximal est obtenu par l'animal et le minimal par le personnage.

Les résultats ne sont pas significatifs pour la question 8 pour chacun des compagnons, mais en les prenant dans l'ensemble, il s'avère que les élèves pensent que leurs compagnons savent faire des calculs mentaux (4, X^2=11,7596, p=0,019). Une exception est fait pour l'école 1, qui considère de façon significative que le personnage ne sait pas faire de calcul mental (1, X^2=11,5652, p=0,021). Le score maximal est obtenu par l'ordinateur, tandis que le minimal est obtenu par l'animal.

Les élèves considèrent que les compagnons ont compris leurs réponses (dans l'ensemble : 4, X^2=115,894, p=0000). Les résultats sont les mêmes pour tous les compagnons et pour les deux écoles indépendamment, sauf pour la deuxième école qui ne montre aucun résultat significatif pour l'ordinateur. Le résultat le plus fort est obtenu par l'animal, tandis que le résultat le plus faible est obtenu par le robot.

Concernant la question 10, dans l'ensemble, les élèves ne pensent pas que le compagnon peut voir (0, X^2=31,2788, p=0,000). Dans le détail, ces résultats ne sont significatifs que pour l'ordinateur et le robot. Le score le plus élevé est obtenu par le personnage tandis que le score le plus faible est obtenu par le robot.

Les élèves pensent dans l'ensemble que le compagnon peut les entendre (4, X^2=22,5411, p=0,000). Individuellement, les résultats ne sont pas significatifs, même si la réponse Énormément est choisie majoritairement pour tous les compagnons. Le score le plus élevé est obtenu par le robot et le plus faible par l'animal.

Concernant la question 12, il n'y a pas de résultats significatifs, ni dans l'ensemble, ni dans le détail. Les réponses sont très mitigées. Dans l'ensemble, les cinq réponses ont quasiment toutes obtenues le même nombre de votes, même si l'ordinateur et le personnage semble donner d'avantage l'impression d'aimer les élèves que le robot et l'animal. Le score maximal est obtenu par le personnage et le score minimal par le robot.

Les élèves pensent que le compagnon est content de jouer avec eux (Dans l'ensemble : 4, X^2=56,2609, p=0,000, Ordinateur : 3, X^2=213,1923, p=0,010, Personnage : 4, X^2=17,8077, p=0,001, Robot : 4, X^2=13,4118, p=0,009, Animal : 4, X^2=20,1154, p=0,000). Les statistiques les plus significatives sont celles de l'animal, suivies par le

personnage, puis le robot et l'ordinateur. Le score le plus élevé est obtenu par l'animal et le score le plus faible est obtenu par l'ordinateur.

Pour toutes les questions, aucune différence significative entre les compagnons n'est apparue.

Le tableau ci-dessous répertorie le nombre de fois que chaque compagnon a obtenu le score maximal (m1) et le score minimal (m2). Le score total est la somme de tous scores obtenus auquel on a soustrait le score de la question 6 qui est négative.

Compagnon	**m1**	**m2**	**Score total**
Ordinateur	3	2	1450
Personnage	2	3	1432
Robot	1	6	1411
Animal	7	2	1483

Table 1. Score obtenu au questionnaire par les compagnons.

Concernant l'évaluation du jeu de cartes, les élèves ont énormément aimé jouer avec (4, X^2=86,8462, p=0,000, Figure 7). Ils aimeraient d'ailleurs l'avoir à la maison (4, X^2=61,8462, p=0,000, Figure 7). Les règles du jeu ont été jugées très faciles (4, X^2=53,5769, Figure 7). Concernant le scénario du jeu, les élèves ont estimé que les interlocuteurs sont intervenus au bon moment (4, X^2=21,2549, p=0,000, Figure 7). Les élèves ont majoritairement exprimé leur envie de pouvoir faire intervenir leur compagnon quand ils veulent et comme ils veulent (4, X^2=31,6538, p=0,000, Figure 7). Il faut cependant noter que cette question est contrastée car la deuxième réponse la plus formulée est le *Pas du tout*. Pour finir, les élèves aimeraient que leur compagnon soit personnalisé (4, X^2=69,7308, p=0,000, Figure 7).

Concernant le choix du compagnon préféré des enfants, le robot et l'animal arrive tous les deux en tête avec 39,22% des voix. Le système robotique totalise donc 78,43% des voix. Le troisième compagnon est le personnage avec 11,76% des voix, puis finalement l'ordinateur avec 9,80%. Significativement, l'ordinateur n'est pas le compagnon choisi par les enfants (X^2=16,5294, p=0,001, Figure 8).

L'ordre de préférence donné par les enfants indique que le robot a été le plus de fois cité en premier (46,15%, X^2=21,2308, p=0,011, Figure 9). Il n'a jamais été cité en dernier. Le deuxième compagnon le plus cité est également le robot avec 36,54% des voix, suivi de près de l'animal avec 30,77% des voix. Significativement, l'ordinateur est le moins cité en deuxième position (5,77%, X^2=11,2308, p=0,011, Figure 9). Le troisième compagnon est le personnage (42,31%, X^2=10,3077, p=0,016, Figure 9) et enfin le dernier compagnon est l'ordinateur (55,77%, X^2=33,0769, p=0,000, Figure 9). L'animal est plus cité en quatrième position qu'en troisième position, à hauteur de 21,15%.

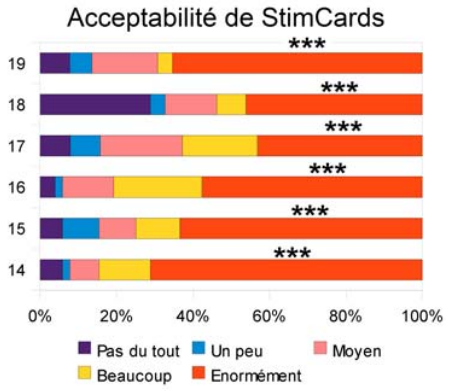


Figure 7. Résultats de l'étude de l'acceptabilité de StimCards

Niveau de significativité : *p<0,05, **p<0,01, ***p<0,001 (Test du Chi-deux)

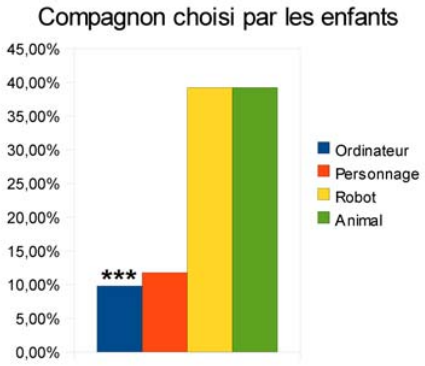


Figure 8. Résultat du compagnon choisi par les enfants

Niveau de significativité : *p<0,05, **p<0,01, ***p<0,001 (Test du Chi-deux)

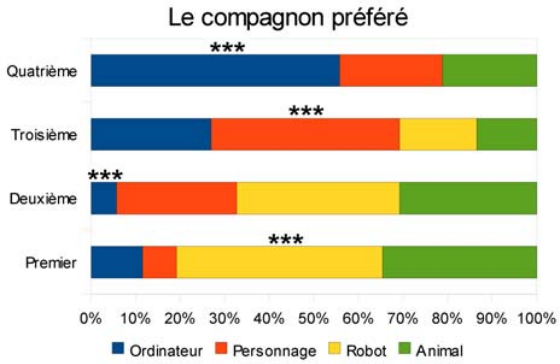


Figure 9. Résultat du compagnon choisi en fonction de l'ordre de préférence

Niveau de significativité : *p<0,05, **p<0,01, ***p<0,001 (Test du Chi-deux)

DISCUSSION ET CONCLUSION

Cette étude a montré que ce nouveau jeu de cartes est apprécié et accepté par une majorité d'enfants. L'activité qui était proposée ne plaisait pas à la majorité puisqu'il s'agissait de calculs mentaux. Et malgré cela, les enfants ont aimé jouer avec le jeu. Cela montre que StimCards stimule et qu'il pourrait être un bonne aide pour la stimulation cognitive des personnes plus âgées.

De plus, nous avons cherché à accompagner ce jeu d'un compagnon. Quatre partenaires potentiels ont été testés : l'ordinateur (le jeu lui même), le personnage (un personnage virtuel), le robot et l'animal (le robot déguisé en habit de peluche). L'objectif était de définir quel était celui que les enfants préféraient. Les résultats n'ont pas révélé un choix précis. Nous avons noté une différence de préférence entre les deux écoles testées. Par exemple, pour la première question (As-tu trouvé les exercices faciles ?) les exercices n'ont pas semblé plus faciles en fonction des partenaires. Mais en détaillant au niveau des écoles, il apparaît que les exercices ont été jugés plus faciles avec l'animal et moins facile avec l'ordinateur pour l'école 1. Le contraire est apparu pour l'école 2. Sachant que dans l'école 1 l'animal est passé en premier et l'ordinateur en dernier et que l'ordre était inversé dans l'école 2. L'ordre de passage des partenaires semble avoir guidé les enfants dans leur jugement. Et en finalité, les résultats des deux écoles s'annihilent et aucun partenaire n'a été au dessus des autres de façon significative.

L'étude des questions 1-13 a permis de montrer que la forme du compagnon a peu d'importance. En effet, les quatre partenaires ont été dépersonnalisés, ils ont tous agi de la même façon, employé les même mots, avec la même voix. Seule leur expression non verbale était différente. Et pourtant, aucun n'a remporté l'intérêt massif, comme si les enfants pensaient avoir interagi quatre fois avec le même partenaire (réflexions faites par plusieurs enfants : « c'est toujours la même chose »). Si chaque partenaire avait eu sa propre personnalité et si l'interaction s'était adaptée aux élèves, un compagnon aurait pu, plus facilement, être choisi par une majorité. En effet, les enfants ne semblent pas avoir bien vécu le fait de rester à un niveau difficile alors qu'ils ne réussissaient jamais à répondre à la bonne question. Le niveau aurait du redescendre pour laisser l'enfant dans une situation confortable. Le fait que les enfants n'arrivent pas à évaluer si leur partenaire les aiment bien confirme cette hypothèse et indique que les partenaires ont manqué d'empathie, d'émotion et n'ont pas pris en compte l'être humain. Ces résultats nous confortent dans l'idée qu'une interaction doit être personnelle et personnalisé.

Même si significativement, aucun partenaire n'émerge dans le questionnaire, les enfants semblent quand même avoir une préférence nette pour les robots qu'ils soient habillés ou non (Robot et Animal). En effet, lorsque nous avons demandé aux enfants leur ordre de préférence des partenaires, l'ensemble des robots a totalisé 80,77% de la première place et 67,31% de la deuxième place. Le robot est arrivé en tête à ces questions tandis que l'Animal a totalisé le score maximal dans le questionnaire. Il n'est pas possible de départager le robot et l'animal mais on en retient que la présence physique d'un partenaire semble être importante. Ainsi, StimCards est apprécié et accepté. Il est difficile de départager les différents compagnons. Cela permet de se demander s'il faut chercher une forme de représentation du compagnon. En fonction des gens, le partenaire préféré est différent. Le compagnon ne peut pas être un objet unique, apprécié de tous, mais un ensemble d'objets avec lesquels l'être humain interagit et qu'il a l'habitude de « côtoyer ».

REMERCIEMENTS

Ce travail est financé par l'agence nationale de la recherche (ANR) à travers le programme TecSan (projet robadom n°ANR-09-TECS-012). Les partenaires de ce projet sont l'hôpital Broca à Paris, le laboratoire ISIR à Paris et l'entreprise Robosoft à Bidart.

RÉFÉRENCES

1. Abreu, P.F., Werneck, V.M.B., Costa, R.M.E. and Carvalho, L.A.V., Employing Multi-agents in 3-D Game for Cognitive Stimulation, In Sympsium on Virtual Reality (SVR), 2011 XIII, IEEE, (Brésil 23-26 mai 2011), 73-78

2. Baudouin-Laffont, M., Lessons learned from the WILD room, a multisurface interactive environment, 23ème Conférence Francophone Sur l'Interaction Homme-Machine, IHM'11, ACM, (Sophia Antipolis, 24-27 octobre 2011) 2011, 18

3. Berenbaum, R., Lange, Y. and Abramowitz, L., Augmentative alternative communication for Alzheimer's patients and families' using SAVION, In Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA'11, ACM, (Grèce 25-27 mai 2011) 2011, 46

4. Brandão,A., Brandão, L., Nascimento, G., Moreira, B., Vasconcelos, C.N., and Clua, E., Jecripe: stimulating cognitive abilities of children with down syndrome in pre-scholar age using a game approach, In Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology, ACE'10, ACM, (Taiwan 17-19 novembre 2010), 2010, 15-18

5. Chilukoti, N., Early, K., Sandhu, S., Riley-Doucet, C. and Debnath, D., Assistive technology for promoting physical and mental exercise to delay progression of cognitive degeneration in patients with dementia, Biomedical Circuits and Systems Conference, BIOCAS, IEEE, (Canada 27-30 novembre 2007), 2007, 235-238

6. Francone, J., Nigay, L., Using the Users Point of View for Interaction on Mobile Devices, 23ème Conférence Francophone Sur l'Interaction

Homme-Machine, IHM'11, ACM, (Sophia Antipolis, 24-27 octobre 2011) 2011

7. Freeman, D., Chevalier, F., Westecott, E., Duffield, K., Hartman, K. and Reilly, D., Tweetris: play with me, In Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction, TEI'12, ACM, (Canada 19-22 février 2012), 2012, 319-320

8. Goth, G., Brave NUI world, in Communications of the ACM, Vol. 54, No 12, décembre 2011, 14-16

9. Heerink, M., Krose, B., Evers, V., Wielinga, B , The influence of a Robot's Social Abilities on Acceptance by Elderly Users, Robot and Human Interactive Communication, on The 15th IEEE International Symposium, ROMAN 2006. , (2006), 521-52

10. Hussaan, A.M., Sehaba, K. and Mille, A., Tailoring Serious Games with Adaptive Pedagogical Scenarios: A Serious Game for Persons with Cognitive Disabilities, 11th IEEE International Conference on Advanced Learning Technologies (ICALT), (Etats-Unis 06-08 juillet 2011), 2011, 486-490

11. Imbeault, F., Bouchard, B. and Bouzouane, A., Serious Games in Cognitive Training for Alzheimer's Patients, 1st International Conference on Serious Games and Applications for Health (SeGAH), IEEE, 2011

12. Jain, J., Lund, A. and Wixon, D., The future of natural user interfaces, In proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, CHI'11, ACM (Canada 7-12 Mai 2011) 2011, 211-214

13. Jost, C., Le Pévédic, B., Duhaut, D., Creating Interaction Scenarios With a New Graphical User Interface, in 5th International Workshop on Intelligent Interfaces for Human-Computer Interaction, IIHCI 2012, (Palerme, 4-6 Juillet 2012), 2012

14. Lopez-Martinez, A., Santiago-Ramajo, S., Caracuel, A., Valls-Serrano, C., Hornos, M.J. and Rodriguez-Fortiz, M.J., Game of gifts purchase: Computer-based training of executive functions for the elderly, 1st International Conference on Serious Games and Applications for Health, IEEE, (Portugal 16-18 novembre 2011), 2011

15. Makedon, F., Zhang, R., Alexandrakis, G., Owen, C.B., Huang, H. and Saykin, A.J., An interactive user interface system for Alzheimer's intervention, In Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments, PETRA'10, ACM, (Grèce 23-25 juin 2010), 2010, 35

16. Morelli, T., Foley, J., Columna, L., Lieberman, L. and Folmer, E., VI-Tennis: a vibrotactile/audio exergame for players who are visually impaired, Proceedings of the Fifth International Conference on the Foundations of Digital Games, FDG'10, ACM, (Etats-Unis 19-21 juin 2010), 2010, 147-154

17. Scheible, J., Tuulos, V.H. and Ojala, T., Story Mashup: design and evaluation of novel interactive storytelling game for mobile and web users, Proceedings of the 6th international conference on Mobile and ubiquitous multimedia, MUM07, ACM,(Finlande 12-14 décembre 2007), 2007, 139-148

18. Scoditti, A., Vincent, T., Coutaz, J., Blanch, R., and Mandran, N., TouchOver: Decoupling Positioning from Selection on Touch-based Handheld Devices, 23ème Conférence Francophone Sur l'Interaction Homme-Machine, IHM'11, ACM, (Sophia Antipolis, 24-27 octobre 2011) 2011, 6

19. Tapus, A., Tapus, C., and Mataric, M., The role of physical embodiment of a therapist robot for individuals with cognitive impairments, The 18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2009, IEEE, (Japon 27 septembre-02 octobre 2009), 2009, 103-107

20. To, L., Thompson, B., Blum, JR., Maehara, G., Hess; RF., Cooperstock, JR., A game platform for treatment of amblyopia. IEEE Trans Neural Syst Rehabil Eng 2011; 19: 280–289.

21. Vanden Abeele, V.A., Van Rompaey, V., Introducing Human-Centered Research to Game Design: Designing Game Concepts for and with Senior Citizens, CHI'06 extended abstracts on Human factors in computing systems, ACM, (2006), 1469-1474.

22. Vo, D.B., Bailly, G., Lecolinet, E., and Guiard, Y., Un espace de caractérisation de la télécommande dans le contexte de la télévision interactive, 23ème Conférence Francophone Sur l'Interaction Homme-Machine, IHM'11, ACM, (Sophia Antipolis, 24-27 octobre 2011) 2011, 17

23. Wei, Z.,Li, L., Zhu, J. and Chi, Z., Games for service science education, International Conference on Service Sciences (ICSS), IEEE, (Chine 13-14 mai 2010), 2010, 237-241

24. Wu, Y.H., Chetouani, M., Cristancho-Lacroix, V., Le Maitre, J., Jost, C., Le Pevedic, B., Duhaut, D., Granata, C., Rigaud, A.S., ROBADOM : The Impact of A Domestic Robot on Psychological and Cognitive State of the Elderly with Mild Cognitive Impairment, in 5th CRI (Companion Robotics Institute) Workshop AAL User-Centric Companion Robotics Experimentoria, Supporting Socio-ethically Intelligent Assistive Technologies Adoption, 2011

25. Zapirain, B.G., Zorrilla, A.M., and Larrañaga, S., Psycho-stimulation for elderly people using puzzle game, Games Innovations Conference (ICE-GIC), 2010 International IEEE Consumer Electronics Society's, (Chine 21-23 décembre 2010), 2010, 1-8

Approche Comparative des Modèles Informatiques des Émotions pour l'Animation Faciale en Situation d'Interaction

M.Courgeon[1] courgeon@limsi.fr
C. Clavel[1, 2] clavel@limsi.fr
J-C. Martin[1, 2] martin@limsi.fr

[1]Laboratoire d'Informatique pour la Mécanique et les Science de l'Ingénieur
Bâtiment 508, 91403 Orsay CEDEX, France

[2]Université Paris Sud
15 rue Georges Clémenceau 91405 Orsay cedex

Résumé :
Les émotions et leurs expressions par des agents virtuels sont deux enjeux importants pour les interfaces homme-machine affectives à venir. En effet, les évolutions récentes des travaux en psychologie des émotions, ainsi que la progression des techniques de l'informatique graphique, permettent aujourd'hui d'animer des personnages virtuels réalistes et capables d'exprimer leurs émotions via de plusieurs modalités. Si plusieurs systèmes d'agents virtuels existent, ils restent encore limités par la diversité des modèles d'émotions utilisés, par leur niveau de réalisme, et par leurs capacités d'interaction temps réel. Dans cet article, nous présentons quatre modélisations complémentaires des émotions en informatique orientées vers l'animation faciale interactive. Ces modèles ont été intégrés au sein de la plateforme d'agents virtuels expressive MARC.
Mots-clés : Informatique affective, animation faciale, modèles émotionnels

Abstract:
Emotions and their expressions by virtual characters are two important issues for future affective human-machine interfaces. Recent advances in psychology of emotions as well as recent progress in computer graphics allow us to animate virtual characters that are capable of expressing emotions in a realistic way through various modalities. Existing virtual agent systems are often limited in terms of underlying emotional models, visual realism, and real-time interaction capabilities. In this paper, we present four computational models of emotions oriented toward interactive facial animation. These models have been integrated into the platform MARC expressive virtual agents.
Keywords: Affective computing, Facial animation, Computational models of emotion

1 Introduction

La notion de réalisme d'un agent virtuel est souvent associée à son apparence graphique. Pourtant, le réalisme d'un agent virtuel repose également sur sa capacité à exprimer des émotions en situation d'interaction avec l'utilisateur. On parle de « réalisme comportemental ».
Cependant, doter un agent de réactions expressives pertinentes et dynamiques nécessite une modélisation informatique des émotions et de leurs réactions expressives faciales associées. Ces travaux sont donc à la croisée des recherches en psychologie des émotions et des recherches en informatique affective.
Dans cet article, nous présentons nos travaux de modélisation de ces processus émotionnels et de leur animation faciale. L'ensemble des développements logiciels associés ont été intégrés dans la plateforme MARC, dédiée à l'animation multimodale d'agents virtuels expressifs.

2 Etude de l'existant

En psychologie, la définition de ce qu'est une émotion ne fait pas consensus (Russell, 2012). Différentes approches coexistent et s'influencent et plusieurs théories des émotions tentent de formaliser ce qu'est une émotion. Ces approches ne sont pas totalement exclusives, ainsi certaines théories relèvent de plusieurs approches (Gross et Feldman-Barrett, 2011). La plupart des théories supposent l'existence d'un processus cognitif sous-jacent aux émotions. Pourtant, certaines théories considèrent l'existence d'un processus cognitif dédié à chaque émotion (Ekman et Friesen, 1975, Tomkins, 1984), alors que d'autres considèrent un processus cognitif unique, commun à toutes les émotions (Scherer, 1984).
Le traitement informatique des émotions se heurte donc au problème de devoir sélectionner une ou plusieurs théories parmi l'ensemble des théories formulées en psychologie. Cette sélection doit être effectuée en fonction de

l'application ciblée. Plusieurs types d'application ont été explorés (Niewiadomski et al., 2009, Kopp et Jung, 2000, Swartout et al., 2010, Leite et al., 2009) pour plusieurs de ces approches. Dans le domaine de l'informatique affective, plusieurs modèles computationnels ont été proposés pour augmenter le réalisme comportemental des agents virtuels (Cassell et al., 2001, Kopp et al., 2007, Marsella et Gratch, 2006).

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

3 Modèles émotionnels pour l'animation faciale interactive

3.1 Objectifs de recherche

L'objectif principal de nos travaux est de modéliser, implémenter, et évaluer différentes approches des émotions et de l'animation faciale temps-réel afin de contribuer à l'amélioration de l'interaction temps-réel entre les agents expressifs et l'utilisateur. Nous avons donc proposé différents modèles émotionnels informatiques, inspirés de différentes approches des émotions issues de la psychologie. Pour chacun des modèles proposés, nous avons exploré ses liens avec l'animation faciale.

Tous les travaux que nous avons effectués ont été validés par des études perceptives. En effet, comme Wallraven *et al.* (2005) l'ont montré, les agents virtuels peuvent être utilisés pour étudier la perception humaine, et réciproquement, la perception humaine peut être utilisée pour évaluer, étudier et améliorer les modèles et les rendus d'agents virtuels. Les résultats de ces études ont ainsi contribué à l'élaboration des modèles successifs que nous avons proposés.

3.2 Modèles proposés

3.2.1 Approche catégorielle

La première approche que nous avons explorée est l'approche catégorielle des émotions. Après avoir conçu un système d'animation faciale permettant de contrôler l'expressivité de l'agent par des labels émotionnels, nous avons mené plusieurs études perceptives.

Nous avons commencé par évaluer la reconnaissance catégorielle des expressions des émotions de base mise en place avec notre système d'animation. Pour cela, nous avons présenté des stimuli animés non interactifs à un ensemble de sujet. Ces stimuli se composaient chacun d'une animation faciale d'une émotion sur le modèle onset-apex-offset. Nous avons montré que ces expressions étaient correctement reconnues par les participants (Courgeon et al, 2009), avec des taux similaires à ceux de la littérature en perception des expressions humaines, issus de la liste des études établie par Russell (1994).

En utilisant une approche similaire, nous avons également mené une étude sur la perception de la dynamique des expressions faciales. Nous avons montré que les sujets anticipent la dynamique expressive du visage selon une courbe composée d'un pic expressif suivi d'un déclin d'intensité. Nous avons ainsi observé le phénomène d'anticipation suggéré par Thornton et al. (1998) et que nous avons appelé le *moment émotionnel* (Courgeon et al, 2010).

Ces différentes études nous ont permis de nous confronter à un certain nombre de limitations de l'approche catégorielle. Par exemple, le modèle catégoriel impose de spécifier chaque état émotionnel, ainsi que la ou les expressions faciales associées. De plus, il n'établit pas de relation entre les émotions. Les émotions sont considérées comme des états indépendants et possédant leurs propres mécanismes neuronaux. Cette approche limite donc les raisonnements de plus haut niveau sur les émotions, puisque chaque émotion doit avoir ses propres conditions d'apparition.

3.2.2 Approche dimensionnelle

Notre démarche itérative nous a ensuite conduits à aborder l'approche dimensionnelle des émotions. Dans le modèle informatique à trois dimensions, inspiré du modèle P.A.D. (Pleasure Arousal Dominance), (Russell et Mehrabian, 1977, Broekens, 2012) que nous avons proposé, nous obtenons un espace continu qui nous permet de créer des relations de continuité entre les émotions. Ces relations imposent un certain nombre de contraintes implicites. Par exemple, l'état affectif de l'agent doit conserver une trajectoire continue dans l'espace, et ne peut passer d'une position à l'autre de manière discontinue (Courgeon et al, 2008).

En utilisant notre modèle informatique, nous avons mis en place un système de profils expressifs individuels permettant de moduler l'expressivité du personnage en fonction de paramètres liées aux dimensions P.A.D. Ainsi, la manipulation par l'utilisateur de l'expression faciale du personnage était modulée par des paramètres expressifs liés au modèle émotionnel dimensionnel. Pour évaluer ce modèle, nous avons mis en place une étude dans laquelle le participant devait contrôler manuellement l'expression du personnage virtuel l'aide d'une souris 3D, en manipulant un curseur dans l'espace P.A.D. L'expressivité étant modulée par le profil expressif, le sujet avait pour consigne de décrire ce qu'il percevait de ce profil. Cette étude nous a permis de montrer d'une part que l'utilisation d'un espace dimensionnel des émotions permet la manipulation de l'expression d'un agent virtuel, et d'autre part que les profils expressifs modulant l'expression sont bien perçus par les sujets manipulant l'agent virtuel.

Cependant, l'approche dimensionnelle ne nous a pas semblé suffisamment complexe pour permettre à l'agent de manifester un comportement autonome, car elle ne modélise pas le traitement de l'information ni le processus de prise de décision, et ne permet donc pas de déterminer la réaction émotionnelle dans un contexte dynamique.

3.2.3 Approche cognitive

Pour aborder l'approche cognitive des émotions, nous avons choisi de nous inspirer du modèle CPM (*Componential Process Model*, Scherer, 2001) et de proposer un modèle dynamique de l'émotion. Ce modèle nous a semblé pertinent pour créer une interaction avec un agent virtuel autonome car il simule une évaluation cognitive de la situation. Ainsi, le système est capable de gérer de manière continue l'état émotionnel et l'expressivité de l'agent virtuel durant une interaction avec l'utilisateur. Pour spécifier notre module, nous avons effectué un certain nombre de choix d'implémentations. De plus, dans notre étude, nous avons limité le contexte d'interaction à un jeu de plateau de société (Fig 1), afin de mieux contrôler les différentes situations possibles durant l'interaction (Courgeon et al, 2009).

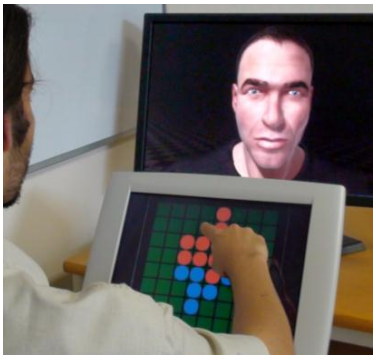


Fig. 1- Jeu Othello pour l'Interaction avec le Modèle Cognitif

Ce système nous a permis lors d'une étude expérimentale avec des utilisateurs de comparer l'expressivité générée par le modèle cognitif proposé avec l'expressivité issue de l'approche catégorielle des émotions. En effet, le modèle cognitif propose une animation faciale séquentielle de l'évaluation cognitive. La comparaison avec l'approche catégorielle a été possible en masquant cette dynamique issue de l'évaluation cognitive. Ainsi, nous avons pu montrer que l'utilisation d'un modèle cognitif modifie la perception que les utilisateurs ont de l'agent. En effet, nos résultats montrent que les sujets attribuent plus d'états mentaux à l'agent lorsqu'il exprime son évaluation de la situation à travers

ses expressions faciales. De plus, l'agent est perçu comme étant plus expressif avec le modèle cognitif qu'avec le modèle catégoriel. Pour finir, nos résultats montrent que le modèle cognitif modifie le comportement de l'utilisateur. En effet, en mode cognitif, les utilisateurs ont passé plus de temps à jouer, mais ont gagné plus souvent que dans le mode catégoriel.

3.2.4 Approche cognitive et sociale[1]

Considérant les limites de notre modèle cognitif, nous avons cherché à prendre en compte les aspects sociaux du processus cognitif émotionnel. Pour cela, nous avons étudié un phénomène social particulier nommé le *social appraisal*. Le social appraisal, ou évaluation cognitive sociale, est défini par Manstead et Fisher (2001) comme la prise en compte de l'évaluation cognitive d'autrui dans notre propre évaluation d'un événement. Nous avons donc proposé un modèle de réévaluation des stimuli cognitifs, basé sur la prise en compte de l'évaluation d'un second personnage virtuel (agent observé). Ce modèle s'appuie sur notre modèle cognitif pour réaliser l'animation faciale.

Aucune donnée n'étant disponible sur les mécanismes responsables de l'influence sociale sur le processus d'évaluation cognitive, nous avons proposé deux approches. La première consiste à copier le résultat de l'un des critères d'évaluation de l'agent observé. La seconde méthode consiste à utiliser un ensemble de règles logiques pour établir la seconde réaction de l'agent social, en fonction des premières évaluations des deux agents. Ces deux approches ont été comparées lors d'une expérimentation perceptive. Afin de limiter la complexité des règles logiques, nous avons limité notre modèle à quatre émotions : Peur, Joie, Colère et Tristesse.

Dans cette étude, nous avons montré que l'utilisation d'un modèle de réévaluation sociale permet d'augmenter la perception de l'expressivité de l'agent dans le contexte social. Cependant, nous n'avons pas pu déterminer quelle méthode de simulation du processus social était la plus efficace. Pour cela, d'autres modèles devront être proposés et évalués. Néanmoins, les sujets semblent percevoir la communication émotionnelle entre les deux agents. L'agent doté d'une réaction sociale est en effet perçu comme s'adaptant mieux à l'autre agent que l'inverse.

Malgré ses limites, notre modèle expérimental de *social appraisal* semble donc permettre la prise en compte automatique de réactions sociales dans l'évaluation cognitive. Ces résultats sont encourageants, et d'autres travaux devront être effectués.

3.3 Modèles proposés

L'ensemble des modèles présentés ci-dessus ont été intégrés dans la plateforme d'agents virtuels expressifs MARC. L'architecture (Fig. 2) est donc composée d'une boucle d'interaction temps réelle permettant l'utilisation de plusieurs dispositifs d'interaction et d'immersion, ainsi que l'utilisation des quatre approches émotionnelles présentées ci-dessus.

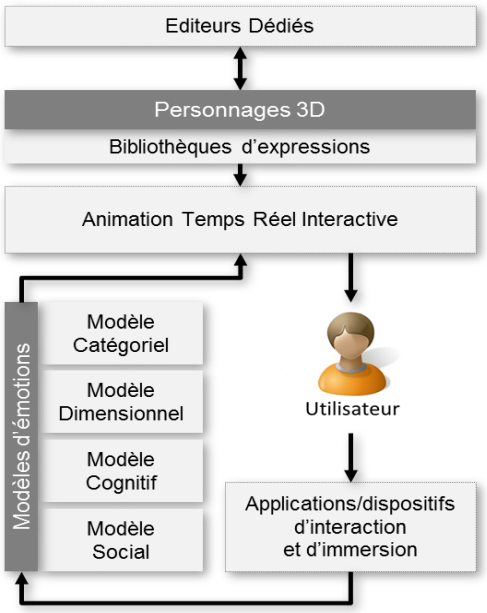


Fig. 2 - Architecture de la plateforme MARC (v11)

4 Perspectives de recherche

4.1 Approche Catégorielle

Les perspectives ouvertes par nos travaux sur l'approche catégorielle concernent principale-

[1] Collaboration avec David Sanders

ment l'animation faciale. Pour commencer, plusieurs questions restent ouvertes concernant la gestion de la dynamique des émotions et de leurs expressions faciales. En effet, si plusieurs travaux adressent ce problème (Pelachaud et al, 2005), le phénomène du *moment émotionnel* que nous avons mesuré pourrait permettre de proposer de nouvelles approches sur la gestion de la dynamique des émotions et de leurs expressions faciales.

Nos travaux n'abordent que partiellement les mélanges d'émotions. Pourtant, ces phénomènes sont régulièrement observés dans l'interaction humaine (Scherer, 1997, Abrillian et al. 2005), et des phénomènes tels que le masquage d'expressions, la politesse et plus généralement le contexte social, ont une grande influence sur nos émotions et la façon de les exprimer. L'approche catégorielle des émotions, par sa simplicité théorique, permet d'explorer ces phénomènes émotionnels et leurs expressions faciales.

4.2 Approche Dimensionnelle

Notre modèle dimensionnel inspiré de P.A.D. pourrait être étendu de plusieurs façons. D'une part, nous avons effectué un certain nombre de simplifications, comme le placement des émotions aux coins du cube. En utilisant un plus grand nombre d'émotions, localisées différemment dans l'espace dimensionnel, nous pourrions obtenir une expressivité plus fine. De plus, nos travaux sur les profils expressifs individuels montrent que l'approche dimensionnelle peut être mise en relation avec des paramètres individuels. Nos profils expressifs continus pourraient donc être étendus à des phénomènes affectifs à plus long terme, tels que les humeurs, où la personnalité.

4.3 Approche Cognitive

Notre premier modèle cognitif inspiré du modèle CPM, associé à l'animation faciale des évaluations séquentielles, nous a donc permis de montrer l'apport d'un modèle cognitif sur l'animation faciale en situation d'interaction. Cependant, pour permettre une interaction plus riche et à plus long terme, le modèle doit être étendu pour prendre en compte l'historique de l'interaction. De plus, pour permettre la simulation de raisonnements cognitifs plus complexes, le modèle nécessiterait une représentation plus détaillée de la situation et de l'état interne de l'agent virtuel. Par exemple, l'intégration d'un modèle de type *Belief-Desire-Intention* (Ochs et al, 2005, Rivière et al., 2011) permettrait de modéliser l'état interne de l'agent et d'apporter des informations importantes pour l'évaluation cognitive des événements. L'intégration d'un modèle BDI permettrait donc de rendre notre modèle cognitif plus complet et applicable à d'autres types d'applications.

Notre modèle cognitif pourrait également bénéficier de la prise en compte de l'état émotionnel de l'utilisateur. L'ajout de systèmes de capture d'expressions faciales et de capteurs physiologiques (Knapp et al., 2011) permettrait d'améliorer notre système cognitif et ainsi de créer une boucle affective complète.

4.4 Approche Sociale

En ce qui concerne notre modèle d'évaluation cognitive sociale, nos travaux sur le *social appraisal* ne sont que préliminaires. Le modèle que nous avons proposé est empirique et l'étude effectuée nécessite d'être étendue à d'autres émotions, et d'être menée sur un plus grand nombre de sujets. Cependant, à l'instar des travaux de Mumenthaler et Sander (2010), nos premiers résultats sont encourageants et suggèrent que notre approche est pertinente.

Pour étendre ce modèle, l'une de nos perspectives est d'utiliser l'utilisateur comme source de l'influence sociale. Ainsi, l'agent aurait un comportement social vis-à-vis de l'utilisateur. En effet, en plus de la prise en compte de l'état émotionnel de l'utilisateur dans le traitement, rendre l'agent capable d'interpréter les réactions faciales de l'utilisateur pour effectuer une seconde évaluation cognitive et sociale permettrait une plus grande adaptabilité de l'agent.

L'utilisation de deux agents pour l'interaction sociale ouvre également la voie à une autre problématique, celles des relations sociales et hiérarchiques. En effet, nous posons l'hypothèse que la relation de statut entre les deux personnages influence l'effet de l'évaluation cognitive sociale. Si l'un des agents

est le supérieur hiérarchique de l'autre, son influence sera probablement différente que si les deux agents ont le même statut. La modélisation de ces relations permettrait donc d'obtenir une évaluation sociale plus complexe, basée sur des informations sociales plus larges.
Une autre extension possible et ambitieuse de nos travaux serait d'intégrer notre modèle de réévaluation sociale dans les systèmes de simulations sociales multi-agents à plus large échelle. Ainsi, il serait possible d'augmenter le nombre d'agents pris en compte dans l'évaluation sociale. Nous pourrions ainsi simuler des évaluations cognitives sociales dans un groupe d'agents virtuels, par exemple, dans le cas d'interaction en réalité virtuelle, propice à l'utilisation de plusieurs agents virtuels (Fig 3).

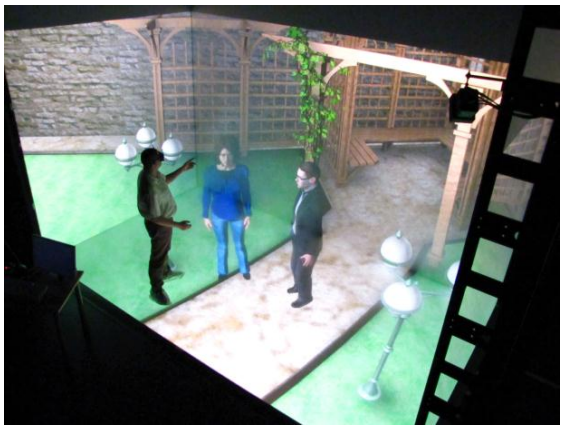


Fig. 3- MARC et Réalité Virtuelle dans EVE (Martin et al., 2011) pour l'interaction sociale avec plusieurs agents virtuels.

4.5 Perspectives générales

La plateforme logicielle MARC que nous avons développée permet d'explorer différentes approches des émotions. Si nos travaux se sont focalisés sur quatre d'entre elles, d'autres approches pourraient être pertinentes pour l'animation faciale interactive temps réel. Par exemple, dans la classification de Scherer (2010), on trouve les approches motivationnelles et adaptatives, que nous n'avons pas abordées.
De plus, nous n'avons pas étudié la relation entre les différentes approches des émotions afin de déterminer si elles sont ou non exclusives. En effet, le continuum proposé par Gross et Feldman-Barrett (2011) suggère que les différentes approches partagent certaines caractéristiques et qu'elles ne sont donc pas opposées. Du point de vue des modèles informatiques, nos travaux sur l'approche sociale mélangent l'approche cognitive et l'approche sociale. De plus, certains travaux (Becker-Asano et Wachsmuth, 2008) présentent également des systèmes tirant partie de plusieurs approches simultanément. Ainsi, si nos travaux présentent un certain nombre d'informations importantes pour choisir entre différentes approches des émotions, de nombreuses études sont encore nécessaires pour établir clairement les apports et les limites de chacune, ainsi que la manière de les combiner, pour se diriger vers un modèle intégratif, capable de combiner simultanément plusieurs approches des émotions.
De plus, nous n'avons pas considéré les phénomènes affectifs de plus longue durée, telles que l'humeur ou la personnalité. Pourtant, ces phénomènes sont importants pour la modélisation des comportements affectifs d'agent virtuels (André et al., 2000) et des agents relationnels (Bickmore et al., 2011). Ils impactent à la fois l'évaluation cognitive à l'origine de l'émotion et la manière d'exprimer l'émotion. Nos modèles cognitifs semblent pertinents pour permettre de prendre en compte ces phénomènes. En effet, il serait ainsi possible de moduler l'expressivité de l'agent et d'influencer la simulation du processus émotionnel en fonction de paramètres de personnalité (Clavel et Martin, 2009). Cependant, cela soulève de nombreuses problématiques, en particulier : comment modéliser et représenter la personnalité de l'agent ? Comment la personnalité de l'agent impacte son évaluation cognitive ? Pour la gestion des interactions prolongées avec l'agent, il semble également pertinent de considérer l'utilisation de la mémoire autobiographique (Ho et Watson, 2006), et de la mémoire émotionnelle (Kensinger et Corkin, 2004). Le modèle FATIMA (Dias et Païva, 2005) utilise par exemple une combinaison du modèle cognitif OCC et d'une mémoire autobiographique.

Le traitement des informations à plus long terme semble donc pouvoir apporter une cohérence dans une interaction à long terme avec l'utilisateur en évitant que l'agent ne devienne répétitif, et que l'utilisateur se désengage de l'interaction.

5 Conclusion

Les agents virtuels expressifs sont à l'intersection de nombreux domaines de recherche en informatique et en sciences humaines et sociales. Nos travaux se sont focalisés uniquement sur l'informatique affective (et plus particulièrement la simulation des émotions), et sur l'animation faciale interactive. Pourtant, les agents virtuels sont également liés à d'autres domaines, telles que la synthèse et la reconnaissance de la parole, l'intelligence artificielle, le traitement automatique des langues, la représentation des connaissances, la psychologie, les sciences sociales, etc. Ainsi, les agents virtuels sont un carrefour interdisciplinaire où de nombreuses collaborations scientifiques sont possibles. Les travaux que nous avons présentés nous ont permis de répondre à certaines questions de recherches, mais nous sommes encore loin de savoir simuler le comportement affectif humain dans toute sa complexité. Cet objectif ne peut être atteint que par une collaboration interdisciplinaire forte.

De plus, les agents virtuels expressifs peuvent servir d'outils pour étudier la perception humaine et la communication émotionnelle. En retour, ces études permettent de contribuer à l'amélioration des systèmes d'agents virtuels expressifs en fournissant des informations importantes et des règles de conception basées sur des validations perceptives. En appliquant ce principe d'enrichissement réciproque, nous espérons que nos travaux, ainsi que la plateforme MARC qui en résulte, contribueront à mieux comprendre et à améliorer l'interaction avec les agents virtuels expressifs réalistes et temps-réel.

6 Références

[1] Russell, J. (2012) Introduction to Special Section: On Defining Emotion, in Emotion Review, Vol 4 (4) pp 337-338

[2] Gross, J., & Feldman Barrett, L. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. Emotion review, 3(1), 8-16.

[3] Ekman, P., & Friesen, W. (1975). Unmasking the Face. A guide to recognizing facial clues. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[4] Tomkins, S. S. (1984). Affect theory. Dans K. Scherer, & P. Ekman, Approaches to emotion (pp. 163-195). Hillsdale, NJ: Erlbaum.

[5] Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. Approaches to emotion, 293-317.

[6] Niewiadomski, R., Bevacqua, E., Mancini, M., & Pelachaud, C. (2009). Greta: an interactive expressive ECA system. Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, (pp. 1399-1400).

[7] Kopp, S., & Jung, B. (2000). An Anthropomorphic Assistant for Virtual Assembly: MAX. Workshop on Communicative Agents in Intelligent Virtuel Environments.

[8] Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., & Bronnenkant, K. (2010). Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. Intelligent Virtual Agents, (pp. 286-300).

[9] Leite, I., Castellano, G., Pereira, A., & P., M. (2009). Designing a Game Companion for Long-Term Social Interaction. International Conference on Multimodal Interaction AFFINE Workshop.

[10] Cassell, J., Vilhjàlmsson, H., & Bickmore, T. (2001). BEAT: the behavior expression animation toolkit. Proceedings of the 28th annual conference on Computer graphics and interactive techniques, (pp. 477--486).

[11] Kopp, S., Stocksmeier, T., & Gibbon, D. (2007). Incremental multimodal feedback for conversational agents. Intelligent Virtual Agents, (pp. 139-146).

[12] Wallraven, C., Breidt, M., Cunningham, D., & Bülthoff, H. H. (2005). Psychophysical evaluation of animated facial expressions. 2nd Symposium on Applied Perception in Graphics and Visualization (pp. 17-24). ACM Press, New York, NY.

[13] Courgeon, M., Buisine, S., Martin, J-C. (2009) Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions, In: Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA 09) (pp 201-214), Amsterdam, The Netherlands, 10-12 septembre 2009

[14] Russell, J. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. Psychological Bulletin, 115(1), 102-114.

[15] Thornton, I. (1998). The Perception of Dynamic Human Faces. PhD Thesis, University of Oregon.

[16] Courgeon, M., Amorim, M-A., Giroux, C., Martin, J-C. (2010), Do Users Anticipate Emotion Dynamics in Facial Expressions of a Virtual Character?, in: Proceedings of the 23rd International Conference on Computer Animation and Social Agents (CASA 2010), Saint Malo, France, 31 mai - 2 juin 2010

[17] Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. Research on Personality 11(3), 273-294.

[18] Broekens, J. (2012). In Defense of Dominance: PAD Usage in Computational Representations of Affect. International Journal of Synthetic Emotions (IJSE), 3(1), pp 33-42

[19] Courgeon, M., Martin, J-C., Jacquemin, C. (2008) User's Gestural Exploration of Different Virtual Agents' Expressive Profiles, in: Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 08), vol 3 (pp 1237-1240), Estoril, Portugal, 12-16 mai 2008

[20] Scherer, K. R. (2001). Appraisals Considered as a Process of Multilevel Sequential Process. Emotion: Theory, Methods, Research, 92-120.

[21] M. Courgeon, C. Clavel, J-C. Martin, (2009) Appraising Emotional Events during a Real-time Interactive Game, in: Proceedings of the ICMI 2009 Workshop on Affective Computing (AFFINE), Cambrigde, U.S.A., 1-6 novembre 2009

[22] Manstead, A. S., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. Dans K. R. Scherer, A. Schorr, & T. Johnstone, Appraisal processes in emotion: Theory, methods, research. (pp. 221-232).

[23] Marsella, S., & Gratch, J. (2006). EMA: A computational model of appraisal dynamics. Agent Construction and Emotions.

[24] Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. ACM international conference on Multimedia, (pp. 683--689).

[25] Scherer, K. R., & Ceschi, G. (1997). Lost Luggage: A Field Study of Emotion–Antecedent Appraisal. Motication and Emotion, 21(3), 211-235.

[26] Abrilian, S., Devillers, L., Buisine, S., & Martin, J.-C. (2005). EmoTV: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces. HCI Internationnal.

[27] Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2005). Intelligent Expressions of Emotions. Affective Computing and Intelligent Interaction , (pp. 707-714).

[28] Riviere, J., Adam, C., Pesty, S., Pelachaud, C., Guiraud, N., Lorini, E., et al. (2011). Expressive Multimodal Conversational Acts for SAIBA agent. Intelligent Virtual Agents, (pp. 316–323).

[29] R., Kim, J., & André, E. (2011). Physiological Signals and Their Use in Augmenting Emotion Recognition for Human--Machine Interaction. Emotion-Oriented Systems, 133-159.

[30] Mumenthaler, C., & Sander, D. (2010). Social Appraisal, how the Evaluation of Others Influences our Own Perception of Emotional Facial Expressions. XVII Annual Cognitive Neuroscience Society Meeting, (pp. 17-20).

[31] Devillers, L., Vidrascu, L., & Lamel, L. (2005). Emotion detection in real-life spoken dialogs recorded in call center. Journal of Neural Networks, Emotion and Brain, 18(4), 407-422.

[32] Kulms, P., Krämer, N. C., Gratch, J., & Kang, S. H. (2011). It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze. Intelligent Virtual Agents, 80–92.

[33] Schroder, M., Burkhardt, F., & Krstulovic, S. (2010). Synthesis of Emotional Speech. Blueprint for affective computing, 222-231.

[34] Scherer, K. R. (2010). The component process model: a blueprint for a comprehensive computational model of emotion. Dans K. Scherer, T. Bänziger, & E. Roesch, A Blueprint for Affective Computing: A sourcebook and manual.

[35] Becker-Asano, C., & Wachsmuth, I. (2008). Affect Simulation with Primary and Secondary Emotions. Intelligent Virtual Agent, (pp. 15-28).

[36] Martin, P., Bourdot, P. ; Touraine, D. (2011) A reconfigurable architecture for multimodal and collaborative interactions in Virtual Environments, in: 3D User Interfaces (3DUI) 2011, pp 11-14

[37] André, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. Affective interactions, 150-165.

[38] Bickmore, T., Pfeifer, L., & Schulman, D. (2011). Relational Agents Improve Engagement and Learning in Science Museum Visitors. Intelligent Virtual Agents, (pp. 55-67). Reyjavik, Iceland.

[39] Clavel, C., & Martin, J. C. (2009). PERMUTATION: A Corpus-Based Approach for Modeling Personality and Multimodal Expression of Affects in Virtual Characters. Digital Human Modeling, 211-220.

[40] Ho, W., & Watson, S. (2006). Autobiographic knowledge for believable virtual characters. Intelligent Virtual Agents, 383-394.

[41] Kensinger, E., & Corkin, S. (2004). Two Routes to Emotional Memory: Distinct Neural Processes for Valence and Arousal. National Academy of Sciences of the United States of America, (pp. 3310-3320).

[42] Dias, J., & Païva, A. (2005). Feeling and reasoning: a computational model for emotional agents. EPIA, 127-140.

Interpersonal stance recognition using non-verbal signals on several time windows

Mathieu Chollet, Magalie Ochs, Catherine Pelachaud
{mchollet, mochs, cpelachaud}@telecom-paristech.fr

CNRS-LTCI, Telecom Paristech
75013 PARIS – FRANCE

Abstract:
We present a computational model for interpreting non-verbal signals of a user during an interaction with a virtual character in order to obtain a representation of his interpersonal stance. Our model starts, on the one hand, from the analysis of multimodal signals. On the other hand, it takes into account the temporal patterns of the interactants behaviors. That is it analyses signals and reactions to signals in their immediate context, as well as features of signal production patterns and reaction patterns on different time windows : signal reaction, sentence reaction, conversation topic, whole interaction. In this paper, we propose a first model parameterized using data obtained from the literature on the expressions of stances through interpersonal behavior.

Keywords: Interpersonal stance, non-verbal behavior interpretation, Social Signal Processing

1 Introduction

The last two decades have seen a surge of interest in the field of Human-Computer Interaction for the introduction of Embodied Conversational Agents (ECA) in various application domains, such as interactive storytelling [8], virtual learning environments [17], healthy behaviour promotion [7], or museum guides[13]. One of the major reasons behind this strong movement is that some studies found that using ECAs improved the experience of human-computer interaction, by making learning activities easier to follow [17] or by enhancing the degree of trust users had in relationship with their computer [21].

Moreover, several researchers have recently focused on Social Signal Processing (SSP). The objective of SSP is to allow computers to recognize social information, such as boredom, politeness, or interpersonal stances [22].

One of the ways to make ECAs more believable when they interact with a user, is to give them the capability to adapt themselves to that user's interpersonal stance. For example, in the context or a virtual learning environment, it would be useful for a virtual teacher to detect when a learning user feels embarassed, as it would allow the ECA to adapt its behavior and its teaching strategy. Social Signal Processing provides valuable tools to build ECAs that are capable of reacting to a user.

The work presented in this paper is part of TARDIS, a FP7 funded project whose objective is to help with inclusion of the increasing number of young Europeans not in employment, education or training. The vision of the TARDIS project is to give young people a tool to train their social skills : a serious game of job interviews simulation, that will help them improve their chances of getting a job. One of the research challenges we have, in the context of TARDIS, is the recognition of interpersonal stances. Indeed, in a job interview, recruiters try to assess social skills of candidates by judging their interpersonal dispositions and social attitudes : to improve their performance, candidates thus have to adapt their behavior by strategically adopting the appropriate stance at every moment. For example, when discussing management skills, a job candidate may want to appear dominant, and when discussing team-working abilities, a job candidate may want to appear friendly and not too dominant.

In the TARDIS platform, the recruiter will be a virtual recruiter enacted by an ECA and the recognition of social signals will be automated, using sensors such as webcams and microphones. Therefore, the TARDIS project is a perfect example of the combination of Social Signal Processing and Embodied Conversation Agents : we have to detect the user's interpersonal stance in real-time and we have to know how the virtual recruiter should react to it. One way to react is to express a particular interpersonal stance : for instance the virtual recruiter may decide to express coldness to a dominant user. The context of the interaction should be considered both for detecting the user's interpersonal stance and for deciding which interpersonal stance the agent should express.

In Social Signal Processing, there is still significant progress to be made for social stance recognition. As Pantic recently states in [15] :

> *« despite a significant progress in automatic recognition of audiovisual behavioural cues underlying the manifestation of various social signals, most of the present approaches to machine analysis of human behaviour are neither multimodal, nor context-sensitive, nor suitable for handling longer time scales. In turn, most of the social signal recognition methods reported so far are single-modal, contextinsensitive and unable to handle long-time recordings of the target phenomena. »*

This paper proposes a model for interpersonal stance recognition aiming at tackling these issues, namely analysing multimodal social signals on several temporal scales, taking the context into account. For the temporal issue, we propose to use different time windows of analysis. Signals do not necessarily convey the same information in every time window. For instance, a smiling person might be interpreted as friendly, whereas someone who smiles in response to criticism can be seen as arrogant. We consider signals from different modalities [1].

The rest of the paper is organised as follows. Section 2 presents related works on perception of ECA's interpersonal stances, multimodal social signal recognition and human-ECA interaction driven by users' affect recognition. In section 3, we introduce the definitions we use in our model for the notions of interpersonal stance, social and verbal signals, reactions, features, and time windows. In Section 4 is proposed a first version of our model based on Social and Human Sciences studies. In Section 5, we conclude and discuss future works.

2 Related Work

2.1 Perception of interpersonal stances in agents

In order to study perception of social attitudes, some researchers generated different ECAs behavior expressions showing different interpersonal stances. Users then had to rate how they perceived the agent.

For instance, Fukayama *et al.* [10] have proposed a gaze movement model for embodied agents based on three parameters : the amount of gaze directed at the interlocutor, the mean duration of gaze directed at the user, and the gaze points while averting gaze. They found that variation of these three parameters allowed their agents to convey different impressions of dominance and friendliness to users.

Bee *et al.* [4] studied the relationsip between signals expressed on several different modalities and the perception of social dominance of an ECA. They analysed the relationship between different facial expressions of emotions (joy, fear, anger, surprise, disgust, neutral), different head and gaze orientations, and how users perceived the dominance of the resulting face. They showed that variations of gaze and head orientations do not always have the same effects depending on the displayed emotion. In [5], they looked at the relationship between the head, gaze orientations and parameters of sentence generation (more or less extraverted or agreeable), and found that both the verbal and non-verbal modalities have an effect on the perception of the ECA's dominance.

In [2], Arya *et al.* studied the effect of facial expressions of ECAs on the perception of their interpersonal stance, by displaying videos of an ECA displaying a specific expression with a certain speed. They then asked users to choose an adjective from a list that suited best the expression. The list of adjectives only contained words characteristic to a specific region of the interpersonal circumplex, a bidimensional representation of interpersonal stances (See §3.1 for more details). As a result, they were able to link the facial expressions to specific points on the interpersonal circumplex, thus providing a direct mapping from behavior to interpersonal stance.

These works highlight the fact that ECAs are capable to convey different interpersonal stances through non-verbal behavior. In the next section, we present existing works on multimodal social signal recognition.

2.2 Multimodal social signals recognition

Wagner *et al.* [23] proposed a framework called *SSI* (Social Signal Interpretation), for designing online recognition systems. This framework supports inputs from a variety of sensors and is equipped with algorithms to perform multimodal fusion. In a sample application [23], *SSI* was plugged in with the *Alfred* agent [4]. The agent mirrors the user's emotional state by using

1. However multimodality (i.e. combinations of signals meaning more than just a juxtaposition of signals) will only be considered in ulterior versions of the model.

appropriate facial expressions. The recognition of the user's emotional state is based on both audio and video signals, and yields a dimensional representation of the user's affect in terms of pleasure and arousal [23].

Few attempts have been made for estimation of the most dominant person in a small group meeting [18] [12]. However those works are offline methods for groups of people and might not be applicable in our setting, i.e. real-time human-machine interaction. They still bear some insight as to what nonverbal signals are the most relevant in assessing perception of dominance. The strongest cues were found in most cases to simply be the total speaking time of the participants.

As shown in the above presented works, systems for affect recognition have been proposed. However, the recognition of interpersonal stances has yet to be attempted during a real-time interaction. In the next section, we present works where users' social signals are used to drive the interaction with ECAs.

2.3 Interactions using social signals

Some recent systems have used users' social signals to drive human-ECA interactions. Cavazza *et al.* [9] use emotional speech to drive an interactive narrative taking place within an adaptation of Flaubert's *Madame Bovary*. Emotional features of the user's voice are recognised by the system and are used as part of the scenario planning. One of the main advantages of their approach is that the interaction is driven without any verbal recognition or semantic interpretation, which allows for completely free speech from the users, while still allowing for variability in the scenarii.

As part of the SEMAINE project [20], an integrated platform of Sensitive Artificial Listeners was developped. It consists of affect recognition modules (video, audio inputs) that are fed to a listener model developped by Bevacqua *et al.* [6]. Different listeners with different personalities were implemented. The rules of signal production are dependent on the ECA personality : the enthusiastic and cheerful Poppy will often mimic the user's behavior and will produce a lot of backchannels, while the hostile Spike will display social signals that are contrary to those expressed by the user.

For now, most works in analysis of social signals in interaction have focused on recognition of users' emotions. In contrast, interpersonal stances have not received much attention. However, endowing agents with the capability of detecting interpersonal stances, and colouring their behavior with interpersonal stances, would enable enhancing human-ECA interactions. This paper provides a computational model for users' interpersonal stance recognition in human-ECA interaction.

In the next section, before diving into the details of our model, we introduce definitions for some of its central notions.

3 Definitions

3.1 Interpersonal stance

In [19], Scherer provides a specification for the attributes that differentiate the types of affective phenomena : emotions, moods, attitudes, preferences, affect dispositions, and interpersonal stances. For him, the specificity of interpersonal stances is that

> *« it is characteristic of an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons, coloring the interpersonal exchange in that situation (e.g. being polite, distant, cold, warm, supportive, contemptuous). »*

Attitudes towards others are mapped by Argyle [1] on two dimensions : Dominant/Submissive and Friendly/Hostile. This is in line with works on interpersonal behavior that consistently found these two axes accounted for most of the non-verbal behavior variations, such as the Interpersonal Circumplex proposed by Wiggins [24] (See Fig.1).

Based on Argyle's attitude dimensions and Wiggin's interpersonal circumplex axes, we propose to use two dimensions to represent users' interpersonal stances : *friendliness* (also called warmth or affiliation) and *dominance* (also called agency).

A user can express a general stance in the whole interaction, a stance when discussing a certain topic, and a stance in reaction to specific signals or sentences. For instance, in a job interview, a candidate might be embarassed by a question on a specific skill he does not

have, even though he might have appeared as confident on the overall topic of discussing his skills. Therefore we want to recognize the interpersonal stance a user U expresses in reaction to a signal ($SignalStance_U$), to a sentence ($SentenceStance_U$), during the time a specific topic is discussed ($TopicStance_U$), and on a whole interaction ($InteractionStance_U$).

For an agent U (virtual or human), all of these stances are formally represented as a combination of a *dominance* and a *friendliness* :

$$Stance_U = \{Dom_U, Frnd_U\}$$

with $Dom_U, Frnd_U \in [-1, 1]$ representing respectively the *dominance* and the *friendliness* expressed by an agent U. The more Dom_U (resp. $Frnd_U$) is close to 1, the more dominant (resp. friendly) is the agent's interpersonal stance. The more Dom_U (resp. $Frnd_U$) is close to -1, the more submissive (resp. hostile) is the agent's interpersonal stance.

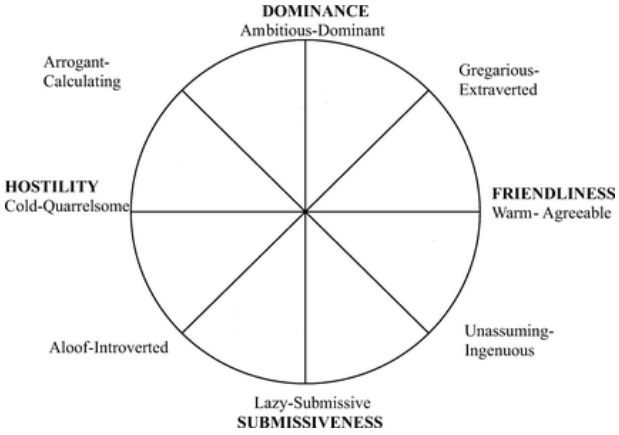


FIGURE 1 – An interpersonal Circumplex with prototypical interpersonal stances every 45°

3.2 Social signals

The notion of signal varies a lot depending on the domain of study. In Social Signal Processing (SSP), a general consensus over the definition of what is a social signal is hard to find. Based on a study of the different definitions of a signal in different disciplines, Vinciarelli *et al.* [22] propose the following definition of social signals :

> *« A Social signal is a communicative or informative signal that, either directly or indirectly, provides information about social facts, namely social interactions, social emotions, social attitudes, or social relations. »*

In our case, we rely on external software that sends messages when it detects non-verbal signals. We thus consider that a non-verbal signal is an input message characterized by a starting time t_{start}, an end time t_{end}, and a non-verbal body modality (e.g. gaze, facial expression, voice). For each of these modalities, a number of additional relevant variables are defined. For instance, a signal from the gaze modality also contains two angles, one used to know the direction of gaze aversion (angle around the head front axis), and one to know how much gaze is averted (angle between the head front axis and the gaze direction axis).

In our model, we propose to classify these non-verbal messages in specific types depending on the values of their variables. For instance, when the user is not looking at the ECA, the user's gaze signal is classified as a *gazeAway* signal. Research showed that gaze [10], head orientation [4], smiles [14] and speech [18] were found to reflect users' interpersonal stances. Although other non-verbal signals can be related to it, we choose as a first step to consider those modalities in our model :

- Gaze :
  - *gazeFront*, when the user's gaze is directed at the ECA
  - *gazeAway*, when the user's gaze is not directed at the ECA
- Head orientation :
  - *headFront*, when the user's head is directed at the ECA
  - *headUp*, when the user's head is directed upwards
  - *headDown*, when the user's head is directed downwards
  - *headSide*, when the user's head is directed sideways
- Facial expressions : *smile*, when the user is smiling [2]
- Voice : *speech*, when the user is speaking [2]

3.3 Verbal signals

In the context of the TARDIS platform, keyword spotting is implemented but this is not sufficient for speech recognition and understanding of the user. However, in the context of an interaction with a virtual recruiter, we know in advance the sentences it utters as they are formally represented in a dialogue manager, Disco [16].

2. In a later version, we will define more types using other variables such as voice pitch, smile intensity. For the sake of simplicity we consider very few types.

The types of these sentences (praise, criticisms, etc...) should be considered to analyse the user's reactions. Indeed, a signal may be interpreted differently depending in reaction to what it is expressed. For instance, a smile can be considered as arrogance when it is expressed in reaction to criticism, whereas a smile in response to praise might be interpreted as pride.

To categorize the ECA's sentences, we base ourselves on the typology proposed by Bales in his Interaction Process Analysis (IPA) Theory [3] (See Table 1). In our model, we consider three categories : sentences can either be socio-emotional positive (noted SEpos), socio-emotional negative (noted SEneg), or task-oriented (noted TO) when the sentence is a question or an answer [3].

Type	Categories
Socio-Emotional Positive *SEpos*	1. Seems friendly 2. Tension release 3. Agrees.
Attempted Answers *A*	4. Gives suggestion 5. Gives opinion 6. Gives information
Questions *Q*	7. Asks for information 8. Asks for opinion 9. Asks for suggestion
Socio-Emotional Negative *SEneg*	10. Disagrees 11. Shows tension 12. Seems unfriendly

TABLE 1 – Bales IPA categories [3]

3.4 Reactions

The signals we described in previous section can be in isolation, that is they are displayed spontaneously by the person with no direct relation to the other interactant's behavior. But signals can also be expressed in direct reaction to the other interactant's behavior.

An isolated signal is noted simply as : $s_{isolated}$. A reaction R is another kind of signal that is expressed in reaction to another signal s_{origin}. This relationship is noted in the following manner :

$$R \leftarrow s_{origin}$$

Determining if a signal is isolated or in reaction to another signal is a hard problem. In a simplifying assumption, we consider that if a signal s_{origin} is sent by the person A, then any signal $s_{reaction}$ sent by B during the $\Delta_{REACTION}$ time window of length δ (See next section) is a reaction to s_{origin}. In a more formal notation, we have :

$$\begin{aligned} if \quad & s_{reaction} : t_{start} \in [s_{origin} : t_{start} + \delta, \\ & \qquad\qquad s_{origin} : t_{end} + \delta] \\ then \quad & s_{reaction} \leftarrow s_{origin} \end{aligned}$$

In the next section, we explore more particularly the features of signals and of reactions produced that are relevant in assessing interpersonal stance.

3.5 Features

The analysis of affective phenomena has to be done at different temporal levels depending on the type of phenomena considered. An emotion is a strong local phenomenum (even though it can last), and considering signals on a short time window might be enough to detect them. Interpersonal stance, on the other hand, is inferred on longer temporal scales, by analysing reccuring tendencies in behavior, and not only single signals occurences at a particular point in time.

For every type of signals (e.g. smiles or gaze aversion) we define relevant features to assess users' interpersonal stance. These features are used to evaluate more stance characteristics of the signal productions : for instance, if the user responds to a smile of the agent by another smile, it can be considered as a sign of friendliness but it is not sufficient to infer that the user has a general friendly stance. On the other hand, the amount of smile reactions the user has produced in reaction to agent smiles on a longer time scale gives us additional information.

In our model, we consider the following kinds of features :

- amount of a signal type (e.g. percentage of $gazeFront$, or number of smiles)
- mean duration of a signal type (in seconds)
- amount of reactions of a type after the agent utters a *Socio-emotional positive* sentence
- amount of reactions of a type after the agent utters a *Socio-emotional negative* sentence
- amount of reactions of a type after the agent utters a *Task-Oriented* sentence

3. We propose to simplify Bales' typology by merging the Questions and Attempted Answers categories in a single Task-Oriented category

3.6 Time windows

Non-verbal signals give out cues about the mental state of the person that displays them. For instance, seeing a person suddenly frown their brows, clench their fists and raise their voice energy are cues that hint this person is angry at this precise moment.

However, as Scherer points out [19], all kinds of affect don't happen in the same span of time. For instance, emotions have a very short duration, and to assess a person's emotion, one should only look at this person's very recent displays of emotion in their non verbal behavior. For moods, one has to look at a person's non-verbal behavior on a longer time span. It might get even longer to get a good sense of someone's interpersonal stance.

Therefore, to recognize interpersonal stances, we have to consider non-verbal behavior on different time spans. For this purpose, we define four time windows of analysis.

Signal reaction window. The signal reaction window, noted Δ_{SIGNAL}, aims at detecting reactions from the user to *non-verbal signals* of the ECA in order to compute the user's $SignalStance_U$ for this signal. For a signal S expressed by the ECA, the Δ_{SIGNAL} window is very short, starting from the signal's start time and lasting for a small constant δ. δ is the length of the time frame in which we can consider that an interlocutor's signal is still in reaction to S. In this time window, we are interested in the types of user signals expressed in reaction to S.

Sentence reaction window. The $\Delta_{SENTENCE}$ window aims at detecting reactions from the user to *sentences* uttered by the ECA in order to compute the user's $SentenceStance_U$ for this signal. Its starting time is the point where the ECA starts the sentence, and it lasts until either the user takes the floor and finishes talking or the agent starts another sentence. In this time window, we are interested in the types of user signals expressed in reaction to the ECA's sentence types : socio-emotional positive, socio-emotional negative, or task-oriented (See Section 3.3).

Dialogue topic window. Users can have a specific interpersonal stance regarding a particular discussed topic. For instance, someone can be embarassed when discussing personal matters in an official context. Therefore, we consider a specific window for every topic discussed, and we use it to compute the user's $TopicStance_U$. The Δ_{TOPIC} window starts when a new topic is being discussed. To represent the topic discussed during the dialog, we use a dialogue model based on hierarchical task networks. We consider that this window begins when a new top-level task (e.g. greetings, discuss resume, discuss job experience) starts, and ends when another top-level task starts. The features used in this time window are described in the Section 3.5.

Interaction window. Finally, the user's $InteractionStance_U$ represents the global stance that a user has expressed through an interaction. It is computed on the $\Delta_{INTERACTION}$ time window, that spans from the beginning of the interaction to its end. The features used in this time window are the same as the ones used to compute the user's stance towards the topic, and are described in Section 3.5.

This section has introduced the notions that are used in our model. In the next section, we present how stance is computed.

4 Computation of interpersonal stance

4.1 Problem definition

Our model aims at computing the stance of a user in reaction to a specific signal ($SignalStance$), to a sentence ($SentenceStance$), within a discussion topic ($TopicStance$) or in an entire interaction ($InteractionStance$).

In essence, our problem is to compute for one kind of stance, both the *dominance* (Dom_U) and *friendliness* ($Frnd_U$), from a set of input variables $X = \{x_i; 1 \leq i \leq n\}$, where each x_i is one of the n features used for that stance (see Section 3.6). We want to find the functions D and F such that $Dom_U = D(X)$ and $Frnd_U = F(X)$.

In a simplifying assumption, we suppose that the input variables are independent, which allows us to split the problem of finding the functions D and F into smaller problems of finding the relationship between an input variable and *dominance* and *friendliness* independently of the others. Specifically, we suppose

that $D(X) = \sum_{i=1}^{n} Dw_i * D_i(x_i)$, where each D_i models the relationship between the variable x_i and dominance independently of other signals, and Dw_i is a weighting factor. The same supposition is made for friendliness, so we have $F(X) = \sum_{i=1}^{n} Fw_i * F_i(x_i)$

4.2 Relationships between input variables and stance

In our case, the relationship between dominance or friendliness and the non-verbal behavior is not always close to linear. For instance, studies on gaze and mutual gaze [10] have shown that a medium to high amount of gaze are rated neutral or slightly positively on the friendliness scale, but a low or very high amount of gaze are rated as negative on the same scale.

The psychological litterature provides good insights about the general properties of the relationship between interpersonal stance and non-verbal behavior. However, precise mappings between these signal patterns and the interpersonal stance dimensions are hard to find.

Considering this, it is hard to make strong assumptions concerning the precise shape of the relationship between patterns of non-verbal signals and perception of interpersonal stance (e.g. logarithmic vs exponential...). Then, in order to use this knowledge while refraining from making too strong assumptions on the shape of these functions, we decide to adopt piece-wise linear function shapes. That is, we consider that the functions that map features to dominance and friendliness are linear on intervals. Using data from reports such as [10], we can find appropriate intervals and slopes for these functions. For instance, in [10], dominance is rated from -1 to -1 for amounts of gaze in $\{25\%, 50\%, 75\%, 100\%\}$. We can then draw a piece-wise linear function that passes through those points (See Fig. 2).

4.3 Tuning the weights of stance equations

Once the shape of the functions D_i and F_i have been found, the only thing remaining is to adjust the corresponding weights (the Dw_i and Fw_i) to reflect the contribution of every input variable with respect to the stance perception.

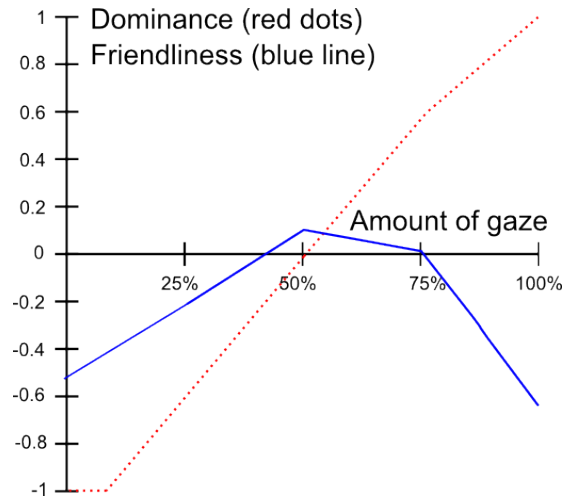


FIGURE 2 – Example of piece-wise linear function of dominance (red dots) and friendliness (blue line), based on [10].

However, as we have not gathered data yet, we rely once again on psychological knowledge to tune the first version of our model. More specifically, in [11], Gifford computes correlations between specific non-verbal modalities occurences and perceptions of interpersonal stance. The more correlated is the non-verbal behavior with dominance or friendliness, the strongest weight we assign to it.

Once those two steps are done, the model can be used online to compute the perceived interpersonal stance of the user, in reaction to a signal or a sentence, and within a certain topic or an entire interaction.

5 Conclusion

We have presented a computational model for interpersonal stance recognition. This model takes into account the interactional nature of a conversation, by considering that a spontaneous signal gives different information than a signal in reaction to an other person's behaviour. It also analyses behavior on different temporal patterns, by using several time windows.

In this first version we tuned the system using data from psychological litterature. In a next step, we plan on learning the parameters of the model using real data. In the TARDIS project, enacments of job interviews have been simulated and videos of these interviews have been recorded. To achieve that goal we will annotate these videos with occurences of non-verbal si-

gnals and interpersonal stances ratings, and then use them as the data for learning the parameters of our model.

We also want to tackle the issue of multimodality : the combination of non-verbal signals can mean something different than just the sum of them. For instance, clenching one's fist can mean anger, and smiling can indicate friendliness. However, the combination of both is used when celebrating success.

6 Acknowledgement

This research has been partially supported by the European Community Seventh Framework Program (FP7/2007-2013), under grant agreements no. 231287 (SSPNet) and 288578 (TARDIS.

References

[1] M. Argyle. *Bodily Communication*. London : Methuen, 2nd edition, 1988.

[2] A. Arya, L. N. Jefferies, J. T. Enns, and S. DiPaola. Facial actions as visual cues for personality. *Computer Animation and Virtual Worlds*, 17(3-4) :371–382, 2006.

[3] R.F. Bales. *A Set of Categories for the Analysis of Small Group Interaction.Channels of Communication in Small Groups*. Bobbs-Merrill, 1950.

[4] N. Bee, S. Franke, and E. Andrea. Relations between facial display, eye gaze and head tilt : Dominance perception variations of virtual agents. *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7, 2009.

[5] N. Bee, C. Pollock, E. André, and M. Walker. Bossy or Wimpy : Expressing Social Dominance by Combining Gaze and Linguistic Behaviors. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271. Springer Berlin / Heidelberg, 2010.

[6] E. Bevacqua, E. De Sevin, S. J. Hyniewska, and C. Pelachaud. A listener model : introducing personality traits. *Journal on Multimodal User Interfaces, special issue Interacting ECAs*, 2012.

[7] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions On Computer-Human Interaction*, 12(2) :293–327, June 2005.

[8] M. Cavazza, F. Charles, and S. J. Mead. Character-based interactive storytelling. *IEEE Intelligent Systems*, 17(4) :17–24, July 2002.

[9] M. Cavazza, D. Pizzi, F. Charles, and E. Vogt, T.and André. Emotional input for character-based interactive storytelling. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '09, pages 313–320, Richland, SC, 2009.

[10] A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents — impression management with agent's gaze. *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, (4) :41–48, 2002.

[11] R. Gifford. Mapping nonverbal behavior on the interpersonal circle. *Journal of Personality and Social Psychology*, 61(2) :279–288, 1991.

[12] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *Transactions on Audio, Speech and Language Processing*, 17(3) :501–513, March 2009.

[13] S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth. A conversational agent as museum guide : design and evaluation of a real-world application. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, editors, *Lecture Notes in Computer Science*, pages 329–343, London, UK, 2005. Springer-Verlag.

[14] E. Krumhuber, A. S. R. Manstead, and A. Kappas. Temporal aspects of facial displays in person and expression perception. the effects of smile dynamics, head-tilt and gender. *Journal of Nonverbal Behavior*, 31 :39–56, 2007.

[15] M. Pantic, R. Cowie, F. D'Errico, D. Heylen, M. Mehu, C. Pelachaud, I. Poggi, M. Schröder, and A. Vinciarelli. *Social Signal Processing : The Research Agenda*, pages 511–538. Springer, London, 2011.

[16] C. Rich and C. L. Sidner. Procedural dialogue authoring with hierarchical task networks and dialogue trees. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents*, IVA'12, 2012.

[17] J. Rickel and W. Lewis Johnson. Animated agents for procedural training in virtual reality : Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13 :343–382, 1998.

[18] R. J. Rienks and D. Heylen. Automatic dominance detection in meetings using easily detectable features. In *Workshop on Machine Learning for Multimodal Interaction*, Edinburgh, U.K, 2005.

[19] K. R. Scherer. What are emotions ? and how can they be measured ? *Social Science Information*, 44 :695–729, 2005.

[20] M. Schröder. The SEMAINE API : Towards a Standards-Based Framework for Building Emotion-Oriented Systems. *Advances in Human-Computer Interaction*, 2010 :1–21, 2010.

[21] X. Van Mulken, E. André, and J. Müller. The persona effect : How substantial is it ? In H. Johnson, L. Nigay, and C. Roast, editors, *People and Computers XIII, Proceedings of HCI '98*, pages 53–66. Springer, 1998.

[22] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schröder. Bridging the gap between social animal and unsocial machine : A survey of social signal processing. *IEEE Transactions on Affective Computing*, 3 :69–87, 2012.

[23] J. Wagner, F. Lingenfelser, N. Bee, and E. André. The social signal interpretation framework (SSI) for real time signal processing and recognition. *Proceedings of Interspeech 2011*, 2011.

[24] J. S. Wiggins. *Paradigms of Personality Assessment*. New York : Guilford, 2003.

Un modèle affectif pour un recruteur virtuel dans le contexte de simulation d'entretiens d'embauches

H. Jones⋆ hazael.jones@lip6.fr
N. Sabouret† nicolas.sabouret@limsi.fr
M. Smail Gondré⋆ melissa.smail.gondre@gmail.com

⋆LIP6 - Laboratoire d'Informatique de Paris 6
4 place Jussieu
75005 PARIS – FRANCE

†LIMSI - Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
91403 ORSAY – FRANCE

Résumé :
Le nombre de jeunes sans emploi et sans formation augmente en Europe. Ces jeunes manquent souvent de confiance en eux et ne maîtrisent pas les compétences sociales requises pour trouver du travail (le « savoir être » en entretien d'embauche). Le projet TARDIS [1] a pour objectif de réaliser un jeu sérieux de simulation d'entretien d'embauche à la première personne, pour permettre à ces jeunes de pratiquer et d'améliorer leurs compétences.
Dans cet article, nous nous intéressons plus particulièrement au modèle affectif de l'agent virtuel qui joue le rôle du recruteur. Notre objectif est de définir un modèle réaliste dans le cadre d'un entretien d'embauche. Il est composé d'émotions, d'humeur et d'attitudes sociales avec des dynamiques à court terme et à moyen terme afin de créer un recruteur virtuel émotionnellement réaliste. Nous présentons les différentes dimensions considérées et nous montrons ses propriétés sur quelques exemples simples. Ce modèle est actuellement en cours d'intégration dans le moteur de simulation du projet TARDIS.

Mots-clés : Modèle affectif, Émotions, Humeurs, Attitudes sociales, Entretien d'embauche.

Abstract:
The number of young people not in employment, education or training is increasing across Europe. These youngsters often lack self-confidence and the essential social skills needed to seek and secure employment. The TARDIS project aims to build a scenario-based serious-game simulation platform for young people at risk of exclusion to improve their social skills.
This paper presents a model for a socio-emotionally realistic virtual agent in the context of job interview simulations. Our model of affects is composed of emotions, moods, social attitudes and personality in order to create a realistic virtual recruiter.

Keywords: Affective model, Emotions, Moods, Social attitudes, Job interview.

1 Introduction

The number of NEETs [2] is increasing across Europe. According to Eurostat, in march 2012, 5.5 million of European youngster (16 to 25 years old) were unemployed meaning that 22.6% of the youngster global population in European union is unemployed. This unemployment percentage is 10 points superior to the whole population showing that the employment of NEETs is a real problem in Europe.

Current research reveals that NEETs often lack self-confidence and the essential social skills needed to seek and secure employment [3]. To help those young people to access jobs, youth inclusion organisations across Europe provide social coaching programmes, in order to help young people acquire and improve their social competencies, especially in the context of job interviews. The TARDIS project, funded by FP7, aims at building a serious game for NEETs and employment/inclusion organisations which supports social training and coaching in the context of job interviews. Youngsters (aged 18 to 25) will be able to explore, practice and improve their social skills in a diverse range of possible interview situations. Using serious gaming for job interview simulations shows two advantages : 1) repeatable experience can be modulated to suit the individual needs and ; 2) technologies are intrinsically motivating for the young [14] and shall help to remove the many barriers that real-life situations may pose, in particular the stress associated with engaging in unfamiliar interactions with others.

In the TARDIS project, the youngster faces a virtual agent acting as a recruiter. This paper presents an Artificial Intelligence model for such socio-emotionally realistic virtual agents. Our model is used to decide which attitude, emotion and mood should be displayed by the virtual agent, and to control the selection of relevant responses, using an internal representation of the user and the recruiter's mental states.

Indeed, it has been proven that the socio-emotional aspect is one of the key feature that

1. TARDIS stands for Training young Adult's Regulation of emotions and Development of social Interaction Skills. url : www.tardis-project.eu

2. NEET is a government acronym for young people not in employment, education or training.

distinguishes a machine from a believable agent [20]. Based on that, numerous tutor applications based on educational agents have been proposed [9, 18, 19] and this research domain, called *Affective Computing* [21], is still in expansion. One core issue in this domain is to build agents that react in a coherent manner : based on the non-verbal inputs (smiles, emotion expressions, body movements), the agent must select relevant verbal and non-verbal responses. The model presented in this paper tries to consider all the different dimensions of the socio-affective interaction, in the context of the job interview situation.

Many job interviews focus on the personality of the applicant. This has been encouraged by the fact that some personality traits predict job performance [2]. However, in a face to face job interview, the personality of the applicant is inferred by the recruiter according to the mood, the emotions and the social attitudes he expressed [8]. Furthermore, it has been proven that visual and vocal perceptions affect interviewers' judgements during an employment interview [5]. For these reasons, the affective model of the youngster in our work is composed of emotions, moods and social attitudes to evaluate the quality of the applicant performance. In order to have a believable simulation, the virtual recruiter must have a credible way to interact with the applicant. The use of affects in the model of our virtual recruiter allows this credibility.

This paper is organised as follows. Section 2 presents existing cognitive architectures in Affective Computing related to our goal and gives the motivation of our work. Section 3 briefly describes the architecture of the TARDIS affective model and its relation to the other project components. Section 4 detailed the affective model of the virtual recruiter and section 5 shows how theses values evolve over time and influence the agent's behaviour. Section 6 illustrates the model on a job interview scenario. The last section concludes the paper by presenting the project's next stages.

2 State of the art

In [23], a study shows that people who tried to suppress or hide negative emotions during a job interview are considered more competent by evaluators. Thus, emotion regulation is a key element to obtain a job. Emotions expression are regulated by situative norms according to social display rules [6]. Similarly, Tiedens [25] shows that anger and sadness play an important role in the job interview.

Several models have been proposed in the domain of affective computing [3] to build credible virtual human based on cognitive models of emotions [17, 9], personality [22] and social relations [16]. However, to our best knowledge, no computational model of social attitude have been proposed. Social attitudes are the expression of the personality of an agent through its behaviour and its emotional expressions, in the context of social norms. For example, in the context of a job interview, the social attitudes tells the recruiter a lot about the interviewee's personality and feelings about the job. This information will influence the way of leading the interview for the virtual recruiter and might decide for a *yes* or a *no* at the end. In that sense, it raises questions that are being studied in Theory of Mind [13] and reverse appraisal [8].

In our model, we use an emotion appraisal model based on OCC [17]. As will be shown in section 4, we only consider a limited subset of emotions that are relevant in the context of job interviews and that are compatible with the TARDIS emotion recognition system.

Baron showed the importance of the interviewer's mood during the interview and its impact to the applicant [1]. The evaluation of the applicant is also influenced by interviewer's mood. In the TARDIS project, we want learners to be able to detect these change of moods and to adapt their social attitude accordingly. For this reason, we require an accurate model of mood-behaviour influence. Our model for moods is based on the ALMA model [7]. According to [15], emotions are one of the factor that is able to change moods in human.

For the personality of the virtual recruiter, we rely on the big five model [10] that considers 5 quantitative dimensions (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) because some links between 5-factors personality and moods have ever been identified by Mehrabian [11].

In the TARDIS project, we require a computational model of social attitudes for the virtual recruiter. This model must encompasses most (if not all) dimensions mentioned above. The next sections present this model.

3. See the Humaine project : `emotion-research.net`

3 Architecture overview

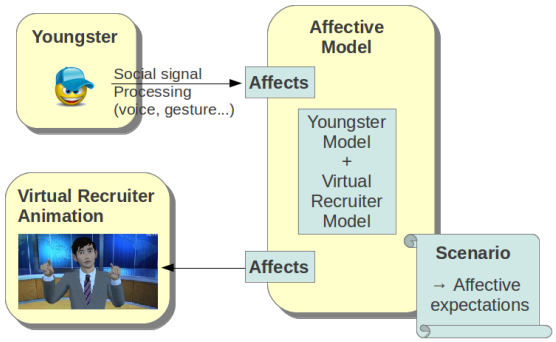


FIGURE 1 – Global architecture

The TARDIS architecture considers four main components :

- The Social Signal Interpretation module provides the affective model with information about the youngster's emotions and social attitude that are detected by the system.
- The Scenario module tells the virtual recruiter the expectation in terms of emotions and attitudes, depending on the interview progress. In the current version of TARDIS, the agent has no understanding of the youngster's actual answers to the questions. It simply follows a predetermined scenario, while focusing on the affective recognition and adaptation.
- The Animation module is responsible for expressing the virtual recruiter's affective state through its behaviour.
- The Affective Model, which is presented in this paper, is responsible for determining the agent's internal state (output) depending on the recognized affects and scenario expectations (inputs).

Figure 1 gives an overview of this architecture.

The Affective Model has two main computation functions :

- It will periodically compute the new affective states for the Virtual Recruited Model, based on the perceptions, expectations from the scenario and current affective states. The differences between expected affective states and expressed affective states from the youngster are the key element in the update of the virtual recruiter affective state (see next section).
- It will select actions in the scenario. This part is not presented in this paper.

4 Virtual recruiter affective Model

Our affective model is based on the youngster detected and expected affects and the internal states of the virtual recruiter. Emotions, moods and social attitudes are defined on $[0, 1]$ interval.

4.1 Youngster detected affects

This module relies on the affective outputs given by the real-time social signal interpretation of the youngster. We focus on three affective categories : emotions, moods and attitudes that will be given with a level of confidence. These detected youngster affects will be denoted as $E_d(emotion)$ for emotions, $A_d(attitude)$ for attitudes and $M_d(mood)$ for moods ("d" stands for *detected*). Table 1 summarize relevant affects in the context of a job interview. Note that emotions, moods and attitudes are organized as positive and negative. This does not necessarily mean that Distress, Anxiety or Agitation are not good in the context of a job interview, but it depends on the context (given by the scenario's expectations).

We note Ev^+ the set of positive affects (joy, focused, etc) and Ev^- the set of negative ones (distress, anxious, etc).

	Positive	Negative
Emotions	Joy	Distress
		Anger
Moods	Relaxed	Anxious
		Exuberant
		Bored
Attitudes	Focused	Inattentive
	Calm	Agitated
		Aggressive

TABLE 1 – Youngster affects

4.2 Youngster expected affects

These states will be given by the scenario context, each question in the scenario is marked with expectations about the impact of the question on youngster emotions and attitudes. As our model rely on the comparison of detected affects and expected affects, expected affects stick to the list given by social signal interpretation (table 1). If this list increases in the future, the new affects will be added in the expected list and will be considered by the scenario. Expected emotions and attitudes will be denoted as

$E_e(emotion)$ and $A_e(attitude)$ with the same set of emotions and attitudes than for the detected affect ("e" stands for *expected*).

4.3 Internal affective model

The internal affective model contains emotions, moods, personality and attitudes of the virtual recruiter. Personality is static and will not evolve during the simulation but it will influence the dynamics of other affects. We model emotions as a short-term timing affect and mood as a middle-term timing affect.

The affects of the virtual recruiter (i.e. felt emotions, attitude and mood) will be denoted as E_f, A_f and M_f ("f" stands for *felt*). The personality of the virtual recruiter is not dynamic and will be denoted as $P_f(personality)$.

	Positive	Negative
Emotions	Joy	Distress
	Relief	Disappointment
	Admiration	Anger
	Hope	Fear
Moods	Relaxed	Hostile
	Exuberant	Bored
		Disdainful
Attitudes	Friendly	Aggressive
	Supportive	
		Dominant
	Attentive	Inattentive
		Gossip

TABLE 2 – Recruiter affects

Moods are affects with medium term evolution. Our model for moods is based on the ALMA model [7].

Attitudes of the recruiter will be determined by its actual moods and personality and will be initialised by its personality. Table 2 summarize relevant emotions, moods and attitudes for the virtual recruiter.

5 Dynamics of the affective core

Our dynamics follow this principle : we compute emotions by comparing youngster detected affects and youngster expected affects. Then, we compute moods on the base of computed virtual recruiter emotions. Finally, considering agent's personality and actual moods, we compute the social attitudes of the virtual recruiter.

5.1 Dynamics of virtual recruiter's emotions

Computation of emotions is based on OCC [17]. In OCC, events of the simulation allow the computation of emotions. Events are perceptions from the virtual agent. In our simulation these events are related to affective expressions of the youngster detected by the virtual recruiter : $E_d(emotion)$, $A_d(attitude)$ and $M_d(mood)$. The perception of events lasts during the full time of the youngster's answer. Lets detail the different computations.

Joy and distress. Following OCC [17], joy is the occurrence of a desirable event. In the current version of our model, we do not consider the semantic context of the interaction in the job interview. For this reason, we simply assumed that youngster's detected positive affects (Ev^+) increase the joy of the recruiter whereas detected negative affects (Ev^-) decrease it.

In order to balance between short-time emotions and mid-term moods, we compare all affects to decide the overall expression of the youngster. Let us denote Δ_d the difference between positive detected affects and negative ones.

$$\Delta_d = \sum_{a \in Ev^+} E_d(a) - \sum_{a \in Ev^-} E_d(a)$$

and let us define $norm$ the normalization function between 0 and 1 :

$$norm(x) = \begin{cases} 1 & \text{if } x > 1 \\ 0 & \text{if } x < 0 \\ x & \text{otherwise} \end{cases}$$

The intensity of joy felt by the recruiter is then defined by :

$$E_f(joy) = norm(\Delta_d)$$

Similarly, the $distress$ is the occurrence of an undesirable event, *i.e.* negative expressed affects by the youngster :

$$E_f(distress) = norm(-\Delta_d)$$

Hope and fear. Following OCC [17], $hope$ is the expectation of a desirable event, and $fear$ corresponds to undesirable events. Similarly to joy and distress, we define Δ_e the difference

between positive expected affects and negative ones :

$$\Delta_e = \sum_{a \in Ev^+} E_e(a) - \sum_{a \in Ev^-} E_e(a)$$

The intensity of $hope$ and $fear$ is then defined by :

$$E_f(hope) = norm(\Delta_e)$$

$$E_f(fear) = norm(-\Delta_e)$$

Disappointment, admiration, relief and anger. $Disappointment$ happens if a desirable event does not occur, *i.e.* when the agent is in a state such that $E_f(hope) > 0$ and the desirable events (detection of positive emotions) do not occur with an intensity as high as expected. Concretely, if $E_f(hope) > 0$:

$$E_f(disap.) = norm(max_{a \in Ev^+}(E_e(a) - E_d(a)))$$

Note that $E_f(disap.) = 0$ when $E_f(hope) = 0$.

Reciprocally, admiration occurs when the detected positive emotions are bigger than expected. Concretely, when $E_f(hope) > 0$:

$$E_f(admir.) = norm(max_{a \in Ev^+}(E_d(a) - E_e(a)))$$

Similarly, relief occurs when undesirable events do not occur with the expected intensity : when $E_f(fear) > 0$,

$$E_f(relief) = norm(max_{a \in Ev^-}(E_e(a) - E_d(a)))$$

Finally, anger is triggered by higly detected undesirable events. However, we also use the current agressivity of the recruiter to increase the intensity of the felt anger (the more the recruiter is agressive, the more it wil get angry). Concretely, when $E_f(fear) > 0$,

$$E_f(anger) = norm\Big((1 + A_f(agress.)) \times$$

$$max_{a \in Ev^-}(E_e(a) - E_d(a))\Big)$$

Based on this emotions (computed through expectations and perceptions of the youngster), the next section presents how we compute the mood of the recruiter.

5.2 Virtual recruiter moods

The computation of moods is based on emotions following the ALMA [7] : the mood is a point in the PAD (Pleasure, Arousal, Dominance) space proposed by Mehrabian [12]. Based on these models, we propose a mapping of OCC emotions into PAD space that will be used to compute virtual recruiter moods.

In the context of a job interview, the recruiter is always in a dominant position considering its status. As a consequence, the D parameter of the PAD space is never negative changing the mapping of emotions and mood as shown in table 3. According to the intensity of the emotion, the arousal can be positive or negative for some emotions and it will trigger different moods. For example, if a joy is intensive (positive arousal), it will lead to an exuberant mood. But, if the joy intensity is weak, the agent will just become relaxed. However, some emotions have always big intensity : anger will always have a positive arousal. In the same way, a disappointment can imply disdainful attitude if the dominance is important and bored attitude for a low dominance.

Emotion	P	A	D	Mood
Joy	+	+/-	+	Exuberant, Relaxed
Distress	-	+/-	0/+	Hostile, Disdainful, Bored
Hope	+	+/-	0	Exuberant, Relaxed
Fear	-	+	0	Hostile
Relief	+	-	+	Relaxed
Disappointment	-	+/-	0/+	Hostile, Disdainful, Bored
Admiration	+	+	0	Exuberant
Anger	-	+	+	Hostile

TABLE 3 – Mapping of emotions into PAD space and corresponding moods

We can notice that, usually, emotions $Admiration$, $Disappointment$, $Distress$, $Fear$ and $Hope$ are associated to a negative dominance. However, as we discussed above, the context of a job interview confers the virtual recruiter a dominant status and we consider that these emotions will tend to a low dominance which is still positive.

As proposed by [7], mood is initialised with the personality of the virtual recruiter. The personality is defined using the OCEAN model and we transform this into affective values and then into a point in the PAD space following the proposition of [16]. Then, all along the simulation, felt

emotions modify the initial agent's mood by attracting the 3D PAD point to the new PAD center of the future mood. The new PAD center is computed according to emotions felt during a certain period thanks to the PAD mapping given in table 3.

In the context of our job interview simulation, the period is determined by the number of cycle question/answer. Each answer slightly influences recruiter's mood. The base for calibration is the following, after 5 cycles of a specific emotion (anger for example), the virtual recruiter will be in the corresponding mood (hostile).

5.3 Virtual recruiter attitudes

[24] has shown relation between attitudes and personality and [26] exhibits some relations between moods and attitudes. Based on that, our computation of attitudes is based on two main parameters : the actual mood of the virtual recruiter (which evolves during the simulation according to new emotions) and the initial personality of the virtual recruiter (which will remain in the same state during the simulation). In order to analyse the mood and personality, we compare them to a threshold θ, which was set to 0.5 in our experiments in the next section.

The intensity of $friendly$ attitude ($A_f(friendly)$) is defined in our model as the combination of the personality trait $agreableness$ (A) and the degree of mood $exuberant$:

If $(P_f(A) > \theta) \vee (M_f(exub.) > \theta)$, then :

$$A_f(friendly) = max(M_f(exub.), P_f(A))$$

The intensity $A_f(aggressive)$ depends on $agreeableness$, $neuroticism$ (N) and the $hostile$ mood :

If $\Big((P_f(A) < \theta) \wedge P_f(N) > \theta\Big) \vee (M_f(hostile) > \theta)$, then :

$$A_f(aggr.) = max(M_f(hostile), P_f(N), 1 - P_f(A))$$

The intensity $A_f(dominant)$ is based on $extraversion$ (E), $neuroticism$ and the $hostile$ mood :

If $\Big((P_f(E) > \theta) \wedge P_f(N) > \theta\Big) \vee (M_f(hostile) > \theta)$, then :

$$A_f(dominant) = max(M_f(hostile), P_f(N), P_f(E))$$

The intensity $A_f(supportive)$ is based on $agreeableness$, $extraversion$ and the $relaxed$ mood of the agent :

If $\Big((P_f(E) > \theta) \wedge P_f(A) > \theta\Big) \vee (M_f(relax) > \theta)$, then :

$$A_f(compr.) = max(M_f(relax), P_f(A), P_f(E))$$

The intensity $A_f(inattentive)$ is based on $conscientiousness$ and the $disdainful$ mood :

If $(P_f(C) < \theta) \vee (M_f(disd.) > \theta)$, then :

$$A_f(inatt.) = max(M_f(disd.), 1 - P_f(C))$$

Similarly, the intensity $A_f(attentive)$ depends on $conscientiousness$ and $relaxed$:

If $(P_f(C) > \theta) \vee (M_f(relax) > \theta)$, then :

$$A_f(att.) = max(M_f(relax), P_f(C))$$

Last, $A_f(gossip)$ is based on $extraversion$ and the $exuberant$ attitude :

If $(P_f(E) > \theta) \vee (M_f(exub.) > \theta)$, then :

$$A_f(gossip) = max(M_f(exub.), P_f(E))$$

The way we compute attitudes follow this principle : an agent can adopt an attitude according to its personality or according to its actual mood. For example, someone who is not aggressive due to his personality can become aggressive if its mood is very hostile. The mood compensate the personality and vice versa.

6 Simulation

This section details a concrete scenario inspired by video of job interview taken with the Tardis users. The scenario is a succession of 6

questions/answers during the job interview for a bus driver position. At the beginning of the interview, the recruiter asks the youngster to talk about himself and his professional career. The youngster has no experience in this domain. Quickly, he seems to be stressed and does not find relevant arguments. The succession of questions/answers is the following :

Q1 - Recruiter : What is the customer looking for when he takes the bus ?
A1 - Youngster : Uncomfortable, doesn't find many relevant arguments.
Q2 - Recruiter : What about the journey ? Could it be long ?
A2 - Youngster : Comfortable, he is flexible about commuting and work hours.
Q3 - Recruiter : How do you see your career in 10 years ?
A3 - Youngster : Little bit more at ease, he expects an evolution in the company but remains vague. He doesn't know enough about career advancement.
Q4 - Recruiter : Do you have plans to evolve in this activity ?
A4 - Youngster : As she ignores the professional perspectives, she simply answers : "Being responsible".
Q5 - Recruiter : Let's consider a practical case, how do you manage a complicated situation ? Find one and explain me your solution.
A5 - Youngster : Has trouble to find relevant arguments. Many hesitations.
Q6 - Recruiter : Can you list your main qualities and drawbacks ?
A6 - Youngster : She whispers, not convincing and uncomfortable.

In this scenario, the youngster is often in difficulty during the interview. He expresses many hesitations and negative affects that we annotate with socio-cognitive specialists. With these data, we want to see if our model can answer in a realistic way to youngster reactions. At the beginning of the simulation, the recruiter is relaxed. Our first simulation consider the youngster affective reactions.

First let's analyse some affective answer of the virtual recruiter : Expected affective answer for question 3 is $[Joy = 0.5, Anxious = 0.8, Agitated = 0.7, Focus = 1]$. Affective answer to question 3 is quite positive : $[Anxious = 0.6, Agitated = 0, Focus = 1, Calm = 0.7]$. The recruiter was expected negative affect (agitated), for this reason, he was feeling fear. Since its fear is not confirmed, he is relieved $[Relief = 0.7]$ but a bit disappointed $[Disappointment = 0.5]$ because he was expecting joy. The emotional answer of the recruiter is coherent with the video expectations.

At each question, every emotion will give new influence to the PAD center according to emotions triggered influencing the recruiter's mood. Figure 2 shows this evolution. In this figure the mood of the recruiter moves between slightly relaxed to slightly exuberant but its level of pleasure depends of the applicant answers. We can see the correlation between good answers and pleasure increase and bad answers and pleasure decrease. Its main attitude during all the interview is attentive.

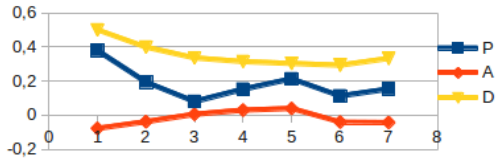


FIGURE 2 – PAD evolution for a relaxed youngster

In order to see another reaction of our virtual recruiter, let's consider the same scenario with an aggressive youngster. Reaction will be different for the recruiter and will lead to the PAD evolution exposed in figure 3. Because of the aggressivity of the youngster, the pleasure decrease more and the intensity of emotions is more important (in particular because of anger emotion). Recruiter's mood will tend to be hostile and as a consequence, its attitude will then become aggressive.

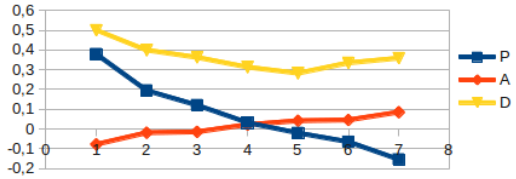


FIGURE 3 – PAD evolution for an aggressive youngster

This example shows that our agent is able to interact in a realistic emotional way in a job interview situation because emotions, moods and attitudes triggered are coherent with socio-cognitive specialists annotations. Next step will be to test this evolution based on Social Signal Interpretation in real time and to follow by evaluation.

7 Conclusion

In this paper, we have presented an affective model for a virtual recruiter in job interview simulations. This model is based on emotions, moods and attitudes in both inputs (recognized affects) and outputs (expressed affects). We illustrated on a scripted example scenario the results of our model and we showed that we could achieve variability in the agent's attitude.

However, our work is still in a preliminary stage. Our first goal is to include this model in the TARDIS platform and to validate it through user experimentations (winter 2013). One core issue that has to be dealt with is the imprecisions and errors in the social signal interpretation. We are currently considering how our model can be extended to consider probabilistic or fuzzy theories.

The next step in our research is to build an affective representation of the interaction from the recruiter's point of view, so that action selection based on virtual recruiter internal affective states and the scenario considers also the strategic intentions and the goals of the virtual recruiter in the decision process. Furthermore, the recruiter shall adapt its vocabulary and its level of politeness according to its social attitudes towards the youngster. Previous work [4] has proven that the language level could be adapted using a simple affective model.

Acknowledgment

The research leading to this paper has received funding from the European Union Information Society and Media Seventh Framework Programme FP7-ICT-2011-7 under grant agreement 288578.

Références

[1] R A Baron. Interviewer's Moods and Reactions to Job Applicants : The Influence of Affective States on Applied Social Judgments. *Journal of Applied Social Psychology*, 17(10) :911–926, 1987.

[2] M.R. Barrick and M.K. Mount. The big five personality dimensions and job performance : A MetaAnalysis. *Personnel psychology*, 44(1) :1–26, 1991.

[3] John Bynner and Samantha Parsons. Social Exclusion and the Transition from School to Work : The Case of Young People Not in Education, Employment, or Training (NEET). *Journal of Vocational Behavior*, 60(2) :289–309, April 2002.

[4] Sabrina Campano and Nicolas Sabouret. A socio-emotional model of impoliteness for Non-Player Characters. *Affective Computing and Intelligent*, pages 1123–1124, 2009.

[5] Timothy DeGroot and Stephan J Motowidlo. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, 84(6) :986–993, 1999.

[6] P Ekman and W V Friesen. *Unmasking the face : A guide to recognizing emotions from facial clues*. Number 1968. Prentice Hall, 1975.

[7] Patrick Gebhard. ALMA - A Layered Model of Affect. *Artificial Intelligence*, pages 0–7, 2005.

[8] Shlomo Hareli and Ursula Hess. What emotional reactions can tell us about the nature of others : An appraisal perspective on person perception. *Cognition & Emotion*, 24(1) :128–140, 2010.

[9] Stacy Marsella, Jonathan Gratch, and J Rickel. Expressive behaviors for virtual worlds. *Lifelike Characters Tools Affective Functions and Applications*, pages 317–360, 2003.

[10] R. R McCrae and O. P John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60 :175–215, 1992.

[11] Albert Mehrabian. Analysis of the big-five personality factors in terms of the pad temperament model. *Australian Journal of Psychology*, 48(2) :86–92, 1996.

[12] Albert Mehrabian. Pleasure-arousal-dominance : A general framework for describing and measuring individual.. *Current Psychology*, 14(4) :261, 1996.

[13] A N Meltzoff. Understanding the Intentions of Others : Re-Enactment of Intended Acts by 18-Month-Old Children. *Developmental Psychology*, 31(5) :838–850, 1995.

[14] Alice Mitchell and Carol Savill-Smith. The use of computer and video games for learning : A review of the literature, 2004.

[15] Paula P. Morris, William N. ; Schnurr. *Mood : The frame of mind*. New York, NY, USA, springer-v edition, 1989.

[16] Magalie Ochs, Nicolas Sabouret, and Vincent Corruble. Simulation of the Dynamics of Non-Player Characters' Emotions and Social Relations in Games. *IEEE Transactions on Computational Intelligence and AI in Games*, 1 :4 :281–297, 2010.

[17] Andrew Ortony, Gerald L Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, July 1988.

[18] A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods, C. Zoll, and L. Hall. Caring for agents and agents that care : Building empathic relations with synthetic agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 194–201, Washington, DC, USA, 2004. IEEE Computer Society.

[19] L Pareto, DL Schwartz, and Lars Svensson. Learning by guiding a teachable agent to play an educational game. *in Education Building Learning*, pages 1–3, 2009.

[20] Lynellen D S Perry. Emotionware. *Crossroads*, 3(1) :5–6, September 1996.

[21] R W Picard. Affective Computing. *Emotion*, TR 221(321) :97–97, 1995.

[22] Helmut Prendinger and Mitsuru Ishizuka. Social role awareness in animated agents. *Proceedings of the fifth international conference on Autonomous agents AGENTS 01*, pages 270–277, 2001.

[23] Monika Sieverding. 'Be Cool !' : Emotional costs of hiding feelings in a job interview. *International Journal of Selection and Assessment*, 17(4), 2009.

[24] Mark Snyder. The influence of individuals on situations : Implications for understanding the links between personality and social behavior. *Journal of Personality*, 51(3) :497–516, 1983.

[25] L Z Tiedens. Anger and advancement versus sadness and subjugation : The effect of negative emotion expressions on social status conferral. *Journal of personality and social psychology*, 80(1) :86—-94, 2001.

[26] Duane T Wegener, Richard E Petty, and David J Klein. Effects of mood on high elaboration attitude change : The mediating role of likelihood judgments. *European Journal of Social Psychology*, 24(1) :25–43, 1994.

Des micro-expressions au service de la macro-communication pour le robot compagnon EMOX

Y. Sasa[1] V. Aubergé[1] P. Franck[2] L. Guillaume[2] S. Moujtahid[2]

Yuko.Sasa@e.u-grenoble3.fr Veronique.Auberge@imag.fr pascal.franck@awabot.com leslie.guillaume@awabot.com salma.moujtahid@gmail.com

1 LIG, CNRS, Grenoble, France

[2]AWABOT S.A.S., 11 avenue Albert Einstein, 69100 Villeurbanne

Résumé :
Cette étude préliminaire vise à doter EMOX, un robot-compagnon, de « bruits de bouche » esthétiquement modifiés, dont la prosodie globale est préservée. Notre enquête cherchera à savoir si la valeur émotionnelle de ces sons, associés à des comportements supposés expressifs du robot, est préservée ou non. Le but sera également de comprendre les besoins et les attentes actuels des futurs utilisateurs, face à ce robot-compagnon souhaitant créer un véritable lien social avec l'humain.
Mots-clés : robot-compagnon, microexpressions vocales, prosodie, social, émotions

Abstract:
This preliminary study is an attempt to make a companion robot, EMOX, produces aesthetically modified "mouth sounds", while maintaining their overall prosody. An investigation will determine if or not the emotional information of these sounds associated with expressive behaviors is preserved. The goal will be to understand, the future users current needs and expectations toward this companion robot who wants to create a real social bound with humans.
Keywords: companion robot, vocal microexpressions, prosody, social, emotions

1 Introduction

Le robot anthropomorphique était et reste une expression de l'humain désirant se connaître, se compléter, se suppléer et communiquer. Cet objet est en constante évolution depuis les automates des temps antiques jusqu'à nos jours avec l'avènement de l'électronique [9]. Dès que le robot est introduit dans le quotidien vernaculaire de l'humain, quelque soit son usage spécifique, l'aspect socio-affectif et relationnel est une réelle tendance, voire une nécessité, jusqu'à l'émergence de robots quasi-dédiés à cette relation, le robot compagnon [8]. Lorsque l'aspect du robot n'est pas explicitement humanoïde, c'est à travers le statut de « *pet* » que l'humain crée sa relation socio-affective au robot. Ainsi, les enjeux actuels sont notamment leur expressivité affective, leur personnalité ainsi que leur rôle sociétal [3]. Mais comment communiquer l'émotion à un humain ? Alors que les aspects visuels de cette communication commencent à être bien abordés [11, 6], les études sur la production vocale de ces robots sont encore naissantes. L'enjeu de ce travail s'appuie sur une hypothèse forte qui est que toute l'essence informationnelle, communicationnelle, et en même temps intrinsèquement socio-affective de la communication vocale est contenue déjà dans un répertoire de micro-expressions non verbales [*]. Le pari est pris ici que doter le robot Emox de telles micro-expressions (sur une gamme de voix de la plus androïde vers la plus robotique, en passant par des voix typiquement « pet ») donnera à l'usager humain la liberté du *bootstrap* de la relation socio-affective avec Emox.
Nous présentons ainsi ici, sous forme d'une enquête, une étude préliminaire (comportement/personnalité perçue, attentes, besoins, motivations), les réactions de sujets naïfs face à ces productions vocales par EMOX.

2 L'entreprise Awabot et son robot EMOX

Awabot, créée en 2010 par Bruno Bonnell (Robopolis, Infogrames, Infonie), conçoit, développe et produit des robots à

destination du grand-public et du monde de l'éducation. La société emploie 15 personnes et collabore avec une vingtaine de partenaires industriels et académiques. Sa mission est de rendre accessible au plus grand nombre les objets connectés et intelligents ainsi que de faciliter leurs interactions avec les humains. La première gamme de robots EMOX sera lancée en 2012, à destination des développeurs et des formateurs en robotiques avec un ensemble d'applications logicielles visant à faciliter l'apprentissage et la programmation robotique. Cependant, le but ultime est son intégration dans les foyers comme robot compagnon (2015).

L'objectif des travaux actuels est donc de doter Emox d'une intelligence dite « émotionnelle ». Awabot souhaite qu'il puisse interagir avec l'homme de la façon qui soit la plus naturelle possible. C'est ainsi, que pour nous, l'expression d'émotions devient essentielle. Son absence pourrait être interprétée comme de l'indifférence et freiner la communication, souhaitée socio-affective. La communication analogique véhiculant des informations sur la personnalité et l'état émotionnel, nous désirons en doter ce robot. Un travail de recherche sur la prosodie (intonation, débit de parole, pauses et silences…) semble alors être un point important pour son développement.

FIG. 1 – Robot EMOX

Ci-dessus en FIG. 1 le prototype EMOX utilisé dans notre étude. Le design et la couleur ont été choisis pour être le plus neutre possible.

3 Les micro-expressions vocales

La prosodie évoquée dans le § 3 est prégnante dans le langage humain [10], elle est première acquise en perception et production par l'enfant, et dernière acquise en langue seconde. Si le langage humain est doublement articulé (une suite de phonèmes qui accède au lexique et à la morpho-syntaxe), il existe des micro-sons, éventuellement même non phonologiques, faits de « pure prosodie » qui pourtant sont informativement, communicationnellement et affectivement hautement significatifs [*]. Ces objets acoustiques ont été étudiés sous divers aspects et sous plusieurs de leurs fonctions communicatives, telles que : *affect bursts* [2], *non-verbal emotional vocalizations* [13], *mind markers* [12], *fillers* dans le cadre d'apprentissage de langues étrangères par exemple [7], *interjections* [16], ou encore *conversational grunts* [17]. Ce sont donc des productions spontanées, non lexicales, apparaissant pendant ou entre des tours de parole. A corréler et étudier parallèlement à des expressions faciales et autres indices visuels, ces productions sont construites sur des sons tels que des bruits non linguistiques, parfois des sons phonétiques voire même des phonèmes (interjections, onomatopées). De tels stimuli ont été recueillis, analysés, et étiquetés, en passant d'abord par l'élaboration d'un protocole de type magicien d'Oz et un recueil des données [1], l'étiquetage des icônes gestuelles [5] et phonétiques [6], une étude sur la prosodie culturelle de ces objets [14], suivie d'un travail sur la perception de la valeur informative des données chez les Français [4], puis sur le contraste informatif franco-japonais [**mémoire master Sasa, 2012]. C'est de cette base de données ainsi élaborée qu'une petite sélection de « bruits de bouche » a permis la présente étude.

4 Matériel testé

Pour modifier l'esthétique des bruits de bouche de notre étude, nous avons utilisé un outil permettant de modifier des paramètres prosodiques du signal acoustique. Voici donc une brève description de ce programme puis les

modifications véritablement effectuées sur nos signaux.

4.1 Outil de modification acoustique

Le programme utilisé a été implémenté au sein d'Awabot afin de modifier la prosodie des sons qui sont pris en entrée. Le but étant de synthétiser à partir de sons « neutres », des sons exprimant une émotion.
Il permet donc de modifier 5 paramètres : la valeur globale de la fréquence fondamentale (*pitch*), le contour du *pitch*, l'intensité globale, le contour d'intensité et la durée des segments audio (en format *wave*). Le contour d'un paramètre correspond à l'évolution du paramètre en question au cours du temps. La modification de ses paramètres se fait de manière indépendante. Ainsi un changement de valeur sur la durée n'altère pas la fréquence fondamentale du son (et vice versa). Cela est rendu possible grâce aux techniques de *Pitch Shifting* et *Time Stretching*, fournis par la librairie *Dirac LE*[1].

La modification du *pitch* se fait de manière relativement rudimentaire grâce à un facteur de modification par demi-ton : facteur pitch = $2^{x/12}$

Pour x = 0 il n'y a pas de modifications,
Pour x = 12, la fréquence est augmentée d'une octave, plus aigue,
Pour x = -12, la fréquence est diminuée d'une octave, plus grave.

Au delà d'une octave, en-dessous et au-dessus, le son est très altéré, en créant un effet « robotisé ».

Pour notre étude, nous avons modifié le pitch de nos signaux par pas de 1 sur x, avec une étendue de deux octaves soit une échelle allant de -24 à + 24.

4.2 Traitement acoustique des signaux

A partir d'une première sélection de « bruits de bouche » réalisée sur des sons vocalisés, nous avons choisi 15 stimuli de 6 locuteurs français, équitablement produits par 3 hommes et 3 femmes, les mieux reconnus dans les précédentes études [4]. Puis nous avons modifié les valeurs de la fréquence fondamentale (F0) de ces sons afin de transformer leur esthétique acoustique, sans altérer la prosodie globale de la production, tout cela via le programme décrit dans le § 4.1.

Par ailleurs, une valeur émotionnelle, expérimentalement validée [1, 5,6] avait été fournie sous forme d'auto-annotations à chacun des stimuli. Voici donc la liste des étiquettes affectives portées par les sons que nous avons sélectionnés: amusement, plaisanterie, déception, angoisse, agacé, surprise, irrité, stress, inquiétude, incertitude, déception et concentration.

Enfin, une analyse comparative de spectrogrammes a brièvement été réalisée pour vérifier que les contours prosodiques globaux n'ont pas été affectés par le programme. Cette observation s'est appuyée sur les figures du son d'origine, d'un son modifié dans les aigus puis d'un son modifié dans les graves, pour chaque « bruits de bouche ». Le résultat n'a montré aucune modification majeure du contour prosodique des signaux traités.

4.3 Construction des comportements

A partir de 980 stimuli créés par la modification de la fréquence fondamentale, 69 ont été sélectionnés sur leur originalité et leur qualité acoustique. Ces sons ont alors été associés à 15 comportements physiques du robot, élaborés de manière intuitive et se voulant les plus illustratifs possibles vis-à-vis des étiquettes d'auto-annotation associées à chacun des sons d'origine. Cela fait donc un total de 84 stimuli testés, en audiovisuel puis en visuel seul.

Les 15 mouvements que nous avons par la suite filmés, correspondent à des rotations et/ou des inclinaisons de haut en bas de la tête et à des mouvements d'avancée ou de recul sur des trajectoires plus ou moins droites d'EMOX. Ces comportements été parfois couplés à d'autres objets. Par exemple l'incertitude a été illustrée par un robot placé entre deux cubes faisant un choix pour l'un deux, cela se mimant avec des mouvements de tête de gauche à droite du robot

[1] http://dirac.dspdimension.com/Dirac3_Technology_Home_Page/Dirac3_Technology.html

qui avance en même temps vers les cubes.

Les sons étant à l'origine très courts, les mouvements associés sont eux-mêmes très courts puisqu'ils sont synchronisés aux productions acoustiques. Au final, nous avons réalisé un montage vidéo de ces comportements avec sons (son d'origine et sons modifiés) ou sans sons (condition visuelle seule) pour tester nos 84 conditions (69 en audiovisuel et 15 en visuel seul), chaque vidéo durant entre 3 à 5 secondes.

5 Protocole expérimental

L'enquête s'est déroulée en deux étapes. La première fut une étude préliminaire de perception portant sur les productions comportementales du robot en audiovisuel puis visuel. Elle s'est appuyée sur une grille de questions à choix multiples (QCM) portant sur chacun des 84 stimuli. Cette étape était facultative car longue et contraignante mais une alternative fut proposée, le cas échéant, pour permettre aux sujets qui ne souhaitaient pas remplir la grille, de participer à la seconde étape du protocole. Cette alternative consistait au visionnage de 15 vidéos sélectionnées aléatoirement parmi les 69 stimuli audiovisuels de l'expérience.

La seconde étape de l'enquête quant à elle fut un questionnaire portant sur la personnalité et les caractéristiques attribuées au robot.

5.1 Les participants

Au total, 21 sujets (10 hommes et 11 femmes) ont participé à l'enquête et au questionnaire. Pour le test de perception 16 sujets, dont 9 hommes et 7 femmes, ont rempli la grille de QCM, les 5 restants ayant uniquement visionné 15 vidéos. Ce panel, âgé de 16 à 91 ans, reste majoritairement composé de jeunes adultes de 20 à 35 ans. Ces personnes sont des Français ou des francophones étrangers bien établis en France qui sont d'un niveau socioculturel relativement élevé. Par ailleurs, 4 profils de sujets se dessinent en fonction de leur sensibilité aux technologies dans leur quotidien :

1) des personnes portant de l'intérêt mais n'utilisant pas et ne connaissant pas les nouvelles technologies,

2) des personnes ayant de l'intérêt mais avec un usage restreint de ces technologies (utilisation ponctuelle d'ordinateur),

3) des utilisateurs (notamment d'informatique au quotidien) mais ayant un intérêt limité vis-à-vis des technologies,

4) des utilisateurs confirmés portant un intérêt certain aux technologies actuels en cours de développement, ce dernier cas étant majoritaire.

5.2 Le test de perception

Cette première étape de l'étude se découpe en trois phases. Une première phase d'habituation permet préparer le sujet à mettre en contexte les comportements du robot. En effet, comme nos stimuli sont très courts et non scénarisés, nous avons présenté 4 situations illustrant de petites scènes de jeu du robot, utilisées dans son *teaser*, afin de permettre une meilleure appréhension des comportementaux à évaluer par les sujets. La seconde phase consiste en une présentation aléatoire de tous les stimuli audiovisuels (les 69 stimuli). Et enfin, la troisième phase consiste à la présentation aléatoire des extraits en visuel seul (les 15 comportements physiques).

La grille de QCM, que les sujets remplissent pour chacun des stimuli, comporte 3 parties où le sujet :

1) choisit un ou plusieurs contextes parmi les 4 situations présentés dans la première phase du test. Le sujet a également la possibilité de cocher une case « autre » s'il imagine le robot dans un autre contexte ou encore de répondre qu'aucun contexte n'est attribuable, notamment s'il n'arrive pas à donner de signification à l'extrait observé,

2) attribue une ou plusieurs valeurs émotionnelles parmi les 12 étiquettes d'auto-annotation citées dans le § 4.2. Le participant a là encore la possibilité de rajouter des termes qui lui sont propres pour qualifier l'attitude qu'il a pu percevoir de la vidéo,

3) confirme ou non la cohérence du comportement observé et l'adéquation de ce

comportement sur un robot tel qu'EMOX.

5.3 Le questionnaire

Après réalisation de cette première étape de perception, les sujets répondent à une dizaine de questions portant sur : la première impression qu'ils ont eu en regardant et en écoutant EMOX, avec notamment ses aspects positifs et négatifs, les usages qu'ils en feraient aussi bien de manière générale que pour une utilisation personnelle et quotidienne, tout en soulignant les capacités qu'ils perçoivent chez ce robot. Les questions portaient également sur le sexe, l'âge et l'éventuel nom qu'ils lui attribueraient, qui sont autant d'éléments révélateurs sur une partie de la personnalité et les caractéristiques saillants du robot. Enfin, ils étaient interrogés sur leur intérêt à faire acquisition d'un tel robot. Pour finir, un temps de commentaire libre leur était accordé pour clore l'entretien.

6 Résultats

6.1 Observations des grilles de perception

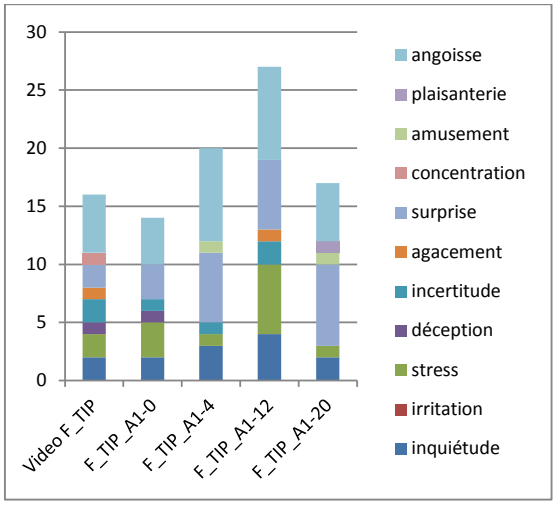


FIG. 2 – Histogramme des valeurs perçues par les sujets sur un signal d'angoisse de femme avec ses variantes aigües de pitch.
(Abscisse : nom des stimuli,
ordonnée : nombre de réponses)

Bien que les stimuli soient courts, non lexicaux et hors contextes, les sujets arrivent à trouver une signification et à projeter le robot dans une situation puisque généralement, ils attribuent des contextes à chaque stimulus. Donc l'association comportement physique et « bruit de bouche » reste fortement informatif même après modification de l'esthétique acoustique des signaux.

Ceci est également confirmé par le choix des auto-annotations sur chacun des stimuli. En effet, nous voyons sur la FIG. 2 que les valeurs émotionnelles attribuées à un bruit de bouche donné restent globalement les mêmes pour toutes les variantes du son (le 0 indique le son non modifié et les derniers chiffres des noms de stimuli indiquent le coefficient x de modification de la F0 / cf. § 4.1).

De plus, la situation en visuel seul (notée « Video » sur la FIG. 2) ne donne pas de réponses significatives sur le choix d'étiquette, donc c'est véritablement le son qui permet de désambiguïser et rajouter de la valeur informative au comportement du robot.

Par ailleurs nous avons constaté que les sons graves obtenus par la diminution de la valeur du pitch avaient tendance à introduire une valence négative sur le comportement et étaient globalement perçus comme incohérents et inadaptés à ce robot.

La FIG. 3 ci-dessous récapitule les valeurs émotionnelles globalement choisies pour les 15 signaux sélectionnées. La première colonne indique les étiquettes d'auto-annotation associées aux stimuli d'origine, testés en perception chez les Français [4]. La dernière colonne indique les termes rajoutés par les sujets (cf. §5.2.).

Stimuli	Etiquette attendue	Tendance générale	Nuances
P_SC _amusmt	amusement	amusement	rire franc…
P_SC _plstr	plaisanterie	amusement plaisanterie ++	moquerie farceur…
R_TIP	Rire	amusement	timide charmeur
ROB _surprise	Surprise	amusement plaisanterie	moquerie ironie…
ROB _agace	Agacé	déception	bouder
MAR _irrite	Irrité	surprise	curiosité
RS _MAR	Stress	déception	triste, penaud, grondé
H_JEA _inqtd	inquiétude	incertitude	hésitation réflexion / agacement

H_JEA _incert_NIC	incertitude	incertitude	hésitation tristesse
H_JEA _decept_TIP	Déception	incertitude	hésitation lassitude /flemme
H_JEA _conc_SAB	concentration	incertitude	indifférence ras le bol
R_NIC	déception	déception	attitudes enfantines
F_TIP	angoisse	angoisse surprise	peur
S_ROB	_ (foulala)	agacement pour le son original diffus sur les sons modifiés	1) débousso-lé/perdu 2) gêne/ embarrassé 3) ennui, sommeil, flemme 4) dépit, dommage, perdu, triste
ROB_12	_ (son insolite avec friction des lèvres)	diffus mais incertitude concentration déception	pet ronflement

FIG. 3 – Tableau récapitulatif des réponses sur la valeur informative perçue par les sujets

Les quatre premiers éléments du tableau sont des productions que nous pouvons tous qualifier de « rires ». Or nous voyons que les subtilités affectives sont très bien perçues par les sujets donc les « bruits de bouche » permettent d'illustrer des informations émotionnelles relativement précises.

Il en est de même pour les signaux annotés H_JEA. Ces quatre stimuli, malgré des informations affectives différentes, avaient pour point commun d'être des interjections de type « heu » ou « bah » que nous associons généralement à de l'hésitation en français. En intégrant le robot dans une situation où il mime un choix entre deux cubes, les sujets reconnaissent sans difficulté de l'incertitude, notant une différence dans toutes ces productions d'incertitude.

Les signaux ROB_agace et MAR_irrite ont eux été testés avec plusieurs comportements qui n'étaient pas forcément en adéquation avec les étiquettes attendues. Les résultats montrent que nous avons une modification de la nature de l'information affective par rapport à celle attendue, ce qui illustre bien le fait que la manière dont on met en contexte les sons à l'aide du comportement physique influe énormément sur notre perception des signaux.

Au contraire, quand nous avons une parfaite adéquation entre comportement physique et bruit de bouche, comme pour R_NIC et F_TIP, l'information émotionnelle est très bien reconnue. De plus, nous avons constaté que les taux de reconnaissance des étiquettes, quel que soit le bruit de bouche, étaient particulièrement bons pour certaines valeurs de F0. Nous voyons cette particularité pour les signaux F_TIP-A1-4 et F_TIP-A1-12 de la FIG. 2 par exemple.

Enfin certains signaux (S_ROB et ROB_12) sont très ambigus et leur modification aboutit parfois à des sons très insolites qui sont tout de même intéressants de remarquer car ils mettent en évidence des types de productions que les sujets pensent pouvoir être générés par le robot.

6.2 Les réponses du questionnaire

6.2.1 Avis généraux

Dans un premier temps, les sujets ont remarqué l'expressivité émotionnelle d'Emox, avec des « nuances variées ». Ils ont par ailleurs remarqué une intelligence notable à travers la moquerie qui serait, d'après eux, une aptitude à comprendre la plaisanterie. A cela, s'ajoute leur envie d'interaction avec EMOX alors même que dans notre étude, le robot bouge très peu, produit de très brefs sons et n'apparaît que sous forme de vidéos.

La communication est alors envisagée sous deux formes par les sujets : soit par le biais d'une voix humaine qui nécessite une synthèse vocale de très bonne qualité, soit par des bruits de bouche complétés par un autre type de médiation comme par exemple la projection d'images ou de mots, de définitions, de mode d'emploi, ce second cas suscitant beaucoup plus d'intérêts que le premier.

Dans un second temps, l'esthétique et la forme du robot sont évoquées aussi bien positivement que négativement. Notamment, sa petite taille semble être un atout car cela évite l'encombrement et le rend mignon. Or cette taille serait également un handicap au niveau de ses capacités physiques, limitant la fluidité, la force, la dextérité et la vivacité attendue par les sujets.

Enfin, l'amalgame entre le petit animal de compagnie, les caractéristiques humaines et la « haute technologie » est apprécié. En effet ce côté donnerait une simplicité et une naïveté à EMOX qui permettrait une meilleure appréhension du robot et une intégration plus facile dans le quotidien des sujets.

6.2.2 Âge, sexe et personnalité d'EMOX

La grande majorité des personnes trouvent le robot asexué. Il est parfois perçu comme un animal, dans quel cas aucun sexe ne lui est attribué. En revanche, il est également perçu comme un humain qui suggère alors aussi bien un féminin que masculin. Ce serait alors l'esthétique de la voix et la personnalité choisie qui définirait le sexe du robot ainsi que son rôle.

En termes d'âges, EMOX est essentiellement perçu comme un enfant de 3 à 11 ans ou à un jeune adulte d'une vingtaine d'années. D'après les sujets, ce serait son apparence qui lui donne un aspect enfantin alors que ses réactions évoqueraient plus de maturité.

Pour finir, concernant les noms, une grande majorité de personnes attribue au robot un nom de petit garçon. Certains l'ont également considéré comme un objet futuriste, doté d'un certain potentiel ou ont vu des caractéristiques affectives, intellectuelles ou physiques particulières. Parfois assimilé à des robots existants tels que R2D2 ou Wall-E, d'autres encore n'ont pas voulus lui mettre de nom car attendent de le voir évoluer au quotidien.

6.2.3 Attentes et besoins

Tout d'abord, les gens voient un lien avec la surveillance du domicile et le domaine de la domotique, notamment en l'envisageant comme un outil pour la commande vocale allumant des appareils, la lumière ou fermant des portes. Certains vont même jusqu'à même lui attribuer des tâches ménagères.

L'intelligence du robot est elle couplée avec l'informatique et le traitement automatique. En effet des personnes le voit accéder à des bases de données pour aller recueillir des définitions de mots, une page encyclopédique d'un concept ou le voit encore doter d'un outil de traduction gérant de multiples langues étrangères.

Ces aspects soulignent un fait largement évoqué: les gens veulent dialoguer, pas seulement se faire comprendre mais pouvoir demander des choses et obtenir une réponse ou une réaction en retour. De plus, le côté affectif, semble aider les sujets à se projeter dans une relation d'amusement et de compagnie au quotidien. Les actions qu'EMOX pourrait alors faire seraient entre autre de danser, jouer de la musique, réveiller les personnes. Par ailleurs, avec des fonctions plus orientées, les personnes voient une capacité d'aide pour des publics spécifiques tels que les enfants, les personnes âgées ou encore les handicapés.

Enfin le marché de ces robots compagnons est une réalité puisque la quasi-totalité des sujets (19 sur 21) aimerait acquérir EMOX à condition d'avoir une souplesse d'accessibilité économique et une réponse réelle à leurs attentes.

7 Conclusion

L'usage de micro-expressions, « bruits de bouche », se confirme être une issue intéressante comme « glue relationnelle » avec des robots compagnons de type « *pet* » tel que EMOX.. La modification rudimentaire de la voix par des paramètres prosodiques très globaux, sans toucher directement au spectre vocal, semble conserver, c'est étonnant, une très grande résistance informative du bruit de bouche..

De plus, la contextualisation des comportements du robot et la cohérence des productions ressort comme un point essentiel de cette étude. De fait, l'esthétique, la forme, l'expression émotionnelle et l'intelligence du robot converge pour une évolution mettant en jeu une véritable interaction s'intégrant dans le quotidien des personnes, ce qui sera particulièrement facilité par les compétences interactionnelles spécifiques et totalement ouvertes de EMOX (cf. http://www.emox-robot.com/).

8 Perspectives

Le corpus considéré étant majoritairement de valence négative, il est absolument nécessaire d'étendre les stimuli sur des valences positives par le biais d'un recueil adapté aux besoins du robot. La modification seule de la F0 reste limitée quand au rendu de l'esthétique acoustique des sons, donc le couplage à d'autres techniques de traitement de signal serait à envisager. Le nouveau matériel sonore ajouté à de nouveaux comportements plus scénarisés et contextualisés du robot devra alors être étudié dans une véritable situation d'interaction avec l'homme, le but à terme étant de permettre une automatisation de la synthèse de ces signaux répondant intelligemment aux attentes des utilisateurs. Plus spécifiquement, ce robot EMOX semble particulièrement adaptable à la médiation dans un environnement domotique, notamment celui qui est adressé aux personnes âgées et à l'éducation.

Références

[*] V. Aubergé, Prosody as the architect of language: from the involuntary to the voluntary control of affects, *Spoken Corpora and Linguistic Studies, Raso & Pettorino Ed, John Benjamins Pub*, (to be published).

[1] V. Aubergé, A. Rillard, N. Audibert, De E-Wiz à E-Clone : méthodologie expérimentale pour la modélisation des émotions et affects authentiques, *Workshop sur les Agents Conversationnels Animés*, Grenoble, 2005.

[2] R. Banse, K. Scherer, Acoustic profiles in vocal emotion expression, *Journal of Personality and Social Psychology*, **Vol. 70(3)**, pp. 614-636, 1996.

[3] K. Dautenhahn, S. Woods, C. Kaouri, M. Walters, K. Koay & I. Werry, What is a Robot Companion - Friend, Assistant or Butler? *Proceedings of IEEE IRS/RSJ Int. Conference on Intelligent Robots and Systems Edmonton, Alberta, Canada*, pp. 1488-1493, 2005.

[4] G. De Biasi, V. Aubergé, L. Granjon, & A. Vanpé, Perception of social affects from non lexical sounds. *In Proceedings of VII GSCP International Conference: Speech and Corpora, Brazil*, 2012.

[5] F. Loyau, V. Aubergé, Expressions outside the talk turn: ethograms of the Feeling of Thinking, *5th LREC*, pp. 47-50, 2006.

[6] T. Hashimoto, S. Hiramatsu, T. Tsuji & H. Kobayashi, Development of the Face Robot SAYA for Rich Facial Expressions, *SICEICASE Int. Joint Conf.*, pp. 5423-5428, 2006.

[7] N. Iwasaki, Filling social space with fillers: gains in social dimension after studying abroad, *Japanese Language and Literature*, **Vol. 45**, pp. 169-193, 2011.

[8] F. Kaplan, Un robot peut-il être notre ami?, *In Orlarey, Y., editor, L'Art, la pensée, les émotions, Grame*, pp. 99-106, 2001.

[9] N. Lazaric, Origines et développement de la robotique, *Revue d'Économie Industrielle, Programme National Persée*, **Vol. 61(1)**, pp. 54-67, 1992.

[10] B. Lindblom, On the notion of 'possible speech sound', Journal of Phonetics **Vol.18**, pp.135-152, 1990:

[11] C. Pelachaud, & I. Poggi, Subtleties of facial expressions in embodied agents, *J. Visual. Comput. Animat.*, **Vol. 13**, pp. 301-312, 2002.

[12] I. Poggi, Mind markers, *In NTM Rector, I. Poggi, editor, Gestures. Meaning and use. University Fernando Pessoa Press, Oporto, Portugal*, 2003.

[13] D. Sauter, F. Eisner., P. Ekman, & S.K. Scott, Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, **Vol.107(6)**, 2408-2412, 2010.

[14] R. Signorello, V. Aubergé, A. Vanpé, L.

Grandjon, N. Audibert, A la recherche d’indices de culture et/ou de langue dans les micro-événements audio-visuels de l’interaction face à face, *Proceedings of WACA 2010, Lille, France,* pp.69-76, 2010.

[15] A. Vanpé, Expressions et micro-expressions spontanées de la face et de la voix en Interaction Homme-Machine : esquisse d'un modèle du Feeling of Thinking, *PhD thesis, Grenoble University*, 2011.

[16] T. Wharton, Interjections, language and the 'showing'-'saying' continuum, *Pragmatics and Cognition*, **Vol.11(1)**, pp. 39-91, 2003.

[17] N. Ward, Non-lexical conversational sounds in American English, *In Pragmatics & Cognition*, **Vol.14(1)**, pp. 129-182, 2006.

Il était une fois... un robot compagnon qui racontait des histoires

Carole Adam[1]⋆
carole.adam@imag.fr

⋆Laboratoire d'Informatique de Grenoble
Université Joseph Fourier, Grenoble, France

Résumé :
Dans cet article nous proposons un langage de balisage pour représenter une histoire afin d'autoriser un compagnon artificiel (ACA ou robot) à la raconter de manière personnalisée pour son auditeur.

Mots-clés : Interactive Storytelling, compagnon artificiel, personnalisation

1 Introduction

Les êtres humains se sont toujours raconté des histoires. Certains sont de bons conteurs, capables de captiver, ou engager, leur audience dans l'histoire qu'ils racontent. Puis les chercheurs en Intelligence Artificielle ont commencé à créer des "conteurs virtuels", des agents artificiels capables de raconter des histoires [5, 24]. Mais si ces conteurs virtuels savent raconter des histoires, leur permettre de le faire de manière engageante reste un défi non encore résolu.

Le concept d'engagement a été étudié dans diverses situations (travail, interaction, jeux, sport...), ce qui a conduit à diverses caractérisations d'une situation engageante. Ainsi [4] se basent sur le modèle de Karasek de l'engagement au travail pour identifier trois exigences pour une interaction homme-machine agréable : l'utilisateur doit se sentir en **contrôle** de l'interaction (personnalisation, *feedback*, *etc.*) ; les **demandes** envers l'utilisateur doivent être adaptées à ses capacités (*i.e.* interaction stimulante et surprenante mais pas écrasante) ; et le système doit offrir un **support** à l'interaction sociale (*i.e.* ne pas isoler l'utilisateur). Dans le cadre du développement d'un conteur virtuel engageant, on peut jouer sur trois paramètres : l'agent, l'histoire et la narration, qui doivent tous trois être engageants.

Concernant les **histoires engageantes**, les chercheurs dans le domaine de la narration interactive (*virtual storytelling*) ont développé des systèmes où l'utilisateur incarne un personnage et donne des commandes pour faire progresser l'histoire comme il le souhaite. Il a été montré que ces histoires interactives sont très engageantes car elles procurent au joueur un sentiment de contrôle ou d'agentivité (*agency*, [21]). Cependant les systèmes de narration interactive soulèvent aussi un certain nombre de problèmes. Tout d'abord, donner le contrôle à l'utilisateur signifie que le système doit savoir réagir à des commandes inattendues, qui pourraient diriger l'histoire dans une direction indésirable (problème des limites, ou *boundary problem* [18]). Les systèmes permettent en général toutes les actions, soit en gérant leurs conséquences a posteriori, soit en assurant une *médiation* de sorte que celles qui menacent le bon déroulement de l'histoire échouent (*e.g.* Mimesis [11]) ; une solution originale a été proposée dans [8] où des techniques de persuasion sont utilisées pour influencer l'action de l'utilisateur. Un deuxième problème est que l'utilisateur doit généralement fournir des commandes explicites pour faire avancer l'histoire, ce qui peut nuire à son immersion dans celle-ci ; [9] ont développé le système PINTER qui exploite des données physiologiques passives de l'utilisateur pour orienter le dérouement d'une histoire jouée comme une animation 3D.

Concernant les **agents engageants**, le domaine des Agents Conversationnels Animés (ACA) s'intéresse depuis longtemps au développement d'agents pouvant interagir naturellement avec les humains, exprimer des émotions appropriées, ou créer et maintenir des relations à long-terme avec leur utilisateur. De tels agents sont appelés agents relationnels [3], ou compagnons artificiels [2]. Les agents autonomes sont déjà utilisés comme personnages dans les histoires interactives, poursuivant leurs propres buts et interagissant entre eux et avec le joueur humain [22] ; certains sont dotés d'émotions pour les rendre plus crédibles et améliorer l'immersion, par exemple dans un scénario de brimades à l'école [7]. Nous pensons que ces agents relationnels ou compagnons pourraient être exploités non seulement comme personnages mais aussi comme conteurs virtuels engageants.

Enfin, il existe très peu de travaux concernant la **narration engageante** d'une histoire par un agent. Par contre les conteurs humains peuvent apprendre diverses techniques pour engager leur

audience dans l'histoire [10] ; en particulier les bons conteurs s'adaptent aux caractéristiques de leur auditoire (âge, centres d'intérêt...) [12], c'est-à-dire qu'ils personnalisent la narration.

Nous souhaitons permettre à un compagnon artificiel de raconter des histoires à partir de textes existants (contes, articles de journal...). Rendre ces compagnons engageants fait déjà l'objet de travaux dans le domaine des ACA, et écrire des textes engageants est le travail de leurs auteurs. Dans cet article nous nous intéressons donc à la narration engageante, et en particulier à la **narration personnalisée**.

2 Narration personnalisée par un compagnon artificiel

2.1 Notre approche

La narration personnalisée consiste à raconter une histoire d'une manière qui est non seulement adaptée à l'auditeur, à ses spécificités (âge, personnalité...) et à ses préférences, mais aussi personnalisée grâce à l'insertion de commentaires faisant référence à des aspects de sa vie.

Intérêt des compagnons artificiels. Les compagnons artificiels ont d'abord été vus comme des assistants, utiles de par leur capacité à offrir un certain nombre de services, comme par exemple de gérer l'agenda d'une personne pour lui rappeler ses rendez-vous. Le rôle de narrateur virtuel capable de raconter des histoires ou de lire le journal est donc très intéressant pour ces compagnons artificiels en leur permettant de fournir un service supplémentaire.

Mais une des principales qualités d'un compagnon artificiel est d'engager son utilisateur à court et à long terme. Or pour atteindre cet engagement il a été montré qu'il était important de développer une relation avec l'utilisateur, en particulier en apprenant à le connaître lors de discussions ouvertes (bavardage) et en utilisant les informations récoltées pour personnaliser progressivement l'interaction et les services offerts. Un compagnon artificiel sera donc particulièrement apte à personnaliser la narration en mettant à profit son profil de l'utilisateur.

Annotation vs génération. Les divers outils de création d'histoires digitales ne sont pas encore très utilisés par des écrivains ; [25, p.2] passent en revue divers systèmes de narration digitale avant de remarquer que "beaucoup (sinon toutes) les histoires mentionnées ci-dessus n'ont pas été créées par des écrivains professionnels, mais par les concepteurs du système". En effet les systèmes existants créent en général l'histoire à partir d'une spécification dans un langage informatique : un ensemble d'unités atomiques à combiner pour former l'histoire, ou encore des plans et des règles permettant de générer automatiquement l'histoire. Cette spécification est facilement exploitable par le système de narration (*i.e.* par le conteur virtuel) mais plus difficile à créer pour les auteurs qui ne sont pas familiers de ce format.

C'est pourquoi plutôt que d'exiger des auteurs qu'ils créent leur histoire dans un format informatique qui restreint leur créativité et peut limiter l'utilisation du système par de vrais écrivains professionnels, nous adoptons une approche différente. Nous souhaitons pouvoir partir d'histoires déjà écrites (contes de fées, histoires pour enfants, mais aussi articles de journaux...) et les enrichir *a posteriori* avec les informations nécessaires à leur narration personnalisée. C'est ce que permettent les langages de balisage de type XML, qui sont simples d'utilisation et ont en plus l'avantage de pouvoir être étendus facilement si nécessaire avec de nouvelles balises.

Notre approche On remarquera que le conteur virtuel a besoin de plusieurs types d'informations pour raconter un texte de manière personnalisée : des informations sur l'utilisateur et sur le contexte d'utilisation, dont les agents compagnons disposent dans le profil de l'utilisateur qu'ils maintiennent, et des informations sur l'histoire racontée, qui seront annotées dans le texte à raconter. Notre approche consiste donc à **annoter des textes existants** avec un certain nombre d'informations sur l'histoire, puis à **équiper un agent compagnon d'un moteur de narration** capable d'exploiter ces informations, ainsi que celles dont il dispose sur l'utilisateur, pour personnaliser sa narration du texte.

2.2 Différences avec les travaux existants

Sabouret *et al.*. ont développé le *virtual storyteller* [20], un agent capable d'enrichir une histoire avec des éléments engageants, mais sans la personnaliser ; de plus l'histoire étant générée à partir d'un scénario formel écrit par des informaticiens, il n'est pas possible d'enrichir un texte existant (conte, article de journal) pour permettre à un compagnon de le lire.

MPISTE [6] est un environnement de narration interactive personnalisée mobile ; le conteur virtuel raconte des histoires sur les endroits visités, de son propre point de vue et avec sa personnalité propre. Cependant ces histoires sont générées à partir de fragments créés par les développeurs de l'application. Par ailleurs le domaine d'application (guide de musée [14]) ne permet pas à l'agent d'apprendre à connaître les visiteurs suffisamment pour personnaliser l'histoire pour chacun, contrairement à un compagnon qui interagit régulièrement avec le même utilisateur pendant longtemps.

[16] se sont intéressés à la personnalisation narrative dans les jeux videos éducatifs ; ils génèrent automatiquement une histoire à partir d'un ensemble d'unités atomiques formelles en adaptant la séquence pédagogique aux progrès de l'utilisateur. Il ne s'agit donc toujous par d'annoter un texte existant, et la personnalisation ne prend en compte que le niveau de compétence de l'utilisateur.

Enfin PAROS [15] est un système qui fournit de l'information personnalisée (par exemple pour un guide virtuel de musée) en se basant sur un modèle de l'utilisateur comprenant des caractéristiques comme son âge, sa langue, ou son niveau d'éducation. Il s'agit donc d'une catégorisation peu fine qui ne tient en particulier pas compte des préférences de l'utilisateur ou d'aspects de sa vie auxquels un agent compagnon pourrait faire référence pour personnaliser la narration à un individu spécifique plutôt qu'à un type d'individus.

2.3 Stratégies de personnalisation

Ci-dessous nous listons quelques stratégies dont nous souhaitons doter notre narrateur virtuel afin de lui permettre de personnaliser la narration d'une histoire à un enfant. Ces stratégies de personnalisation sont spécifiques aux agents compagnons et exploitent notamment les informations dont ils disposent sur l'utilisateur. Leur réalisation nécessite aussi des informations sur l'histoire, que nous discutons pour chacune.

Adaptation du vocabulaire à l'âge Le vocabulaire utilisé dans l'histoire doit être adapté à l'âge de l'enfant, et il peut être nécessaire d'insérer des définitions de certains mots ou de les remplacer par des synonymes plus simples. Les mots potentiellement difficiles doivent donc être marqués dans l'histoire.

Intelligence émotionnelle Le conteur virtuel doit disposer de capacités émotionnelles, en particulier l'expression d'émotions cohérentes avec le contenu de l'histoire, et la reconnaissance des émotions déclenchées chez l'enfant. La plupart des ACA sont capables d'exprimer multimodalement une émotion ; il existe aussi des modèles formels pour la déduction d'émotions [1]. Le conteur virtuel a néanmoins besoin de savoir quels mots ou passages de l'histoire doivent être évalués émotionnellement, de son point de vue (pour synchroniser l'expression multimodale d'une émotion cohérente avec le texte lu) et du point de vue de l'enfant (pour deviner son émotion en réaction à l'histoire). Il doit aussi disposer de commentaires scriptés lui permettant d'exprimer son émotion et/ou de montrer qu'il a bien reconnu celle de l'enfant.

Variation aléatoire Un compagnon doit pouvoir être utilisé régulièrement et pendant une longue période ; il est donc important qu'il puisse raconter la même histoire plusieurs fois, mais avec quelques variations pour éviter la lassitude. Certains éléments peu importants de l'histoire peuvent donc être modifiés aléatoirement (par exemple certaines descriptions, la météo, la couleur d'un objet...). Le conteur virtuel pourra aussi utiliser des synonymes, ou changer la formulation de certaines phrases. Pour cela il faut lui fournir les différentes options disponibles pour ces passages variables.

Personnalisation Le conteur virtuel doit exploiter les informations dont il dispose sur l'enfant (dans le profil) et sur le contexte de l'interaction afin d'adapter et de personnaliser sa narration, notamment en insérant des commentaires faisant référence au contexte ou à ces informations personnelles. Par exemple le compagnon peut lier l'histoire qu'il raconte avec ce que l'enfant apprend à l'école, les sports ou activités qu'il pratique. Pour cela le compagnon doit avoir des informations sur les mots de l'histoire qui sont potentiellement déclencheurs de tels commentaires, et disposer de commentaires scriptés.

Jeux interactifs, traduction L'interactivité est très importante pour l'engagement. Le conteur virtuel peut donc proposer un certain nombre de jeux interactifs comme la traduction de certains mots dans une autre langue, soit en tant que simple anecdote, soit pour aider l'enfant à pratiquer un langage qu'il apprend. Pour cela l'agent doit disposer de traductions de certains mots dans différentes langues.

Déviations Toujours dans le but de rendre l'histoire plus interactive et de donner à l'utilisateur un sentiment de contrôle positif pour l'engagement, on peut lui permettre de choisir entre plusieurs options à certains points. Cependant afin d'éviter les problèmes créés par une trop grande liberté du joueur on se limitera ici à des déviations courtes et sans influence sur le cours de l'histoire, qui reprend donc normalement après. Le conteur virtuel doit être informé de ces points de contrôle, et disposer de la question scriptée, de la liste des réponses attendues, et du script des déviations correspondant à chacune.

Diversions Afin d'empêcher l'enfant de s'ennuyer, le conteur virtuel peut insérer des diversions dans l'histoire, comme des blagues ou des anecdotes. Celles-ci doivent être cohérentes avec le contexte courant. Il faut donc donner au conteur virtuel des informations sur les mots potentiellement déclencheurs de diversions dans l'histoire, et lui fournir un ensemble de scripts de diversions possibles, chacune correspondant à un ou plusieurs déclencheurs possibles.

Blocs Toutes les stratégies ci-dessus vont interrompre l'histoire pour y insérer du contenu supplémentaire (commentaire personnel, anecdote, reconnaissance d'une émotion...). Mais de telles diversions peuvent être assez pertubatrices si elles arrivent au mauvais moment, interrompant par exemple un passage de grande intensité narrative et dérangeant l'immersion de l'auditeur dans l'histoire. Un problème important est donc de placer ces diversions au moment opportun, ou au moins d'éviter les moments inopportuns. Pour cela il faut prévenir le conteur virtuel quand un certain passage de l'histoire ne doit pas être interrompu, en particulier quand c'est une étape de grande intensité (*climax*).

3 Notre langage de balisage SMILE

3.1 Limites des langages existants

Divers langages de balisage sont utilisés dans les agents virtuels : AIML permet de définir des modèles de réponses scriptées correspondant à des motifs d'entrées utilisateur attendues, et est utilisé pour programmer des *chatbots* (par exemple ALICE) ; le standard SAIBA pour les ACA utilise deux langages de balisage [26] pour transférer les informations entre les trois modules de l'architecture interne de l'agent : FML (Function Markup Language) exprime l'intention communicative de l'agent, alors que BML (Behaviour Markup Language) exprime la réalisation multimodale de cette intention.

Il existe aussi de nombreux langages permettant d'annoter les histoires dans différents buts : SML [17] ou ICML sont utilisés pour décrire la structure du scénario à partir duquel l'histoire interactive sera générée à l'exécution en fonction des actions du joueur ; [19] attachent diverses méta-données à des histoires racontées par des nonnes pour faciliter la recherche, la comparaison et le partage de ces histoires entre chercheurs ; StoryML [13] fournissent une description abstraite de l'histoire indépendante du dispositif de narration, afin de permettre une narration distribuée sur plusieurs dispositifs.

Cependant aucun de ces langages ne permet d'annoter une histoire avec des méta-informations utiles pour sa narration personnalisée. Nous proposons donc ici notre propre langage, SMILE (Story Meta-Information LanguagE), composé des balises nécessaires à la réalisation des stratégies présentées précédemment (Section 2.3). Notez que la liste de balises ci-dessous (Section 3.2) n'est pas exhaustive puisque ce langage est extensible, et que l'on peut aussi intégrer ces balises dans un langage existant. Nous discuterons dans un deuxième temps de la possibilité d'automatisation de l'annotation de textes avec ces balises (Section 3.3).

3.2 Liste des balises de SMILE

Mots difficiles La balise `<hardword>` marque un mot difficile et contient des balises `<syn age=`*age*`>`*synonyme*`</syn>` et `<def age="`*age*`">`*def*`</def>` fournissant des synonymes et des définitions de ce mot pour différentes catégories d'âge.

Mot émotionnel La balise `<emoword>` permet de marquer les mots à connotation émotionnelle, et contient une ou des balises `<comment emo="`*emo_detect*`">`*comment*`</comment>` fournissant des scripts de commentaires correspondant à différentes émotions de l'auditeur.

Variabilité aléatoire La balise `<alt>` permet de spécifier des alternatives : elle contient des balises `<option>`$option_i$`</option>` fournissant différentes options pour ce passage (synonymes, reformulations d'une phrase) afin d'introduire de la variabilité dans la narration.

Commentaires Une manière plus générique de faire des commentaires contextuels personnalisés (similaires aux commentaires émotionnels) est la balise <perso>*texte*</perso>, qui encadre un morceau de texte commentable, et contient une ou plusieurs balises <comm cdt="*cond*">*commentaire*</comm> fournissant les commentaires appropriés sous différentes conditions (sur le profil ou contexte).

Traduction en langue étrangère La balise <lg>*mot*</lg> permet d'insérer des traductions d'un mot dans différentes langues, en y imbriquant des balises <tr lg="*langue*">*traduction*</tr>.

Déviations La balise <choice> contient les informations nécessaires pour laisser le joueur décider d'une déviation : une balise <question>*question*</question> fournissant la question à choix multiples à lui poser, et plusieurs balises <alt input="$reponse_i$">$suite_i$</alt> fournissant les scripts alternatifs correspondant à chaque réponse attendue à cette question.

Diversions Des diversions (blagues, anecdotes...) peuvent être insérées pour relancer l'intérêt dans les passages peu intenses. La balise <divs> permet de marquer un mot comme déclencheur possible d'une telle diversion, et contient une balise <anecdote>*texte*</anecdote> fournissant le script de l'anecdote à insérer.

Blocs Les blocs de l'histoire qui ne doivent absolument pas être interrompus seront marqués avec la balise <focusblock>.

3.3 Annotation des histoires avec SMILE

Il existe déjà des outils pour annoter **automatiquement** le thème d'une phrase ainsi que son ton émotionnel (analyse de sentiment). Par ailleurs des ressources en ligne comme Wordnet peuvent être utilisées pour récupérer automatiquement des synonymes, des définitions ou des traductions dans d'autres langues, voire la connotation émotionnelle (WordNet Affect).

Mais comme l'ont noté [23, p.1], l'adaptation d'histoires pour les médias interactifs est une tâche difficile, et les contenus interactifs "ne peuvent pas être dérivés seulement d'un récit écrit et nécessitent la contribution des créateurs de l'histoire pour fonctionner de manière interactive". Les informations concernant les variantes de l'histoire, ou encore les commentaires et questions possibles, devront donc être annotées **manuellement**, soit par les auteurs de l'histoire, soit par un tiers *a posteriori*.

Enfin d'autres recherches seront nécessaires pour **déterminer si et comment** certaines annotations peuvent être automatisées. Par exemple les blocs non interruptibles pourraient être annotés manuellement par l'auteur de l'histoire, ou être créés automatiquement en reposant sur une division simpliste de l'histoire en étapes, ou en phases d'intensité croissante ; mais il serait aussi intéressant d'élaborer un processus de raisonnement plus complexe par lequel le narrateur virtuel pourrait dynamiquement évaluer l'état courant de l'histoire, afin de détecter des opportunités de diversions, ou au contraire des moments à ne surtout pas interrompre.

Pour l'instant et afin de valider l'intérêt de notre langage, nous avons nous-mêmes annoté manuellement quelques extraits du Petit Chaperon Rouge et montrons des exemples de narration que nous souhaitons générer en les exploitant.

4 Application

4.1 Exemples d'illustration des stratégies

Adaptation du vocabulaire à l'âge Dans l'extrait suivant, le mot "velours" est annoté comme étant difficile, et le conteur virtuel dispose d'un synonyme plus simple ("tissu") pour les enfants de moins de 8 ans, et de la définition du mot pour les enfants de moins de 11 ans (ou ceux qui demandent).

> " Il était une fois une petite fille que tout le monde aimait bien, surtout sa grand-mère. Elle ne savait qu'entreprendre pour lui faire plaisir. Un jour, elle lui offrit un petit bonnet de <hardword> velours <synonym age="8">tissu</synonym> <def age="11">Le velours est une sorte de tissu très doux qu'on peut utiliser pour fabriquer des vêtements.</def> </hardword> rouge, qui lui allait si bien qu'elle ne voulut plus en porter d'autre... "

Intelligence émotionnelle Dans l'extrait suivant le mot "loup" est annoté comme émotionnel. Des commentaires sont fournis pour deux émotions possibles : la peur et l'excitation. Le profil de l'enfant permet au conteur virtuel de déduire son émotion concernant les loups et d'insérer le commentaire correspondant dans l'histoire. Par exemple si l'enfant est plutôt peureux, le conteur dira *"...rencontra le loup. Heureusement qu'il n'y a pas de loups par ici n'est-ce-pas ? Mais il ne savait..."*), alors que si l'enfant aime les loups, le commentaire sera différent.

> " Lorsque le Petit Chaperon Rouge arriva dans le bois, il rencontra le <emoword>loup <comm emo="peur">Heureusement qu'il n'y a pas de loups par ici n'est-ce-pas ?</comm> <comm emo="excitation">Tu aimes bien les animaux effrayants, non ?</comm> </emoword>. Mais il ne savait pas que c'était une vilaine bête... "

Variation aléatoire L'extrait suivant montre deux exemples de variations aléatoires dans l'histoire : à chaque lecture, une formulation de la première phrase, et un type de fleur, seront choisis au hasard parmi les options disponibles.

> " Le Petit Chaperon Rouge ouvrit les yeux et lorsqu'elle vit <alt> <option>comment les rayons du soleil dansaient de-ci, de-là à travers les arbres</option> <option>comment le soleil brillait au-dessus de la forêt</option> </alt>, et combien tout était plein de <alt> <option>fleurs</option> <option>pissenlits</option> <option>marguerites</option> <option>muguet</option> </alt>, elle pensa... "

Personnalisation Dans l'extrait ci-dessous, le mot "courait" est annoté avec deux commentaires personnels conditionnels. Par exemple un enfant inscrit dans un club d'athlétisme vérifie la condition du premier commentaire (u représente l'utilisateur, et la formule signifie que u pratique le sport "course", information stockée dans son profil) et entendra donc *"Le loup lui, courait tout droit vers la maison. Tu dois courir super vite toi aussi ! Il frappa à la porte."* alors qu'un enfant qui déteste courir (dont le profil contiendra donc l'information $\neg like(u, running)$) entendra plutôt le deuxième commentaire ; si aucune information n'est disponible sur les goûts de l'enfant pour la course, aucun commentaire ne sera inséré.

> " Le Loup lui, <perso> courait <comment cdt= "sport(u, running)">Tu dois courir super vite toi aussi !</comment> <comment cdt= ¬ like(u,running)">Je crois que tu ne courrais pas très loin...</comment></perso> tout droit vers la maison de la grand-mère. Il frappa à la porte. "

Traduction Dans cet extrait le mot "forêt" est marqué avec la balise <lg> et sa traduction est donnée dans plusieurs langues, permettant au conteur virtuel d'insérer des commentaires ("Savais-tu que forêt se dit mori en japonais ?") ou des questions ("Peux-tu me dire comment on dit forêt en anglais ?") dont le script à trous devra bien sûr lui être fourni.

> " Elle quitta le chemin, pénétra dans la <lg>forêt <tr lg="anglais">forest</tr><tr lg="espagnol">bosque</tr><tr lg="japonais">mori</tr></lg> et cueillit des fleurs. "

Choix d'action Dans cet exemple l'enfant peut choisir si le Petit Chaperon Rouge s'aventure dans les bois ou se dirige directement vers la maison de sa grand-mère. Seule une phrase de l'histoire est modifiée, après quoi l'histoire reprend normalement.

> " <choice><question>Est-ce que tu penses qu'elle devrait aller se promener dans la forêt ou aller directement chez sa grand-mère ?</question> <alt input="grand-mère">Comme c'était une petite fille très sage elle obéit à sa mère et se dirigea tranquillement vers la maison de sa grand-mère.</alt> <alt input="forêt">Elle quitta le chemin, pénétra dans le bois et cueillit des fleurs.</alt> </choix> Le Loup lui..."

Diversions et blocs L'extrait suivant contient une anecdote sur les fleurs, qui pourra être insérée après avoir mentionné les fleurs dans l'histoire, selon la stratégie du conteur virtuel.

> " ...le petit Chaperon Rouge avait fait la chasse aux <divs>fleurs <anecdote>Sais-tu que certaines espèces de fleurs sont vénéneuses ?</anecdote></divs>.
> Lorsque la fillette en eut tant qu'elle pouvait à peine les porter... "

Blocs Un bloc assure que l'histoire ne soit pas interrompue pendant un moment clé (que ce soit par le conteur virtuel ou par l'enfant) ; ici il s'agit de la conclusion de l'histoire.

> " ...et reprit la route pour se rendre auprès d'elle. <focusblock>Elle fut très étonnée de voir la porte ouverte. [...] À peine le Loup eut-il prononcé ces mots, qu'il bondit hors du lit et avala le pauvre Petit Chaperon Rouge. </focusblock> "

5 Conclusion

Dans cet article nous avons exploré une nouvelle direction de recherche à l'intersection entre les domaines des ACAs et de la Narration Interactive. Les capacités des ACA, et plus généralement des compagnons artificiels (personnages virtuels ou robots), peuvent être exploitées pour créer de bons narrateurs virtuels, à condition de leur fournir les informations utiles à la personnalisation de la narration. Nous avons pour cela présenté un langage de balisage qui permet d'attacher ces informations à des textes sous la forme d'annotations. Cette approche a l'avantage de permettre d'exploiter des textes déjà écrits (contes de fées, articles de journal). Cela permet d'appliquer ce conteur virtuel à différents domaines, que ce soit comme dans cet article un jouet compagnon pour enfants, qui peut leur lire des contes de fées et leur proposer des jeux sur l'histoire, ou bien encore un compagnon pour adultes capable de leur lire le journal de manière plus agréable et personnalisée. Dans tous les cas, il est indispensable de développer une interface permettant à un écrivain d'annoter une histoire sans parler couramment le XML, et sans brider sa créativité par la nécessité d'écrire dans un langage très formel.

Ces travaux encore préliminaires soulèvent un certain nombre de questions, notamment concernant l'automatisation du processus d'annotation (discutée ci-dessus), mais aussi concernant la stratégie de narration d'un conteur virtuel exploitant ces informations. En particulier à quel moment et à quelle fréquence est-il opportun d'insérer des diversions ? Comment juger l'efficacité de différentes stratégies sur différents utilisateurs, selon le type d'histoire, ou le moment de l'histoire, afin d'adapter dynamiquement le choix parmi plusieurs stratégies possibles ? Comment reprendre l'histoire après une diversion, faut-il résumer, répéter ? Le conteur peut-il profiter de l'histoire pour essayer de rassembler de l'information sur l'auditeur via des questions personnelles en lien avec le thème ? Des expérimentations avec des utilisateurs seront utiles pour apporter une première réponse empirique à ces questions.

Dans des travaux futurs, il s'agira donc d'implémenter un module de narration virtuelle capable d'interpréter les annotations ci-dessus pour raconter une histoire de manière personnalisée, et d'y tester plusieurs stratégies de personnalisation afin de trouver comment engager l'utilisateur. En rendant ce module compatible avec le standard SAIBA, on permettra à n'importe quel agent (virtuel ou robotique) de raconter des histoires personnalisées à leur utilisateur.

Remerciements

L'auteur remercie le LIG pour avoir financé son séjour au RMIT (Melbourne) qui a permis à ces travaux de voir le jour, ainsi que Lawrence Cavedon et Wilson Wong du RMIT pour leurs idées et leur participation.

Références

[1] C. Adam, A. Herzig, and D. Longin. A logical formalisation of the occ theory of emotions. *Synthese*, 168(2) :201–248, 2009.

[2] D. Benyon and O. Mival. Introducing the companions project : intelligent, persistent, personalised interfaces to the internet. In *BCS HCI*, 2007.

[3] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interactions*, 12(2) :293–327, 2005.

[4] P. B. Brandtzaeg, A. Folstad, and J. Heim. Enjoyment : Lessons from karasek. In *Funology - From Usability to Enjoyment*, volume 3 of *Human-Computer Interaction*, pages 55–65. Springer, 2005.

[5] M. Cavazza and S. Donokian, editors. *International Conference on Virtual Storytelling (ICVS)*, volume 4871 of *LNCS*. Springer, 2007.

[6] C. Delgado-Mata, R. Velazquez, R. Pooley, R. Aylett, and J. Robertson. MPISTE : A Mobile, Personalised, Interactive Story Telling Environment. In *Electronics, Robotics and Automotive Mechanics Conference*, pages 243–248. IEEE, 2010.

[7] J. Dias and A. Paiva. Agents with emotional intelligence for storytelling. In *ACII*, volume 6974 of *LNCS*. Springer, 2011.

[8] R. Figueiredo and A. Paiva. "i want to slay that dragon !" - influencing choice in interactive storytelling. In *ICIDS*, 2010.

[9] S. W. Gilroy, J. Porteous, F. Charles, and M. Cavazza. Exploring passive user interaction for adaptive narratives. In *IUI*, 2012. full paper.

[10] E. Greene. *Storytelling : Art and Technique*. Libraries Unlimited, 1996.

[11] J. Harris and R. Young. Proactive mediation in plan-based narrative environments. *IEEE transactions on computational intelligence and AI in games*, 2010.

[12] P. Hostmeyer and M. A. Kinsella. *Storytelling & QAR Strategies*. Libr. Ultd, 2011.

[13] J. Hu and L. Feijs. An adaptive architecture for presenting interactive media onto distributed interfaces. In *IASTED Int. Conf. on Applied Informatics*, 2003.

[14] J. Ibanez, R. Aylett, and R. Ruiz-Rodarte. Storytelling in virtual environments from a virtual guide perspective. *Virtual Reality*, 7 :30–42, 2003.

[15] Y. Ioannidis and *et al.* Profiling attitudes for personalized information provision. *IEEE Data Engineering Bulletin*, 34(2) :35–40, 2011.

[16] M. D. Kickmeier-Rust, T. Augustin, and D. Albert. Personalized storytelling for educational computer games. In *Serious Games Development and Applications*, volume 6944 of *Lecture Notes in Computer Science*, pages 13–22. Springer, 2011.

[17] H. Koenitz. Extensible tools for practical experiments in idn : the advanced stories authoring and presentation system. In *ICIDS*, volume 7069 of *LNCS*, 2011.

[18] B. Magerko. Evaluating preemptive story direction in the interactive drama architecture. *Journal of Game Development*, 2(3) :25–52, 2007.

[19] R. McDaniel and N. Underberg. eXeMbeLishment : using the eXtensible Markup Language as a tool for storytelling researchers. In *iDMAa 2006 : CODE*, 2006.

[20] R. Miletitch, N. Sabouret, and M. Ochs. Susciter l'émotion dans la narration automatique. *Technique et Science Informatiques*, 31(4) :477–501, 2012.

[21] J. H. Murray. *Hamlet on the holodeck : the future of narrative in cyberspace*. The MIT press, 1998.

[22] R. Paul, D. Charles, M. McNeill, and D. McSherry. Mist : An interactive storytelling system with variable character behavior. In *ICIDS*, pages 4–15, 2010.

[23] U. Spierling and S. Hoffmann. Exploring narrative interpretation and adaptation for interactive story creation. In *ICIDS*, pages 50–61, 2010.

[24] G. Subsol, editor. *Int. Conf. on Virtual Storytelling*, volume 3805 of *LNCS*, 2005.

[25] N. Szilas, M. Axelrad, and U. Richle. Propositions for innovative forms of digital interactive storytelling based on narrative theories and practices. In *Transactions on Edutainment*, volume 7145 of *LNCS*, pages 161–179. Springer, 2012.

[26] H. H. Vilhjalmsson. Representing communicative function and behavior in multimodal communication. In *COST 2102 School*, volume 5398 of *LNCS*, 2009.

Vers un modèle computationnel de l'influence de la personnalité d'un ACA sur son comportement

Florian Pecune pecune@telecom-paristech.fr
Magalie Ochs ochs@telecom-paristech.fr
Catherine Pelachaud pelachaud@telecom-paristech.fr

Laboratoire Traitement et Communication de l'Information
Telecom Paristech - 37/39 rue Dareau
75014 Paris – FRANCE

Résumé :
Dans cet article, nous présentons un modèle computationnel permettant de simuler l'influence de la personnalité d'un ACA sur son comportement. Ce modèle vise à enrichir le modèle computationnel d'émotion FAtiMA en considérant l'influence de la personnalité sur l'humeur, les émotions et les besoins de l'ACA. Ce modèle a été implémenté et testé à travers un scénario mettent en scène un concessionnaire automobile et son client.
Mots-clés : agent conversationnel, émotion, personnalité, humeur

Abstract:
In this paper, we present a computational model that allows us to simulate the influence of personality on an ACA's behavior. This model aims to enhance the computational model of emotions called FAtiMA considering the influence of personality on mood, emotions and agent's needs. This model has been implemented and tested through a scenario depicting a dialogue between a car salesman and his client.
Keywords: conversational agent, emotion, personality, mood

1 Introduction

Les machines occupent de nos jours une place prépondérante dans notre quotidien. L'étude présentée en [24] démontre ainsi que les règles sociales régissant les interactions Humain-Machine sont identiques à celles entre deux humains. Si le rêve de Bledsoe [6] de voir des ordinateurs se lier d'amitié avec des individus est encore loin d'être une réalité, de nombreux efforts sont cependant fait pour attribuer à ces machines qui nous entourent une apparence et des propriétés typiquement humaines. Ce phénomène s'appelle la personnification. Deux éléments clés sont ainsi à prendre en compte : l'émotion et la personnalité. Le rôle des émotions dans la création d'agents virtuels n'est désormais plus à prouver depuis le livre fondateur de Picard sur l'Informatique Affective [22]. La personnalité, quant à elle, peut être définie comme "*l'ensemble des attributs, qualités et caractéristiques qui distinguent le comportement, les pensées et les sentiments des individus"* [26]. Selon Chuck Jones, "character designer" chez Disney, c'est cette personnalité qui donne vie à ses personnages et les rend crédibles aux yeux des spectateurs [4].

Les recherches exposées dans [18] montrent par ailleurs qu'il est tout à fait possible de simuler de manière perceptible le concept de personnalité au sein d'un simple programme informatique. Miller et al. [17] montrent que la combinaison de la personnalité et des émotions mais aussi de la motivation et de la culture joue un rôle important dans le comportement en influençant les décisions.

Nous nous attacherons donc, dans cet article, à proposer un modèle permettant de simuler les liens qui existent entre la personnalité et les émotions ressenties par un agent virtuel. Nous mettrons également en exergue l'influence de cette personnalité sur les besoins de l'agent, et donc sur ses décisions et son comportement. Pour cela, nous présentons tout d'abord, dans la section 2, des agents virtuels existants dotés d'émotions et d'une personnalité. Nous introduisons la méthode que nous avons utilisée pour modéliser cette personnalité dans la section 3. La section 4 est quant à elle consacrée à l'implémentation et à l'illustration du modèle à travers un scénario de test.

2 Etat de l'art

Il existe de nombreux modèles computationnels simulant à la fois les émotions et la personnalité. C'est le cas de SCREAM [23] où la personnalité va impacter la régulation des émotions, en particulier la vitesse à laquelle elles décroissent. André et al. [1] présentent quant à eux divers agents capables d'exprimer des émotions et dotés de personnalité. Cette personnalité intervient dans le choix des réponses de l'agent, et le ton donné à la parole. Un agent colérique aura ainsi tendance à choisir des réponses sèches, données sur un ton agressif. Cet impact sur les choix et les décisions de l'agent est également prouvé dans un scénario militaire mettant en scène une équipe de deux soldats aux personnalités distinctes. Leurs réactions face à un évènement identique sont alors complètement différentes : en cas de tirs ennemis, l'un va reculer en restant à couvert pendant que son coéquipier restera figé [11].

L'état de l'art présenté dans [12] souligne que la majorité des modèles existants sont fondés sur le modèle OCC des émotions [20] et sur le modèle OCEAN [15] pour modéliser la personnalité. Dans ce modèle, chaque individu est caractérisé par cinq traits distincts : *Openness, Conscientiousness, Extraversion, Agreeableness* et *Neuroticism*. Dans ce même article, McCrae et Costa s'appliquent également à démontrer l'influence de ces cinq traits sur les émotions, exprimant par exemple le lien entre le trait *neuroticism* et la propension d'un individu à ressentir des émotions négatives, telles que l'anxiété, la colère ou la dépression.

Parmi les modèles computationnels d'émotions et de personnalité existants, deux d'entre eux ont particulièrement attirés notre attention. Le premier, *Alma* [10], a l'avantage de retranscrire de manière détaillée les relations entre les émotions, l'humeur et la personnalité. La représentation proposée des émotions sur un axe tridimensionnel permet en outre de bien observer l'évolution de l'état émotionnel et les dynamiques mises en œuvre. Le second modèle computationnel sur lequel nous focaliserons nos recherches est *FAtiMA* [8]. *FAtiMA* s'articule autour d'une architecture modulaire permettant d'ajouter ou de modifier des phénomènes tels que la culture ou l'empathie dans le processus de déclenchement des émotions de l'agent. Il permet par ailleurs la gestion des buts et des stratégies de réactions en se basant sur les besoins intrinsèques de l'agent.

Concernant l'agent conversationnel GRETA [21], des travaux ont par ailleurs déjà été réalisés pour tenter de modéliser la personnalité. Ces recherches visent à améliorer le comportement d'écoute de l'agent [5] et plus précisément les rétroactions effectuées par ce dernier [7]. Un lien est ainsi fait entre la fréquence des rétroactions et le trait *extraversion* mais également entre le trait *neuroticism* et le type de rétroactions effectuées. Un agent exubérant agira ainsi plus souvent et aura tendance à imiter les gestes de l'utilisateur avec lequel il dialogue. Fondés sur le modèle PEN d'Eysenck [9], ces travaux s'intéressent plus précisément à l'expressivité de l'agent virtuel et non a l'influence de la personnalité sur le déclenchement des émotions.

En conclusion, bien que de nombreux modèles démontrent l'influence de la personnalité sur des points précis, il n'y en a pas, à notre connaissance, qui modélise à la fois l'impact de la personnalité sur les émotions, l'humeur et les besoins d'un agent. Grâce à sa grande modularité et à la complexité des processus qu'il met en œuvre, le modèle FAtiMA nous a fourni les outils nécessaires à la modélisation d'une telle influence.

3 Personnification d'agent

Dans cette section, nous présentons en détail la méthode utilisée pour modéliser l'influence de la personnalité d'un agent sur ses décisions et sur son comportement. Pour ce faire, nous proposons de modéliser l'influence de la personnalité sur l'humeur, les émotions et sur les besoins.

Ce modèle vise à s'intégrer dans l'architecture FAtiMA déjà existante. Nous présentons donc tout d'abord, dans la section suivante, la

manière dont est modélisé le déclenchement des émotions afin de mettre en évidence les points d'influence de la personnalité.

3.1 Modélisation du processus de déclenchement des émotions d'un agent dans FAtiMA

Dans cette section, nous présentons plus en détails l'architecture du modèle FATIMA et en particulier comment le processus de déclenchement d'émotion est modélisé (Fig. 1). La modélisation de ce processus est fondée sur le modèle OCC présenté dans [20].

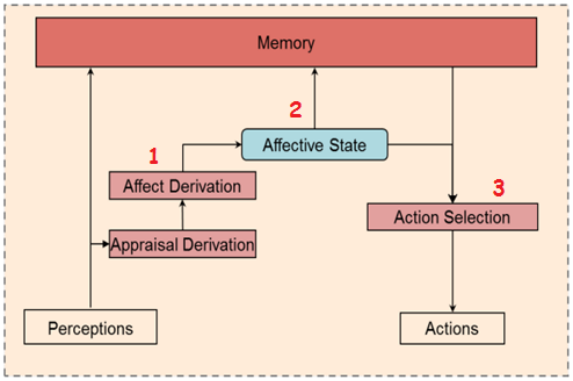


FIG 1- Architecture de FAtiMA

Comme décrit dans la figure 1, dès l'instant où un évènement est perçu par l'agent, *l'appraisal derivation* va associer à cet évènement des valeurs de variables telles que la désirabilité (*desirability)* ou le mérite (*praiseworthiness).* Par exemple, l'évènement "il fait beau" est perçu comme désirable. Ces variables sont ensuite utilisées par *l'affect derivation* pour calculer les émotions qui viendront s'ajouter à l'état émotionnel de l'agent. Par exemple, un évènement va déclencher de la joie chez l'agent. L'état émotionnel, représente l'ensemble des émotions "ressenties" par l'agent à un instant t. Pour finir, en fonction de cet état émotionnel, de ses buts et de ses besoins, l'agent entreprendra des décisions particulières pour faire face à la situation.

A ce jour, FAtiMA ne permet pas de modéliser efficacement l'influence de la personnalité de l'agent sur le déclenchement de ses émotions, ni sur ses décisions. Or, la personnalité de l'agent va influencer les trois processus indiqués sur la figure 1. Elle va en effet (1) modifier le seuil d'intensité nécessaire à une émotion pour être ressentie [23], (2) définir l'humeur par défaut de l'agent [10] et (3) pondérer les besoins qui déterminent sa décision finale [2]. Dans cet article, nous proposons donc un modèle permettant de simuler l'influence de la personnalité sur ces trois sous-processus (voir FIG. 3). L'impact de la personnalité sur les actions en elles-mêmes n'est pas traité ici, et fera l'objet de travaux futurs.

3.2 Modélisation de la personnalité de l'agent

Un de nos objectifs est de permettre de facilement personnifier un agent virtuel. La plupart des agents sont aujourd'hui personnifiés à travers les traits de personnalité OCEAN [15]. Or, ces traits de personnalité sont difficilement appréhendable par des non-experts en psychologie. Notre objectif est donc de fournir un ensemble de termes affectifs permettant à un "concepteur d'agent" de déterminer la personnalité de l'agent durant une interaction.

Pour déterminer ces termes affectifs, nous nous appuyons sur les travaux de Gebhard [10] qui définit huit "attitudes" (voir tableau 1) en fonction de leurs positions sur un axe tridimensionnel PAD (Pleasure, Arousal, Dominance) [16]. Notre modèle reprend donc ces huit attitudes et calcule automatiquement leur correspondance avec les traits de personnalité du modèle OCEAN selon le mappage de Mehrabian cité en [10].

TAB 1 - Correspondance Attitudes / Modèle OCEAN

	P	A	D	O	C	E	A	N
Anxieux	-0,5	0,49	-0,5	2,8	3	1,2	2,6	1
Dédaigneux	-0,49	-0,49	0,5	1,2	5	3,8	1	3,2
Dépendant	0,49	0,5	-0,5	4,2	1,2	2,4	5	2,6
Docile	0,5	-0,49	-0,5	1,6	2,8	2,6	4,2	5
Ennuyé	-0,51	-0,5	-0,5	1	4,8	1	1,8	3,6
Exubérant	0,49	0,5	0,51	4,2	2,4	5	4,2	2,2
Hostile	-0,5	0,48	0,5	5	3	3,2	1,8	1,2
Relaxé	0,5	-0,49	0,5	3,2	3	4,8	3,4	5

3.3 Impact de la personnalité sur l'humeur

Nous avons vu dans [10] que la personnalité de l'agent virtuel influence son humeur initiale. Cette humeur représente un état affectif global qui se distingue des émotions par sa plus faible

intensité mais sa plus longue durée [27]. Elle est représentée dans FAtiMA par une valence dont la valeur varie de -10 pour une humeur négative à 10 pour une humeur positive. Or, à ce jour, l'humeur initiale de l'agent est neutre. Pour modéliser l'influence de la personnalité p sur cette humeur initiale, nous nous sommes donc inspirés des travaux de Russell et al. présentés en [25]. Ces auteurs ont en effet montré que la dimension de plaisir mais également de dominance et d'excitation (*arousal*) déterminent l'humeur d'un individu. L'article fait par ailleurs état d'une influence plus importante du plaisir par rapport à l'excitation, la dominance ne jouant qu'un rôle mineur. Pour représenter ce phénomène dans notre modèle, nous proposons donc la formule suivante :

$$Mood(p) = \alpha * valence(p) + \beta * arousal(p) + \gamma * dominance(p)$$

$$avec\ \alpha > \beta > \gamma$$

Afin d'obtenir les humeurs indiquées dans le tableau 2, nous avons fixé les valeurs suivantes (La pondération accordée à la dimension de plaisir est supérieure à celle d'excitation, elle même supérieure à celle de dominance) : α = 8, β = 4 et γ = 2[1]. Un agent exubérant est donc doté d'une humeur initiale fortement positive, contrairement à un agent dédaigneux. L'humeur va ensuite fluctuer au cours de l'interaction en fonction de la charge émotionnelle des évènements perçus par l'agent. Ce processus est d'ores et déjà modélisé dans FAtiMA.

Nous proposons de plus dans notre modèle de modifier la fonction de décroissance de l'humeur de telle sorte à ce qu'elle revienne naturellement à sa valeur par défaut si aucun évènement venant modifier l'état émotionnel n'apparait dans l'environnement. Pour cela, nous nous sommes une nouvelle fois appuyés sur le modèle ALMA en proposant une fonction de décroissance similaire [10].

[1] Ces valeurs ont été choisies de manière heuristique. D'autres valeurs pourraient convenir pour spécifier des humeurs différentes.

TAB 2 - Correspondance personnalité / humeur par défaut

Attitude	*Humeur*
Anxieux	-3
Dédaigneux	-4,9
Dépendant	4,9
Docile	1
Ennuyé	-7,1
Exubérant	6,9
Hostile	-1,1
Relaxé	3

3.4 Impact de la personnalité sur les émotions

FAtiMA est fondé sur la théorie OCC [20] pour calculer les émotions "ressenties" par l'agent. 22 émotions sont définies et caractérisées par des valeurs de seuil (*treshold)* et de décroissance (*decay)* allant de 0 à 10. La valeur de seuil caractérise l'intensité minimale qu'une émotion doit atteindre pour être ressentie. La valeur de décroissance définit la vitesse à laquelle l'intensité de l'émotion décroit.

Nous proposons donc de définir les valeurs de décroissance et de seuil de chaque émotion en fonction de la distance entre sa position dans l'axe PAD et la personnalité de l'agent. Correspondant à des points dans l'espace PAD, il est donc possible d'évaluer la distance euclidienne séparant une émotion *e* d'une personnalité *p* en suivant la formule suivante :

$$f(e) = \frac{\sqrt{(x_e - x_p)^2 + (y_e - y_p)^2 + (z_e - z_p)^2}}{distanceMaxPossible(e,p)} * 10$$

FIG 2 – Calcul du seuil et de la décroissance d'une émotion.

Cette distance calculée à partir de la fonction f correspond aux valeurs de seuil et de décroissance de l'émotion. Ainsi, plus une émotion est éloignée de la personnalité de l'agent, plus il aura de difficulté à « ressentir » cette émotion et plus vite elle s'estompera. Un agent anxieux est par exemple beaucoup plus prompt à ressentir la peur ou le désarroi que des émotions telles que le soulagement ou la satisfaction, ce qui correspond aux résultats observés en [15].

Afin de bien visualiser cet impact de la personnalité sur les émotions, le couplage de FAtiMA avec l'agent Greta a été affiné afin de prendre en compte l'intensité des émotions. L'intensité calculée viendra ainsi moduler les paramètres d'expressivité de l'agent. Ces expressions faciales constituent alors une sorte de feedback pour l'utilisateur, que peut voir en temps réel l'impact de ses actions sur l'agent, appréhender au mieux la personnalité de ce dernier et s'adapter en conséquence.

3.5 Impact de la personnalité sur les besoins

Dans FAtiMA, les besoins de l'agent sont représentés sous la forme de cinq besoins primaires issus de la théorie PSI de Dörner (cité dans [3]). Ces besoins jouent un rôle important dans le choix des buts et la sélection des actions, mais ils influencent également le calcul de l'intensité des émotions. Ces besoins sont représentés par trois caractéristiques (*weight, decay* et *default*) qui expriment respectivement l'importance accordée par l'agent à ce besoin, la vitesse à laquelle il s'accroit et sa valeur par défaut. Pour modéliser l'influence de la personnalité sur les besoins de l'agent, nous nous appuyons sur les travaux de Bach [2] afin de définir l'influence de chacun des traits du modèle OCEAN sur trois de ces besoins[2] : (1) l'affiliation qui représente le besoin de relations sociales de l'agent, (2) la compétence qui correspond au besoin qu'il 'ressent' de réussir ses actions et (3) la certitude définissant le besoin de l'agent d'accroître ses connaissances sur l'environnement dans lequel il évolue.

Les tableaux 3 et 4 illustrent l'influence des traits de personnalité sur l'importance et l'accroissement des besoins. Ce traits peuvent avoir une influence élevée (++), faible (+) ou aucune influence du tout sur ces paramètres (n.i). Par exemple, le tableau 4 montre que le trait *Conscientiousness* a un fort impact sur le besoin de compétence, un faible impact sur la certitude mais aucun impact sur l'affiliation.

[2] Dans Fatima, cinq besoins comparables à certains désirs du modèle BDI sont définis. Au regard des travaux de Bach, nous ne considérons pas l'énergie et l'intégrité comme étant impactés par la personnalité.

TAB 3 - Influence personnalité / vitesse d'accroissement des besoins

	Affil	*Compet*	*Certain*
Openness	n.i	++	++
Conscientousness	n.i	++	+
Extraversion	++	n.i	n.i
Agreeableness	++	n.i	n.i
Nevrotism	n.i	++	++

TAB 4 - Influence personnalité / importance des besoins

	Affil	*Compet*	*Certain*
Openness	n.i	++	+
Conscientousness	n.i	++	++
Extraversion	++	++	n.i
Agreeableness	+	+	+

Pour modéliser l'influence de la personnalité illustrée dans les tableaux 3 et 4, nous introduisons une méthode de calcul. Dans [13] et [19], une méthode par règles est proposée pour mapper la personnalité aux besoins. Cette méthode présente néanmoins certaine limites. Les règles permettant de définir l'influence des traits de personnalité ne sont pas assez clairement définies pour être implémentées dans notre modèle et les effets modélisés ne correspondent pas à ceux attendus au regard des travaux (décrits dans [2]) sur lesquels notre modèle est fondé. Nous proposons donc une méthode de calcul prenant notamment en compte les exigences de FAtiMA en termes d'intervalle de valeurs des paramètres. Les fonctions suivantes sont ainsi définies :

$$\text{decay(b)} = \frac{\sum \text{ImpHaut}(t_h, b) + \sum \text{ImpBas}(t_b, b)}{n} * 10$$

$$\text{poids(b)} = \frac{\sum \text{ImpHaut}(t_h, b) + \sum \text{ImpBas}(t_b, b)}{n}$$

$$ImpHaut(t_h, b) = \frac{t_h}{5} \quad et \quad ImpBas(t_b, b) = \frac{t_b}{10}$$

$$default(b) = 1 - poids(b) * 10$$

ImpHaut(t_h,d) représente l'impact élevé (représenté dans les tableaux par ++) du trait t_h sur le besoin b, ImpBas(t_b,d) l'impact faible du trait t_b sur ce même besoin et *n* le nombre de

besoins intervenant dans le calcul. Par exemple, pour calculer l'importance accordée à la certitude par un agent anxieux, nous nous appuyons sur les tableaux 1 et 4 pour obtenir les résultats suivants :

$$\text{ImpBas}(\text{Openness}, \text{Certainty}) = 2.8/10$$
$$\text{ImpHaut}(\text{Conscientiousness}, \text{Certainty}) = 3/5$$
$$\text{ImpBas}(\text{Agreeableness}, \text{Certainty}) = 2.6/10$$

$$\Downarrow$$

$$\text{poids}(\text{Certainty}) = (0.28 + 0.6 + 0.26)/\,3 = \mathbf{0.38}$$

Ainsi, notre modèle permet de simuler que la forte importance accordée à la compétence par un agent relaxé, combinée aux vitesses élevées d'accroissement d'affiliation et de certitude, lui confèrent un caractère aventureux. L'agent n'hésitera par exemple aucunement à engager la conversation avec d'autres personnes. Un autre agent ennuyé cherchera par contre à éviter les interactions et ne cherchera pas à effectuer de nouvelles actions (i.e. qu'il n'a jamais réalisé et dont il ne connait pas le résultat).

Pour conclure, la figure 3 représente le déroulement d'une interaction telle qu'il est désormais modélisé. En bleu apparaissent les processus déjà existants dans FAtiMA et en rouge les influences de la personnalité que nous avons introduit.

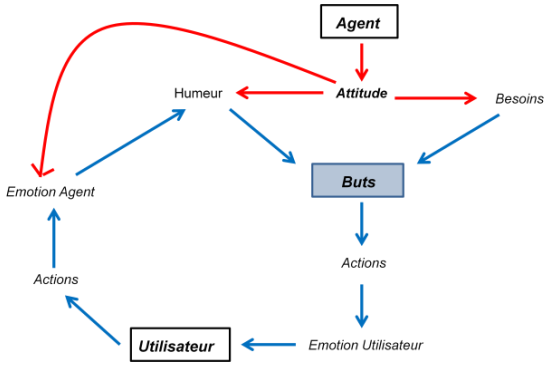


FIG 3 - Déroulement d'une interaction

4 Implémentation et scénario de test

Le modèle proposé a été intégré dans le système FATIMA. Une interface permettant au "concepteur d'agent" d'indiquer la personnalité de l'agent et de visualiser les influences sur l'humeur, les émotions, et les besoins, ainsi que les correspondances en termes des traits de personnalité OCEAN a été développée.

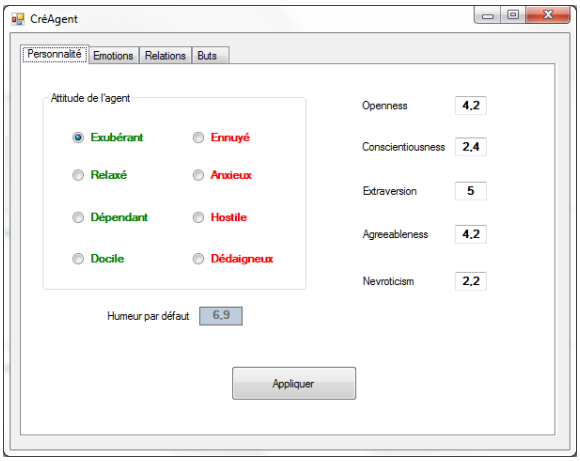


FIG 4 - Interface de création d'agent

Suivant la personnalité de l'agent définie, le comportement de celui-ci est différent. Nous illustrons l'effet de la personnalité sur le comportement de l'agent à travers un scénario mettant en scène des agents dotés de personnalités bien distinctes.

Ce scénario (fourni par Dialonics) met en scène un concessionnaire automobile (l'agent) et son client (l'utilisateur). Il se divise en trois parties distinctes : (1) Une phase d'introduction au cours de laquelle l'utilisateur se présente et se voit proposer l'une des deux gammes de voiture, (2) une phase de discussion durant laquelle ce dernier peut demander à changer de gamme de voiture selon ses désirs et (3) une phase de négociation qui lui permet de tirer les prix vers le bas. A certains moments, l'utilisateur se verra offrir la possibilité de quitter la conversation, mettant ainsi fin au scénario s'il le désire.

Le scénario comporte plusieurs dialogues, qui se déroulent selon un schéma précis. L'utilisateur a en effet à sa disposition deux[3] possibilités de dialogue (en bleu sur la fig.5) : la première polie et la seconde hostile. Cette action déclenchera une émotion particulière chez l'agent, qui viendra modifier son humeur. L'agent aura alors le choix entre plusieurs

[3] Il est prévu d'améliorer le scénario afin d'étoffer le nombre de possibilités offertes à l'utilisateur mais également d'affiner les réponses de l'agent.

réponses (en rouge sur la fig.5) en fonction de cette humeur et de ses besoins. Une humeur négative induira toujours une réponse hostile. Une humeur positive pourra engendrer une réponse exubérante si l'agent accorde une importance suffisante à son besoin d'affiliation ou simplement polie dans le cas contraire.

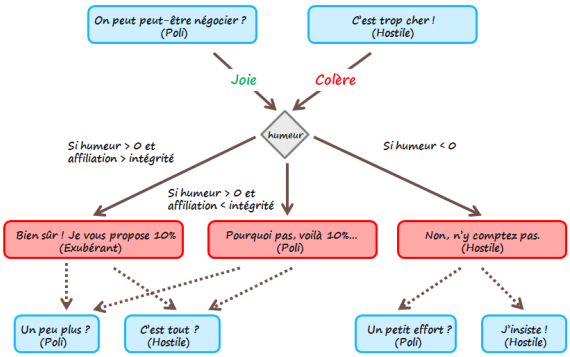


FIG 5 - Exemple de dialogue

L'objectif de l'utilisateur est d'obtenir le meilleur prix possible et cela passe par l'obtention de réductions auprès du concessionnaire. Afin de bien observer l'influence de la personnalité sur les décisions de l'agent, et donc sur la finalité du scénario, deux cas d'utilisation ont été définis. Pour chacun d'eux, l'utilisateur réalise **exactement** les mêmes actions, alternant les phrases polies et les phrases hostiles.

Dans le premier cas, l'utilisateur fait face à un agent défini comme hostile. Ce dernier, d'humeur négative, n'accède pas favorablement aux différentes requêtes de l'utilisateur, notamment concernant une quelconque réduction. Ses réponses sont désagréables tout au long du scénario et il ne sourit presque jamais, même en réponse à des actions polies.

Dans le second cas, l'utilisateur dialogue avec un agent exubérant et donc de bonne humeur. Au terme d'une discussion très agréable, durant lequel l'agent exprime régulièrement sa joie en souriant aux actions polies de l'utilisateur, ce dernier se voit octroyer deux réductions successives, baissant ainsi le prix de départ de la voiture de 20%.

5 Conclusion et perspectives

Dans cet article, nous avons proposé un modèle computationnel permettant de simuler l'influence de la personnalité sur l'humeur, les émotions et les besoins de l'agent Ce modèle a été développé de manière à permettre à des non-experts de déterminer simplement la personnalité d'un agent sans connaissances préalables en psychologie. Ce modèle s'intègre dans l'architecture FAtiMA permettant plus globalement de simuler le processus émotionnel d'un agent. Il a par ailleurs été implémenté et testé à travers un scénario mettant en scène des agents avec différentes personnalités.

De nombreuses améliorations peuvent cependant être envisagées. Par exemple, l'ajout d'un module d'intelligence émotionnelle à l'architecture permettrait à l'agent de rester poli bien qu'étant de mauvaise humeur afin de se soumettre aux normes socioculturelles. La poursuite des travaux entamés en [14] est également envisagée, afin de modéliser l'impact de la personnalité sur les différents paramètres d'expressivité de l'agent.

Remerciements

Les travaux de recherche présentés dans cet article ont été supportés par la société Dialonics. Nous remercions particulièrement Vincent Louis et François Klein concernant les nombreux points sur lesquels nous avons débattus.

Références

[1] E. André, M. Klesen, P. Gebhard, S. Allen et T. Rist, «Integrating models of personality and emotions into lifelike characters,» chez *Proceedings of the workshop on Affect in Interactions - Towards a new Generation of Interfaces in conjunction with the 3rd i3 Annual Conference*, Sienna, Italy, 1999.

[2] J. Bach, «Functional Modeling of Personality Properties Based on

Motivational Traits,» chez *Proceedings of International Conference on Cognitive Modeling (ICCM-7)*, Berlin, Germany, 2012.

[3] J. Bach, «The MicroPsi Agent Architecture,» chez *Proceedings of International Conference on Cognitive Modeling (ICCM-5)*, Bamberg, Germany, 2003.

[4] J. Bates, «The Role of Emotion in Believable Agents,» chez *Communications of the ACM, Special Issue on Agents*, 1994.

[5] E. Bevacqua, E. de Sevin, C. Pelachaud, M. McRorie et I. Sneddon, «Building Credible Agents : Behavior Influenced by Personality and Emotional Traits,» chez *Proceedings of the International Conference on Kansei Engineering and Emotion Research*, Paris, France, 2010.

[6] W. Bledsoe, «I Had a Dream: AAAI Presidential Address,19 August 1985,» *AI Magazine vol.7, n°1*, pp. 57-61, 1986.

[7] E. de Sevin, S. Hyniewska et C. Pelachaud, «Influence des Traits de Personnalité sur la Sélection des Rétroactions,» chez *Quatrième Workshop sur les Agents Conversationnels Animés (WACA '10)*, Lille, France, 2010.

[8] J. Dias, J. Mascarenhas et A. Paiva, «FAtiMA Modular : Towards an Agent Architecture with a Generic Appraisal Framework,» 2011.

[9] H. Eysenck, «Four Ways Five Factors are not Basic,» *Personality and individual differences vol.13, n°6*, pp. 667-673, 1992.

[10] P. Gebhard, «ALMA - A layered model of affect,» chez *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, Utrecht, 2005.

[11] A. Henninger, R. Jones et E. Chown, «Behaviors that emerge from emotion and cognition: implementation and evaluation of a symbolic-connectionist architecture,» chez *Proceedings of the 2nd international joint conference on Autonomous agents and multiagent systems (AAMAS'03)* , 2003.

[12] Z. Kasap et N. Magnenat-Thalmann, «Intelligent virtual humans with autonomy and personality: State-of-the-art,» *Intelligent Decision Technologies vol.1*, pp. 3-15, 2007.

[13] M. Y. Lim, J. Dias, R. Aylett et A. Paiva, «Creating Adaptive Affective Autonomous NPCs,» *Autonomous Agents and Multi-Agent Systems vol.24 n°2*, pp. 287-311, 2012.

[14] M. Mancini et C. Pelachaud, «Generating distinctive behavior for Embodied Conversational Agents,» *Journal on Multimodal User Interfaces vol.3, n°4*, pp. 249-261, 2010.

[15] R. McCrae et P. Costa, «Validation of the Five Factor Model of Personality Across Instruments and Observers,» *Journal of Personality and Social Psychology vol.52, n°1*, pp. 81-90, 1987.

[16] A. Mehrabian, «Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament,» *Current Psychology vol.14, n°4*, pp. 261-292, 1996.

[17] L. Miller, S. Read, W. Zachary et A. Rosoff, «Modeling the Impact of Motivation, Personality, and Emotion on Social Behavior,» *Proceedings of the 3rd international conference on Social Computing, Behavioral Modeling, and Prediction (SBP'10)*, pp. 298-305, 2010.

[18] C. Nass, Y. Moon, B. Fogg, B. Reeves et D. Dryer, «Can Computer Personalities Be Human Personalities?,» *International Journal of Human-Computer Studies vol.43, n°2* , pp. 223-239, 1995.

[19] A. Nazir, S. Enz, M. Y. Lim, R. Aylett et A. Cawsey, «Culture-personality based

affective model,» *AI & Society, vol.24, n°3,* pp. 281-293, 2009.

[20] A. Ortony, G. L. Clore et A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, 1988.

[21] C. Pelachaud, «Modelling Multimodal Expression of Emotion in a Virtual Agent,» *Philosophical Transactions of Royal Society B Biological Science vol.364,* pp. 3539-3548, 2009.

[22] R. Picard, «Affective Computing,» *M.I.T Media Laboratory Perceptual Computing Section Technical Report n°321,* 1995.

[23] H. Prendinger, S. Descamps et M. Ishizuka, «Scripting Affective Communication with Life-Like Characters,» *Applied Artificial Intelligence Journal,* 2002.

[24] B. Reeves et C. Nass, The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press, 1996.

[25] J. Russell, A. Weiss et G. Mendelsohn, «Affect Grid: A Single-Item Scale of Pleasure and Arousal,» *Journal of Personality and Social Psychology vol.57, n°3,* pp. 493-502, 1989.

[26] G. Saucier et L. Goldberg, «Personnalité, caractère et tempérament : la structure translinguistique des traits,» *Psychologie française vol.51,* pp. 265-284, 2006.

[27] K. Scherer, «Emotion,» chez *Introduction to social psychology: A European perspective (3rd edition),* Blackwell, 2000, pp. 151-191.

Sélection d'un vocabulaire commun: une étude autour de l'énaction dans l'interaction entre agents

K. Prepin*
ken.prepin@telecom-paristech.fr

N. Sabouret**
Nicolas.Sabouret@limsi.fr

*LTCI/TSI, Telecom-ParisTech/CNRS,
37-39 rue Dareau,
75014, Paris, France

**LIMSI-CNRS,
BP133,
91403 Orsay Cedex, France

Abstract:
Dans cet article, nous présentons une étude autour de la synchronie dans les interactions entre agents qui vise à reproduire un mécanisme évolutif pour la sélection d'un vocabulaire commun. Notre modèle s'inspire des travaux sur l'énaction [14] tout en l'élargissant au cadre de la communication dialogique. Nous utilisons un modèle simple d'affect attaché aux concepts et nous étudions différentes conditions expérimentales pour la construction de la synchronie. Nous discutons ensuite de l'extension de ces travaux préliminaires vers des modèles plus complexes qui pourraient prendre en compte la sémantique des éléments.

Keywords: Interaction, Alignement et synchronie, enaction, sémantique

1 Introduction

L'étude du langage, malgré son ancienneté - sept siècle avant notre ère, le pharaon Psammétique en proposait déjà une approche scientifique [1] -, que ce soit en philosophie depuis des siècles ou en linguistique depuis deux-cents ans, a longtemps abordé la langue comme une entité autonome, ayant une structure logique intrinsèque, clé de la transmission de sens.

Ce n'est que récemment, ces quarante dernières années, que l'aspect dyadique de la communication par le langage est apparu comme essentiel. En 1966, les annotations d'interactions enregistrées sur vidéos par Condon et Ogston, suggèrent qu'il existe des corrélations temporelles entre les comportements de deux personnes qui discutent [6] : Condon défini en 1976 l'hétérosynchronie comme la synchronie entre partenaires de l'interaction [5]. À peu près à la même période, en linguistique Grice pose pour la première fois la question de l'interprétation du sens de la part de l'auditeur [10], et ouvre la voie à la *théorie de la pertinence* [21] : Sperber et Wilson définissent la *pertinence* d'une interprétation d'un énoncé par le rapport entre son *effet cognitif* (changement de croyance, de but...) et l'*effort cognitif* qu'elle nécessite (hypothèses, inférences...) ; l'interprétation de l'énoncé a alors pour principe que le locuteur cherche à être le plus *pertinent* possible pour l'auditeur. Ainsi, le sens n'est plus défini comme inhérent à l'énoncé, il est considéré comme intrinsèquement dyadique : le sens est lié à la *pertinence* d'une interprétation dans une dyade locuteur/auditeur donnée.

Ainsi depuis les années 80 ce sont les mécanismes collaboratifs qui caractérisent la communication qui sont abordés : d'une part, les interlocuteurs s'alignent sur le plan verbal, c'est à dire qu'ils font en permanence référence aux énoncés de l'autre dans leurs propres énoncés [4] ; d'autre part, les agents d'une interaction, s'alignent sur le plan non-verbal en fonction de la qualité de l'interaction : meilleurs est la communication plus les agents synchronisent leurs comportements [15, 16, 1, 7] et plus ils s'imitent réciproquement [2, 13]. Ainsi, ces dernières années, les Comportement Non-Verbaux (CNVs) apparaissent comme essentiels à la communication : à la fois ils facilitent l'interaction et valident l'échange verbal.

En général, les CNVs sont considérés comme traduisant des *intentions communicatives* [17]. Ils sont de ce fait considérés comme une succession d'éléments sémantiques qui vient s'ajouter au discours au même titre que les éléments du langage [9, 11] pour en modifier, valider, mod-

1. « Psammétique fit remettre à un berger deux nouveaux-nés, des enfants du commun, à élever dans ses étables dans les conditions suivantes : personne, ordonna-t-il, ne devait prononcer le moindre mot devant eux ; [...] Psammétique voulait surprendre le premier mot que prononceraient les enfants. [...] Pendant deux ans le berger s'acquitta de sa tâche, puis un jour, quand il ouvrit la porte et entra dans la cabane, les enfants se traînèrent vers lui et prononcèrent le mot bécos, en lui tendant les mains. [...] [Psammétique] découvrit ainsi que bécos est, chez les Phrygiens, le nom du pain. C'est ainsi que les Egyptiens [...] reconnurent que les Phrygiens étaient plus anciens qu'eux. » [12]

uler le sens.

Prepin et Pelachaud [20] proposent une approche différente pour modéliser l'alignement non-verbal de deux agents. Ils considèrent l'alignement non-verbal, et plus précisément la synchronie, comme un effet de bord d'une bonne communication [1, 20, 19] : la synchronie n'est pas un état que les agents cherchent à atteindre pour montrer qu'ils « sont impliqués, co-opératifs et se comprennent bien », c'est un état de la dyade qui *émerge si et seulement si* les deux agents « sont impliqués, co-opératifs et se comprennent bien » ; à ce titre la synchronie est un marqueur de la qualité de l'interaction. Prepin et Pelachaud proposent un modèle faisant émerger une synchronisation des CNVs de deux agents virtuels si et seulement si les éléments de discours qu'ils échangent ont une *pertinence*, au sens de Sperber et Wilson, similaire pour les deux agents [20, 19].

Nous proposons dans cet article un modèle dual dont l'objectif est de faire émerger la synchronie en prenant pour indice l'alignement *verbal* des agents. Notre objectif est de prendre en compte la pragmatique du dialogue comme élément responsable de l'émergence de la synchronie. Pour cela, nous reprenons le principe de l'attribution d'une valeur de *pertinence* propre à chaque agent, pour chaque élément du discours puis, en nous basant sur les résultats de [20, 19], nous simulons le fait que la synchronie des agents émerge de leur alignement verbal. La valeur de synchronie entre agents est calculée directement à partir de la différence de *pertinence* du discours pour les agents. Les agents se basent sur cette valeur de synchronie pour sélectionner le vocabulaire qu'ils utilisent.

Notre proposition n'est donc pas la construction d'un vocabulaire, comme dans les *talking heads* de Steels [22], à partir d'éléments de base. Au contraire, nous nous plaçons dans une vision dialogique de la synchronie et nous faisons l'hypothèse que les agents savent communiquer. De même, nous ne questionnons pas l'intercompréhension, comme l'ont fait Brassac et Pesty [3]. Nous supposons que les agents se comprennent correctement. L'objectif est alors de sélectionner le vocabulaire commun qui permet de maximiser la synchronie, c'est-à-dire de co-construire un sous-ensemble de sujets d'intérêt partagés et motivants. C'est en cette qualité de marqueur de l'alignement verbal que nous utilisons la valeur de synchronie pour que les agents alignent le vocabulaire qu'ils utilisent.

Nous présenterons en détail une première version du modèle que nous proposons dans la Section 2. Puis nous exposerons les premiers tests et résultats que ce modèle nous permet d'obtenir dans la Section 3. En guise de conclusion, nous discuterons les résultats obtenus avec ce premier modèle Section 4.

2 Modèle

Dans cette section, nous présentons notre modèle d'interaction, support de notre étude sur la sélection de vocabulaire commun fondé sur la synchronie. Nous présentons tout d'abord le modèle de donnée. Nous décrivons notre algorithme de sélection de concepts. Enfin, nous discutons les critères d'évaluation de la convergence.

2.1 Modèle de données

Soit $\mathcal{A}$ l'ensemble des agents et $\mathcal{C}$ l'ensemble des concepts que les agents peuvent utiliser. Le terme *concept* est à prendre au sens large de l'élément de sens, comme les *sinsets* de WordNet [8]. Par exemple, « il faut beau ce matin » peut être vu dans notre modèle comme une unité atomique, un seul élément de sens. D'autre part, ces *concepts* intègrent la notion de contexte : deux ensembles de termes énoncés dans des contextes différents correspondront à des concepts différents. Pour simplifier, nous supposerons que tous les agents ont accès à tous les concepts et nous ne discuterons pas ici le lien entre les concepts manipulés et leur représentation locutoire sous forme de mots.

Pour tout couple $(a, c) \in \mathcal{A} \times \mathcal{C}$, nous notons $E_a(c) \in [0, 1]$ l'*effet cognitif* du concept c pour l'agent a, c'est-à-dire l'intérêt suscitée par l'évocation de c chez l'agent a. Cette valeur représente la pertinence du concept ou son degré de compréhension au sens de Sperber et Wilson [21]. Plus un concept « fait sens » chez l'agent, plus il aura un effet cognitif élevé. La *pertinence* d'un concept vaudra ici 0 si le concept ne fait aucun sens, n'évoque rien de particulier, ne porte aucun intérêt de communication pour l'agent et 1 s'il est au contraire complètement pertinent. De manière générale, les agents chercheront à éviter l'utilisation des concepts associés à une pertinence faible et sélectionneront les concepts ayant une pertinence élevée.

Dans ce travail, nous nous intéresserons à l'étude de la synchronie, c'est-à-dire l'inter-

compréhension des concepts entre deux agents. Cette situation se produit lorsque les agents partagent pour un ensemble de concepts des pertinences élevées. Pour notre étude, nous nous sommes concentrés sur l'interaction entre deux agents a_1 et a_2. Pour simplifier, nous assimilons $E_{a_1}(c)$ à l'effet cognitif du concept c sur l'agent a_1 *dans le contexte de son interaction avec* a_2 (et réciproquement). Nous discutons dans la section 4 du lien entre l'effet cognitif individuel d'un concept et son effet dyadique, dans le contexte d'une interaction donnée.

Modèle d'interaction. Les interactions entre les agents se font par sélection et envoi d'une phrase : à tour de rôle, les agents choisissent un ensemble de concepts (en fonction de leurs préférences) et l'envoie à leur interlocuteur. Dans ce modèle, une phrase p est un multi-ensemble de n concepts $c_1, ..., c_n$ (un même concept peut apparaître plusieurs fois et l'ordre n'est pas important). Nous notons alors $E_a(p)$ l'effet cognitif de la phrase p sur l'agent a, qui est défini par :

$$E_a(p) = \sum_{c \in p} E_a(c) \quad (1)$$

La valeur de synchronie (dans $[0, 1]$) issue de l'interaction pour la dyade (a_1, a_2) est définie par le rapport entre les deux valeurs :

$$synch(a_1, a_2, p) = \frac{min\{E_{a_1}(p), E_{a_2}(p)\}}{max\{E_{a_1}(p), E_{a_2}(p)\}} \quad (2)$$

Notons ici que cette valeur de synchronie est un paramètre émergent de l'interaction. Ce que l'agent perçoit dans l'interaction, n'est pas l'effet cognitif de la phrase sur l'autre (ce qui nécessiterait un modèle de l'autre), il perçoit uniquement la valeur de synchronie, indice de la qualité de l'interaction. Par comparaison entre ses propres réactions et les réactions qu'il perçoit de l'autre, chacun des agents a accès à ce paramètre de synchronie [18].

Nous définissons un seuil de synchronie noté σ tel que l'interaction est considérée comme réussie si et seulement si :

$$synch(a_1, a_2, p) > \sigma \quad (3)$$

Dans ce cas, les agents vont renforcer les $E_a(c)$ pour les c utilisés. Dans le cas contraire, ceux-ci seront diminués de manière à être moins utilisés par la suite.

2.2 Sélection des concepts

Pour tout agent $a \in \mathcal{A}$, nous notons $C_a \subset \mathcal{C}$ l'ensemble des concepts dont la pertinence est supérieure au seuil σ : $c \in C_a \Leftrightarrow E_a(c) > \sigma$. Puisque les pertinences $E_a(c)$ évoluent au fil des interactions, l'ensemble C_a est dynamique.

Pour construire une phrase, chaque agent choisit uniquement des concepts dans cet ensemble C_a. Une phrase de l'agent a est donc définie par :

$$p = \{c_1, ..., c_n\} \text{ tel que } \forall i \in [1, n], c_i \in C_a \quad (4)$$

Autrement dit, chaque terme de la phrase est un terme « préféré » de l'agent orateur. Tous les termes « préférés » ont la même probabilité d'être sélectionné.

Notons que si l'on sélectionne les termes avec des probabilités variables selon leur pertinence, cela conduit à des convergences totales : tous les agents finissent par s'entendre sur l'ensemble du vocabulaire, sans sélection. Statistiquement, chaque $E_a(p)$ va être proche de la moyenne $moy_{c \in \mathcal{C}}\{E_a(c)\}$ et donc la valeur de synchronie $synch(a_1, a_2, p)$ va être proche de 1 (et donc supérieure à σ). Donc tout se renforce [2].

Définition du vocabulaire commun. Pour que la convergence soit possible, il est nécessaire que les agents aient un minimum de vocabulaire commun initial. C'est ce vocabulaire initial qui permet, lors des interactions entre humains, de « parler de la pluie et du beau temps » afin d'assurer une bonne synchronie au départ de l'interaction (on parle de *small talk* en anglais). Dans notre modèle, cela signifie qu'il existe un ensemble de concepts $C_{st} \subset \mathcal{C}$ tel que $\forall a \in \mathcal{A}, \forall c \in C_{st}, E_a(v) > \sigma$: c'est un ensemble de concepts pertinents pour tous les agents. Par conséquent, dans notre étude avec seulement deux agents en interaction, $C_{st} \subset (C_{a_1} \cap C_{a_2})$: il peut exister des concepts pertinents pour les deux agents qui ne sont pas du *smalltalk*.

Nous notons $ratio_{st}$ la proportion de concepts « communs » (au sens du *smalltalk*) dans $\mathcal{C}$:

$$ratio_{st} = \frac{|C_{st}|}{|\mathcal{C}|} \quad (5)$$

La section3 présente une étude du rôle de cette valeur $ratio_{st}$ et de son lien avec la taille des

2. Nous avons essayé...

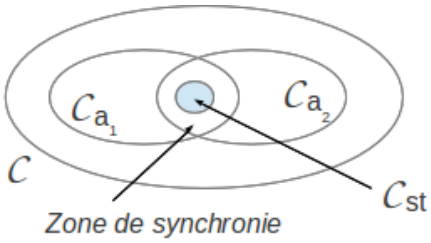


FIGURE 1 – Concepts communs, smalltalk et synchronie.

phrases n dans la convergence vers un vocabulaire partagé. Notre objectif est en effet que les agents puissent s'enrichir au cours de l'interaction. Nous montrerons que, pour atteindre cet objectif, n doit être d'un ordre comparable à $|C_{st}|$. En effet, si $n << |C_{st}|$, les agents risquent de sélectionner uniquement des concepts de type *smalltalk* dans les interactions réussies et le vocabulaire ne sera pas enrichi. Au contraire, si $n >> |C_{st}|$, les phrases contiendront statistiquement trop de vocabulaire disjoint (en terme de pertinence) et les interactions échoueront systématiquement.

Mise à jour des concepts. Lors d'une interaction, les deux agents (orateur et auditeur) mettent à jour leurs pertinences en fonction du succès ou de l'échec de l'interaction : lorsque l'interaction est réussie, les concepts de la phrase seront renforcés. Lorsque l'interaction échoue, leur pertinence sera diminuée. [3]

Toutefois, nous ne voulons pas que tous les concepts soient modifiés de manière proportionnelle. En effet, les concepts « forts » (qui sont les concepts préférés d'un agent, $E_a(c) > \sigma$) devraient être moins pénalisés en cas d'interaction échouée que des concepts déjà faibles. Réciproquement, afin d'arriver à un vocabulaire commun fort, nous avons choisi de renforcer faiblement les concepts déjà « faibles » en cas d'interaction réussie (par symétrie, nous avons choisi $E_a(c) < 1 - \sigma$ comme limite des mots « faibles »).

Pour faire cela, nous avons utilisé une fonction sigmoïde ($f(x) = \frac{1}{1+e^{-\lambda x}}$) (plutôt qu'une fonction linéaire) qui joue le rôle de séparateur à la valeur $1 - \sigma$ pour le renforcement dans les interactions réussies et σ pour la pénalité dans les interactions échouées (voir figure2). Ainsi :

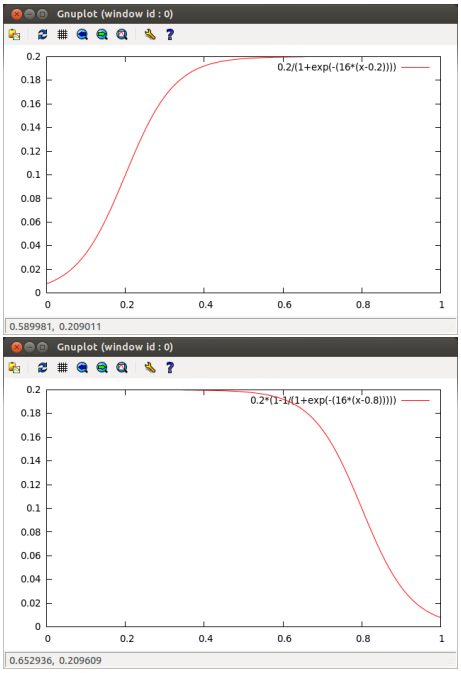


FIGURE 2 – Les sigmoïdes de renforcement (en haut) et pénalité (en bas) pour les concepts lors des interactions.

$$E_a(c)_{t+1} = E_a(c)_t + \Delta_t(c) \tag{6}$$

avec $\Delta_t(c)$ défini, si l'interaction est réussie, par :

$$\Delta_t(c) = \epsilon \times \frac{1}{1 + e^{-16*(E_a(c)_t - 1 + \sigma)}} \tag{7}$$

et si l'interaction échoue, par :

$$\Delta_t(c) = \epsilon \times (1 - \frac{1}{1 + e^{-16*(E_a(c)_t - \sigma)}}) \tag{8}$$

Nous avons choisi $\epsilon = 0,2$ (*cf.* section 3.4), ce qui signifie qu'à chaque interaction, un terme est renforcé ou pénalisé au plus de $0,2$ point.

3. Nous avons fait le choix, dans notre modèle, d'avoir une approche symétrique pour l'orateur et l'auditeur : écouter a la même influence sur les concepts que parler. On pourrait au contraire considérer que l'auditeur n'est influencé que lors d'interactions réussies, et introduire ainsi une dissymétrie. Ceci dit, compte tenu de l'alternance des rôles dans notre algorithme, nous pensons que cela n'introduit pas de différence majeure dans les résultats. De plus, nous ne voulions pas introduire de connaissance supplémentaire dans l'interaction.

2.3 Étude de la convergence

Nous nous intéressons à la construction d'un vocabulaire commun, c'est-à-dire que $C_{a_1} = C_{a_2}$:

$$\forall c \in \mathcal{C}, (E_{a_1}(c) > \sigma \Leftrightarrow E_{a_2}(c) > \sigma) \quad (9)$$

Et on note $C_{a_1,a_2} = C_{a_1} = C_{a_2}$ ce vocabulaire commun. Autrement dit, les mots du vocabulaire commun et uniquement ceux-ci sont préférés pour une interaction entre ces deux agents.

L'étude de la convergence (c'est-à-dire de la sélection d'un tel vocabulaire commun) peut se faire sous plusieurs angles. Dans la section suivante, nous montrerons l'influence de différents paramètres (la taille des phrases n, la proportion de smalltalk $ratio_{st}$, etc) mais il nous semble intéressant de distinguer de prime abord deux dynamiques :
– La dynamique à court terme (au bout de 2 ou 3 échanges) qui correspond à la construction de la synchronie ;
– La dynamique à long terme (lorsque le vocabulaire est stable).
Cette deuxième condition, si elle est intéressante du point de vue informatique, ne correspond pas forcément à une réalité. On imagine mal, en effet, deux agents discuter *ad vitam eternam* entre eux, sans aucune influence extérieure, jusqu'à arriver à un ensemble de concepts partagés. C'est pourquoi nous avons choisi ϵ relativement grand pour observer des convergences rapides.

Ainsi, il ne s'agit pas forcément d'une convergence mais d'un renforcement du vocabulaire, que nous discutons dans la section suivante.

3 Étude du modèle

La question essentielle qui se pose pour l'étude de notre modèle est celle des conditions à la fois de convergence et d'enrichissement du vocabulaire commun. Les paramètres sur lesquels nous allons pouvoir jouer sont la taille n des phrases p et la proportion $P_{st}(p)$ de vocabulaire commun ($\in C_{st}$) dans la phrase p.

3.1 Synchronie et proportion de vocabulaire commun

Lorsque le locuteur ($locut.$) construit une phrase p, par définition (cf. Section 2.2), il ne sélectionne des concepts qu'au sein de $C_{locut.}$. Parmi les concepts de p, certains appartiennent à C_{st} et les autres à $C_{locut.} \cap \overline{C_{st}}$. Par définition, les concepts de C_{st} ont une pertinence forte pour l'auditeur, tandis que les autres ont une pertinence aléatoire pour l'auditeur, inconnue *a priori*. Ainsi, dans p, la proportion de concepts de C_{st} que nous noterons $P_{st}(p)$ est déterminante dans la synchronisation des agents et donc dans le renforcement éventuel des concepts utilisés. Autrement dit, que doit valoir P_{st} (on le fixera pour toute phrase p) pour assurer la synchronisation avec leur partenaire tout en introduisant des concepts nouveaux dans la discussion ?

L'effet cognitif d'une phrase p sur un agent a vérifie l'équation suivante :

$$E_a(p) = E_a(p_{st}) + E_a(p_{\overline{st}}) \quad (10)$$

où p_{st} est la partie de la phrase choisie dans C_{st}, et $p_{\overline{st}}$ est la partie de la phrase choisie dans $\overline{C_{st}}$. Les pertinences de concepts de C_{st} étant proches de 1 et rapidement renforcées jusqu'à 1 si les agents commencent leur interaction par du *smalltalk* (cf. Sec.3.3), nous faisons l'approximation que pour la partie de la phrase p_{st} choisie dans C_{st}, $E_{a_1}(p_{st}) = E_{a_2}(p_{st}) = 1$. D'où, pour un agent a et une phrase p de longueur n,

$$E_a(p) = n \cdot P_{st} + E_a(p_{\overline{st}}) \quad (11)$$

avec $0 \leq E_a(p_{\overline{st}}) \leq n \cdot (1 - P_{st})$.

Dans la description de notre modèle (section 2), nous avons proposé que le locuteur choisisse uniquement des concepts ayant une forte pertinence pour construire ses phrases. Ainsi nous pouvons faire l'approximation que pour le locuteur $E_{locut.}(p_{\overline{st}}) = n \cdot (1 - P_{st})$

Finalement, en reprenant l'équation 11, l'effet cognitif de la phrase sur le locuteur est égale à la longueur de la phrase :

$$E_{locut.}(p) = n \quad (12)$$

Ainsi, le renforcement des concepts émis par le locuteur dépend de la synchronie et donc de la proportion P_{st} de concepts commun selon l'équation suivante :

$$synch(a_1, a_2, p) = \frac{n \cdot P_{st} + E_a(p_{\overline{st}})}{n} \quad (13)$$

avec $0 \leq E_a(p_{\overline{st}}) \leq n \cdot (1 - P_{st})$ soit :

$$P_{st} \leq synch(a_1, a_2, p) \leq 1 \quad (14)$$

3.2 Définition de P_{st}

La proportion P_{st} de concepts communs au sein des phrases sera influencée de deux manières différentes selon que l'on suppose ou non que les agents ont une connaissance de C_{st}.

Stratégie. Dans ce premier cas, les agents connaissent C_{st}, ils ont la possibilité d'avoir une *stratégie* de construction des phrases consistant à contrôler P_{st}. Ainsi, d'après (14) le locuteur en maîtrisant la valeur de P_{st} maîtrise la valeur de synchronie.

Paramètres globaux. Dans le second cas, les agents ne connaissent pas C_{st} *a priori*, ils construisent leurs phrases en choisissant aléatoirement les concepts dans leur vocabulaire pertinent. P_{st} dépend alors des *paramètres globaux* du modèle tels que $ratio_{st}$ (taille de C_{st} au regard de C) et n (longueur des phrases).

D'après la définition de notre modèle, chaque agent a à un vocabulaire composé d'une part du vocabulaire commun C_{st} de taille $|C_{st}| = |C|ratio_{st}$ (cf. (5)), et d'autre part des $|\overline{C_{st}}| = |C|(1 - ratio_{st})$ concepts restant, dont la pertinence à été attribuée aléatoirement. Dans $\overline{C_{st}}$, les pertinences étant tirées uniformément, la proportion de concepts dont les pertinences seront au-dessus de σ sera de $(1 - \sigma)$. Ces concepts seront donc au nombre de $|C|(1 - ratio_{st})(1 - \sigma)$. Finalement pour un agent a, la taille du vocabulaire à forte pertinence vérifie l'équation suivante :

$$|C_a| = |C|ratio_{st} + |C|(1-\sigma)(1-ratio_{st}) \quad (15)$$

Soit, avec un vocabulaire de 20 concepts et une proportion de concepts communs imposés $ratio_{st} = 0.2$, nous avons :

$$\begin{aligned} |C_a| &\simeq 20 \cdot 0.2 + 20 \cdot 0,2 \cdot 0,8 \\ &\simeq 4 + 3 \\ &\simeq 7 \end{aligned} \quad (16)$$

Dans ce contexte, si l'agent construit une phrase en tirant aléatoirement des éléments de C_a, si la phrase est *suffisamment* longue, la proportion P_{st} de concept issus de C_{st} vérifiera :

$$\begin{aligned} P_{st} &= \frac{|C_{st}|}{|C_a|} \\ &= \frac{ratio_{st}}{(1-\sigma)(1-ratio_{st}) + ratio_{st}} \end{aligned} \quad (17)$$

Ce paramètre va directement influer sur la longueur des phrases nécessaire à ce que les agents puissent se synchroniser régulièrement. En effet, les phrases étant composées par définition de concepts de C_a, une phrase trop courte pourra ne contenir que des concepts de $C_a \cap \overline{C_{st}}$ et donc ne pas assurer une synchronisation, tandis qu'une phrase suffisamment longue aura toute les chances de contenir des concepts de C_{st}.
P_{st} va influer sur la longueur minimale des phrases et à σ fixé, et le choix de sa valeur va imposer le choix de $ratio_{st}$ selon l'équation suivante :

$$ratio_{st} = \frac{(1-\sigma)P_{st}}{1 - \sigma P_{st}} \quad (18)$$

3.3 $P_{st} = 1$: le *smalltalk*

La stratégie la plus simple, et aussi la plus efficace pour que les agents s'alignent au niveau non-verbal, est l'utilisation du *smalltalk*, c'est-à-dire que les agents se borne à un ensemble d'éléments sémantiques qui est culturellement connu de tous les agents, C_{st}.

Si les agents utilisent le *smalltalk*, c'est à dire si $P_{st} = 1$, alors les *pertinences* des concepts de C_{st} sont rapidement renforcées jusqu'à 1 (il suffit que les concepts apparaissent dans une phrase) tandis qu'aucune des *pertinences* des autres concepts n'est modifiée.

Dans le contexte où les agents connaissent C_{st} et peuvent choisir une stratégie, ils peuvent choisir $P_{st} = 1$. Par contre dans le contexte où les agents ne connaissent pas C_{st}, en reprenant (18), $P_{st} = 1$ si et seulement si $ratio_{st} = 1$.
Et donc les agents ne feront du *smalltalk* que si ils partagent l'ensemble du vocabulaire.

3.4 P_{st} assurant la convergence

Pour éviter absolument de rompre la synchronie tout en utilisant des concepts en dehors de C_{st}, nous pouvons considérer le pire des cas, celui où tous les concepts extérieurs à C_{st} ont une pertinence nulle pour l'auditeur. Dans ce cas $E_{audit.}(p_{\overline{st}}) = 0$, soit en reprenant l'équation 11 :

$$E_{audit.}(p) = n \cdot P_{st} + 0 \quad (19)$$

Ainsi, dans le pire des cas, d'après (13) nous

avons :

$$synch(a_1, a_2, p) = \frac{n \cdot P_{st}}{n} = P_{st} \qquad (20)$$

Pour que un renforcement soit systématiquement assuré, selon (3) et (20) nous devons avoir :

$$P_{st} \geq \sigma \qquad (21)$$

Ainsi tous les concepts pertinents pour le locuteur mais extérieurs à C_{st} seront renforcés : les agents devraient rapidement converger vers un vocabulaire commun élargi à leurs vocabulaires pertinents à chacun.

Dans notre cas $\sigma = 0,8$, et donc la proportion de concepts d'une phrase devant appartenir à C_{st} est $P_{st} = 0,8$.

Dans le contexte où les agents connaissent C_{st} et peuvent choisir une stratégie, ils peuvent choisir $P_{st} = 0,8$. Par contre dans le contexte où les agents ne connaissent pas C_{st}, la solution pour que $P_{st} = 0.8$ est donnée par (18) :

$$ratio_{st} = 0.44 \qquad (22)$$

3.5 P_{st} plus risquée

Qu'en est-il de cette convergente si, pour fixer la proportion de concepts parfaitement partagés P_{st}, nous ne prenons plus cette fois le pire des cas mais le cas moyen ?

En effet, la pertinence des concepts extérieurs à C_{st} étant fixée aléatoirement, ces concepts ont statistiquement une pertinence de $1/2$ pour l'auditeur. Ainsi, pour des pertinences fixées, N étant le nombre de phrases, en reprenant (11) la pertinence moyenne des phrases vérifie :

$$\lim_{N\to\infty} \frac{1}{N} \sum_{p}^{N} E_{audit.}(p) = nP_{st} + \frac{n}{2}(1 - P_{st}) \qquad (23)$$

soit :

$$\lim_{N\to\infty} \frac{1}{N} \sum_{p}^{N} E_{audit.}(p) = \frac{n}{2}(1 + P_{st}) \qquad (24)$$

La synchronie moyenne des agents vérifie alors l'équation suivante :

$$\overline{synch}(locut., audit., p) = \frac{1 + P_{st}}{2} \qquad (25)$$

Une stratégie *raisonnable* consisterait alors à ne plus considérer le pire cas comme dans la section précédente, mais à considérer le cas moyen. Le renforcement est alors assuré en moyenne si l'équation suivante est vérifiée :

$$\frac{1 + P_{st}}{2} \geq \sigma \qquad (26)$$

soit :

$$P_{st} \geq 2\sigma - 1 \qquad (27)$$

Ainsi avec les paramètres choisis pour notre modèle, $\sigma = 0,8$, une stratégie raisonnable consisterait à choisir $P_{st} \geq 0.6$.

Dans le contexte où les agents connaissent C_{st} et peuvent choisir une stratégie, ils peuvent choisir $P_{st} = 0,6$. Par contre dans le contexte où les agents ne connaissent pas C_{st}, la solution pour que $P_{st} = 0.6$ est encore une fois donnée par (18) :

$$ratio_{st} = 0.23 \qquad (28)$$

En fait cette stratégie qui pourrait sembler raisonnable apparait comme risquée à l'usage, dans la mesure où les concepts choisis lors des premières phrases si ils entraînent un renforcement négatif peuvent fortement diminuer le vocabulaire commun.

4 Discussion et perspectives

Les travaux présentés dans cet article constitue le stade préliminaire de notre étude sur la synchronie et les interactions entre agents et nous avons fait plusieurs hypothèses qui sont sujet à discussion.

4.1 Effect cognitif local et global

Notre étude s'est concentrée sur l'interaction entre deux agents, en dehors de tout autre contexte. Or dans les interactions entre humains, il existe certainement une influence réciproque entre l'intérêt qu'on a pour un concept dans l'absolu, en dehors de toute interaction, et l'intérêt dans le contexte d'une interaction donnée avec une personne définie. Ainsi, il serait plus correct de distinguer d'un côté les $E_a(c)$ et de l'autre des $E_a^{(b)}(c)$ qui seraient l'intérêt suscité

chez l'agent a par l'évocation de c lors de ses interactions avec l'agent b. On peut admettre que, dans l'état initial, $\forall (a,b) \in \mathcal{A}^2, E_a^{(b)}(c) = E_a(c)$ et utiliser l'algorithme décrit dans cet article pour étudier l'évolution des $E_a^{(b)}(c)$. Ainsi, chaque agent se construit un vocabulaire spécifique en fonction de son interlocuteur. Il se pose alors la question de l'influence inverse des $E_a^{(b)}(c)$ sur les $E_a(c)$: dans quelle mesure les échanges que j'ai avec une personne changent mes opinions en général, ou mon intérêt pour certains concepts.

Une première expérience possible est d'utiliser la différence entre $E_a^{(b)}(c)$ et $E_a(c)$ comme information pour l'agent : si l'intérêt d'un concept a beaucoup évolué (en positif ou en négatif), il peut être intéressant de le « tester » dans les prochaines interactions avec d'autre agents pour voir s'il conduit à de bonnes valeurs de synchronie ou non. C'est donc un concept intéressant !

4.2 Sens de la valeur de synchronie

Notre étude a montré une influence forte de la taille des phrases n et de la valeur de $ratio_{st}$ sur la construction du vocabulaire commun. Certains concept « disparaissent » du vocabulaire d'un agent (au sens où il ne les emploie plus, parce que leur effet cognitif est passé sous le seuil σ), d'autres deviennent du vocabulaire commun, mais ces concepts changent selon n et $ratio_{st}$.

Cela provient du fait que la valeur de synchronie n'est pas simplement une valeur de renforcement. Elle représente à la fois l'intérêt porté au concept dans la diade (selon sa position par rapport à σ) et la tendance à utiliser le concept. Ainsi, il se produit que des agents ont des valeurs de synchronies élevées sur des phrases utilisant des concepts à valence faible : les agents sont « d'accord pour ne pas être excités » par ces concepts. Mais justement, parce qu'ils sont en synchronie, cela va les conduire à augmenter leur intérêt pour ces concepts qui les a rapproché.

C'est pourquoi nous pensons qu'il est nécessaire d'étudier plus finement la dynamique de ces valeurs de synchronie.

4.3 Vers plus de dynamique

Une première limite du modèle est que les agents n'ont pas conscience de leur vocabulaire commun. Or au fil des interactions, les termes qui tendent vers $E_a(c) = 1$ sont justement ces termes de forte synchronie. Nous pensons que les agents devraient utiliser cette connaissance pour découvrir d'autres termes du vocabulaire. En effet, dans la situation actuelle, nos agents apprennent à utiliser un sous-ensemble de $C_{a_1} \cup C_{a_2}$. Or dans une interaction entre humains, la connaissance des mots porteurs d'intérêt chez soi et chez l'autre est utilisée pour élargir le vocabulaire commun : découvrir les intérêts de l'autre, ou amener l'autre à partager nos propres intérêts.

C'est pourquoi nous proposons d'étudier un autre mécanisme qui, partant du *smalltalk*, utiliserait les valeurs de synchronie des phrases pour élargir le vocabulaire aux mots « moints intéressants ». Un accumulateur de synchronie, qu'on chercherait à maintenir autour d'une valeur intermédiaire, permettrait de rechercher les échanges synchrones et dignes d'intérêt, tout en utilisant parfois des mots moins porteurs pour le locuteur. Initialement, les agents vont uniquement piocher dans les mots qu'ils savent communs mais dès qu'ils sont « trop » synchrones, ils prendraient des mots plus faibles.

4.4 Vers plus de sémantique

Une deuxième piste que nous souhaitons explorer, est l'utilisation de la sémantique dans notre modèle d'interaction. Actuellement, les concepts ne sont pas reliés et leur sélection n'est pas porteuse de sens. Au contraire, dans un dialogue entre humains, les différents concepts évoqués sont cohérents entre eux, à travers certaines relations. Une première approche qu'il nous semble intéressant d'étudier est de définir une ontologie dont les concepts et les relations seraient les éléments de $\mathcal{C}$. Les concepts seraient alors sélectionnés en tenant compte de cette ontologie (pour avoir un ensemble de concepts sémantiquement cohérents). Mais réciproquement, comme le suggère l'enaction, l'évocation d'un concept influerait aussi sur les concepts voisins (au sens de l'ontologie), en tenant compte du contexte, modifiant ainsi non seulement les préférences des agents (E_a) mais aussi leurs ontologies.

Références

[1] M. Auvray, C. Lenay, and J. Stewart. Perceptual interactions in a minimalist virtual environment. *New ideas in psychology*, 27 :32–47, 2009.

[2] E. Bevacqua, S. Hyniewska, and C. Pelachaud. Positive influence of smile backchannels in ecas. In *International Workshop on Interacting with ECAs as Virtual Characters (AAMAS 2010)*, Toronto, Canada, Oct. 2010.

[3] C. Brassac and S. Pesty. *Analyse et Simulation de conversations : De la théorie des actes de discours aux systèmes multiagents*, chapter Simuler la conversation : un défi pour les systèmes multiagents, pages 317–345. L'interdisciplinaire de Lyon, 1999.

[4] H. H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1) :1–39, 1986.

[5] W. S. Condon. An analysis of behavioral organisation. *Sign Language Studies*, 13 :285–318, 1976.

[6] W. S. Condon and W. D. Ogston. Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, 143 :338–347, 1966.

[7] G. Dumas, J. Nadel, R. Soussignan, J. Martinerie, and L. Garnero. Inter-brain synchonization during social interaction. *PLoS One*, 5(8) :e12166, 2010.

[8] C. Fellbaum, editor. *WordNet, An Electronic Lexical Database* . MIT-Press, 1998.

[9] S. Garrod and M. J. Pickering. Why is conversation so easy ? *TRENDS in Cognitve Sciences*, 8(1) :8–11, 2004.

[10] H. Grice. Logic and conversation. In *Syntax and Semantics*. Academic Press, 1975.

[11] K. Hadelich, H. Branigan, M. Pickering, and M. Crocker. *Alignment in Dialogue : Effects of Visual versus Verbal-feedback*. 2004.

[12] Hérodote. *L'Enquête, II, 1 (-460) ; Œuvres complètes, trad. A Barguet*. Gallimard, 1964.

[13] L. Huang, L.-P. Morency, and J. G. V. R. . I. . 68-79. Virtual rapport 2.0. In *Proceedings of 11th International Conference on Intelligent Virtual Agents :IVA'11*, pages 68–79, Sept 2011.

[14] P. D. Loor, K. Manac'h, and J. Tisseau. Enaction-based artificial intelligence : Toward co-evolution with humans in the loop. *Minds and Machine*, 19(3) :319–343, 2009.

[15] L. Murray and C. Trevarthen. Emotional regulation of interactions vetween two-month-olds and their mothers. *Social perception in infants*, pages 101–125, 1985.

[16] J. Nadel and H. Tremblay-Leveau. *Early social cognition*, chapter Early perception of social contingencies and interpersonal intentionality : dyadic and triadic paradigms, pages 189–212. Lawrence Erlbaum Associates, 1999.

[17] I. Poggi and C. Pelachaud. Emotional meaning and expression in animated faces. *Lecture Notes in Computer Science*, pages 182–195, 2000.

[18] K. Prepin and P. Gaussier. How an agent can detect and use synchrony parameter of its own interaction with a human ? In A. E. et al., editor, *COST Action2102, Int. Traning School 2009, Active Listening and Synchrony. LNCS 5967*, pages 50–65. Springer-Verlag, Berlin Heidelberg, 2010.

[19] K. Prepin, M. Ochs, and C. Pelachaud. Mutual stance building in dyad of virtual agents : Smile alignement and synchronisation. In *Proceedings of International Workshop on Exploring Stances in Interactions, ASE/IEEE International Conference on Social Computing*, pages 938–943, 2012.

[20] K. Prepin and C. Pelachaud. Basics of intersubjectivity dynamics : Model of synchrony emergence when dialog partners understand each other. In J. Filipe and A. Fred, editors, *ICAART 2011, Revised Selected Papers, Series : Communications in Computer and Information Science (CCIS)*, volume 271, pages 302–318. Springer-Verlag, 2012.

[21] D. Sperber and D. Wilson. *Relevance : Communication and Cognition*. Blackwell Publishing, Oxford, UK, 1986.

[22] L. Steels. The Talking Heads Experiment. Technical report, 1999.

Acceptabilité et relations Humain-Compagnons artificiels

D. Duhaut[1] S. Pesty[2]
Dominique.Duhaut@ubs.fr Sylvie.Pesty@imag.fr

[1]Laboratoire LabSTICC
Vannes – France

[2]Laboratoire LIG
Grenoble – France

Résumé :

Nous proposons dans cet article quelques éléments de réflexion sur l'acceptabilité des dispositifs ACA et robots socio-cognitifs, de type compagnons artificiels, et sur le concept de relation entre l'Humain et ces dispositifs. Nous présentons un modèle de relation issu d'un travail pluridisciplinaire et proposons la définition d'un plan des relations Humain-Compagnon Artificiel. Enfin nous nous interrogeons sur la relation entre l'Humain et un collectif de compagnons artificiels.

Mots-clés : Compagnon artificiel, personnage virtuel, robot socio-cognitif, relation Humain-Compagnon Artificiel

1 Introduction

La diversité des outils numériques utilisés dans le cadre de notre vie quotidienne, leur nombre, leur degré de sophistication sont en croissance perpétuelle. Ces dispositifs (terminal de poche à écran tactile, tablettes numériques,… ou même robots domestiques spécialisés pour des tâches de nettoyage par exemple) accompagnent de plus en plus notre quotidien. Ce sont des outils d'assistance, de service, de communication et d'information, devenus désormais indispensables voire incontournables pour beaucoup d'entre nous. De nouveaux dispositifs, les *compagnons artificiels*, c'est-à-dire des dispositifs autonomes, interactifs et intelligents, robots socio-cognitifs et personnages virtuels (ou ACA – Agent Conversationnel Animé), sont amenés à entrer également dans notre quotidien. Pour preuve, de plus en plus de personnages virtuels sont expérimentés ou en application sur des sites web par exemple et du côté de la robotique la société Aldebaran Robotics envisage de commercialiser son robot humanoïde Nao en version grand public fin 2012.

Mais que faut-il pour que ces nouveaux dispositifs entrent réellement dans notre quotidien et soient *acceptables*? Une des pistes explorée est de doter l'entité artificielle de capacités à établir et maintenir une *relation à long terme* avec son usager (Bickmore & Picard, 2005) (Pfeifer Vardoulakis L., Ring L., Barry B., Sidner L.S., & Bickmore T., 2012)

Nous présentons dans cet article quelques éléments de réflexions autour de ces questions. Dans un premier temps nous revenons sur les travaux de *l'acceptabilité* des technologies et montrons comment la problématique du compagnon artificiel complexifie ce sujet. Ensuite nous présentons un résultat du groupe de travail pluridisciplinaire MIAC[1] de l'Université Européenne de Bretagne, qui a initié un travail de modélisation de la notion de *relation*. Ensuite, nous proposons la définition d'un *plan des relations* Humain-Compagnon Artificiel selon deux dimensions : *proximité* et *ascendance*. Ce plan des relations permet de définir neuf catégories de compagnons artificiels dont nous donnons quelques caractéristiques. Enfin nous posons le problème de diverses perceptions par l'homme d'une relation avec ses compagnons artificiels.

[1] http://www-valoria.univ-ubs.fr/miac/

2 L'acceptabilité des compagnons artificiels

Si la question de l'acceptabilité des technologies est étudiée depuis les années 1980, l'approche a été pour beaucoup fonctionnelle et a permis de distinguer clairement les deux notions importantes que sont l'*utilité* et l'*utilisabilité* d'un système (Davis F.D., 1989) (Nielsen, 1993). Certains travaux ont mis en avant la propriété d'*adaptabilité* au contexte et aux préférences de l'utilisateur traduisant la capacité du système à être flexible, par exemple en fonction des habitudes d'un usager. Bien d'autres critères interviennent également dans le fait qu'un usager va ou non utiliser un dispositif et de nombreux chercheurs en Sciences Humaines et Sociales et en Ergonomie Cognitive ont développé une approche de *l'acceptabilité sociale* qui met en avant l'importance des facteurs comme les normes sociales, l'image de soi, etc. allant jusqu'au concept de symbiose humain-technologie (Brangier, 2003) (Brangier, Hammes-Adelé & Bastien, 2010).

Les dispositifs qui nous intéressent ici, les compagnons artificiels, sont différents des dispositifs jusqu'à présent étudiés : le fait qu'il s'agisse d'*entités d'intelligence artificielle* avec notamment des capacités de décision, d'autonomie, mais aussi des capacités d'expression verbale et non verbale, qu'ils soient *incarnés* et pour les robots, *immergés* dans l'espace physique de l'usager, cela pose de nouvelles questions sur leur acceptabilité dans nos vies de tous les jours.

Dans le domaine des ACA, la question de l'acceptabilité a le plus souvent été appréhendée sous celle de l'*expressivité* (expression de l'émotion en particulier, faciale, mais aussi gestuelle) et celle de la *crédibilité* des personnages, deux caractéristiques considérées comme essentielles (Bates, 1994). L'acceptabilité de ces nouveaux dispositifs est aussi étudiée dans son acception la plus large ou dans des expérimentations sur des publics ciblés de (Klamer & Ben Allouch, 2010) (Odetti, et al., 2007) (Pesty & Duhaut, 2011).

Plus récemment, c'est la notion de *relation,* voire de *relation à long terme* qui retient l'attention des chercheurs. Dans l'article très complet de Bickmore et Picard (Bickmore & Picard, 2005), des *agents relationnels* sont définis comme étant des ACA capables de construire et d'entretenir une relation sociale et émotionnelle à long-terme avec l'utilisateur. Les auteurs s'appuient plus particulièrement sur les comportements identifiés chez les humains pour construire et entretenir leurs relations, et énumèrent ainsi les nombreux points à considérer pour modéliser un agent relationnel : l'empathie, la communication relationnelle et méta-relationnelle, les émotions, l'humour, les références à des connaissances mutuelles, aux valeurs, à l'histoire de l'interaction, etc. D'autres éléments comme la synthèse vocale (qualité de la voix, prosodie), la justesse du langage (choix lexical adapté à l'utilisateur) sont également à prendre en considération. L'aspect visuel est aussi de toute évidence important, même s'il peut induire des réactions allant jusqu'au rejet instinctif du dispositif par l'utilisateur (phénomène de l'« *uncanny valley* »[2], réaction des utilisateurs à l'anthropomorphisme d'un robot). Mais la question « Comment nos dispositifs que nous appelons compagnons artificiels peuvent-ils établir puis maintenir une relation avec un humain ? » est encore loin d'être résolue.

Pour clarifier la notion de relation Humain-Compagnon Artificiel, nous proposons, dans la continuité des travaux du projet MIAC (Modélisation interdisciplinaire de l'acceptabilité et de l'intercompréhension dans les interactions)[3] de l'Université Européenne de Bretagne, d'approfondir la notion de relation Humain-Compagnon artificiel et définissons un *plan des relations* selon deux dimensions, « *proximité*» et « *ascendance* » permettant de positionner neuf catégories de compagnons artificiels.

[2] ou « vallée dérangeante » défini par le roboticien japonais Masahiro Mori en 1970

[3] http://www-valoria.univ-ubs.fr/miac/

3 Modèle de la dynamique de la relation

Entre 2010 et 2012 le projet MIAC a réuni des chercheurs de disciplines différentes, Sciences du Vivant, Sciences Humaines et Sociales et Informatique, dans l'objectif de croiser les regards sur les interactions et relations entre 3 types d'entités : les hommes, les animaux et les machines.
Ce projet a conduit à proposer un modèle décrivant la relation et la dynamique de la constitution d'une relation entre deux entités intra ou inter spécifiques (Grandgeorge & Duhaut, 2011) (Le Pevedic, Grandgeorge, M., & Pugnière, 2012). Une *relation* entité-entité est définie comme le résultat de *mises en présence* répétées de deux entités caractérisées chacune par leur *identité,* ces mises en présence se déroulant dans différents *contextes*. La relation est donc une co-construction puisque le temps et la double présence sont nécessaires.

Ce modèle est résumé par la figure ci-dessous (figure 1) et nous en décrivons les principaux constituants :

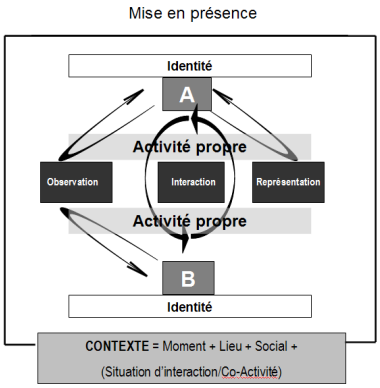


FIG. 1 – Modèle de la dynamique de constitution de la relation

Identité

La première notion proposée par la définition de la relation est celle d'Identité. Cette identité se décline selon trois dimensions :

- Dimension de la *personne* : cela représente les caractéristiques personnelles, c'est-à-dire le biologique, les aptitudes, la personnalité, les capacités, les besoins
- Dimension du *savoir* : cela représente les acquis relativement stables, c'est-à-dire les connaissances, croyances, normes, valeurs
- Dimension du *style* : cela représente les acquis en évolution rapide, c'est-à-dire les habitudes, besoins, motivations, émotions

Cette notion d'identité est définie pour l'Humain et les dimensions ne sont pas toutes pertinentes pour des dispositifs numériques amenés à être des compagnons artificiels. En particulier les aspects relatifs à la biologie, aux habitudes, aux besoins ne semble pas avoir de d'utilité. D'autres traits sont en revanches nécessaires et complexes. Les **capacités** sont bien évidement ce qui fait l'intérêt premier du compagnon artificiel puisqu'il est sensé fournir des services; ce sont les aspects d'*utilité* et d'*utilisabilité* des modèles d'acceptabilité. Ses capacités doivent donc inclure la sécurité, la fiabilité et l'affordance nécessaire pour un usage simple. La **connaissance** se décline pour un compagnon artificiel suivant trois dimensions (Duhaut, D, 2010) : les connaissances universelles, les connaissances sociales sur le groupe auquel il appartient et enfin des connaissances propres comme ses goûts, ses préférences, etc. Ces trois niveaux de connaissances doivent être associées à un référentiel de **valeurs** pour que le compagnon artificiel puisse savoir ce qui est bien/mal, beau/laid, positif/négatif. Les derniers aspects de l'identité qui semblent être pertinents pour le compagnon artificiel sont la **personnalité** et les **émotions**.

Contexte

Le contexte permet de décrire la situation dans laquelle se trouvent les deux identités en présence. Il se décline autour de quatre dimensions :

- Le *moment* correspond à la donnée temporelle du contexte. La situation peut-être répétitive (tous les matins à la même heure) inattendue, le matin, pendant l'hiver …
- Le *lieu* correspond au lieu géographique de la situation. Les comportements varient en fonction des lieux. Par exemple une église,

une plage ou une école sous-tendent des comportements différents.

- Le *social* correspond à l'environnement social d'une situation. Un humain avec son compagnon artificiel ne se comportera pas nécessairement de la même façon s'il est seul avec lui ou entouré d'autres personnes. L'environnement social peut être anonyme comme une foule dans le métro ou chargé d'un message comme dans un cortège pour un mariage.
- Enfin la *situation d'interaction* permet de décrire ce que les deux entités en présence font comme tâche dans l'environnement. En particulier dans la réalisation commune d'une tâche se pose des problèmes de répartition d'autorité : qui dirige, comment s'organise la planification des tâches, la définition des buts, la vérification de la progression de la tâche ?...

Mise en présence

La mise en présence est un élément essentiel puisque c'est à travers elle que la relation se construit, se développe, évolue. Dans le cas très général d'une relation entité-entité, la mise en présence se réalise de trois façons : soit par des *interactions* directes (les entités échangent des signaux à destination de l'autre), soit par des *observations* (une entité observe l'autre à son insu) ou encore par des *représentations* (les entités n'ont pas d'accès direct à l'autre mais uniquement au travers de leurs productions, des productions artistiques par exemple).

Dans le cas qui nous préoccupe, celui d'une relation Humain-Compagnon artificiel, seule la mise en présence par interactions directes nous intéresse. En effet, l'observation et la représentation sont des mises en présence pour lesquelles le compagnon artificiel est regardé ou évalué par l'Humain, et donc qui ne pose pas de problème particulier. Lors d'une mise en présence par interactions, l'Humain et le compagnon artificiel échangent, créant le lien ; ces échangent peuvent être de deux natures, communication ou co-activité.

La communication va être contrainte par l'incarnation du compagnon artificiel (robot ou personnage virtuel) car dépendante des capteurs et effecteurs communicationnelles. Ainsi tel personnage virtuel disposera d'une sortie audio et pourra communiquer verbalement, alors que tel robot ne pourra communiquer qu'en se mouvant dans l'espace par exemple. La question du langage de communication est fondamentale. Faut-il reproduire les phénomènes conversationnels humains et communiquer en langue naturelle avec un compagnon artificiel ? De nombreux travaux vont en ce sens et les systèmes de dialogue oraux progressent même si les difficultés restent nombreuses tout au long de la chaîne de traitement : en reconnaissance vocale, en compréhension de parole, en gestion des tours de paroles, en gestion du dialogue et en génération de la réponse. Mais ne faudrait-il pas communiquer autrement ? Faudrait-il inventer un « espéranto » par exemple ?

Un humain et un compagnon artificiel peuvent également être en présence pour co-agir. La co-activité consiste à réaliser une tâche ensemble, préparer une recette de cuisine ou transporter une table à deux par exemple. Cette activité s'accompagne de deux grands problèmes que le compagnon artificiel doit savoir résoudre : la planification de la tâche et sa réalisation conjointe. Ces problèmes ne sont pas simples à résoudre ; ils posent les questions du contrat passé entre l'homme et son compagnon, de la capacité à prédire un comportement ou à le comprendre. G. Klien et ses collaborateurs proposent dix challenges à relever qui sont pertinents pour cette problématique (Klien, Woods, Bradshaw, Hoffman, & Feltovich, 2004).

Ce modèle proposé par le groupe de travail MIAC propose donc une définition de la relation entre deux entités, intra ou inter spécifiques. Elle ne propose pas en revanche de la catégoriser ; c'est ce que nous proposons de développer au paragraphe suivant.

4 Plan des relations et catégories de compagnons artificiels

Si l'on se réfère aux relations interpersonnelles,

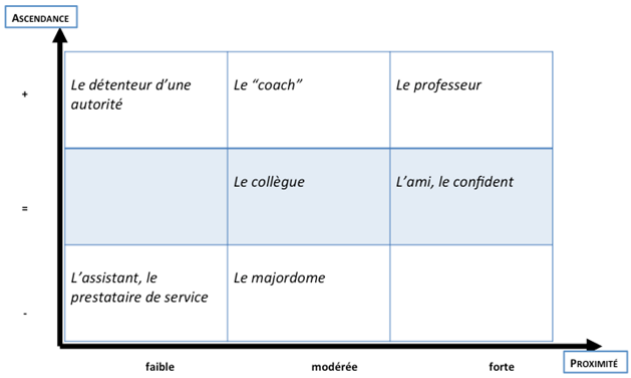
on ne peut que constater qu'elles sont extrêmement complexes et de nature très variée. Entre la relation de travail entre collègues et la relation d'intimité avec un ami, un frère ou un conjoint, il existe une infinité de possibilités. Chaque nature de relation implique un comportement adapté : on ne se comporte pas de la même façon avec ses proches qu'avec ses collègues. Le contexte dans lequel se déroule l'échange est également déterminant ; entre amis par exemple, on se comportera différemment à la mairie lors d'un mariage ou dans une boîte de nuit. Il y a donc des codes, des rituels, des niveaux de langages, des tenues vestimentaires,… que l'on adopte en fonction de l'autre et de la situation. De plus, du fait de la diversité des personnalités, la variété des relations s'en trouve augmentée. Les recherches en psychologie et psychologie sociale sont très fécondes sur ces sujets et apportent de nombreux éclairages.

En ce qui concerne nos relations Humain-Compagnon artificiel, peut-on tenter de catégoriser les compagnons et les relier à un type de relation ?

Nous proposons ici différents prototypes de compagnons artificiels que nous positionnons dans un espace des relations défini par deux dimensions. La première dimension représente l'axe que nous appelons « *proximité* ». En effet, la nature des relations est en grande partie fonction de la distance qui existe entre les personnes (Marc & Picard, 2000). La distance qui nous intéresse ici est la distance dite « psychique » et nous mettons de côté la distance physique entre l'Humain et le Compagnon artificiel qui pourtant est une donnée importante entre deux personnes. Cette distance caractérise d'une part la proximité intellectuelle entre les protagonistes de la relation, c'est-à-dire les connaissances, valeurs, croyances partagées et d'autre part le niveau d'affection qui les unit.

La deuxième dimension qui nous semble pertinente de retenir représente l'axe que nous appelons « *ascendance* ». Elle représente le rapport de subordination qui s'établit dans la relation en fonction de la *position* prise par chaque protagoniste. On parle également de *structure de la relation,* les protagonistes d'une relation pouvant se situer dans un rapport de symétrie ou d'asymétrie. Par exemple on aura un rapport hiérarchique chef/employé, ou un rapport égalitaire entre deux amis, un rapport complémentaire professeur/élève, un rapport d'autorité entre parent/enfant… Notons que ce rapport est très souvent dicté par l'aspect institutionnel qui implique des rôles sociaux préétablis. La similarité des positions, ou au contraire la dissemblance des positions, détermine la nature et la forme des échanges, mais ne détermine pas la valeur (bonne ou mauvaise) de la relation.

Ces deux dimensions *proximité* et *ascendance* définissent notre plan des relations Humain-Compagnon artificiel (cf. fig2). Il est à rapprocher des nombreux travaux en psychologie, initiés par Timothy Leary (Leary, 1957) autour du concept de « cercle interpersonnel » (ou interpersonal circumplex)



pour positionner les comportements interpersonnels.

FIG. 2 – Le plan des relations

Ce plan des relations permet de définir neuf catégories de relation, chaque catégorie étant représentée par un compagnon artificiel prototypique, membre le plus représentatif de la catégorie. Il faut noter que certaines relations peuvent ne pas être pertinentes entre un Humain et un Compagnon artificiel.

A titre d'exemple, nous donnons ici la description brève de quatre des compagnons

artificiels prototypiques :

Le compagnon assistant, prestataire de service (relation asymétrique - ascendance de l'humain, proximité faible à modérée) : il est sous les ordres de l'Humain et la relation est donc asymétrique, dans un rapport de subordination. Il ne donne des informations que sur demande, il coopère, aide à faire une tâche, s'il est sollicité. Il est au service de l'Humain. En termes de personnalité, il est coopératif mais se montre modéré et posé dans ses échanges avec l'Humain.

Le compagnon détenteur d'une autorité (relation asymétrique - ascendance du compagnon, proximité faible) : il est garant du bon fonctionnement de la situation et veille à la bonne exécution d'une tâche. Il est légitime dans la situation. Il se positionne comme détenteur de l'autorité et intervient fermement quand les choses ne vont pas dans le sens voulu. En termes de personnalité, il se montre le plus souvent distant pour asseoir son autorité.

Le compagnon « coach » (relation asymétrique - ascendance du compagnon, proximité moyenne) : il soutient la vie de tous les jours en fournissant des rappels de tâches à faire ou prend l'initiative de donner une information qu'il estime pertinente. Il sait encourager. Il se positionne presque sur un plan d'égalité, même si la relation reste asymétrique. En termes de personnalité, il se montre le plus souvent souriant, chaleureux et agréable.

Le compagnon ami, confident (relation symétrique, proximité élevée) : il comprend l'Humain, il est toujours là pour lui, il lui apporte un soutien psychologique et affectif. Il est coopératif et adopte facilement les attitudes de l'Humain. Il est à égalité avec l'Humain, il est doté d'empathie.

5 De *un* à *des* compagnons artificiels

Dans la mesure où les compagnons artificiels de notre quotidien seront de plus en plus nombreux à accompagner notre vie de tous les jours et que nous les utiliserons fréquemment, une question se pose alors, celle de la *représentation* que nous aurons de ce collectif de compagnons artificiels. Il y a de bonnes raisons de penser que nous aurons une personnification « forte » de ces dispositifs puisque l'humain en est déjà coutumier (Byron, Reeves, Clifford, & Nass, 1998). Si plusieurs compagnons artificiels doivent nous assister pour des actions personnalisées alors il est probable que les compagnons artificiels partageront entre eux un certain nombre d'informations qui pourront nous donner l'impression qu'ils collaborent, qu'ils forment un collectif. Dans cette mesure, la collaboration entre les entités fera peut-être émerger chez l'homme un sentiment d'une unique présence malgré les différents dispositifs ACA et robots. Cette impression de n'avoir qu'*un* compagnon, relayé par de multiples dispositifs faisant partie de « sa famille », impression qui est souvent qualifiée négativement de « big brother » ou positivement de « big mother », doit-elle être un objectif recherché lors de la conception des compagnons artificiels ? La figure 3 illustre cette question.

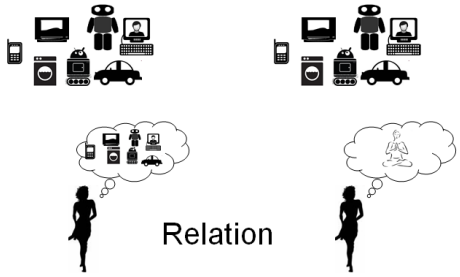


FIG. 3 – Comment perçoit-on ses compagnons artificiels ?

L'homme entretiendra-t-il des relations multiples et différentes avec ses dispositifs ou aura-t-il le sentiment d'être en relation avec une entité unique abstraite ? Selon ce second point de vue, les dispositifs ACA et robots (appelons-les, les *compagnons physiques*) deviennent des incarnations partielles de l'entité abstraite supérieure. Chaque compagnon physique est doté de ses capacités propres, certaines étant communes à d'autres. Ainsi le compagnon artificiel serait cette entité abstraite qui supervise le collectif de compagnons physiques

et qui choisit la meilleure incarnation à un instant donné par rapport à une tâche donnée.
Cette approche présente de notre point de vue trois avantages :

- elle propose à l'homme une interface unique pour le pilotage de l'ensemble des dispositifs ACA et robots qui l'entourent
- elle permet à des prestataires de service de « programmer » à travers cet unique compagnon abstrait des services complexes utilisant plusieurs des dispositifs
- elle permet, en gardant une structure d'apparence identique, d'être compatible avec l'évolution et la variété des dispositifs ACA et robots qui seront disponibles sur le marché

Interface unique

Même dématérialisé le compagnon artificiel abstrait est toujours présent et peut s'incarner à la demande de l'usager dans les compagnons physiques qui l'entourent. Le type d'incarnation dépend des propriétés de chaque compagnon physique. La question qui se pose alors est : comment l'homme peut-il avoir l'impression de retrouver « son » compagnon abstrait à travers chacun de ses compagnons physiques ? C'est sur ce point que la personnalité et les émotions du compagnon abstrait vont jouer un rôle important puisque ce sont ses traits de personnalité, son identité que l'homme devra retrouver à travers le compagnon physique avec lequel il interagit.

Prestataires et scripts

La capacité à disposer d'un collectif de compagnons physiques en relation les uns avec les autres, incarnant un seul compagnon abstrait, permet d'envisager des utilisations et des usages non prévus initialement par les concepteurs des dispositifs. Par exemple, la collaboration entre plusieurs compagnons physiques peut permettre de détecter la chute d'un objet ou d'une personne alors qu'un seul dispositif ne pourrait à lui seul le faire. Ainsi, un prestataire de service peut, via l'écriture de scripts par exemple, confier au compagnon abstrait le soin de coordonner lui-même les actions des différents compagnons physiques, réalisant ainsi des fonctionnalités plus complexes.

Evolution du marché

Le dernier point est la probable forte dynamique du marché des dispositifs ACA et robots. De nouveaux dispositifs viendront remplacer, substituer ou introduire des fonctionnalités nouvelles. L'intégration d'un nouveau compagnon physique dans « sa » famille doit se faire sans perte d'acquis et l'humain doit avoir le sentiment de retrouver immédiatement « son » compagnon artificiel abstrait derrière ce nouveau dispositif tout juste arrivé.

6 Conclusion

Dans cet article, nous avons proposé une réflexion sur l'acceptabilité des dispositifs ACA et robots, de type compagnons artificiels, et sur le concept de relation entre l'Humain et ces dispositifs. En particulier nous avons d'abord montré que ces nouveaux dispositifs ouvrent de nouvelles questions autour de l'acceptabilité des technologies. Ensuite nous avons présenté un modèle de relation issu d'un travail pluridisciplinaire et proposé la définition d'un plan des relations Humain-Compagnon Artificiel. Enfin nous nous sommes interrogés sur la multiplicité des compagnons artificiels amenés à accompagner l'Humain dans sa vie quotidienne et la vision que l'Humain pourrait se faire de ce collectif.

Références

Byron, Reeves, Clifford, & Nass. (1998). *The Media Equation How People Treat Computers, Television, and New Media Like Real People and Places.* University of Chicago Press.

Bates, A. (1994). The role of emotion in beleivable agents. *Communications of the ACM , 37* (7), 122-125.

Bickmore, T. W., & Picard, R. (2005). Establishing and Maintaining Long-Term

Human-Computer Relationships. *ACM Transactions on Computer Human Interaction , 12*, 293-327.

Brangier, E. (2003). *Le concept de "symbiose homme-technologie-organisation".* (Vol. 3). (N. Delobbe, G. Karnas, & C. Vandenberg, Éds.) Louvain: Presses universitaires de Louvain.

Brangier, E., Hammes-Adelé, S., & Bastien, J.-M. (2010). Analyse critique des approches de l'acceptation des technologies : de l'utilisabilité à la symbiose humain-technologie-organisation. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology , 60* (2), 129-146.

Duhaut, D. (2010). A way to put empathy in a robot. *Proceedings of the International Conference on Artificial Intelligence (ICAI)*. Las Vegas.

Davis F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology . (M. Quarterly, Éd.) *13*, 319-339.

Grandgeorge, M., & Duhaut, D. (2011). Human-Robot: from Interaction to Relationship. *Proccedings of The 14th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR2011)*. Paris.

Klamer, T., & Ben Allouch, S. (2010). Acceptance and use of a social robot by elderly users in a domestic environment. *Pervasive Computing Technologies for Healthcare (PervasiveHealth)* (pp. 1-8). IEEE.

Klien, G., Woods, D., Bradshaw, J., Hoffman, R., & Feltovich, P. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *Intelligent Systems, IEEE Computer Society , 19* (6), 91-95.

Le Pevedic, B., Grandgeorge, M., & Pugnière. (2012). *Intercompréhensions comparées dans trois types d'interactionsHomme-Homme, Animal-Homme-Machine et Homme-Machine.* . Fernelmont, Belgique: E.M.E Editions.

Leary, T. F. (1957). *Interpersonal Diagnosis of Personality: A Functional Theory and Methodology for Personality Evaluation.* (N. Y. Company, Éd.) Etats-Unis.

Nielsen, J. (1993). *Usability Engineering*. New-York: Academic Press.

Marc, E., & Picard, D. (2000). *Relations et communications interpersonnelles.* Dunod.

Odetti, L., Anerdi, G., Barbieri, M., Mazzei, D., Rizza, E., Dario, P., et al. (2007). Preliminary experiments on the acceptability of animaloid companion robots by older people with early dementia. *Engineering in Medicine and Biology Society. 29th Annual International Conference of the IEEE* (pp. 1816-1819). IEEE.

Pesty, S., & Duhaut, D. (2011). Acceptability in Interaction : From Robots to Embodied Conversational Agents. *GRAPP 2011 International Conference on Computer Graphics Theory and Applications* (pp. 365-370). SciTePress.

Pfeifer Vardoulakis L., Ring L., Barry B., Sidner L.S., & Bickmore T. (2012). Designing relational agents as long term social companions for older adults. *12th International Conference on Virtual Agents*. Springer.

Moods and Moral Values in Blog Posts

M. Généreux† R.P. Evans††
genereux@clul.ul.pt R.P.Evans@brighton.ac.uk

†Centro de Linguística da Universidade de Lisboa
Lisbon, Portugal

††Natural Language Technology Group
Brighton, United Kingdom

Résumé :
Dans cet article, nous utilisons une approche basée sur les corpus pour explorer la relation entre les valeurs morales et le bien-être humain dans les médias sociaux. Dans la foulée, nous examinons s'il y a un effet significatif relié au genre et à l'âge en ce qui concerne l'intérêt porté à la morale et quelles sont les valeurs que l'on priorise. Nos résultats supportent en partie l'idée d'une relation directe entre la moralité et le bien-être ainsi qu'un effet du genre et de l'âge sur l'intérêt des gens vis-à-vis des questions morales, mais aucun effet réel sur la priorité des valeurs.

Mots-clés : blogues, bien-être, moralité

Abstract :
In this paper we use a corpus-based approach on blog posts to investigate the relationship between moral values and human well-being in social media. In the process, we look at whether there is a significant gender and age effect with regards to interest in morality and value priority. We find some support for a direct relationship between morality and well-being and a gender/age effect on the extent to which people are interested in moral questions, but no real effect on value priority.

Keywords : blog posts, well-being, morality

1 Introduction

In his book *The Moral Landscape* [4], Sam Harris defends the idea that human (and animal) well-being is the direct consequence of people living ethical lives by prioritizing specific moral values. In other words, well-being is the direct consequence of what kind of moral principles or actions we apply for ourselves and others. This way of thinking implies that to experience well-being we should prioritize moral values that impact positively on us and others. To put it more simplistically, if we want to be happy, we should be good. This paper looks for supporting evidence for such claims using a corpus-based approach, by looking into the relationship between moods (as a proxy for well-being) and value priorities in blog posts.

We make the assumption that people's mood as well as the type of moral values they prioritize can be detected through their writing, and moreover that moral values and mood expression can be detected through lexical content alone, something which has already been verified empirically for mood [5]. The questions that interest us are to what extent a relationship exists between the morality and mood of a blogger, and to what extent such a relationship varies in bloggers of different age and gender, as reflected through their blog posts.

2 Models for moods and moral values

Our model or typology for moods [1] is shown in figure 1. This typology is based on a model of emotion as a multicomponent process [9]. In this model, the distribution of the affective states is the result of analysing similarity judgements by humans for a set of emotion terms using cluster-analysis and multidimensional scaling techniques to map out the structure as a two-dimensional space. For present purposes, we are primarily interested in the distinction between positive moods, represented by the left half (quadrants 3 and 4), and negative moods, represented by the right half (quadrants 1 and 2) [2]. The positioning of words in this space is somewhat 'fuzzy' ; an affective state such as "angry" to describe facial expression in speech may have a slightly different location than an "angry" blog post. Nevertheless this model has been shown to correlate well with automatic measurement of affect in texts [3], so it is well suited for our purpose.

Our reference model for value types is based on [11], as shown on Figure 2. Schwartz derived ten motivational types of values from the universal requirements of human existence and further research offers considerable evidence to support the comprehensiveness of the model [12, 13, 1]. This model synthesizes the results of

1. We are treating "mood", "affective state" and "emotion" as the same thing.

2. The division between active and passive moods is shown only for completeness, as it is part of Scherer's model.

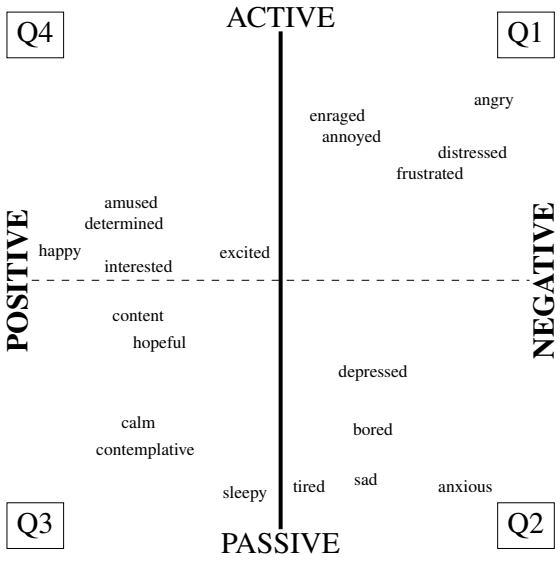


FIGURE 1 – Typology of affective states based on [9].

cross-cultural research in more than 200 samples from over 60 countries. Competing value types emanate in opposing directions from the center ; complementary types are in close proximity going around the circle [3]. Subsequently, Sagiv and Shwartz [8] carried out a study to investigate the relationships of value priorities to measures of subjective well-being (1261 subjects). Their results provided some of the first evidence for what we are trying to show : a direct association between basic value priorities (as measured by lexical frequency) and the affective component of subjective well-being (as measured by self-annotated moods). *Achievement*, *self-direction* and *stimulation* values correlated positively with well-being on the pure affective index, while *tradition*, *conformity* and *security* values correlated negatively. The positive correlations are shown in the dotted pattern in figure 2 and negative correlations in the checkered pattern. One of the three types of measures of subjective well-being used in the study was the *Bradburn affect scale*, a measurement index using an affective vocabulary similar to the moods in figure 1. Our first experiment will try and replicate these findings in blog posts, using Scherer's model to measure well-being. As we already mentioned, earlier studies have verified that this model is a good description of the distribution of affect in a two-dimensional space : in other words, quadrants in the model are clearly separable. For our purpose, it is sufficient that we can separate *positive* (i.e. quadrants 3 and 4) from *negative* (i.e. quadrants 1 and 2) moods.

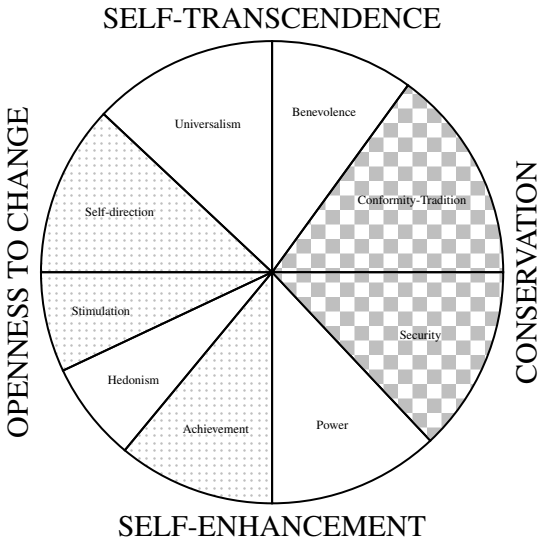


FIGURE 2 – Theoretical model of relations among ten motivational types of values based on [8]. The positive correlations are shown in the dotted pattern and negative correlations in the checkered pattern.

3 Corpora

For our experiments we will use two corpora. The first corpus is called *LiveJournal* and was harvested through the LiveJournal [4] website during the year 2005 for a total of 28,672 blog posts (mood-annotated by authors), belonging to one of the four quadrants from figure 1 [5] :

Quadrant 1 (4777) annoyed (3607) frustrated (445) angry (329) distressed (187) enraged (75) jealous (29) envious (24) bitter (20) disgusted (19) impatient (16) hateful (6) tense (5) alarmed (4) defiant (4) suspicious (3) discontented (2) insulted (1) startled (1)

3. In [8], the reliabilities for the value types across samples were : universalism 0.73 ; benevolence 0.68 ; tradition 0.49 ; conformity 0.64 ; security 0.64 ; power 0.66 ; achievement 0.65 ; hedonism 0.71 ; stimulation 0.61 ; self-direction 0.58.

4. http://www.livejournal.com

5. The location of top 5 (underlined) moods is shown in figure 1 and numbers indicate how many posts in the corpus were tagged with this mood. The LiveJournal corpus was not cleaned and may include ill-formed English (or no English at all), advertising and other boiler-plate material.

Quadrant2 (7971) tired (2372) bored (1360) depressed (907) sad (821) anxious (654) lonely (392) disappointed (355) worried (303) apathetic (273) gloomy (244) uncomfortable (157) embarrassed (102) miserable (14) ashamed (7) desperate (6) dissatisfied (1) doubtful (1) droopy (1) hesitant (1)

Quadrant 3 (7964) content (1838) calm (1401) contemplative (1302) sleepy (1266) hopeful (552) relaxed (354) pleased (324) satisfied (286) pensive (236) peaceful (235) impressed (129) serious (13) glad (9) serene (9) astonished (3) at ease (2) reverent (2) attentive (1) friendly (1) longing (1)

Quadrant4 (7960) happy (3009), amused (2726), excited (1689), determined (480), interested (13), triumphant (13), adventurous (5), ambitious (5), delighted (5), expectant (5), enthusiastic (4), joyous (3), aroused (2), passionate (1)

As self-annotation of moods on LiveJournal was optional, we can have a certain degree of confidence that the mood indicated by the author truly reflects his/her true mood. We filtered out posts with less then 1000 characters, leaving 3139, 5522, 5999 and 5350 for quandrants 1, 2, 3 and 4 respectively. The average post length is 508 words. The corpus is available for download [6].

The second corpus is the *Blog Authorship Corpus (BAC)* [7] [10]. This corpus consists of posts from 19,320 bloggers gathered from `blogger.com` in August 2004. The corpus incorporates a total of 681,288 posts and over 140 million words - or approximately 35 posts and 7250 words per person. Each blog indicates the blogger's self-provided age, so that we have three groups : 8240 "10s" blogs (ages 13-17), 8086 "20s" blogs (ages 23-27) and 2994 "30s" blogs (ages 33-48). For each age group there are an equal number of male and female bloggers. Each blog in the corpus includes at least 200 occurrences of common English words, so each post is more likely to include garbage-free texts than the LiveJournal corpus. This corpus is a good candidate for our study because it is rather clean, balanced and provides a larger sample of posts for each individual blogger than the LiveJournal corpus (for which there is only one post per individual blogger).

6. `http://www.clul.ul.pt/bigfiles/LJcorpus.tar.gz`

7. `http://u.cs.biu.ac.il/~koppel/BlogCorpus.htm`

4 Experiments

In the following two experiments we are interested in two questions : (1) the relationships that may or may not exist between moral values and moods, and by extension, the well-being of people (2) the effect of age and gender on moral values interest and priority.

Experiment 1

To answer the first question, a corpus-based approach, reminiscent of techniques used in sentiment analysis [6], comes down to looking at the distribution of terms associated with a certain moral value (e.g. *benevolence*) in a mood-annotated blog corpus. We investigate this potential relationship for the LiveJournal corpus. In what follows we present the list of terms used to "model" each type of moral value [8] :

POWER : power (3.9), authority (4.1), wealth (9.8)

ACHIEVEMENT : achievement (5), success (3.4), ambition (7.9), influence (5.2)

UNIVERSALISM : universal (5.7), understanding (3.3) appreciation (4.6), tolerance (7.2) wisdom (6.3) equality (6.3)

BENEVOLENCE : kindness (8.2), helpful (5.8), honesty (6.6), forgiving (9.8), loyal (7.9), responsible (3.4)

STIMULATION : stimulation (8.2), excitement (6.4), novelty (8.7), challenge (4.4)

SELF-DIRECTION : autonomy (4.7), creativity (5.0), freedom (4.8), independence (5.0), curiosity (8.0)

TRADITION tradition (5.3), humility (9.1), devotion (6.2), moderation (6.7)

CONFORMITY : conformity (5.6), politeness (8.4), obedience (9.1)

SECURITY : security (3.7), safety (2.7), stability (4.0), clean (3.9)

HEDONISM : pleasure (4.7), enjoyment (5.8), gratification (9.8)

The number associated with each term is the inverse document frequency (idf) [9] value that we used as a weight for each term in the computation of the ten scores per document (one for each

8. From table 1 on p. 179 of [8].

9. Idf values are taken from the MEAD (`http://www.summarization.com/mead/`) summarization system default database.

VALUE	P/N	POS	NEG	St. error
achievement$\surd$	1.41	1.07	0.76	±0.09
stimulation$\surd$	1.29	1.98	1.53	±0.18
power	1.27	6.16	4.84	±0.34
hedonism	1.27	1.70	1.34	±0.16
universalism	1.25	2.07	1.65	±0.13
self-direction$\surd$	1.25	1.69	1.36	±0.12
tradition	1.00	0.54	0.54	±0.09
security$\surd$	0.96	5.19	5.40	±0.26
benevolence	0.89	1.86	2.08	±0.14
conformity$\surd$	0.38	0.06	0.16	±0.04
Average	1.10	2.23	1.97	-

TABLE 1 – Value-bearing content per 10,000 words for each half of the LiveJournal corpus. $\surd$ indicates agreement with figure 2. P/N is number of Positives divided by number of Negatives.

VALUE	10s	20s	30s
power	5.44 ±.20	9.62 ±.30	12.48 ±.64
security	5.00 ±.20	6.54 ±.18	9.29 ±.44
universalism	1.78 ±.09	3.32 ±.11	4.58 ±.36
self-direction	1.92 ±.11	3.60 ±.16	4.43 ±.29
stimulation	1.71 ±.11	3.47 ±.15	3.86 ±.26
benevolence	1.90 ±.10	2.79 ±.11	3.65 ±.23
hedonism	1.26 ±.10	2.56 ±.14	3.32 ±.30
achievement	1.22 ±.08	2.20 ±.11	2.61 ±.17
tradition	0.63 ±.06	1.19 ±.09	2.04 ±.50
conformity	0.20 ±.04	0.54 ±.12	0.40 ±.08
Total	21.06	35.82	46.65

TABLE 2 – Value-bearing content per 10,000 words for the BAC by age.

moral value), to mitigate the effect of chance occurrence. To give an example of the calculation of a score, a blog post with two occurrences of the word "obedience" and one occurrence of the word "security" and "safety" would get a score of 18.2 (2×9.1) for the moral value *conformity*, 6.4 (3.7+2.7) for the moral value *security* and zero for all other eight values. All scores are subsequently normalized to account for the fact that moral values are modelled with a different number of terms, and that documents have different length. This normalization allows for comparison across values but also across blog types (polarity, gender and age). We do not take into account phenomena such as irony, understatement, reported speech as well as lexical variations and negation [10]. Studies like [7] showed that it is indeed difficult to compute the scope of *not*, which is moreover not always negative in context : e.g. *not only*, *not just*, *not to mention*, etc.

Table 1 shows the value-bearing content scores for POSitive and NEGative blog posts from the LiveJournal corpus. Positive scores include all blog posts annotated with a mood from quadrants 3 and 4, while negative scores include posts from quadrants 1 and 2. For example, positive posts have an average score of 6.16 (per 10,000 words) for the value "power". Moral values with the highest ratio Pos/Neg are presented first in the table : a ratio > 1 indicates a correlation of the associated moral value with positive moods. The differences between the means of scores shown here are all significant at $p < .0001$. The results from table 1 are consistent with those of [8] highlighted in figure 2 above. Only *tradition* cannot be classified clearly as it has a ratio of one between positive and negative posts. This suggests that when blog posts are self-annotated for moods, they represent a reliable indicator of a person moral character, as moods and moral values (as found in blog posts) correlate in almost exactly the same way as when they are measured more directly in studies from psychology. We can also see that positive posts tend to make use more frequently of terms linked to morality (2.23 to 1.97), which might suggest that people in a good/positive mood (happy people ?) are more prone to discuss morality. Finally, in [8], the moral values *power*, *hedonism*, *universalism* and *benevolence* could not be conclusively placed in one of the poles : our results suggest that *power*, *hedonism* and *universalism* are favoured by happy (positive) people while *benevolence* is mostly a preoccupation for sad (negative) people.

10. For example, *I hate conformity* or *I do not like conformity* is also counted as one occurrence of *conformity*.

Experiment 2

We now turn to our second research question, which investigates the effect of age and gender on the "moral vocabulary" and moral value priority of bloggers. Previous studies provide some empirical support for the *Gender-as-Culture* hypothesis [2]. The BAC is well-suited for this task, as it is annotated for age and gender, and comprises a few blog post entries for each blogger, which allows for a more complete picture of the blogger personality. The value-bearing content scores for each blogger are computed as previously.

Table 2 shows the value-bearing content scores for each age category, and reveals that older

VALUE	female	male
power	5.92 ±0.20	10.64 ±0.20
security	6.20 ±0.15	6.42±0.17
self-direction	2.56 ±0.12	3.46± 0.12
universalism	2.46± 0.09	3.26 ±0.10
stimulation	2.63 ±0.17	2.93 ±0.17
benevolence	2.40 ±0.10	2.69 ±0.11
achievement	1.43 ±0.10	2.26 ±0.10
hedonism	2.02 ±0.18	2.22 ±0.20
tradition	0.90 ±0.14	1.26 ±0.15
conformity	0.26 ± 0.20	0.50 ±0.20
Total	22.77	35.63

TABLE 3 – Value-bearing content per 10,000 words for the BAC by gender.

people are increasingly preoccupied with moral questions than younger people (from a score of 21.06 to 46.65). For each moral category (with the except of *conformity*), there is a monotonic increase in the score with age. With respect to value priorities, it seems that people retain the same moral views over time. Nevertheless, they become less altruistic (*benevolence* drops two ranks) and embrace a more holistic view of life (*universalism* steps up two ranks).

Table 3 shows the value-bearing content scores by gender. Given the total scores (23 versus 36), the male discourse appears more preoccupied with morality than the female, and this is particularly obvious for *power* and *achievement*, but also *conformity*. Females prioritize *security* and males *power*, while *stimulation* has less priority for males them females. Considering commonsense psychology, these results are not surprising, say archetypal [11].

5 Conclusion

This paper has explored the relationship that may exist between a set of moral values and the well-being of people. We have hypothesized that the moral values talked about by bloggers in their posts is a reliable proxy of their moral character. We have found support for the claim that *achievement*, *stimulation* and *self-direction* are linked to happiness, as opposed to *security* and *conformity*. This is also supported in the psychology literature.

We have also looked into the effect of age and gender on the "moral vocabulary" and moral value priority of bloggers. Our findings are that older people are more preoccupied by morality than younger people, and that men do seem more talkative about these moral issues than women. Despite small differences, value priorities remain fairly stable among "positive" or "negative" people and across gender and age.

Given the relation between the "moral vocabulary" and age or gender, we think that these results can advantageously assist the type of automatic genre and age identification as put forwards in [8] and based on style and content. The relation between well-being and moral values may also be a good starting point to build an automated system to help draw a psychological profile of people according to their writing. Moreover, since the layout of figure 2 appears to mimic quite closely the political divide between left and right, it would be very interesting to see what the relation between left and right blog posts (instead of positive and negative) and moral values is.

11. As pointed out by an anonymous reviewer.

Références

[1] A. Bardi, R.M. Calogero, and B. Mullen. 2008. *A new archival approach to the study of values and value-behavior relations : Validation of the value lexicon*. Journal of Applied Psychology, 93 :483-497.

[2] A. Mulac, J.J. Bradac and P. Gibbons. 2001. *Empirical support for the gender-as-culture hypothesis : An inter-cultural analysis of male/female language differences*. Human Communication Research 27 :121-152.

[3] M. Généreux and R. Evans. 2006. *Towards a validated model for affective classification of texts*. In Sentiment and Subjectivity in Text, Workshop at the Annual Meeting of the Association of Computational Linguistics (ACL 2006), Sydney, Australia, July 22, 2006.

[4] S. Harris. 2010. *The Moral Landscape. How science can determine human values*. Free Press.

[5] G. Mishne. 2005. Experiments with mood classification in blog posts. In Proceedings of the 1st Workshop on Stylistic Analysis Of Text For Information Access (Style 2005), Brasil, 2005.

[6] B. Pang and L. Lee. 2008. *Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval*, 2 :1-135.

[7] L. Jia, C. Yu and W. Meng. 2009. *The effect of negation on sentiment analysis and retrieval effectiveness*. Proceeding of the 18th ACM conference on Information and knowledge management (CIKM), p. 1827-1830, 2009.

[8] L. Sagiv and S. H. Schwartz. 2000. *Value priorities and subjective well-being : direct relations and congruity effects*. European Journal of Social Psychology, 30 :177-198.

[9] K. R. Scherer. 1984. *Emotion as a multicomponent process : A model and some cross-cultural data*. P.

Shaver (Ed.) Review of Personality and Social Psychology, 5 :37-63.

[10] J. Schler, M. Koppel, S. Argamon, and J. Pennebaker. 2006. *Effects of age and gender on blogging*. In Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs.

[11] S.H. Schwartz. 1992. *Universals in the content and structure of values : Theorical advances and empirical tests in 20 countries*. Advances in experimental social psychology, 25 :1-65.

[12] S.H. Schwartz. 1994. *Are there universal aspects in the content and structure of values ?* Journal of Social Issues, 50 :19-46.

[13] S.H. Schwartz. 1995. *Identifying culture-specifics in the content and structure of values*. Journal of Cross-Cultural Psychology, 26 :92-116.

Semantic Propagation on Contextonyms using SentiWordNet

Ovidiu Şerban[⋆,⋆⋆] Alexandre Pauchet[⋆] Alexandrina Rogozan[⋆] Jean-Pierre Pécuchet[⋆]

[⋆]LITIS, INSA de Rouen,
76801 Saint-Étienne-du-Rouvray, France
email: {firstname.surname}@insa-rouen.fr

[⋆⋆]Computer Science Department
"Babeş-Bolyai" University
Cluj-Napoca, Romania

Abstract:
Sentiment analysis and affect detection algorithms are generally based onto annotated data, structured into dictionaries, ontologies or word nets. The focus, so far, has been concentrated on manual annotation of the data, and then, in some situations, a semantic valence propagation is applied. The problem with this approach is that while it is able to build new affective labels through the propagation process, the precision of the decision decreases. Our approach disambiguates through the data, by offering a strong context using a contextonym model, for the usage of a certain term with a valence.

Keywords: Valence Disambiguation, Semantic Valence Propagation, Contextonyms

1 Introduction

In the field of sentiment analysis and emotion detection based on text data, two main directions for research exist : one concentrating on building better annotations of linguistic resources, such as dictionaries or ontologies, and the other on building better classifiers for valence, sentiment or emotion detection [4]. Very often, building a classifier, relies on having good linguistic resources. Improving these dictionaries, by increasing their size and annotation accuracy, is therefore considered mandatory in this field.

Unfortunately, even if recent approaches have increased the size of the dictionaries, the ambiguity of the decision increased as well [19]. Our goal is to improve these dictionaries by preserving, as much as possible, the annotation accuracy. This objective is performed by taking into account the context of the word, and a new linguistic approach to model this relation, called the contextonym model.

WordNet (WN) or variations over it, remain one of the most used linguistic resources, so far [4]. WordNet [10] is a lexical resource, build at Princeton University, which is mainly used in most of the Natural Language Processing (NLP) applications. The concepts in WN are grouped into synonym sets (also called synsets), which are sets of words semantic linked. Each synset may contain its frequency in the dictionary, and a gloss, which is basically a short sentence describing the sense of the synset. Among the basic synonymic relations, the WN contains also some special relations called : hyperonymy, hyponymy or ISA ("is a"). All these links describe generalisation, specialisation or equivalence relationships between some concepts. All these special links are introduced in WordNet 3.0 for a part of the synsets.

As a synset database example, we mention WordNet Affect [26], an extension of the WordNet [17] data set. WordNet Affect is basically a 6 class emotional annotation (i.e. Ekman's basic annotation scheme [8] : Anger, Disgust, Fear, Happiness, Sadness and Surprise) made on a synset level. It contains nouns, adjectives, adverbs and some verbs for the English WordNet 2.0 version.

ConceptNet [14] is another well-known ontology used widely for semantic disambiguation in classification tasks. This database contains assertions of common-sense knowledge encompassing the spatial, physical, social, temporal, and psychological aspects of everyday life. ConceptNet was generated automatically from the Open Mind Common Sense Project [22].

Another database, used especially for opinion and valence classification, is SentiWordNet (SWN) [2]. Valence is represented by the degree of positivity, negativity or neutrality of a certain word or sentence, while opinion represents the general valence over multiple sentences. SWN is the result of a semantic propagation algorithm over all WordNet synsets according to their valence. This resource is presented in more details in the following sections.

Starting from WordNet Affect, Valitutti et al. [28] proposed a simple word presence method

in order to detect emotions, where the emotion of a sentence is given by the dominant word emotions. Ma et al. [15] designed an emotion extractor from chat logs, based on the same simple word presence. SemEval 2007 (task 14) [25] presented a corpus and some methods to evaluate it, most of them based on Latent Semantic Analyser (LSA) [7] and WordNet Affect presence. This corpus is the one used in our experiments, because it offers a consistent annotation with our approach.

In the following sections we present SentiWordNet (Section 2), with the current problems concerning this WordNet (Section 2.2). In the next section, we introduce the contextonym model, by presenting its advantages. In Section 4 our approach, based on a contextonym model to disambiguate conflicts in SentiWordNet, is presented, followed by a usage example (Section 5). Finally, we conclude this article by presenting the benefits of our model and highlighting some ideas for future work.

2 SentiWordNet

Among other WordNet extensions, like MultiWordNet [18], Balkanet [23], EuroWordNet [29] or WordNetAffect [26], SentiWordNet (SWN) [9] has been built as a lexical resource to help the valence prediction of a sentence. Its main field of application is opinion mining or sentiment classification. This resource contains annotation for mainly all the WordNet 3.0 synsets, having for each link a degree of positivity, negativity or objectivity annotated. Each of these valences are defined on a scale from 0.00 to 1.00, with the sum of all three of them being also 1.00.

Figure 1 presents the word **"good"** annotated according to SentiWordNet. For the selected synset, good has a definite positive value, of 1.0. The authors of SWN propose a triangle based visualisation, where each corner represents a different side of the valence : Positive (P), Negative (N) and Objective (O).

2.1 Semantic valence propagation

The method of semantic valence propagation refers to diffusion of a valence through a structured corpus. The spreading is done by respecting the structure links between the words. The structures tested so far are different versions of

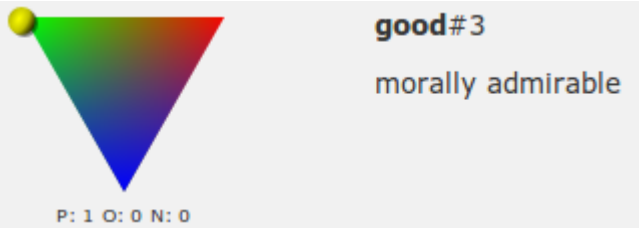


FIGURE 1 – A SentiWordNet [9] example, containing also the visualisation model. P states the positive degree, N the negative and O the objective one

WordNet, using the synsets. In a very frequent scenario, the process starts with a manually annotated set of words (also called "seeds"), and on every iteration the valence of these seeds is spread on the network. Examples of such approach are given by Rao et al. [20], Esuli et al. [9] and Godbole et al. [11]. Christopher Potts gives a more formalised definition of these algorithms in the Sentiment Analysis Tutorial [19], on the Semantic Valence Propagation section.

A more complex approach involves weighted propagation of valences [3], by not only spreading the valences in the neighbourhood of the seed nodes, but also computing a ranking measure attached to a node. This ranking measure is further used as a weight, taking into account the neighbourhood density among with node frequency.

Among others, SentiWordNet is the largest dictionary generated using valence propagation. However, SentiWordNet could not be manually reviewed for a better accuracy, mainly because of its size and the fact that WordNet structure does not disambiguate well between multiple usages. Given so, the valence conflicts remain unfortunately active in the database. More details about this topic will be given in the next section.

2.2 SentiWordNet and context

Christopher Potts [19] conducted an inconsistency level study between several opinion mining resources, and the results are presented in Table 1. This disagreement level between SentiWordNet and other corpora is due to the construction of SentiWordNet (based on automatic semantic propagation) and its size. Compared to the largest manually annotated dictionary, Harvard General Inquirer [24], SentiWordNet has almost 10 times more annotated words, which would give a better coverage over Word-

Net.

Dictionary	Dis. [a]	Annot. [b]	Cnt. [c d]
MPQA [31]	27%	+/-	8,221
Op. Lexicon [13]	25%	+/-	6,789
Gen. Inq. [24]	23%	+/-	11,788
LIWC [27]	25%	Categ. [e]	4,500

a. Disagreement level according to Potts [19]
b. Annotation style available in the corpus
c. Word countm in the corpus
d. SentiWordNet word count is 117,659
e. Words are grouped in several psychometric categories

TABLE 1 – Disagreement level between SentiWordNet and several other corpora

Table 1 presents an average of 25 % disagreement between SentiWordNet and MPQA [31], Opinion Lexicon [13], Harvard General Inquirer [24] or LIWC [27].

MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon is a resource maintained by Theresa Wilson, Janyce Wiebe, and Paul Hoffmann [31] and contains annotations based on the subjectivity level, part of speech and polarity. Polarity corresponds to a discrete valence annotation, having a label for positive or negative.

Opinion Lexicon is maintained by Bing Liu [13] and contains discrete manual annotations for positive and negative words.

Harvard General Inquirer [24] is a lexical resource which is concentrated in attaching syntactic, semantic and pragmatic information to part-of-speech tagged words. It contains positive, negative and hostile labels for most of its containing words.

Linguistic Inquiry and Word Counts (LIWC) [27] is a proprietary database, containing categorised words to their psycho-semantic state, which can be translated into negative or positive labels.

Another aspect of the problem is represented by the conflictual valences which describe the same word. There are two distinct situations :

1. the word has different valences among different synsets,
2. the word has conflictual valences in the same synset.

The first issue is partially solved. Considering that each synset corresponds to a certain usage (context) of that word, then the valence from SentiWordNet corresponds to that context. In practice, finding the proper context, only by using WordNet synsets, is quite challenging.

On the contrary, the second type of conflict is an artefact of semantic propagation algorithm and it is not resolved in SentiWordNet. A term, as part of the same synset, should not have opposing valences because it would lead to ambiguous decisions. In practice, this problem is similar to the first case, because a term with conflictual valences would have two different contexts, even if it is part of the same synset.

By making a short statistical analysis, it could be observed that 10,939 words (out of 117,659) from SentiWordNet carry conflictual valences, while 9,643 words are having conflictual valences in the same synset. These conflicts represent a 9.29 % from the total corpus. Part of these conflicts are highlighted by the disagreement levels, according to Potts [19].

In the following example, we choose the '*heart*' synsets (extracted from SentiWordNet) to present the two situations :

1. spirit#8 **heart#6** : *an inclination or tendency of a certain kind ; "he had a change of heart"*, **+0.5**
2. **heart#1** bosom#5 : *the locus of feelings and intuitions ; "in your heart you know it is true"*, **-0.125**
3. spunk#2 nerve#2 mettle#1 **heart#3** : *the courage to carry on ; "you haven't got the heart for baseball"*, **+0.25 -0.25**

Example 1 and 2 show an inter-synset ambiguity, because the valence of **heart** in example 1 is definitely positive, while in example 2 is negative. In the third example, the ambiguity is showed for the same synset.

Both of the conflictual cases state well that WordNet synsets are not the most adapted solution to describe context.

3 Contextonyms

Contextonyms were introduced by Ji et al. [12] in order to model a more flexible lexical structure, to be used in machine translation. Similar to synonyms, the contextonym model links words, but instead of having an equivalence relation between them, context is modelled by observing the word co-occurrences in a certain

window [1]. A graph-based structure is generated, having the words as nodes and the co-occurring frequencies as edges. In order to model a strong relation between the words, a clique exploration algorithm is usually applied. In the end, the cliques correspond to a strong context, which give a structure called "contextonym" [12].

In graph theory, a clique is represented by a complete sub-graph. In other words, it is a structure where every node is connected to all the other nodes part of this structure. Maximal cliques, represent the largest complete subgraph that could be generated by the selected set of nodes. In information retrieval, cliques represent a strong link between the words that are part of it, and this structure could be exploited as a context [16, 12].

In Figure 2 a full example of the contextonym neighbourhood is given, being centred around the word '*heart*'. The contextonym model has been trained using a subtitle corpus and annotated using the valences from SentiWordNet. The technical details of this approach are given in the following sections.

In order to reduce the noise, several filtering techniques are proposed by Ji et al. [12], which are also considered parameters of the resulting structure :
– a global filter, which eliminates all the nodes that occur very rarely in the corpus
– a local filter, which is applied to every node and remove the neighbours with a low occurrence
– a children filter, which is similar to the local filter but it is applied to the neighbours of every nodes

In our approach, we applied only a global filtering technique, by removing very low occurrences from the graph. The other two filters are integrated in the clique ranking measure, described in more details in the next sections.

3.1 Our solution : Semantic valence propagation and contextonyms

Once the contextonym model is built, the annotated labels could be spread. In our approach, the labels are represented by the valences extracted from SentiWordNet. We consider that each contextonym could not have conflictual valences (multiple values for the same word or opposite valences inside the same contextonym). In the case of a conflict, these are solved by choosing a single value for each conflictual word. In Figure 2, the labels are coloured according to their valence : blue for positive, red for negative, purple for mixed-value and light-grey for neutral.

4 Experiments

Our technique requires a multi-step process, each step assuring the output for the next phase. The first step, also called preprocessing, consists in the filtering and cleaning the text information. After this step a clique exploration is applied using the DDMCE algorithm [21]. This algorithm can be used for clique exploration on large and dynamic data, like semantic approaches and social networks. The third step consists in aligning a certain phrase (consisting in a bag of words) to a set a cliques. In order to keep the alignment process consistent, a ranking measure is proposed.

Building the linguistic model for the contextonyms extraction is the most difficult and important step. In order to keep a link to an actual spoken language, we decided to use a subtitle corpus, collected from multiple sources [2]. Finally, a total of 53,384 unique movie files were kept. Also, in order to keep our linguistic model clean, we have kept the best subtitle file for each movie title, using an author and download count filtering.

Preprocessing Step.
A preliminary step, specific to sentence tokenizing on subtitle files was applied, filtering all the time synchronisation data. Even if the SubRip [3] format is clean and simple, a template validation had to be done, in order to verify the integrity of the data extracted. Recently, most of the subtitles tend to have advertisements included at the beginning or at the end of the file, a brute filtering approach was applied in order to remove them.

During the preprocessing step, we applied on each headline a collection of filters, in order to remove any useless information, such as special characters and punctuation, camel-case

1. The size of the window has been fixed to 5, after applying the filtering process

2. The corpus represents a part of the subtitle database from `http://www.opensubtitles.org/` and `http://www.podnapisi.net/`

3. SubRip (.srt) is a very basic text format, used to encode subtitle files

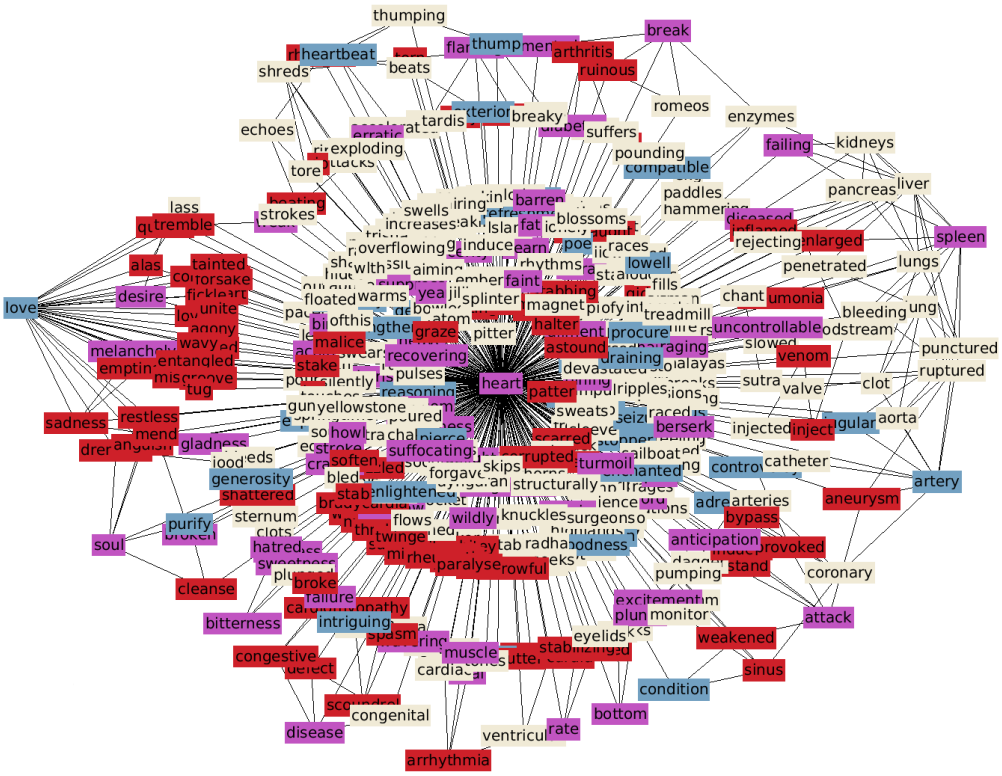


FIGURE 2 – A fully annotated contextonym graph, representing the whole neighbourhood of the word '*heart*'

separators and stop words. We considered as stop words, all prepositions, articles and other short words that do not carry any semantic value. The stop word collection used in our experiment is available at : `http://www.textfixer.com/resources/common-english-words.txt`. From the space reduction perspective, we kept only the words that are considered to carry a strong semantic and emotional value (e.g. nouns, verbs, adverbs and adjectives), as WordNet Affect is suggesting [26]. This technique could be considered as a key word extractor, because of its ability to limit the search space, while it represents quite well the meaning of the original sentence.

The method of stop word filtering offers a good balance between speed and accuracy of the results, compared to other methods like Part of Speech Tagging (POS), which provides comparable results, but tends to be much slower.

Co-occurrence Extraction.
Our context is modelled by observing the word co-occurrences in a certain window. These occurrences are extracted for each central word, by fixing a window size of 5, which means that the co-occurrences were counted for the two words before and two after the central word. In practice, in order to prevent duplicate co-occurrences, a sliding window of size 3 (starting from the central word) could be chosen. This is mainly because, for a given sequence of words, $w_{i-2}, w_{i-1}, w_i, w_{i+1}, w_{i+2}$, the occurrence of $[w_{i-2}, w_i]$ and $[w_i, w_{i-2}]$ should be counted only once and a sliding window over a word sequence would prevent backward duplicates.

Clique Extraction.
After the preprocessing step, 86,276 words (frequency > 0.01 %) and 3,948,359 co-occurrences are maintained after filtering (frequency > 0.01 %). This was done to preserve most of the words collected, by limiting also the size of the problem. On this data, the DDMCE [21] algorithm was applied, obtaining a total of 702,546 cliques, from which 354,109 (50 %) are considered conflictual.

5 Usage example

After building the contextonym model, trained on the subtitle corpus and annotated with the valence values from SentiWordNet, the valence disambiguation can be done on any other

example. The chosen corpus for our disambiguation experiment, is the one from SemEval 2007, task 14 [25], proposed at the conference with the same name. We believe both of the corpora used in our experiments : subtitles corpus and SemEval, are linked with oral data. The data set contains headlines (newspaper titles) from major websites, such as New York Times, CNN, BBC or the search engine Google News. The corpus was manually annotated by 6 different persons. They were instructed to annotate the headlines with emotions according to the presence of affective words or group of words with emotional content.

A valence annotation was carried out. Valence, as used also in psychology, means the intrinsic attractiveness (positive valence) or aversiveness (negative valence) of an event, object, or situation. In SemEval task, the valence is used to describe the intensity of the positive or negative emotion. The valence label ranged from -100 to 100.

From this corpus, we choose the item 24 : "*Hurricane Paul Weakens To Tropical Storm*", as an example for our approach. The complete approach can be observed in Figure 3. After the first parsing step, it could be observed that the word "tropical" is ambiguous, because the SentiWordNet synset containing this word has both negative and mixed valences. Given so, we selected the contextonym for tropical, exemplified in the Figure 4, and we try a best alignment with the existing context.

Given the current set of words, after the filtering, for each ambiguous word, we will align as many words as possible to one of the contextonyms of the ambiguous word. If more contextonyms could be aligned, a ranking is performed, and the best is chosen. The ranking measure is described in the next section.

In our example, the alignment is made between tropical storm and cyclone (hurricane). This is a full alignment, since the word hurricane has the word cyclone as a direct hypernym (a generalisation). Doing this alignment, since both of the words from the aligned contextonyms are negative, the disambiguation can be done, and the value decided for "tropical" would be negative. This disambiguation concerns only the word "tropical".

As for the general valence or the chosen sentence, after the disambiguation of all the words (in our case, just the word "tropical"), any opinion extraction algorithm could be applied [6, 19]. In this particular case, the item has been annotated as a "soft" negative (-1%), mainly because of the influence of the word "weakens" over the phrase, which has the role of decreasing the strong negative valence of the ["hurricane", "tropical", "storm"] contextonym. The purpose of our our work is to aid the word disambiguation, which would increase the accuracy of future classifiers based on our approach.

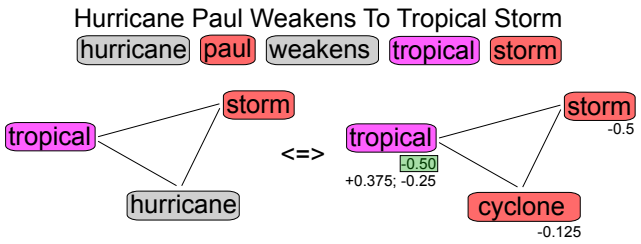


FIGURE 3 – Clique alignment for a SemEval 2007, task 14, item 24. The contextonym on the left is extracted from the given phrase, while the aligned one, on the left, is part of the subtitle contextonym corpus.

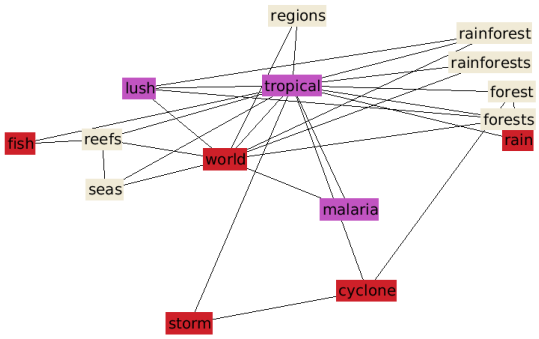


FIGURE 4 – The contextonym of the word tropical, as trained from the subtitle corpus, and annotated with valences from SentiWordNet

5.1 Ranking measure

In the case of multiple contextonym alignment, a ranking should be done in order to choose the best one.

For a given sentence *ph*, corresponding to a set of words, equation 1 describes a ranking measure, used to discriminate partial or full alignments.

$$\forall q \in Q, R(q, ph) = \frac{f(q \cap ph) - f(q \setminus ph)}{f(ph)} \quad (1)$$

Where $f(X)$ represents the combined frequency of the set X. Q is the set of all possible cliques, and q is a clique from this set.

This measure is built to be used for partial alignments, but penalising the ones that are much larger than the actual sentence.

In the context of the previous example (Figure 3), the value of $R(\bullet, \bullet) = 0.3591$

6 Conclusion

In our approach, the context is a key part of the solution of the disambiguation problem. We model the context by using graph-based structures and we extract the strong-context by modelling it into contextonyms. By using these approaches, the disambiguation process is easy and more natural than in any previous work.

For the perspectives, a full validation has to be conducted. So far, only a small manual validation has been done. One of the major obstacles in the way of our validation process is the lack of free annotated resource that could be used.

A second perspective would be the integration of multiple modalities in our approach, like gestures, vocal features and other semantic approaches.

Acknowledgements

The work of this paper would not be possible without the help of the `www.opensubtitles.org` and `www.podnapisi.net` administrators. We thank them for their kindness of sharing a part of the database with us.

Références

[1] C.O. Alm, D. Roth, and R. Sproat. Emotions from text : machine learning for text-based emotion prediction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 579–586. Association for Computational Linguistics, 2005.

[2] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0 : An enhanced lexical resource for sentiment analysis and opinion mining. In *Seventh conference on International Language Resources and Evaluation, Malta. Retrieved May*, volume 25, page 2010, 2010.

[3] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G.A. Reis, and J. Reynar. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*, 2008.

[4] R.A. Calvo and S. D'Mello. Affect detection : An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, pages 18–37, 2010.

[5] T. Danisman and A. Alpkocak. Feeler : Emotion classification of text using vector space model. In *AISB 2008 Convention Communication, Interaction and Social Intelligence*, volume 1, page 53, 2008.

[6] S.K. D'Mello, S.D. Craig, J. Sullins, and A.C. Graesser. Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, 16(1) :3–28, 2006.

[7] S.T. Dumais. Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1) :188–230, 2004.

[8] P. Ekman. Basic emotions. 1999.

[9] A. Esuli and F. Sebastiani. Sentiwordnet : A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.

[10] C. Fellbaum et al. Wordnet and wordnets. *Encyclopedia of Language and Linguistics*, pages 665–670, 2005.

[11] N. Godbole, M. Srinivasaiah, and S. Skiena. Large-scale sentiment analysis for news and blogs. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, volume 2, 2007.

[12] Hyungsuk Ji, Sabine Ploux, and Eric Wehrli. Lexical knowledge representation with contextonyms. In *Proceedings of MT Summit IX, New Orleans, USA*. Association for Machine Translation in the Americas, 2003.

[13] B. Liu. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing,*, pages 627–666, 2010.

[14] H. Liu and P. Singh. ConceptNet-a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4) :211–226, 2004.

[15] C. Ma, H. Prendinger, and M. Ishizuka. A chat system based on emotion estimation from text and embodied conversational messengers. *Entertainment Computing-ICEC 2005*, pages 535–538, 2005.

[16] R. Mihalcea and D. Radev. *Graph-based natural language processing and information retrieval*. Cambridge University Press, 2011.

[17] G.A. Miller. WordNet : a lexical database for English. *Communications of the ACM*, 38(11) :39–41, 1995.

[18] E. Pianta, L. Bentivogli, and C. Girardi. Developing an aligned multilingual database. In *Proc. 1st Int. Conference on Global WordNet*, 2002.

[19] Christopher Potts. Sentiment analysis tutorial, November 2011.

[20] D. Rao and D. Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682. Association for Computational Linguistics, 2009.

[21] O. Serban, A. Pauchet, A. Rogozan, and J.-P. Pecuchet. DDMCE : recherche de cliques maximales dans des graphes dynamiques de grande taille. In *Proceedings of the Journée thématique : Fouille de grands graphes*. Réseaux : Approches Mathématiques et Informatique, 2012.

[22] P. Singh, T. Lin, E. Mueller, G. Lim, T. Perkins, and W. Li Zhu. Open mind common sense : Knowledge acquisition from the general public. *On the Move to Meaningful Internet Systems 2002 : CoopIS, DOA, and ODBASE*, pages 1223–1237, 2002.

[23] S. Stamou, K. Oflazer, K. Pala, D. Christoudoulakis, D. Cristea, D. Tufis, S. Koeva, G. Totkov, D. Dutoit, and M. Grigoriadou. Balkanet a multilingual semantic network for the balkan languages. In *Proceedings of the International Wordnet Conference, Mysore, India*, pages 21–25, 2002.

[24] P.J. Stone, D.C. Dunphy, and M.S. Smith. The general inquirer : A computer approach to content analysis. 1966.

[25] C. Strapparava and R. Mihalcea. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560. ACM, 2008.

[26] C. Strapparava and A. Valitutti. WordNet-Affect : an affective extension of WordNet. In *Proceedings of LREC*, volume 4, pages 1083–1086. Citeseer, 2004.

[27] Y.R. Tausczik and J.W. Pennebaker. The psychological meaning of words : Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1) :24–54, 2010.

[28] A. Valitutti, C. Strapparava, and O. Stock. Lexical resources and semantic similarity for affective evaluative expressions generation. *Affective Computing and Intelligent Interaction*, pages 474–481, 2005.

[29] P. Vossen. *EuroWordNet : a multilingual database with lexical semantic networks*. Kluwer Academic, 1998.

[30] H.G. Wallbott, K.R. Scherer, et al. Emotion and economic development-Data and speculations concerning the relationship between economic factors and emotional experience. *European journal of social psychology*, 18(3) :267–273, 1988.

[31] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2) :165–210, 2005.

A computational model of social attitude effects on the nonverbal behavior for a relational agent

Brian Ravenet
ravenet@telecom-paristech.fr

Magalie Ochs
ochs@telecom-paristech.fr

Catherine Pelachaud
pelachaud@telecom-paristech.fr

Laboratoire Traitement et Communication de l'Information
Télécom Paristech - 37/39 rue Dareau
75014 Paris – FRANCE

Abstract:
The relations we have with others influence in a very subtle way our body gestures. These are cues that are used in an interaction in an unconscious process to communicate and understand an attitude. Depending on the gestures and facial signals one is displaying, a social attitude can be perceived. In this paper, we propose to model socio-emotional agents and how they can express such an attitude in a dyadic interaction. The focus will be on generating the nonverbal behavior of the virtual agent given the social attitude it wants to convey. Depending on the nature of the relation an agent has with someone else, its role and desires, it should display a different attitude and therefore it should display different nonverbal behaviors. In this paper, we propose a computational model based on the findings in Human and Social Sciences on the correspondence between social attitudes and nonverbal behaviors. This computational model is used to select the proper behavior of an agent during an interaction depending on the social attitudes it wants to convey and depending on its gender.

Keywords: virtual agent, relational agent, social attitude, nonverbal behavior

1 Introduction

Embodied conversational agents are used to endow different roles during interactions with users, for instance teachers [32], assistants, guides, coach [7], ally or enemy in video games [22]. In this different roles, agents must be able to express different social attitudes to enhance their believability.
The believability of an agent can be defined as the ability to provide an illusion of life [1]. Different kind of work has been done to increase the believability of agents. It is a difficult task as by doing so we may encounter the uncanny valley phenomenon [28], where we loose all believability of the virtual agent. The first idea is to make a more human-like appearance but we can also create a more human-like behavior. Some looked into how emotions could be conveyed by a virtual agent in order to increase its believability [1]. Starting from this idea, we aim at going a step further by designing agents capable of creating a social relation with users. Relational agents have been demonstrated many times to increase the engagement of the users in the interaction [11]. The model we propose makes the agent able to express different *social attitudes*. For instance, If a virtual agent wants one of its subordinate to do a work which he is not doing, the agent may express a dominant attitude to convey the message. The particular point we focus on, is the expression of these social attitudes through nonverbal behaviors. This paper is organized as follows. In the first section, we present the theories from the literature of psychology that illustrate the influence of social attitude on nonverbal behaviors. Then, in Section 3 we introduce previous research that has been done on relational agents. In Section 4, we present our computational model. Finally, we discuss in Section 5 the limits of our model and the perspectives of our research.

2 Theoretical background

The proposed model is based on studies in Human and Social Sciences that have explored the influence of nonverbal behaviors on social attitude's perception. Social attitude or interpersonal stance is an affective style that can be naturally or strategically employed in an interaction with a person or a group of persons. It consists of conveying a particular feeling in the interaction. For instance being friendly, dominant, hostile or polite [33]. Several studies describe how social relations are perceived through the nonverbal cues. Some studies [13, 20] represent social attitude with two dimensions: a *dominance* dimension (also called *power*, *control* or *agency*) that represents the degree of control one has on another, and a *liking* dimension also called *appreciation*, *affiliation* or *communion*), that represents the degree of appreciation, liking of another. In other studies [9] these dimensions are used among other dimensions like *formality* or *trust*. Dominance and liking correspond to the dimensions of the *Interpersonal Circumplex*, illustrated Figure 1. In [18], Gurtman presents this tool widely used in the social psychology field to describe interpersonal relations.

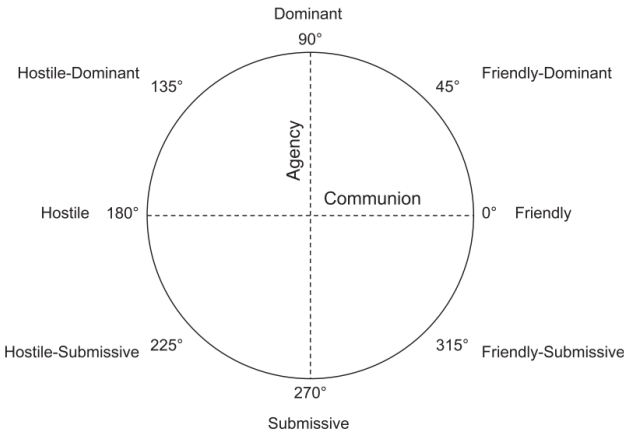


Figure 1: A generic interpersonal circumplex [18]

This tool describes interpersonal relations with the two axes *dominance* and *liking*.
Several studies have explored the expressions of dominance and liking through nonverbal behavior [8, 9, 10, 13, 20, 21, 24, 34]. Moreover some works have highlighted the influence of the gender of the interactants on one's nonverbal behavior [8, 21]. From this literature, we built a table of influence of the social attitudes and of the gender on the nonverbal behavior. We describe this table in Section 2. This table is used in the proposed computational model described in Section 4 to modulate the nonverbal behavior of an agent. Before presenting in details the model we introduce in the next section related works on relational agents.

3 Related Work

One of the challenges of the research in virtual agent is to increase their believability to enhance the human-machine interaction. For this purpose, the capacity to express nonverbal behaviors through different modalities is a key element [15]. In the following section, we present multimodal virtual agents. Few research seem to be done on agents able to express social attitudes. In section 3.2 we present some existing relational agents.

3.1 Multimodal Agents

Max is an example of a multimodal virtual agent able to express emotions through its facial expressions, its gestures and its verbal behaviors[2]. *Rea* [14] is a virtual agent equipped with conversational functions. *Rea* can change her gaze or body posture, interpret those from the user and manage a conversation. Moreover, *Rea* is paying attention to the verbal cues but also to the non-verbal cues. It replies with speech and gestures in order to provide the same kind of nonverbal behaviors to the user. *Greta* [29] is a virtual agent based on SAIBA [25] for the general architecture and the MPEG4 standards [31] for generating the animation parameters. It has previously been integrated into the SEMAINE platform where it was able to show signs of understanding and reply to an user during a conversation [4]. Another framework is *MARC*, also using MPEG-4 standard and BML for the animation of its virtual agents. In [16], *MARC* has been used for displaying appropriate emotions during a game of Reversi against a user.
The presented multimodal agents have a nonverbal behavior that is computed based on communicative intentions and emotions. However, these agents do not consider the influence of the social attitude in their behaviors. Regarding influence on nonverbal behavior, Mancini and Pelachaud have proposed a framework that enables one to describe tendencies for a virtual agent in nonverbal behaviors expression [27]. These tendencies have an influence on the generated nonverbal behaviors and more precisely, they change their expressivity parameters. These parameters are the following[19]:

– *overall activation*: the quantity of movements;

– *spatial extent*: the amplitude of movements;

– *temporal*: the duration of the movements;

– *fluidity*: the continuity of movements;

– *power*: the dynamic of the movements;

– *repetition*: tendency to repeat specific movements;

The influence on the expressivity parameters is described in what they call a *Baseline*, composed of initial values for each parameter, and also in *Dynamic Qualifiers* that specify how these parameters are dynamically affected depending on the emotional state of the agent. One limit of this model is the lack of definition for the expressivity parameters according to the social context. In our model, we propose to go one step further by defining the influence of so-

cial attitudes of certain expressivity parameters. Indeed some works [10, 13] have shown that the social attitude impacts the spatial extend of one's gestures and one's quantity of movement.

3.2 Relational Agents

The influence of social relation means that an agent may not act the same with different persons based on the relation it has with each person. This can be through its actions but it can also be more subtle and be displayed in its nonverbal behavior. This can be very effective to increase the believability of the agent [22, 7]. Usually, an agent is talking and acting the same way with every user. Some research attempts to propose agents that convey different social attitudes during interactions.
Memory can be used to simulate a social bond, for instance in the conversational agent *May* [12]. *May* is reminding the user with previous topic they had. The study shows that the user felt a better connection with the agent because of the shared common ground. Tinker is a virtual museum guide that creates relationships with the visitors [6] by remembering them and recalling past interactions. It can also show empathy and use an appropriate nonverbal behavior to convey it. *Laura* is one of the first relational agents [7]. It displays different nonverbal behavior depending on the state of the relation it has with the user. It expresses different attitudes as the relation is evolving. Actually, Laura is only following the duration of the relation. The more a user interacts with it, the more it will express signs of closeness, no matter what kind of attitude the user has with it. *Rea* was also used in an experiment to make the user trust the agent [5]. It was using a planned dialogue, where small talk was used to gain user's trust and then it could talk about more serious topic after. *Eva* [23] uses the relationships it has with a user, built from the previous interactions, to generate different emotional response, verbal and nonverbal. Its interpersonal relations are also affected by the type of emotion triggered by the interlocutor. In [30], a computational model of the impact of emotions on social relations is also proposed. The virtual agent *Alfred* is able to convey different degrees of dominance by varying its gaze, facial posture and linguistic behaviors [3].
Our model aims at adding to an agent the ability to express different social attitudes. We propose to modulate the selected behavior of the agent depending on its social attitude. Indeed, in our model, the social attitude will inhibit or emphasize some of the agent nonverbal behaviors based on the results from the Human and Social Sciences.

4 Model

4.1 Theoretical Model

Based on the literature in Human and Social Sciences [8, 9, 10, 13, 20, 21, 24, 34] and in virtual agent[3, 7, 22], we have designed a table showing the influence of dominance and liking on the nonverbal behavior depending on the gender of the speaker (Table 1).

This table indicates the influence (inhibition characterized by a $\searrow$ and accentuation characterized by a $\nearrow$) of the dominance and liking on certain nonverbal behaviors. An empty cell means that the dimension considered does not have a significant influence. For instance, in Table 1, 4th row and 3rd column, this cell indicates that a dominant male person tends to have broad gestures. This table is used in the computational model to determine the influence of the social attitude on the nonverbal behavior. We present in more details this model in the next section.

4.2 Computational Model of Social Attitude Influence

In this section, we present a computational model that allows an agent to display, through its nonverbal behavior, the appropriate social attitude it wants to convey to its interlocutor. Our model modulates the probability to perform a specific nonverbal expression and also how it is expressed.
To modulate the expressivity of nonverbal behaviors, we use a similar approach as the work of Mancini and Pelachaud on dynamic behaviors [27]. In our model, we are using the table of influence (Table 1), to change the expressivity parameters of any agent given the social attitude it wants to convey. The functions that modulate the expressivity parameters are described more precisely in Section 4.4.

The social attitude. The virtual agent is characterized both by a *social relation* it has with its interlocutor and a *social attitude* it expresses during the interaction. The social attitude can convey the expressions of a social relation. However, one may decide to express an attitude different from the relation one has depending on

0	1	2	3	4	5	6	
1		Nonverbal Behaviors	Dominance		Liking		References
2			♂	♀	♂	♀	
3	Hand movement	Initiates Hand Shaking	↗	↗			[13]
4	Gesture parameters	Has Broad gestures	↗	↗			[13]
5		High number of gestures	↗	↗	↗		[10, 13]
6	Touch Gestures	Touches other	↗	↗	↗		[9, 13, 22, 34]
7		Self-touches (hands)	↘	↘			[8, 13]
8		Self-touches (head)	↘	↘			[8, 13]
9		Object-Adaptors	↘	↘	↘	↘	[10]
10	Head Movement	Tilts head up	↗	↗	↗	↗	[10, 13]
11		Orients head toward other	↗	↗	↗	↗	[10, 13]
12		Shakes head	↗	↗			[13]
13	Posture	Has Erect posture	↗	↗			[13]
14		Leans forward toward other	↗	↗	↗	↗	[9, 10, 13, 22, 34]
15		Open body position	↗	↗	↗	↗	[9, 13, 22, 34]
16		Orients body toward other	↗	↗	↗	↗	[9, 10, 13, 22, 34]
17	Gaze	Pays attention to other	↘	↘	↗	↗	[9, 13, 22, 34]
18		Glares	↗	↗			[9, 13, 22, 34]
19		Engages in mutual gaze	↗	↗	↗	↗	[9, 13, 22, 34]
20		Gazes for a long time	↗	↗	↘	↘	[9, 13, 22, 34]
21		Averts gaze	↘	↘	↘	↘	[13, 10, 8]
22		Looks while speaking	↗	↗			[13]
23	Face	Expressing face	↗	↗	↗	↗	[13, 10]
24		Self-assured expression	↗	↗		↘	[13, 21]
25		Shows facial fear	↘			↗	[13, 21]
26		Shows facial sadness	↘	↘			[21]
27		Shows facial disgust	↗	↗		↘	[13, 21]
28		Shows facial anger	↗	↘			[13, 21]
29		Smiles				↗	[8, 9, 21, 22, 34]

Table 1: Table of influence of social attitude on nonverbal behaviors

one's goals or one's role. In our model, we focus more particularly on the social attitude. As a first step, we consider that the social attitude the agent wants to express is the exact representation of the social relation it has with the interlocutor. We suppose that the social attitude the agent wants to convey is an input of our model. We propose a model to generate the nonverbal behavior associated to this social attitude. As shown in Section 2, the social attitude influences one's nonverbal behavior. The social attitude may inhibit or emphasize some particular signals [13]. To represent the social attitude, we are using the interpersonal circumplex. Consequently, we consider the two following dimensions: *dominance* and *liking*. We represent formally the social attitude as follows:

Let $A_i(t)$ be the social attitude the agent wants to convey to an interlocutor i at a time t. We define $A_i(t)$ as the pair dominance $D_{A_i}(t)$ and liking $L_{A_i}(t)$ where both $D_{A_i}(t)$ and $L_{A_i}(t)$ take values in the interval $[-1, 1]$.

$$A_i(t) = (D_{A_i}(t), L_{A_i}(t))$$

The more the agent expresses dominance towards another agent i, the closer to 1 is the value of $D_{A_i}(t)$. The more an agent expresses liking towards another person i, the closer to 1 is the value of $L_{A_i}(t)$. When dominance and liking are equal to 0, it means that the agent expresses a neutral attitude. If the value of dominance (resp. liking) is below 0, the agent has a submissive

(resp. hostile) social attitude. In other words, the attitude is a point in *dominance* $\times$ *liking* space [1].

Architecture. We have integrated our model in the context of a SAIBA-like agent [25]. The SAIBA architecture defines three main components. The *Intent Planner* generates the communicative intentions (what the agent intents to communicate). For instance, a communicative intention can be the expression of joy. The *Behavior Planner* transforms these communicative intentions into a set of signals. A signal is a behavior expressed through a modality (e.g. speech, gestures, facial expressions). Multimodal expression (expressing and synchronizing signals on different modalities) is handled by this *Behavior Planner*. For the purpose of our research, we are using an extended version of a SAIBA architecture where signals are also endowed with the expressivity parameters described in section 3.1. Finally the *Behavior Realizer* outputs for each of these signals the animation parameters.
In our architecture, we add three modules to the SAIBA-like agent: the *Social Intention Filter*, the *Adaptor Generator* and the *Social Expressivity Modulator*. The resulting architecture is illustrated in Figure 2. Each component takes as input the social attitude of the agent, its gender and the table of influence. The *Social Intention Filter* take as inputs the potential signals (signals that might be selected to express the communicative intention) and outputs, for each signal that has a corresponding row in the table 1, a new probability to be expressed. The *Adaptor Generator* inserts new signals in the list of signals to be expressed. The *Social Expressivity Modulator* takes as input the list of signals to be expressed and assigns new expressivity parameters to each signals. These modules and their integration in the SAIBA architecture are described in details in the following.

The *Intent Planner* sends the communicative intentions to the *Social Intention Filter* instead of the *Behavior Planner* directly. For each communicative intention, a list of potential signals are available in a behavior set B. To each signal is associated a probability p_{signal} to be performed in order for the behavior planner to decide which signal will be executed. Indeed a communicative intention may be expressed through different signals. For instance, the agent wants to greet. A first possibility is to wave with the arm and the hand and an alternative one could be just a head nod. The *Social Intention Filter* modifies the probability to express a signal depending on the social attitude of the agent. The *Behavior Planner* uses the probability p_{signal} to choose the appropriate signal to generate. In the next section, we describe in details the *Social Intention Filter* and the *Adaptor Generator*.

4.3 The Social Intention Filter and the Adaptor Generator

In our model, we propose to modify the behavior set (both the list of signals and the probability associated) depending on the social attitude and the gender of the agent. To compute the probability of the signals, we define a function g.
For each signal s of B, we search in the table 1 for a corresponding row. If one is found, we represent it in a dominance $\times$ liking space, depending on the gender of the agent. Indeed, we consider $\nearrow$ and $\searrow$ from Table 1 respectively as 1 and -1. An empty cell is considered as 0. This representation in the dominance $\times$ liking space is noted P_s.

$$P_s = (D_{P_s}, L_{P_s})$$

The function g uses an Euclidean distance between P_s and $A_i(t)$ (the current social attitude of the agent). We have choosen the Euclidean distance since our objective is to compute the impact of the agent's social attitude (described in a two-dimension space) based on the difference between the current agent's social attitude and the extracted representation from Table 1. Note that other functions could be used like the Manhattan distance. Let d be the mathematical function calculating the Euclidean distance.

$$d(P_s, A_i(t)) = \sqrt{(D_{P_s} - D_{A_i}(t))^2 + (L_{P_s} - L_{A_i}(t))^2}$$

The higher the distance is, the lower the influence is for this signal. The influence is proportional to $1 - d$. The new probability of the signal corresponds to the mean between the resulting value of $1 - d$ and the original probability p_{signal}.

$$g(P_s, A_i(t), p_{signal}) = \frac{1}{2}(1 - \frac{d(P_s, A_i(t))}{2\sqrt{2}} + p_{signal})$$

1. For now, we are not defining yet how social attitude are evolving through time

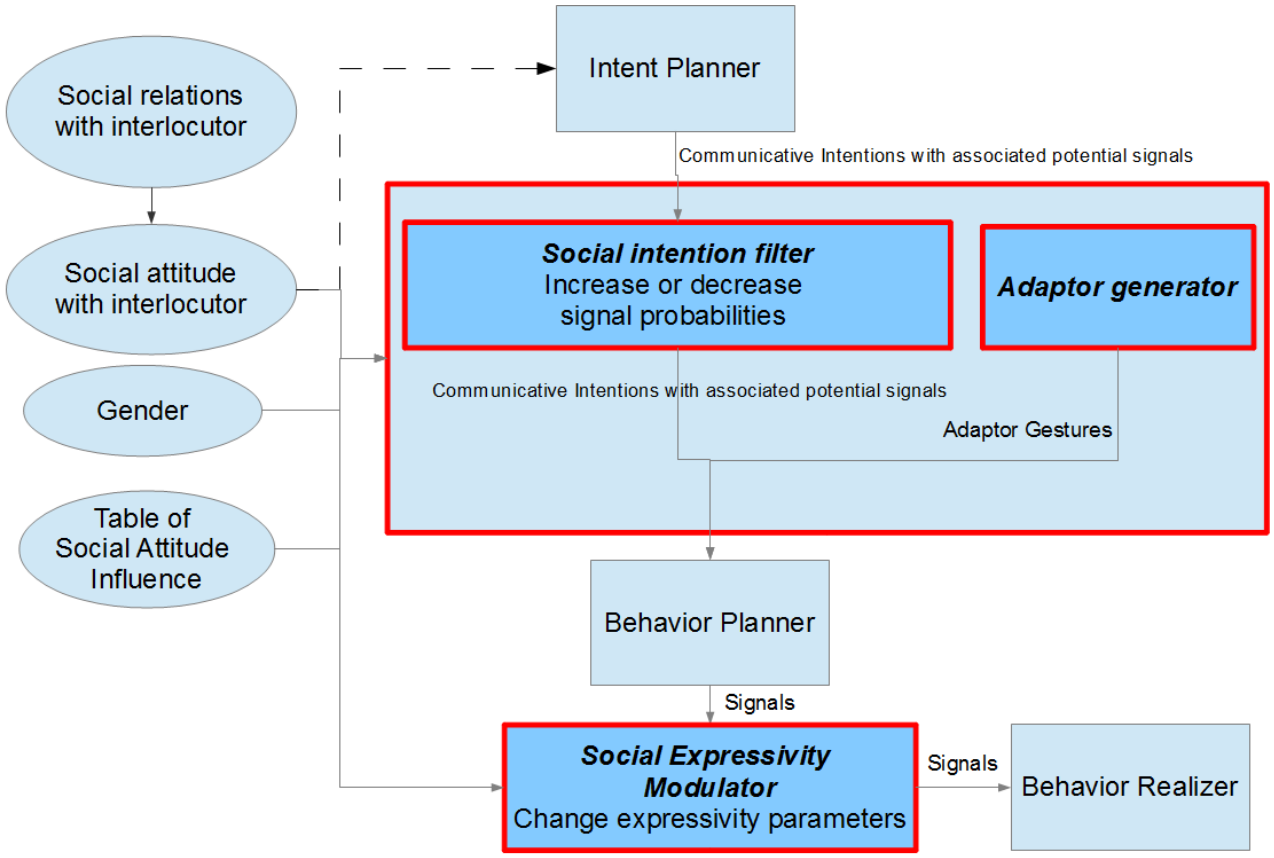


Figure 2: Social Attitude Influence Model

This function returns a value in the interval $[0, 1]$.
The probability of a signal is then computed based on the original probability of the signal, its current social attitude $A_i(t)$, the gender G of the agent and the proposed table of social attitude's influence (Table 1). Let's introduce an example to illustrate the proposed function g. Let's consider a male agent with the communicative intention to express sadness. The social signal has already a probability p_{signal}. The table gives us the point $P_{FSadness}$ of coordinates $(-1, 0)$ (26$^{\text{th}}$ row, 3$^{\text{rd}}$ and 5$^{\text{th}}$ columns of Table 1). The more the agent wants to communicate a dominant attitude, the higher the distance between $P_{FSadness}$ and its attitude $A_i(t)$ is, and the lower the computed output probability is.

The social attitude may lead someone to express specific gestures, for instance, self-touches, or object-touches. these gestures are called adaptor gestures [17]. In our model, we give the capability to an agent to generate new adaptor gestures depending on its social attitude. These adaptor gestures are selected based on Table 1. They correspond to the nonverbal behaviors *self-touches (hands)*, *self-touches (face and head)* and *object-touches*. In the proposed architecture Figure 2, the *adaptor generator* generates the new adaptor gestures in the list of signals to be expressed.
Finally, the *Behavior Planner* module generates a set of gestures to perform depending on the communicative intentions and the probabilities of the associated signals. Then, this set is transferred to the *Social Expressivity Modulator*, instead of the *Behavior Realizer*.

4.4 The Social Expressivity Modulator

The *Social Expressivity Modulator* changes the expressivity parameters of the gestures according to Table 1 to reflect the agent's social attitude. Given Table 1, we consider two expressivity parameters [2]: the *overall activation* parameter and the *spatial* parameter (4$^{\text{th}}$ and 5$^{\text{th}}$ rows of

2. However, other parameters may be impacted as well by the social attitude. A deeper study in the future will enable us to extend this model

Table 1). The expressive parameters take their values in the interval $[0, 1]$.
Let SPC be the mathematical function that computes the *spatial* expressivity parameter from Table 1 and the current attitude $A_i(t)$ of the agent. This function will return a value in the interval $[0, 1]$.

$$SPC(D_{A_i}(t), L_{A_i}(t)) = 1 - \frac{d(P_{BroadGesture}, A_i(t))}{2\sqrt{2}}$$

Let OVA be the mathematical function that computes the *overall activation* expressivity parameter from Table 1 and the current attitude A_i of the agent. This function will return a value in the interval $[0, 1]$.

$$OVA(D_{A_i}(t), L_{A_i}(t)) = 1 - \frac{d(P_{NumberGestures}, A_i(t))}{2\sqrt{2}}$$

These functions are also based on Euclidean distance. If the agent is very dominant, the lower the distance between $P_{BroadGesture}$ (resp. $P_{NumberGestures}$) is, the higher the result of SPC (resp. OVA) is. The obtained values of expressivity parameters are associated to each gestures of B. These gestures are finally sent to the *behavior realizer* that outputs the animation parameters filtered and altered by the agent social attitude.

5 Conclusion and future works

In this paper we have proposed a computational model that enables an agent to convey social attitude through nonverbal behaviors. Based on studies in Human and Social Sciences, functions have been defined to inhibit or emphasize some behaviors and to generate specific gestures according to the agent dominance and friendliness. The computational model has been integrated within a SAIBA-like architecture by introducing three new modules. This model takes as input the communicative intentions, the possible signals associated, a table of influence based on the literature of Human and Social Sciences, the gender and the social attitude of the agent. For now we have considered that the social attitude was an available data but it should be explained in futire works how it is computed based on the goals and the social relations of the agent. The outputs are probabilities and expressivity parameters for the potentials signals. This model is a first step for generating nonverbal behaviors influenced by the social attitude an agent wants to convey. We are placing ourselves in the context of a virtual agent but this model could also be used for a robotic agent as the SAIBA architecture is also used in this domain [26].
The next step is the implementation and the validation of such a model through user perceptive studies. Moreover, our model presents some limits. For instance, the proposed table is built from the studies of psychologies we found so far and therefore does not consider every nonverbal behavior. For instance, there are some important cues generated with speech. Tone, volume, speed of voice can change, depending on the social attitude. Also, mimicry is an important factor to express social attitude that is not considered yet in the proposed model. Moreover, in the presented work, as a first step, we have modeled the results extracted from Human and Social Science with discrete values (-1,0 and 1) to represent the influence of social attitude on nonverbal behaviors. We aim at collecting real data on the user's nonverbal behavior while expressing a social attitude to improve the proposed model and retrieve more precise values.

Acknowledgment

This research has been partially supported by the European Community Seventh Framework Program (FP7/2007-2013) under grant agreement no. 231287 (SSPNet) and for the REVERIE project.

References

[1] J. Bates. The Role of Emotion in Believable Agents. *Communications of the ACM*, 37(7):122–125, 1994.

[2] C. Becker, S. Kopp, and I. Wachsmuth. *Why Emotions should be Integrated into Conversational Agents*, pages 49–67. John Wiley and Sons, Ltd, 2007.

[3] N. Bee, C. Pollock, E. Andrea, and M. Walker. Bossy or wimpy: Expressing social dominance by combining gaze and linguistic behaviors. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271. Springer Berlin / Heidelberg, 2010.

[4] E. Bevacqua, M. Mancini, and C. Pelachaud. A listening agent exhibiting variable behaviour. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA '08, pages 262–269, Berlin, Heidelberg, 2008. Springer-Verlag.

[5] T. Bickmore and J. Cassell. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '01, pages 396–403, New York, NY, USA, 2001. ACM.

[6] T. Bickmore, L. Pfeifer, and D. Schulman. Relational agents improve engagement and learning in science museum visitors. In *Proceedings of the 10th international conference on Intelligent virtual agents*, IVA'11, pages 55–67, Berlin, Heidelberg, 2011. Springer-Verlag.

[7] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2):293–327, June 2005.

[8] N. J. Briton and J. A. Hall. Beliefs about female and male nonverbal communication. *Sex Roles*, 32:79–90, 1995.

[9] J. K. Burgoon, D. B. Buller, J. L. Hale, and M. A. de Turck. Relational Messages Associated with Nonverbal Behaviors. *Human Communication Research*, 10(3):351–378, 1984.

[10] J. K. Burgoon and B. A. Le Poire. Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs*, 66(2):105–124, 1999.

[11] R. H. Campbell, M. Grimshaw, and G. Green. Relational agents: A critical review. *The Open Virtual Reality Journal*, 2009.

[12] J. Campos and A. Paiva. May: My memories are yours. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 406–412. Springer Berlin / Heidelberg, 2010.

[13] D. Carney, J. Hall, and L. LeBeau. Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior*, 29:105–123, 2005.

[14] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 520–527, New York, NY, USA, 1999.

[15] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill. *Embodied Conversational Agents*. MIT Press, 2000.

[16] M. Courgeon, C. Clavel, and J.-C. Martin. Appraising emotional events during a real-time interactive game. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, AFFINE '09, pages 7:1–7:5, New York, NY, USA, 2009.

[17] P. Ekman and W. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, 1969.

[18] M. B. Gurtman. Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass*, 3(4):601–619, 2009.

[19] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud. Design and evaluation of expressive gesture synthesis for embodied conversational agents. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pages 1095–1096, New York, NY, USA, 2005. ACM.

[20] U. Hess, R. Adams, and R. Kleck. Who may frown and who should smile? dominance, affiliation, and the display of happiness and anger. *Cognition and Emotion*, 19(4):515–536, 2005.

[21] U. Hess and P. Thibault. Why the same expression may not mean the same when shown on different faces or seen by different people. In J. Tao and T. Tan, editors, *Affective Information Processing*, pages 145–158. Springer London, 2009.

[22] K. Isbister. *Better Game Characters by Design: A Psychological Approach (The Morgan Kaufmann Series in Interactive 3D Technology)*. Morgan Kaufmann, June 2006.

[23] Z. Kasap, M. B. Moussa, P. Chaudhuri, and N. Magnenat-Thalmann. Making them remember emotional virtual characters with memory. *IEEE Computer Graphics and Applications*, pages 20–29, March 2009.

[24] B. Knutson. Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, 20:165–182, 1996.

[25] S. Kopp, B. Krenn, S. Marsella, A. Marshall, C. Pelachaud, H. Pirker, K. Tharisson, and H. Vilhjalmsson. Towards a common framework for multimodal generation: The behavior markup language. In J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier, editors, *Intelligent Virtual Agents*, volume 4133 of *Lecture Notes in Computer Science*, pages 205–217. Springer Berlin / Heidelberg, 2006.

[26] Q. A. Le, S. Hanoune, and C. Pelachaud. Design and implementation of an expressive gesture model for a humanoid robot. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, pages 134 –140, oct. 2011.

[27] M. Mancini and C. Pelachaud. Dynamic behavior qualifiers for conversational agents. In C. Pelachaud, J.-C. Martin, E. Andre, G. Chollet, K. Karpouzis, and D. Pele, editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, pages 112–124. Springer Berlin / Heidelberg, 2007.

[28] M. Mori. The uncanny valley. *Energy*, 1970.

[29] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. Greta: an interactive expressive eca system. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '09, pages 1399–1400, Richland, SC, 2009. International Foundation for Autonomous Agents and Multiagent Systems.

[30] M. Ochs, N. Sabouret, and V. Corruble. Simulation of the dynamics of nonplayer characters' emotions and social relations in games. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(4):281 –297, dec. 2009.

[31] M. Preda and F. Preteux. Critic review on mpeg-4 face and body animation. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages 505 – 508 vol.3, june 2002.

[32] J. Rickel and W. L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13(4-5):343–382, 1999.

[33] K. Scherer. What are emotions? and how can they be measured? *Social Science Information*, 2005.

[34] Y. Yabar and U. Hess. Display of empathy and perception of out-group members. *New Zealand Journal of Psychology*, 2007.

Eyebrow Motion Synthesis Driven by Speech

Y. DING[1]
ding@telecom-paristech.fr

M. Radenen[2]
mathieu.radenen@lip6.fr

T. Artières[2]
thierry.artieres@lip6.fr

C. Pelachaud[3]
pelachaud@telecom-paristech.fr

1Institut Mines-TELECOM TELECOM ParisTech
37-39 rue Dareau, 75014 Paris – FRANCE

2Computer Science Lab (LIP6)
Pierre and Marie Curie University
4 place jussieu, 75005, Paris – FRANCE

3CNRS – LTCI UMR 5141
Institut Mines-TELECOM TELECOM ParisTech
37-39 rue Dareau, 75014 Paris – France

Abstract:
Natural eyebrow movements during speech are important to model a virtual agent able to sustain a natural and lively conversation with humans. In our work we focus on investigating statistical frameworks for learning the correlation between speech prosody and eyebrow motion features extracted from a video corpus of human faces uttering emotional spoken speech. Such methods may be used to synthesize automatically accurate eyebrow movements from synchronized speech as we demonstrate experimentally.

Keywords: Speech Animation, Nonverbal Behavior Generation, Data-Driven Animation, Eyebrow motion synthesis

1 Introduction

Embodied conversational agents are entities endowed with communicative and expressive capabilities. It has a human-like appearance. It can take various roles, such as serving as a guide or be an online teacher. Natural and lively behaviors are crucial for engaging users in human-computer interactions. Humans are very sensitive to subtle behaviors. They have very good skills to interpret the behaviors they perceive in their interlocutors. Virtual humans ought to be capable of displaying such high quality behaviors.

When communicating we use not only speech but also behaviors. Behaviors may have various meanings. For example, a head nod can convey agreement, mark an emphasis, or be a backchannel signal, etc [13]. Various studies have shown the tight relationship between speech and behaviors production. For example, Bolinger [3] found a strong correlation between the raise of F0 and of eyebrow movements. Verbal and nonverbal behaviors are tied to the meaning of spoken text (what is being said). Moreover, non-verbal behaviors are not only correlated with prosodic and acoustic features, but also with higher level information such as emotional states and attitudes [9].

Kuratate et al. [14] reported very high correlation (r=0.83) on average between head motion and the fundamental frequency (F0); on the other hand, Yehia et al. [25] stated that "*between 80 and 90 of the variance observed in face motion can be accounted for by the speech acoustics*". Computational models such as those proposed by Busso et al. [5] and Hofer et al. [12] rely on such links to learn the relationship between modalities.

In our work we focus on investigating a statistical framework for learning the correlation between speech prosody and eyebrow motion using a video corpus of human faces uttering emotional spoken speech with the aim of learning to synthesize accurate eyebrow movements from speech.

Most existing models of virtual agents' behaviors can be clustered into two main groups. In one group, models are based on experimental data and theoretical models taken from domains such as psychology, emotion studies, linguistic. Examples of such models are works by [1, 6]. On the other hand, statistical

techniques have been applied to learn from data the correlation between speech and multimodal behaviors. These models make use of the tight relationship between acoustic and visual behaviors. Works by [5, 7, 8, 12, 17, 18] belong to this cluster.

Both of these models types have pros and cons. While statically-driven models are more prone to produce natural looking animation, cognitive

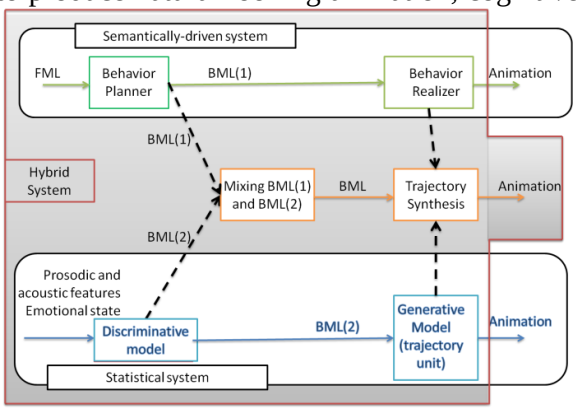


FIG. 1 – Overall Architecture: The aim of this work is to combine a Semantically-driven approach (top system) with a Statistical system (bottom system) by proposing combination schemes at different levels, to infer BML from the inputs, and to synthesize trajectories from the combined BML resulting in a new hybrid system (grey box).

models capture more precisely the semantic-emotional behaviors to communicate. These latter ones are often event-driven; that is they compute a behavior only when a given communicative function is specified. Statically-driven models produce animation continuously that captures the communicative colour of the message to convey but they have difficulty to compute behaviors which have specific meaning. As a result, virtual agents driven by cognitive-like system are able to convey more precise displays while those driven by statistical models look more natural and lively [16].

Figure 1 illustrates the ultimate overall hybrid architecture we want to develop. Our aim is to combine both approaches, embedding in one model the semantically-driven approach and the statistical one allowing us to have an agent able to convey semantic-emotional messages naturally.

Regarding the first model-type, we will make use of an existing platform, Greta, that takes as input a set of intentions and emotions to convey and outputs synchronized multimodal behaviors [1]. Greta follows the SAIBA architecture [22] that proposes a structure to go from intentions to animation. Communicative intentions and emotions are represented with the FLM language while multimodal behaviors with BML [22]. Our research plan is as follows:

1. In a first step, we investigated statistical models to infer the animation motion from speech input. The goal is to build a statistical system which is able to learn from training samples how to generate natural animation motion from speech features. This corresponds to the statistical system in Figure 1.
2. In a second step, relying on the existing GRETA system (the semantically-driven system in Figure 1), we will propose few hybrid schemes to combine the output of both the Greta system and the statistical system for improving speech-based animation. This corresponds to the implementation of combination methods in the hybrid system in Figure 1.

We describe in this paper the work we did up to now in developing a statistical model, corresponding to the first step of the work. More particularly, we have developed three statistical systems that are learned from training samples to generate movements (eyebrow motion) from speech input.

In the remaining of this paper, we present existing works on statistical approach. We then turn our attention to the three statistical approaches we propose.

2 Background and related works

2.1 Related Works

Researchers have presented data-driven approaches to synthesize speech animation, including body and facial animation. Levine et al. [17] and Chiu and Marsella [7] generate

automatically body motions from spoken speech. Given the tight relationship between acoustic phonemes and visual visemes, speech is also used to drive lip motion in [4, 24]. While these examples mainly focus on speech content, other works are particularly interested in synthesizing nonverbal communicative behaviors during speech, such as head and eyebrow motion.

Costa et al. [8] used Gaussian Mixture Model (GMM) to learn the correlation between eyebrow movements and speech features. Busso et al. [5] and Hofer et al. [12] applied HMMs to model the dynamic relationship between speech and motion streams. In the synthesis step, speech features are used to determine the most probable state sequence; then the model of Busso et al. [5] produce the motion mean vector sequence from this state sequence. Finally spherical cubic interpolation is employed to smooth head motion sequences. Such approaches [5, 12] suffer from a tradeoff between the quantization error and the inter-cluster discrimination [5]. At the end, such approaches are not adapted to produce a variety of trajectories but rather produce one motion sequence, the most likely one, which limits the quality of synthesized motion sequence.

Li and Shum [18] stated that synthesizing motion sequence from the most likely state sequence "*obliterates the relationship between vocal and visual signals*". They propose an Input-Output HMM whose emission and transition probabilities are dependent on not only the current state in the Markov chain but also on the current speech feature. Levine et al. [17] applied Conditional Random Field (CRF) [15] to estimate the state probability distribution for relaxing strong independence assumptions made in standard HMM. Xue et al. [24] studied three different structure models: independent HMMs, coupled HMMs and product HMM. The various works [17, 18, 24] mentioned in this subsection propose diverse techniques to avoid one or some limiting assumptions from conventional HMMs.

2.2 Details on the mainstream approach for synthesizing motion from speech

We present in more details the underlying approach that is more or less used in many of the works cited above [5, 12].

Tokuda et al. [21] proposed a procedure to synthesize smooth trajectories from a standard Gaussian HMM (i.e. with Gaussian emission probability function). They distinguish between 2 cases to which we refer in our experiments: a general case (named *case 1* hereafter) consisting in synthesizing an observation sequence from the HMM by integrating over all possible state sequences; a more restricted case (named *case 2*), consisting in synthesizing an observation sequence from the HMM by considering only one state sequence (possibly the most likely one).

Based on Tokuda et al's approach, Hofer et al. [12] proposed a method for synthesizing an observation stream (head motion features) from another observation stream (speech features). The key idea is to design a Gaussian "joint" HMM, named λ, working on concatenated observation vectors, where the observation at time *t* is the concatenation of the first stream observation and of the second stream observation. A key point is that one can define the joint HMM combining a Gaussian HMM for the first stream, named λ^1, and a Gaussian HMM for the second stream, named λ^2.

Then, once the joint HMM is trained, one can synthesize a sequence of observation of the second stream from the sequence of observation of the first stream.

3 Statistical approaches for synthesizing natural motion

This section describes our attempts to design a statistical tool for synthesizing natural motion from speech.

We present three approaches that we developed and experimented. We present them from the simplest to the most sophisticated.

3.1 Contextual HMMs and Fully Parameterized HMM (FPHMM)

3.1.1 Contextual HMMs

Contextual hidden Markov models (CHMMs) have been proposed for taking into account additional contextual variables [23]. Such models have been initially proposed for recognizing gestures with the idea of using contextual information related to the physiology of the person realizing the gesture or the amplitude of the gesture. They have been extended in various ways in [20].

CHMMs are able to handle variability by taking into account external (contextual) variables that modify emission probability distributions. More precisely in CHMMs, the mean $\hat{\mu}^j$ of the Gaussian distribution in state j is defined as a linear function of a vector of contextual (or extern) variables that we note θ (e.g. θ stand for emotional features) it is defined as:

$$\hat{\mu}^j(\theta) = V^j\theta + \bar{\mu}^j$$

where $\bar{\mu}^j$ is an offset vector which may be viewed as an average mean vector (eventually obtained from a traditionally learned HMM) that is modified by the linear transform part V^j (a matrix). The extension in [20] includes parameterization of mean vectors as well as of covariance matrices with external variables θ (note that θ may vary with time).

Using CHMM for modeling eyebrow motion

To design a system with CHMMs one can learn a CHMM with speech features as contextual variables (i.e. emission functions are conditioned not only on current state but also on speech features) and with both motion and speech features as observations.

At the synthesis step, only speech features are first used with the marginalized speech HMM to find the most likely state sequence, then the algorithm described in *case 2* of section 2.2 is used to synthesize a smooth animation along this state sequence with the marginalized motion HMM.

3.1.2 Improved Contextual HMMs: Fully Parameterized HMMs (FPHMMs)

CHMM still suffers from the independence (with respect to speech features) assumption of state transition probabilities. Thereby, we developed a new extension of CHMM, FPHMM, by parameterizing transition probabilities including the initial state distribution with external variables θ. This leads to Fully Parameterized HMM model where transition probabilities, means and covariance matrices depend on external variables.

Let note θ_t the (c-dimensional) vector of contextual variables at time t. Then we define the state transition distribution a_{ij} from i th state to j th state as:

$$a_{ij}(\theta_t) = \frac{\exp(logA_{ij} + W_{ij}\theta_t)}{\Sigma_{j'}\exp(logA_{ij'} + W_{ij'}\theta_t)}$$

where $W_{ij}{}'$ is a c-dimensional vector and A_{ij} may be viewed as an offset value. In this model the transition probabilities change at every time step according to the contextual variables. Such a modeling is interesting in our case when one exploits speech features as contextual variables.

Such a modeling subsumes standard HMM by setting W_{ij} to null vector. Training consists in estimating the parameters contained in CHMM and W_{ij}. To learn a FPHMM we learn first a CHMM with parameterized emission probability distributions. Then we estimate the W_{ij} parameter matrices via a Generalized Estimated Maximization algorithm using gradient ascent.

Using FPHMMs for modeling eyebrow motion

To design a motion from speech synthesis system we learn FPHMM that take speech features as external variables and only motion features as observations. Thereby, speech features influence directly state transition probabilities and emission probability distributions. This model is trained via likelihood maximization.

For synthesis, speech features (external variables) are used to determine a probability distribution over all the hidden states at each time step using only transition probabilities. Then the *case 1* algorithm discussed in section 2.2 is applied to generate the most likely smooth animation using emission probability

density defined on motion features only.

3.2 Combining Fully Parameterized HMMs and Conditional Random Fields (FPHMM&CRF)

We also investigated the combination of Fully Parameterized HMM and of Conditional Random Fields (CRFs) [15] with the same topology structure.

Speech features are taken as external variables and motion features as observations in the Fully Parameterized HMM. Speech features are taken as observation for the CRF which outputs a state sequence or a probability distribution over state sequences, which will be used to synthesize the motion features.

For training, FPHMM are first learned as is done in section 3.1.2; secondly, speech feature (external variables) sequences are used to find the most possible state sequence Q = $(q_1, q_2, \ldots q_T)$. This state sequence and the corresponding speech features are used to train a CRF with the same topology as the FPHMM.

Usage for eyebrow motion synthesis from speech

For synthesis, a speech feature sequence is input to the CRF to get a probability distribution over hidden state sequences. Speech feature are also used in the Fully Parameterized HMM to calculate the mean and covariance matrix of each hidden state at each time step. Given such computed parameters, and the distribution on hidden states output by the CRF, the algorithm described in *case 1* of section 2.2 is used to synthesize a smooth animation along this state sequence with the marginalized motion HMM.

This approach not only overcomes the limitations from the assumptions of standard HMM by Fully Parameterized HMM but also takes the advantages of CRF as a discriminative model for inferring accurate probability distribution over all hidden state sequences.

4 Experiments

4.1 Data Processing

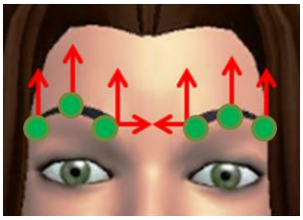


FIG. 2 – Illustration of the eight extracted *facial animation parameters* (red arrows) and of the six *features points* (green circles).

Experiments have been performed on the Biwi 3D Audiovisual Corpus of Affective Communication (B3D(AC)^2) database [11]. 14 subjects were invited to speak loud 40 short English sentences. Each sentence was repeated twice. In total, this corpus includes 1109 sequences, each lasting 4.67s long on average.

Feature Extraction

We used data from three subjects in our experiments to train and test our approaches. For each sequence we obtained a speech stream and an eyebrow motion stream. We detail now the feature extraction process for both streams.

A real-time 3-D scanner was employed to capture at 25Hz the trajectories of 23370 points of the facial mesh of each actor. It also includes 3 head rotations and 3 head translations. These points include all the *feature points* (FPs) defined in the MPEG-4 facial animation standard [19]. In our work we focused only on eyebrow motions and according to MPEG-4 standard we used 8 *facial animation parameters* (FAPs) that control 6 FPs to describe eyebrow animation (see Figure 2). We extracted the trajectories of these 8 FAPs as eyebrow motion features. Since we are only interested in the relative displacements, the neutral pose used as reference is not taken into account. This neutral pose is approximated as the average value of the motion features in the first 5 frames of each sequence. Indeed all the actors were asked to hold the neutral pose before starting talking [11]. For the sake of simplicity, we assume that the right and the left eyebrow motions are

symmetric. So we consider the mean of the right and the left brows as the eyebrow motion features. At the end a signal is transformed in a sequence of *four*-dimensional feature vectors with a rate of 25 frames (i.e. feature vectors) per second (fps).

Concerning the speech features we used prosodic features (pitch and RMS energy) which we extracted with PRAAT software [2] at the same sample rate as for motion feature extraction (25 fps).

Above features are called *static features*, they are representative of what happens locally in the stream. As is traditionally done in speech recognition we may use augmented feature vectors both for motion and for speech streams by adding *dynamic features*, namely first and second order derivatives of static features (i.e. velocity and acceleration noted Δ and $\Delta\Delta$ features). Hence we get 6 dimensional frames for speech and 12 dimensional frames for motion.

Note also that the averages of these static and of the dynamic features of speech over a temporal window are computed at each time step and will be used as dynamic contextual variables θ_t (see details in section 3) in contextual models.

Isolated Action Unit sequences

Facial Action Coding System (FACS) describes the relationship between facial appearances and muscular actions [10]. It categorizes facial behaviors by the underlying muscular contraction. After some attempts to extract automatically AUs using analysis tools that are publically available, we have decided to use manual annotation to ensure a minimum error rate. So, we manually labeled the data of three subjects with respect to four classes corresponding to the combination of Action Units 1+2+4 (oblique and raise eyebrow, eyebrow of fear), the combination of Action Units 1+2 (raise eyebrow, eyebrow of surprise), Action Unit 4 (frown, eyebrow of anger), and the combination of Action Units 1+4 (oblique eyebrow, eyebrow of sadness), plus fifth additional *no motion* class. So far the data was annotated by one coder.

At the end we get 557 of pairs of streams (speech, eyebrow motion), where each sequence comes from a subsequence of an original sequence in the B3D(AC)^2 dataset and corresponds to the realization of a single event from one of the five classes defined above.

4.2 Experimental settings

Note that all following results are averaged results gained through cross validation with random splits of the dataset into 80% for training and 20% for test.

For each of our approaches (CHMM, PFHMM and PFHMM&CRF), we built a model for each motion category / class. The models of all classes are trained together. For evaluating the performance of our approaches for synthesizing motion sequences from new speech, we computed a mean quadratic error between normalized original and synthesized motion sequences.

4.3 Results

Figure 3 shows an example (from the test set) of real trajectories of four motion features together with the synthesized trajectories (by CHMM, PFHMM, PFHMM&CRF and conventional HMM) of these motion features based on speech features only. We can remark that CHMM, PFHMM and PFHMM&CRF provide results that are closer to real eyebrow motion. PFHMM&CRF gives the best performance while conventional HMM the worst one.

Table 1 shows the mean synthesis errors with our three models and with conventional HMMs for various topologies (number of states) with external variables computed base on windows of two different lengths (2 frames and 10 frames). It may be seen in the table that the combination of Fully Parameterized HMM and of CRF performs better than other two approaches and than conventional HMMs.

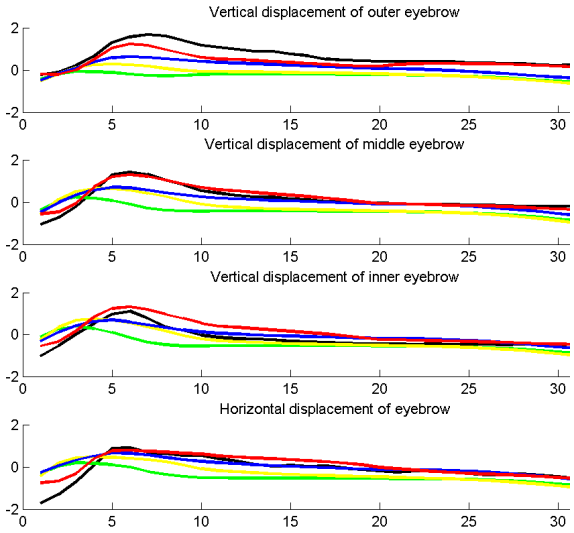


FIG. 3 – Comparison of normalized motion sequence synthesis by HMM (Green curves) and by our three models (CHMM: yellow curves, FPHMM: blue curves, FPHMM+CRF: Red curves) with the normalized original sequence (Black curves). The four boxes correspond to the four motion features. In every box, the five curves show the evolution of the corresponding motion feature (in y-axis) with time (x-axis) when a particular action unit (AU1+AU2) is performed.

Figure 4 shows eight frames extracted from eyebrow motions (AU1+AU2) synthesized by PFHMM&CRF.

5 Conclusion and Perspectives

The work that we have done until now consisted of implementing statistical models for synthesizing eyebrow motion features based on speech features. We investigated few approaches based on contextual extensions of HMMs, including a hybrid system combining conditional random fields and fully parameterized HMMs. To synthesize motion from speech we exploited contextual HMMs that use speech features as external variables, i.e. where HMM's parameters are conditioned on these external variables.

Our results show that contextual models are significantly better than conventional HMM and that our new extension FPHMM outperforms all other methods under investigation. Building on this work we plan to implement a hybrid schema between the existing Greta system and our statistical system.

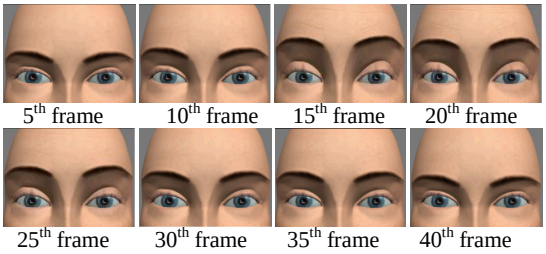


FIG. 4 – Several frames sampled from an example of synthesized eyebrow motions (AU1+AU2)

TAB. 1 – Synthesis error for single Gaussian HMMs and three novel models (CHMM, PFHMM and PFHMM&CRF) as a function of the number of states (S)

Topologies LR	HMM	CHMM	PFHMM	PFHMM& CRF
S=4	0.2276	0.1901	0.1791	0.1581
S=6	0.2105	0.1788	0.1525	0.1462
S=8	0.2034	0.1504	0.1408	0.1352

References

[1] E. Bevacqua, K. Prepin, R. Niewiadomski, E. de Sevin, and C. Pelachaud. Greta: Towards an interactive conversational virtual companion. Artificial Companions in Society: Perspectives on the Present and Future, pages 143-156, 2010.

[2] P. Boersma and D. Weenink. Praat, a system for doing phonetics by computer. *Glot International 5(9/10),* pages 341-345, 2001.

[3] D. Bolinger. *Intonation and its Uses*. Stanford University Press, 1989

[4] M. Brand. Voice puppetry. In SIGGRAPH, 1999.

[5] C. Busso, Z. Deng, U. Neumann, and S. Narayanan. Natural head motion synthesis driven by acoustic prosodic features. *Computer Animation and Virtual Worlds*, 16(3-4): 283-290, 2005.

[6] J. Cassell. Body language: Lessons from the near-human. In J. Riskin, editors, *Genesis Redux: Essays in the History and Philosophy of Artificial Intelligence*. Chicago: University of Chicago Press, pages 346-374, 2007.

[7] C-C. Chiu, S. Marsella. How to train your avatar: A data driven approach to gesture generation. *IVA*, papes 127-140, 2011.

[8] M. Costa, T. Chen, and F. Lavagetto, Visual prosody analysis for realistic motion synthesis of 3D head models. In *Proceedings of ICAV3D01*, pages 343-346, 2001.

[9] P. Ekman. *Emotions revealed*. New York: Times Books (US). London: Weidenfeld & Nicolson (world), 2003.

[10] P. Ekman. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Printice-Hall, 1975.

[11] G. Fanelli, J. Gall, H. Romsdorfer, T. Weise, and L. V. Gool, A 3-D Audio-Visual Corpus of Affective Communication. *IEEE Trans. on Multimedia*, 12(6): 591-598, 2010.

[12] G. Hofer, H, Shimodaria, and J. Yamagishi. Speech driven head motion synthesis based on a trajectory model. In *ACM SIGGRAPH, 2007*.

[13] M. Knapp and J. Hall. *Nonverbal communication in human interaction*. Harcourt Brace, fourth edition, 1997

[14] T. Kuratate, K. G. Munhall, P. Rubin, E. Vatikiotis-Bateson, and H. Yehia. Audio-visual synthesis of talking faces from speech production correlates, *EUROSPEECH*, 1999.

[15] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML*, pages 282-289, 2001.

[16] J. Lee and S. Marsella. Modeling speaker behavior: A comparison of two approaches. In *IVA*, pages 161-174, 2012.

[17] S. Levine, P. Krähenbühl, S. Thrun, and V. Koltun. Gesture controllers. In *SIGGRAPH*, 2010.

[18] Y. Li and H. yeung Shum. Learning dynamic audio-visual mapping with input-output Hidden Markov models. *IEEE Trans.Multimedia*, pages 542-549, 2006.

[19] I. S. Pandzic and R. Forchheimer, editors. MPEG-4 Facial Animation: The Standard, implementations and applications. John Wiley & Sons, Inc., NY, USA, 2003

[20] M. Radenen and T. Artières, Contextual hidden markov models. *ICASSP*, 2012.

[21] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi,T. Kitamura, Speech parameter generation algorithms for HMM-based speech synthesis. *ICASSP*, 2000.

[22] H. H. Vilhjalmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. Van Welbergen, and R. J. van der Werf. The Behavior Markup Language: Recent Developments and Challenges. In *IVA*, pages 99-111, 2007.

[23] A. D. Wilson and A. F. Bobick. Parametric hidden markov models for gesture recognition. *Pattern Analysis and Machine Intelligence, IEEE Transaction on*, 21(9): 884-900, 1999.

[24] J. Xue, J. Borgstrom, J. Jiang, L. Bernstein, and A. Alwan. Acoustically-driven talking face synthesis using dynamic bayesian networks. *ICME* 2006.

[25] H. C. Yehia, T. Kuratate, and E. Vatikiotis-Bateson. Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, 30(3): 555-568, 2002.

Model-based approach for natural language generation from semantic virtual environment

M. Barange P. Chevaillier
barange@enib.fr chevaillier@enib.fr

Lab-STICC, ENIB
Centre Européen de Réalité Virtuelle, 29280 Plouzané, France

Abstract:
Generating natural language (NL) dialogues in a collaborative virtual environment (VE) is a critical task. In such environment, the virtual agent can use semantic model of environment as a knowledge resource for decision-making and for NL generation. Unified Modeling Language (UML) has been used as a modeling paradigm for the conceptual modeling of rich-content VE. However, UML lacks to provide means to associate linguistic features to the model elements of the conceptual model. In this paper, we present a model-driven approach for the generation of NL dialogue, based on the semantic model of the VE. We propose an extension of UML to be able to add linguistic properties to the model elements. Thus, an agent can use these features to generate syntactically correct utterances. Furthermore, we extend the notation of UML associations for the unambiguous interpretation of different types of relationships between concepts reported in the literature. Moreover, we describe a model-driven approach to generate NL dialogues.

Keywords: semantic virtual environment, natural language generation, conversational agents

1 Introduction

Generating natural language (NL) dialogues is a critical task in a collaborative virtual environment (VE) such as VE for training and cultural heritage applications. In such environments, users and virtual agents are supposed to collaborate in order to perform a given task. They need to exchange information about resources they can use, and the activities they have to perform by their own or collectively with other agents (virtual agent or user). Communication in NL is one of the means to exchange information in such VEs. The agent should be able to generate short utterances for descriptions to be interactive and, when necessary, to go further in details, thus building up a structured discourse. Generating grammatically correct and unambiguous utterances by agent is needed for users to engage themselves within the VE.

The agent can use the semantic model of the VE as a knowledge base for decision-making and for communication with other agents [2]. In order to provide the semantic description of VEs, [15] have developed a specific modeling language, while [6] have proposed an extension of the Unified Modeling Language (UML). This paper uses MASCARET a model-based approach, for the design of semantic virtual reality (VR) environments [6]. This approach is grounded on a meta-model which is a specialization, and an extension of UML and which covers all aspects of VE's semantic representation : ontology of the domain, structure of the VE, behavior of entities, both user's and agents' interactions and activities. MASCARET provides a modeling language expressive enough for experts to formulate their knowledge about a specific domain. In this approach, UML is used to provide a knowledge-driven access to the semantic contents of the VE, and not for code generation as in classical software development process. Conceptual modeling defines the ability to design the model of application domain in terms of concepts.

The agents must provide information about concepts in the VE and the collaborative activities that change during the simulation. To ensure the correctness of information, the generation of NL dialogues can use semantic information of VE for the processing and unambiguous interpretation of relationship between concepts in VE. In most of the natural language generation (NLG) approaches [12, 18, 19] they do not take into account these semantic information of VE, and mostly used static predefined text for utterances. In order to generate grammatically correct sentences, we need to enrich modeling elements with linguistic properties. Existing conceptual models of VE fail to include NL properties, as UML does not provide built-in features for specifying linguistic characteristics of concepts.

The concepts in VE are often linked with each other through some relationships. Among other issues, it is necessary to precise the semantics of relationship, e.g, whether all the components of an association should keep their own properties, whether it is possible to manipulate them as whole, or let them behave individually, or can be manipulated within the constraints imposed by semantics of association. In UML, association defines the relationship between two

or more typed instances. Since UML does not define the direction of relationship between the typed instances, it can be interpreted in multiple ways. From the linguistic point of view, the semantics of associations are not sufficient for the unique and unambiguous interpretation of relationships in VE. Since the semantics of associations differ from one domain to another, and because the properties of UML associations do not fit exactly the need of NLG, it makes difficult to conciliate these different point of view in a consistent unified modeling framework.

The aspects of this paper are two folded. First, we bridge the gap between the specification of the VE and the corresponding conceptual model. Second, we describe the model-based approach for the generation of the NL dialogues using the semantic information of VE. The main contributions of the paper are as follows. First, we define a set of heuristics as guideline rules for the mapping of concepts from the specification of the application domain to the corresponding conceptual modeling of VE. These rules can help to choose appropriate name and characteristics for these concepts in conceptual model. Second, we propose a linguistic extension of existing UML meta-model in order to associate linguistic properties with model elements. Third, we propose well-defined semantics for the representation of relationships, up to the granularity of experts, that is needed for the clear and unambiguous interpretation of relationship for the generation of NL dialogues by agents. Finally, we apply model based approach for the generation of NL dialogues from the clear and unambiguous conceptual model of VE. Moreover, the semantic VE is served as a knowledge base for the agent.

In the next section we review the work proposed in the literature. In section 3 we propose the extension of UML meta-model by adding new linguistic elements in it. the section 4 explores the extended semantics of association. Then, in section 5, we propose a model-based approach to generate natural language dialogues from the conceptual model of VE. Discussion and future aspects are given in section 6. The section 7 concludes the paper.

2 Related Work

Identifying linguistic characteristics such as noun, verb adjective etc. for the mapping of linguistic elements to the object-oriented concepts has long been a research focus for the software engineering (SE) and database community. Abbott [1] suggested to use common nouns as data types, proper nouns as objects, adjectives as attributes and verbs to denote operations for the designing of programs from informal english language description. [16] classified the noun into class noun, value noun, attribute noun, and action noun. The author also classified verbs into relational verb, state verb, action verb, and action relational verb. [5, 11] proposed guideline rules for the construction of entity-relationship (ER) models. [7] proposed that the adjective as a noun, can describe one end of a relationship that is the role of the object with respect to the object at another end of the relationship. The results of the empirical study conducted in [13] also confirms that the examined class names are described by nouns or noun phrases, and associations by verbs or verb phrases, and noun can be used as role. All these work suggested how to automate the analysis of linguistic specifications, but provided no evidence to couple these linguistic properties with the generated model, and thus enlarging the gap between specification and model of the system.

A very few attempts are made for the natural language generation (NLG) from the model [13, 3]. [13] generates the NL specifications from class diagram. [3] first transforms the class diagram into an intermediate linguistic model using Grammatical Framework, which is then transformed into NL specifications. [4] constructs the NL from UML and OCL using semantic business vocabulary and business rules (SBVR) that maps the limited set of possible sentence structures.

Being one of the central structural concept in UML, the concept of association has significantly enhanced to increase expressiveness of the language. Unfortunately, these enhancements have introduced many ambiguities for the interpretation of part-whole relationships. [9] identifies semantic problems of lower multiplicity in n-ary associations where multiplicity values provides only partial understanding of the objects structure. [8] deals with the temporal aspect of associations defined in the form of verbs representing the state and the events, but does not deal with part-whole type of associations also called meronymic relationship [17]. For the better understanding of relationship in database design, [17] proposed the taxonomy of relationship based one the work of [20]. In [14], author formally defined the semantics of association

ends that deals with the uniqueness property of association ends and proposed intentional interpretation of association ends. All these work reflects the different approaches for the interpretation of associations, however, they do not address unambiguous and clear interpretation of associations from the linguistic point of view.

3 Linguistic properties of UML model elements

3.1 Understanding UML model elements

In principle, UML has been used to define conceptual model in VR applications. For example, in MASCARET, the static part of the VE such as entities and their relationships are modeled through UML class diagram, whereas the behavior of these entities and of overall system are modeled through state machine and activity diagrams [6]. From the SE point of view, UML can be used to define conceptual model of VE, but from the linguistic point of view, it is lacking in expressiveness of linguistic characteristics of model elements, and also lacking in providing clear and unambiguous semantics of association for the generation of NL dialogues from the model. However, to overcome theses domain specific limitations, UML offers extension mechanisms through stereotypes, tagged values, and constraints to add new kind of modeling concepts.

3.2 Linguistic Annotation

From the literature survey, it is clear that linguistic features have been used for the construction of conceptual model from specification requirements. The focus was on concepts, their attributes, and operations, but not on the linguistic features, and thus, linguistic features have not given attention in the generated conceptual model. It is possible to define the linguistic nature of modeling elements. For example, we restrict the name of the operation be performed on objects to be verbs or verb phrase. Nevertheless, the linguistic nature of an element allows to associate specific information. For example, the name of class is a noun, it has a gender type (e.g. the gender of noun "Ship" is feminine).

To be able to use linguistic properties of concepts for the generation of propositional contents by virtual agents, the designer of the model must associate these linguistic properties to model elements. Therefore, we list some of the lessons learnt from the literature, in the form of heuristics for modeling elements. These heuristics can be served as guidelines for designers to construct semantic model of VE coupled with linguistic information following the semantic restrictions of model elements.

In the context to the modeling of semantic VE, a class corresponds to a physical entity or logical concept in the VE.
Heuristic-1 : The class name should be defined as a common noun.
To form a sentence, common noun helps to determine limiting modifiers such as a, an, every, some, etc. Heuristic-1 does not impose that all common nouns can be assigned to the class name. for example the word “red” is a common noun, but should not be assigned as a class name.

Objects in VE represent instances of concept classes, and the proper noun names and refers to a specific (one of the kind-of) concept.
Heuristic-2 : Object name should be defined as proper noun.

Attributes in the VE refer to abstract characteristics of classes or entities. In general, the attribute name is a noun. The value associated with the attribute can define the state of an instance of the class or entity. Furthermore, as the adjective describes the noun, it can be used to define the boolean type property of the enclosing class [7].
Heuristic-3 : Attribute names should be a common noun or an adjective.

Operation corresponds to the behavior of the enclosing class. It is generally denoted by verb or a verb phrase.
Heuristic-4 : Name of the operations should be a verb or a verb phrase.

An association links a typed instance to another typed instances in the VE, similar to as the verb or verb phrase links noun or noun phrase to another. A transitive verb has two main characteristics. First, it is an action verb [16], and second, it must have a direct object, something or someone who receives the action of the verb. Similarly the action noun represents the action (e.g. transmission, shipment).
Heuristic-5 : The association name should be a Transitive verb or an action noun.

In VR applications, the use of semantic model of VE as linguistic resource for NLG imposes

a number of constraints on the nature of modeling language and properties of modeling elements. We identified these constraints as semantic restrictions and / or additional properties of model elements. Thus, to support linguistic features of model elements, we propose a linguistic extension of the UML meta-model by adding new model elements at the meta level. The reason to add these elements to the meta-model are *(a)* features of newly added model elements are independent to or from the model, and *(b)* it provides semantics for the unique interpretation of associations.

The figure 1 illustrates the extension of UML corresponding to the MASCARET profile that extends and specializes semantics of UML for the linguistic domain modeling. We introduce

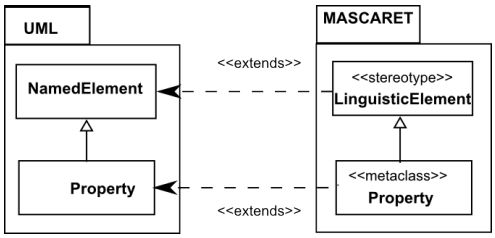


FIGURE 1 – an abstract view of the linguistic extension of UML in MASCARET

a new model element so called "LinguisticElement" that extends the “NamedElement” from UML model in the form of stereotype (see figure 2). Each Linguistic element is characterized by a name, language, regular expression and can have synonyms. A linguistic element can be subdivided into noun, verb, adjective, preposition etc. The stereotype “noun” has a gender, number (singular/plural), and can specify whether it is a proper noun, and whether or not it is countable. The stereotype “verb” is characterized by whether it is transitive, it is pronominal, and by it’s conjugation type. It should be noted that the concept “noun” is used to represent both the noun and noun phrase, and similarly, the concept “verb” is used for the verb and verb phrase both.

These linguistic elements can be linked with corresponding model elements of UML to couple linguistic semantics with elements in conceptual model. As, there is no “automatic” check to respect these linguistic constraints, the designer of the model must associate these linguistic features by choosing appropriate names and properties of modeling elements.

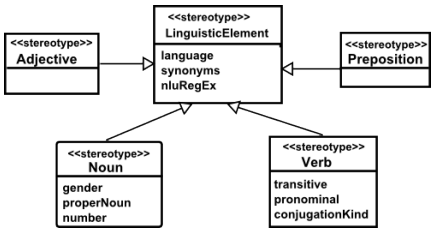


FIGURE 2 – Stereotypes of the meta-model defining linguistic properties of model elements

4 Semantic Relationship

From the linguistic point of view, the existing semantics of association in UML are not clear. First, the UML meta-model does not provide means to specify the direction of relationship between typed instances. Second, semantics of association in UML are not sufficient for unique and unambiguous interpretation of relationships in VE. These limitations make it difficult to understand the relationship between participating typed instances in VE. For example, consider a “part-of” relationship between a helm and a ship, and between a ship and a navy (figure 3). From the existing semantics of UML association, a virtual agent can not clearly determine that the “part-of” relationship between a helm and a ship is actually from “ship to helm” or from “helm to ship”. A syntactically correct UML model containing associations can be differently interpreted by different agents. Furthermore, the virtual agent can not determine whether both “part-of“ relationships represent the same semantics or they are different.

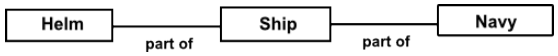


FIGURE 3 – "part-of" relationship

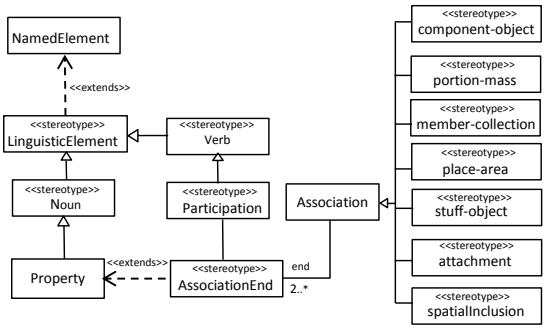


FIGURE 4 – Extract of extended UML meta-model : Association concept and linguistic elements

To overcome these limitations, we propose an extension of the notion of UML association.

The figure 4 shows an extract of extended UML meta-model. Each association has two or more association ends (AssociationEnds). The association end extends the UML property class. The linguistic characteristic of association end is that it is of the type noun. Each association end has one participation. The association end maps typed instance of opposite end to the typed instance at owner end. The association end can represent the semantics of thematic role proposed by [8], and can also have the same semantics as the participation role types in ER models [5, 11]. From linguistic point of view, the name of the association end defines the nature of a typed instance owning the association end. Participation is a named element of the type verb. It defines an occurrence of an action or a state of being for the typed instance at opposite end to the typed instance at owner end of association. The participation name should be defined as transitive verb. The typed instance at owner end plays the role of direct object for the verb denoting participation name. The syntax of the participation end is :

```
name\participationRole= valueSpecification
```

where name is the name of association end, and the participationRole is of the type participation and valueSpecification is the name of the verb.

Heuristic-6 : The name of association end should be a noun.
Heuristic-7 : The participation name should be a transitive verb.

Now, consider again the "part'-of" association between ship and helm as depicted in figure 5. By following the semantics of association ends and participation, the agent can clearly and uniquely determine that *(a)* a helm is a part of ship, or more clearly, a helm which is a steering, is a part of cargo vessel that is a ship, and *(b)* a ship contains a helm, or more clearly, a ship which is cargo vessel, contains a steering that is a helm.

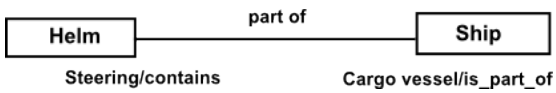


FIGURE 5 – "part-of" association between helm and ship

Moreover, to provide unambiguous interpretation of associations, we follow the work of [17, 20] and define stereotypes of associations. These stereotypes do not add new properties to the definition of association, but they define semantics in the form of constraints on how to use associations for the unambiguous modeling and for interpretation of associations in VE. We enlist the stereotyped associations for which we believe that these associations play important role in the VE.

«Component-Object» : Component-Object relationship exists between the component and the object of which it is a part of. The important characteristic of this relationship is that the component part performs an important function in the object part. Also, the components of the aggregated object can be different in nature. For example, there exist component-object relationship between helm and ship, where helm and ship both have different characteristics.

«Member-Collection» : It defines an association relationship in which a set of members is considered as an object in it. This relationship is also referred to as belongs-to or grouping or partitioning. This relationship differs from component-object relationship as the member does not perform any specific function in the collection.

«Portion-Mass» : The semantic of this relationship is that all parts are similar to other parts and to the whole.

«Place-Area» : This defines the relationship between an area and the specific location within it. For example Paris is located in France. The semantics of this relationship is similar to that of portion-mass relationship, except that place can not be separated from area.

«Stuff-Object» : It is the relationship between an object and its constituents. This relationship differs from the component-object relationship. Unlike in component-object relationship, it is not possible to physically separate stuffs from the whole object. e.g, a ship is partly made of wood.

«Attachment» : This relationship defines that one entity is physically attached or joined to another entity. For example, a handle attached to a cup.

«Spatial Inclusion» : This relationship exists when one object is surrounded by another object. Semantically, this relation differs from component-object relationship as in that components are connected to the object. e.g. a ship is in a port.

Consider again the "part-of" association between Ship and Helm class as depicted in figure

6. Thus, the associated stereotype «component-object» defines the semantics that the helm is a part of ship, and although, both the ship and the helm have different characteristics, helm performs specific function in ship. From the given example, it is clear that the extended UML model now provides appropriate mechanism to associate linguistic features with the model elements, and to define clear and unambiguous semantics for their interpretations.

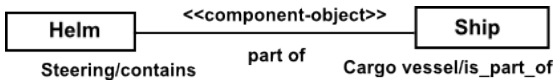


FIGURE 6 – «component-object» type association "part-of" between helm and ship

Nevertheless, it has to be clearly stated that, neither we aim to define the exhaustive classification of relationships, nor do we propose a universal set of characteristics through which any relation can be presented. We proposed to represent a relation with well defined semantics, up to the granularity, that is needed for the clear and unambiguous interpretation of relationship for the generation of NL dialogues by agents in VR applications. Thus, it is necessary for the modeling of semantic VE that the experts not only define the relationships, but also to quantify and precise the semantics of relationships with additional constraint within a given context.

5 Model Driven approach for Natural language generation

Our approach of generating NL dialogues focuses on what should be generated and how to construct dialogues from the model of the semantic VE. We assume that the speaker agent has understood the current context and the semantics of the last utterance from the addressee agent (user or virtual agent). The agent uses the semantics of VE and the task activity, as the knowledge base for the generation of NL dialogues [2]. The agent holds the information about the domain related concepts. In such context, it is also reasonable that the agent holds the weak-belief about other agents or users. Once the contexts is defined, the generation algorithm must determine what information to transfer to the receiver. We observed desirable properties of the agent which are the following :

– Do not deliver all information related to the context at once.
– Do not deliver the information at once so as not to overload the user information.
– Provide the most useful information to the listener.

In a VE, these desirable properties for the NLG also have links to the philosophical background of cooperative principle of conversation that is "Make your conversational contribution such as is required... by the accepted purpose or direction of the talk exchange in which you are engaged" [10]. In order to achieve these desirable properties, the agent uses different strategies for the generation, which are based on the following three axes named as "communication axes" :

– Novelty : indicates that the agent is talking about some concept for the first time.
– Particularity : focuses on properties of the concept that makes it singular.
– Usefulness : defines which behavior of the concept should be known to perform the activity or to achieve their goal.

The generation of the NL dialogue is performed in three phases. In the first phase, the agent reasons about whether he has to answer a question or to ask for some information for the processing of dialogues. The second phase uses the semantic VE as knowledge base, and the belief base of the agent to formalize sentences by introspecting into the model of the VE using the services provided by MASCARET [2, 6]. Based on the three communication axes, the agent chooses the strategy for deciding what should be generated and thus, generates the structure of the sentence in the form of templates. The third phase associates grammatical constructs with the given template values, based on their associated linguistic features e.g, linguistic type noun, verb, and on specific attributes such as gender, transitive verb etc. We use english grammar structure for the generation of dialogues.

The agent determines "what should be generated" by following the information-retrieval protocol :

```
If the Agent is processing the concept for the first time
     focus on Novelty
     focus on Usefulness
else if the Agent has already talked about the concept
     Focus on particularity
else if (! explained(Detailed description))
    generate summary or explanation of concept
```

This protocol governs the generation of dialogues with different levels of description. It reduces the redundancy in the generated dialogues and provides the minimal set of dialogues based on the levels of details needed in the current context of the conversation.

The agent utters different levels of information suitable in different conditions and context. As

the information differs at the class level and at instance level, agent should decide which element of the model should be taken into account for NLG. In the following sections we describe the generation of dialogues from (few) elements of conceptual model (class diagram).

5.1 Generating from Class

It is possible to generate the description of the class using different sequence of patterns. One of the possible patterns is, first class describes generalization relationship, following the class attributes, then operations of the classe are described and then other specific associations can be generated.

```
if concept class has a parent class
     <Class name > <is-a> <Parent Class Name>
For each relavent attribute
     <Class name> <has> <attribute name>
For each relavent operation
     Generate from Operations
For each relavent operation
     Generate from Associatons
```

The agent should describe only those operations of the concept class which are supposed to be used during the activity. The concept class can contain other operations which are not necessarily used by that in the activity, and can be treated as skills of that class.

5.2 Generating from Operation

Operations define behaviors of the owner concept. If the operation name is defined as a transitive verb, one of the parameter of that operation should be defined as direct object for that operation and should add the linguistic element preposition with it. The template for generation is :

```
If Operations is defined by transitive verb
  Class name><Has><Operation name>< Object:preposition>
else
  <Class name> <Has> <Operation name>
```

5.3 Generating from Association

The agent can interpret the association through the help of the associated stereotype that determine the nature of the association, participation ends which specify the thematic role played by the typed instances, and the participation which specify the state of the action for typed instances. The template for the generation from typed instance1 to instance2 is :

```
<q1><inst1>{assoEnd1}<partRole2><q2>{assoEnd2}<inst2>
```

Here, inst1 and inst2 are the typed instances, assoEnd1 and assoEnd2 are the association ends owned by corresponding instances. partRole2 indicates the participation of instance2, and q1 q2 represent the participation cardinality of corresponding typed instances.

6 Discussion and future aspects

We applied the approach to the design of an educational agent for the CORVETTE project [1], which proposes solutions for the cost-efficient development of VE for the learning of maintenance procedures. The agent can provide explanations about the actions learners have to perform. Because the information come right from the conceptual model, it reduces the induced additional cost. The heuristics we proposed allow to generate utterances, not limited to restrictive topics, but with richer content in a progressive manner.

A virtual agent should be able to generate different levels of information suitable to the current context of the conversation in the VE. These should be consistent with the Grice's conversation maxims [10]. Now, one of the research question is "how the Grice's maxims can be corelated with the communication axes (novelty, particularity and usefulness) ?" Also, the generated utterances should be short in length, with a small number of information. We plan to conduct an evaluation to verify up to what extent our approach succeeds to satisfy these maxims. Also we want to analyze what should be the maximum length of a utterance and how much information should be conveyed in a task oriented collaborative VE.

Finally, our ongoing work is focusing on the generation of the NL dialogues from the activity diagram that represents behaviors of agents and entities, and also describes the scenario of the application supposed to be realized by agents in the VE. Our aim is to identify patterns in the scenario that explicitly need that agents should participate in a dialogue conversation in order to achieve the goal in the current context to the task.

7 Conclusion

In this paper we presented a model-based approach for the NLG in collaborative VE. We

1. COllaboRative Virtual Environment for Technical Training and Experiment, http ://corvette.irisa.fr/

proposed a set of heuristics as guideline rules for domain experts to efficiently design conceptual models of VE. To be able to do this, we extended UML meta-model by adding new semantical and linguistic elements. Also, we extended the notion of UML association to provide clear and unambiguous interpretation of relationships between concepts in the VE. The rich-semantic conceptual model is then served as knowledge base for the agent to generate NL dialogue using a model-based approach. We believe that disambiguating the semantics of associations is itself a valuable contribution. Although, our approach imposes an additional burden on domain experts by requiring the definition of linguistic properties of model elements and the use of controlled NL, but the extra information provided is used by the conversational agents to generate grammatically correct NL dialogues.

Acknowledgments

This work was partly supported by the Agence Nationale de la Recherche, (Corvette project ANR-10-CORD-012).

Références

[1] Russell J. Abbott. Program design by informal english descriptions. *Commun. ACM*, 26(11) :882–894, November 1983.

[2] Mukesh Barange, Pierre De Loor, Vincent Louis, Ronan Querrec, Julien Soler, Thanh-Hai Trinh, Eric Maisel, and Pierre Chevaillier. Get involved in an interactive virtual tour of brest harbour : Follow the guide and participate. In *IVA - 11th Int Conf, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proc*, volume 6895 of *LNCS*, pages 93–99. Springer, 2011.

[3] Håkan Burden and Rogardt Heldal. Natural language generation from class diagrams. In *Proc. of the 8th International Workshop on Model-Driven Engineering, Verification and Validation*, NY, USA, 2011. ACM.

[4] Jordi Cabot, Raquel Pau, and Ruth Raventós. From uml/ocl to sbvr specifications : A challenging transformation. *Inf. Syst.*, 35(4) :417–440, 2010.

[5] Peter P. Chen. Entity-relationship diagrams and english sentence structure. In *Proc. of the 1st International Conference on the Entity-Relationship Approach to Systems Analysis and Design*, pages 13–14, The Netherlands, 1980.

[6] Pierre Chevaillier, Thanh-Hai Trinh, Mukesh Barange, Frédéric Devillers, Julien Soler, Pierre De Loor, and Ronan Querrec. Semantic modelling of virtual environments using MASCARET. In *Proc. of SEARIS, in conjunction with IEEE VR 2011*, Singapore, 20 March 2011.

[7] Alistair A.R. Cockburn. Using natural language as a metaphoric base for object-oriented modeling and programming. Technical report, IBM Technical Report TR-36.00021, May 1, 1992.

[8] Joerg Evermann. A cognitive semantics for the association construct. *Requirements Engineering*, 13 :167–186, 2008.

[9] Gonzalo Génova, Juan Llorens, and Paloma Martinez. Semantics of the minimum multiplicity in ternary associations in uml. In *Proc. of the 4th International Conference on The UML, Modeling Languages, Concepts, and Tools*, pages 329–341, London, UK, 2001.

[10] H.P. Grice. *Logic and Conversation*. William James lectures. Harvard Univ., 1970.

[11] Sven Hartmann and Sebastian Link. English sentence structures and eer modeling. In *Proc. of the 4th Asia-Pacific conference on Comceptual modelling - Volume 67*, APCCM '07, pages 27–35, Darlinghurst, Australia, 2007.

[12] Staffan Larsson and David R. Traum. Information state and dialogue management in the trindi dialogue move engine toolkit. *Nat. Lang. Eng.*, 6 :323–340, 2000.

[13] Farid Meziane, Nikos Athanasakis, and Sophia Ananiadou. Generating natural language specifications from uml class diagrams. *Requir. Eng.*, 13(1) :1–18, January 2008.

[14] Dragan Milicev. On the semantics of associations and association ends in uml. *IEEE Trans. Softw. Eng.*, 33(4) :238–251, 2007.

[15] Bram Pellens, Frederic Kleinermann, and Olga De Troyer. Intuitively specifying object dynamics in virtual environments using VR-WISE. In *Proc. of the ACM symposium on VRST'06*, 2006.

[16] Motoshi Saeki, Hisayuki Horai, and Hajime Enomoto. Software development process from natural language specification. In *Proc. of the 11th international conf on Software engineering*, pages 64–73, NY, USA, 1989.

[17] Vede C. Storey. Understanding semantic relationships. *The VLDB Journal*, 2 :455–488, 1993.

[18] David R. Traum, Michael Fleischman, and Eduard Hovy. Nl generation for virtual humans in a complex social environment. In *Working Notes AAAI Spring Symposium on NLG in Spoken and Written Dialogue*, 2003.

[19] R. S. Wallace. *Be Your Own Botmaster*. ALICE A.I. Foundation Inc., 2003.

[20] Morton E. Winston, Roger Chaffin, and Douglas Herrmann. A taxonomy of part-whole relations. *Cognitive Science*, 11(4) :417–444, 1987.

Modeling relational reactions in conversational topics

J-P. Sansonnet
jps@limsi.fr

LIMSI-CNRS BP 133 F-91403 Orsay cedex France

Abstract:
Everyday a lot of subjects of interest (here called topics) are published on the Web and compete to reach an audience and to achieve their communicative goal. We propose an approach to these issues, based on two main propositions: First, each topic is personified by a conversational agent capable of interacting with users about the topic; Second, agent's reactions to users' requests involve traditional rational reactions but also relational reactions taking into account social roles, judgments and feelings.

Keywords: Conversational assistant agents, Personification of topics, Rational and relational reactions.

1 Introduction

Topics Following the development of the Internet culture, general public can now access a plethora of computational entities through Web applications and services. These entities, which we will refer to as *topics* in this article, are data structures (technical, encyclopedic, statistical *etc.*) They appear in particular interactional situations where a so-called *publisher* addresses a targeted audience of so-called *viewers* in order to support a communicative goal (didactic, institutional, social, entertaining, advertising *etc.*).
Because of their abundance on the Internet, topics are in competition in order to: 1) reach their audience (enticement issue) [9]; 2) not to be discarded immediately (ergonomic acceptability issue) [20]; 3) taken seriously (believability issue) [8]; 4) be understood as the publisher considers proper (communicative goal issue) [4].
Unfortunately most topics are published through static pages, often buried in the Web (enticement issue); other topics have a more complex structure but discourage users by their technicalities (acceptability issue); besides the communicative goal, mainly relying on believability, requires that users appropriate the topic before they understand it and subscribe to it, for example while experimenting and/or debating about it.

Proposition To face those issues, we put forward the notion of *conversational topic*, which is based on two main propositions:
1. *Personification of topics*: To each topic is associated a conversational assistant agent [11] in charge of its dialogical mediation with viewers. Roughly speaking, a user can interact with the entity topic/agent "as if it were a person", thus following the current trend to personify technological artifacts, as described by Reeves and Nass [16].
2. *Relational reactions*: typical conversational assistant agents are based on the concept of rational agent [1] and as such, they exhibit *rational reactions* when prompted by users. Rational answers focus on problem solving and don't take into account: a) the social context (publisher and viewers roles, current activity *etc.*) or b) the pragmatic context (publisher and viewers judgments and affects). Consequently they are often criticized for their lack of naturalness and human-likeness by users and poorly rated *wrt* above mentioned issues [21]. This is the reason why we propose to study agents capable of producing *relational reactions i.e.* taking also into account social roles, judgments and feelings.

Outline of the paper The extensive part of the paper is dedicated to the model of conversational topic in Section 2 and the model of relational reaction in Section 3. In Section 4 are sketched considerations about actual implementation, further experiments and related works.

2 Topic model

2.1 General architecture

The general architecture of an application involving a conversational topic is sketched in Figure 1. It is composed of two main parts:
1. A typical application encompasses a publisher $\mathcal{P}$ (an institution or a person) wishing to communicate about a particular subject of interest with a specific set of viewers $\mathcal{V}_i$ (one or more persons, concerned with this subject). The publisher endorses a social position *wrt* the viewers that entails some knowledge about how viewers feel about the subject; moreover the publisher has a communicative intention *wrt* the viewers, mainly expressed in terms of influence over viewers' feelings about the subject. This application-dependent information is described by the publisher into a so-called topic model $\mathcal{T}$,

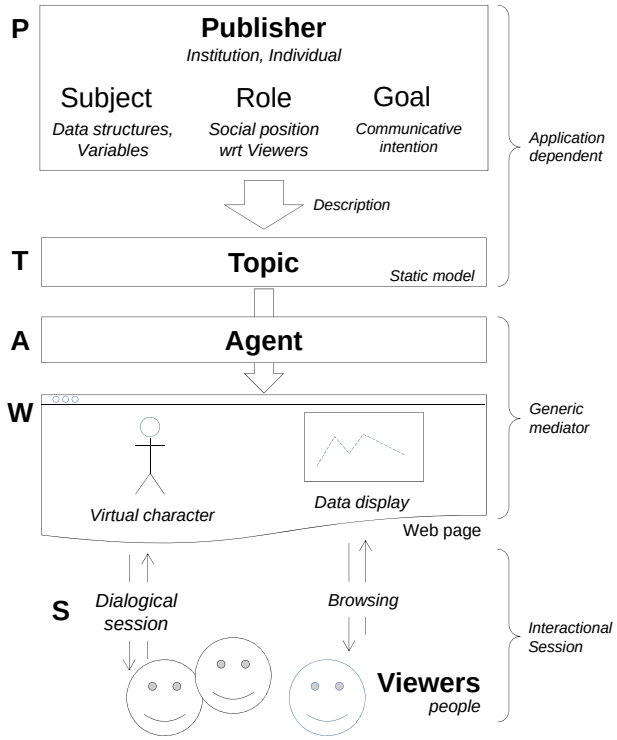


Figure 1: A conversational topic application

which is a static symbolic representation [18].
2. The agent $\mathcal{A}$ is a generic software tool in charge of the dynamic mediation of the topic *wrt* viewers through: a) a Web page $\mathcal{W}$ where data is displayed by the agent (*e.g.* plots as in Figure 2) and browsed by the viewers; and b) a dialogical session $\mathcal{S}$ where viewers can interact in a multimodal way (textual answers, embodied animations, display of pictures, plots *etc.*) with the agent about the topic.

2.2 Topic structure

Syntax. Elements of $\mathcal{T}$ are implemented as a tree where nodes names $n \in \mathbb{N}$ and where leaves are valued with expressions in typical description language $v \in \mathcal{L}$. Three notions are useful:
– a path p starting from tree-root refers to a sub tree, called a *structure* (*e.g.* $p =$ T.viewer.profile.preferences);
– in $p.n_i$, n_i are called *attributes* of structure p;
– a valued path, noted $p \rightarrow value$ where $value \in \mathcal{L}$ is called a *feature* (*e.g.* T.viewer.profile.preferences.color $\rightarrow$ blue).

Definition 1. Variable. A particular class of structure that can be found in a topic is a *variable*, defined by the following attributes:
p.id $\rightarrow label$ short name of the path: $label \equiv p$
p.type $\rightarrow$ var
p.now $\rightarrow v_{t_{now}}$ current value
p.past $\rightarrow [v_{t_0}, v_{t_1} \ldots v_{t_{now-1}}]$ recorded values
p.future $\rightarrow [v_{t_{now+1}}, v_{t_{now+2}}, ..]$ predicted values
etc.

Intuitively, a variable represents an information about the evolution over a time unit of the values of a topic feature. We distinguish past values that are actually recorded (or synthesized from recorded data) from future values that are predicted from past ones or expected, depending on an external source. Present now is defined as the moment of the current turn in session $\mathcal{S}$ (see Definition 6). When no predictions or expectations are available p.future $= \emptyset$.

Given the description of a variable x in T, it is possible for the agent to exploit the information contained in T.x for building a visual version. Typically, it would display on screen a plot showing the evolution of x values, thus prompting the notion of *trend*.

Definition 2. Trends of a variable. Trends of a variable x are defined as a function $trend_x(interval) : \mathcal{I} \longmapsto \mathcal{E}$ where:
– $\mathcal{I}$ is a set of intervals upon the values of x; intervals are often disjoint and often upon time axis. For example, *wrt* variable profit, we can set $\mathcal{I} = \{past, now, future\}$, with $past \Leftrightarrow$ past, *resp.* $future$ and $now = [t_{now-1}, t_{now}]$.
– $\mathcal{E}$ is a set of symbols, representing classes of profiles of evolution for x. Typically, one can define three main classes $\mathcal{E} = \{\nearrow, \longrightarrow, \searrow\}$ where:
$\nearrow$ means that x increases quite monotonously.
$\longrightarrow$ means that x is quite stable.
$\searrow$ means that x decreases quite monotonously.

The computation of trend classes can be achieved through an interpolation over the considered interval. For example, if a simple linear function is used trends can be displayed as lines in the variable plot.

Example 1. Fireworks profits. Let be company FireWorks that endorses the role of publisher about a variable associated with its benefits:

T.name$\rightarrow$ "FireWorks"
T.finance.benefits.id$\rightarrow$ profit
T.profit.label$\rightarrow$ "FireWorks profits"
T.profit.type$\rightarrow$ var
T.profit.now$\rightarrow$ 140
T.profit.past$\rightarrow$ $[123, 130, 145]$
T.profit.future$\rightarrow$ $[135, 120, 125]$
T.profit.future.source$\rightarrow$ "Audit by Standard and Rich"
T.profit.time.unit$\rightarrow$ year
T.profit.value.unit$\rightarrow$ M$

Figure 2: Plot of FireWorks profits.

Given this information, the agent can display on a Web page a plot of variable profit as showed in Figure 2 where trends have been been computed with linear interpolation over past (plain line) and future (dashed line).

2.3 Social relationships

With above definitions, it is possible to build a typical conversational assistant agent (as in previous works [17]), which is capable of rational reasoning about the topic model in order to display information and to answer users' request about it. However, in a conversational topic the function of assistance is carried out in a social context that involves psychological relationships between the publisher and target viewers. In the following, we will consider three notions, defined below: publisher/viewers roles, viewers affects and publisher appraisal.

Definition 3. Roles. A role $\rho(P, V) \in \Re$ describes the application-dependent social relation of the publisher *wrt* a typical viewer. For example, it can be expressed as a position in a two-dimensional space, based on two basic relations upon a set of people $\mathbb{P}$: *familiarity* [12] and *dominance* [2], which are often used in studies about social relationships.
–$\forall x, y \in \mathbb{P}$; fam$(x, y)$ is a relation meaning that x feels affectively close to y (the relation is not symmetric because y can not reciprocate the same feeling);
– $\forall x, y \in \mathbb{P}$; dom$(x, y)$ is an antisymmetric relation meaning that x has some kind of authority (institutional or moral) upon y.
For each dimension, a discrete scale with three values $\{-1, 0, 1\}$ is used, hence resulting in a 9-position space.

Definition 4. Affects about trends. This notion represents the application-dependent [1] impact of the trends of a variable x in $\mathcal{T}$ upon the mental states of the viewers. It is expressed as a function attached to the topic model:
T.x.affect $\rightarrow affect_x(trend) : \mathcal{E} \longmapsto \Phi$
where $\mathcal{E}$ is defined in Definition 2.

A restricted model of affects is sufficient to develop the case-study developed in Section 3. Hence affects $\varphi \in \Phi = \{-, =, +\}$ are represented by a symbolic value over a three-position scale: negative feelings (noted -), *resp.* neutral feelings (=), positive feelings (+). where position negative feelings (-) stands for a class of feelings encompassing: boredom, sorrow, anger, disgust, regret, fear ... *resp.* (+) contains antonym concepts; and (=) contains for example: unprejudiced, dispassionate, calm, *etc.*

Definition 5. Appraisal by the publisher. This notion represents the application-dependent impact of viewers affects upon the mental states of the publisher. It is expressed as a function attached to the topic model:
T.x.appraisal $\rightarrow appraisal(\varphi) : \Phi \longmapsto \Psi$

Similarly, we use a restricted model where appraisals $\psi \in \Psi = \{\ominus, \odot, \oplus\}$ are represented by a symbolic value over a three-position scale, corresponding to the fact that an affect φ is considered desirable *wrt* the publisher (denoted $\oplus$) or not desirable ($\ominus$) or indifferent ($\odot$).

3 Relational reactions

3.1 Dialogical session

Given a topic at hand $\mathcal{T}$, the agent $\mathcal{A}$ is in charge of the dialogical session $\mathcal{S}$ between the viewers and the topic. We will restrict $\mathcal{S}$ to the particular case of 'conversation for information' with simple turns composed of:
1) an informational request [2] (further called a *question*) from a viewer;
2) the reaction of the agent to the question.

Definition 6. Turn. A session $\mathcal{S}$ is defined by a sequence of turns. A turn t_i is composed of a question q_i of the viewer to the agent followed by a reaction r_i of the agent to the viewer.

1. For example, a shareholder is happier when stock shares are increasing whereas a tenant is annoyed when housing rentals are increasing; a doctor is happier when patient temperature is stable *etc.*

2. Such requests exclude operational commands *i.e.* the viewer can't ask the agent to perform an operation altering the structure of $\mathcal{T}$.

$\mathcal{S} = [[t_{i=1,n}]]$ such that $t_i = \langle q_i, r_i \rangle$

In the following, we focus on a specific kind of questions where the viewer asks the agent the current value of a variable of id x within $\mathcal{T}$. Such questions are of the form: $q = Ask\ T.x$ (for example: "Tell me about FireWorks profits").

In this context, we distinguish two main types of agent behaviors *wrt* their reactions to viewers' questions:
1) Rational reaction: the agent performs the rational resolution process of the question: typically, it tries to found an information in $\mathcal{T}$) and reports the result of this execution to the viewer in a non committal manner.
2) Relational reaction: the agent takes into account two things: a) the rational resolution process of the question and b) the psychological impact of this process over the viewer. Note that a) and b) can be inter mixed, hence the rational resolution process can be altered by b) and the reporting of the result of a) can be altered by b).

Definition 7. Rational reaction. A rational reaction of the agent to $q = Ask\ T.x$ is composed of two phases:

1. $\mathbb{Q}(T.x) \Rightarrow \mathsf{rep}[a_i\ v_i]$
2. $\mathbb{E}(T.x, \mathsf{rep}[a_i\ v_i]) \Rightarrow r_i[$ multimodal answer $]$.

where
– $\mathbb{Q}$ is a symbolic query sent by the agent upon the topic model. It returns a report $\mathsf{rep}[a_i\ v_i]$ as a vector of pairs: attribute symbol (a_i), associated value (v_i). Typically, the first attribute of a report is RES with value $\in \{found, \neg found\}$ depending on x was found or not in $\mathcal{T}$.
– $\mathbb{E}$ is the symbolic expression by the agent of the report elements *wrt* the viewer, often multimodal and often using natural language.

Example 2. A rational turn where the viewer asks about profits ($= x$) of FireWorks ($= \mathcal{T}$)

`Viewer`: "Tell me about FireWorks profits"

$$\mathbb{Q}(T.\mathsf{profit}) \Rightarrow \mathsf{rep} \begin{bmatrix} RES & found \\ TYPE & \mathsf{var} \\ VAL & 140 \\ VAL_{time} & now \\ VAL_{unit} & M\$ \end{bmatrix}$$

$\mathbb{E} \Rightarrow$ `Agent`: "Profit is now 140 M$"
(modalities used for $\mathbb{E}$: plain text)

Definition 8. Relational reaction. A relational reaction of the agent to $q = Ask\ T.x$ is composed of five phases:

1. $\mathbb{Q}(T.x) \Rightarrow \mathsf{rep}[a_i\ \ v_i]$
2. $\mathbb{F}(T.x, \mathsf{rep}[a_i\ \ v_i] \Rightarrow \varphi \in \Phi \Rightarrow \psi \in \Psi$
3. $\mathbb{S}(T.x, \psi, \varphi, \rho) \Rightarrow \sigma \in \Sigma. \quad (\rho \in \Re)$
4. $\mathbb{A}(T.x, \mathsf{rep}[a_i\ \ v_i], \sigma) \Rightarrow \mathsf{arg}[\alpha_i]$
5. $\mathbb{E}(T.x, \mathsf{arg}[\alpha_i] \Rightarrow r_i[$ multimodal answer $]$.

where
– $\mathbb{Q}$ the querying phase, is similar to that in rational reaction.
– $\mathbb{F}$ is the computation by the agent of: 1) the affect φ (triggered feeling) in the viewer's state of mind by the rational expression of the report and 2) the appraisal ψ (evaluation of desirability) by the publisher of φ.
– $\mathbb{S}$ is the computation by the agent of a strategy σ for transmitting the information contained in the report obtained in phase 1. The strategy is selected from viewer's feelings, publisher appraisal and also takes into account the social role $\rho \in \Re$, linking the viewer and the publisher.
– $\mathbb{A}$ is the computation by the agent of a set of so-called formal arguments $\mathsf{arg}[\alpha_i]$ to be transmitted to the viewer. For simplicity here, syntax of arg will be similar to rep, that is $\alpha_i \equiv a_i\ v_i$.
– $\mathbb{E}$ is similar to that in rational reaction except it works with arg elements.

In the following sections, phases $\mathbb{F}$, $\mathbb{S}$, $\mathbb{A}$, $\mathbb{E}$ are detailed and illustrated with *Fireworks* example.

3.2 Feelings phase $\mathbb{F}$

During phase $\mathbb{F}$, agent $\mathcal{A}$ performs a look-ahead operation involving two sub phases:

1) *Affect on viewer* The agent retrieves T.x.trend and T.x.affect functions to assess the feeling φ of the viewer's about rep by evaluating
$$\varphi = affect_x \circ trend_x(VAL_{time})\ .$$
2) *Appraisal from publisher* The agent retrieves T.x.appraisal function and computes the desirability ψ of φ *wrt* the publisher by evaluating
$$\psi = appraisal_x(\varphi)\ .$$

Example 3. Let $\mathcal{T}$ be *Fireworks* company as in example 1 with publisher $\mathcal{P}$ its *manager* and viewer $\mathcal{V}$ a typical *shareholder*, thus prompting:

$$\mathsf{T.profit.trend} \rightarrow \begin{cases} past & \nearrow \\ now & \searrow \\ future & \searrow \end{cases}$$

$$\mathsf{T.profit.affect} \rightarrow \begin{cases} \nearrow & + \\ \longrightarrow & = \\ \searrow & - \end{cases} \quad \mathsf{T.profit.appraisal} \rightarrow \begin{cases} + & \oplus \\ = & \odot \\ - & \ominus \end{cases}$$

In the report of phase $\mathbb{Q}$ built in example 2 $VAL_{time} = now$, which entails $\varphi = -$ and $\psi = \ominus$. This means: "the manager will not be satisfied with the expected displeasure of a shareholder receiving the rational information $VAR = 140$". Here, publisher's appraisal mirrors viewer's feelings (see a counterexample in footnote 1).

Table 1: Main strategy operators.

code	Operator	Description
Stating: Operators over report rep[..]		
Neu	neutral	arguments mirror a rational expression of rep[..]
Emb	embellish	arguments provide an embellished version of rep[..]
Dim	diminish	arguments provide a belittled version of rep[..]
Vag	vague	arguments contain vague information about rep[..]
Hid	hide	arguments do not convey all information in rep[..]
Opinion: Operators over viewer's feelings φ		
Agr	agree	builds an argument making explicit that P agrees with viewer's φ
Dis	disagree	builds an argument making explicit that P disagrees with viewer's φ
$\emptyset$	empty	no opinion step is expressed
Justification: Operators over trends of report's variable x		
Def	defend	arguments are synthesized (mainly from x's trends) in favor of x
Cri	criticize	argument are synthesized (mainly from x's trends) in disfavor of x
$\emptyset$	empty	no justification step is expressed

3.3 Strategy phase $\mathbb{S}$

Given a report rep[..] built by phase $\mathbb{Q}$, it can be expressed in a direct, non committal, manner as defined in a rational reaction. However, it can also be expressed in a relational manner, according to various *strategies* that take into account social and psychological relationships.

Definition 9. Strategy. A strategy for the relational expression of a query report, $\sigma \in \Sigma$, is composed of three sequential steps:
$\sigma = \langle \sigma_S; \sigma_O; \sigma_J \rangle$
where:
1. Stating step σ_S defines an operator that controls how the report will be stated to viewers;
2. Opinion step σ_O defines an operator that controls how the publisher expresses its opinion (based on ψ) upon the expected feelings of the viewer (based on φ) about the report;
3. Justification step σ_J defines an operator that controls how the publisher justifies its opinion expressed in step σ_O.
For each step, a set of operators is listed in Table 1 with their general description.

Heuristic 1. Selection of σ. For each step σ_{step}, it is possible to define a *generic* heuristic $\mathcal{H}_{step}$ that selects a proper σ operator. It is represented as a decision tree taking into account, in that order: publisher's appraisal ψ, viewer's affect φ and the Fam and Dom dimensions of publisher/viewer social role $\rho(P, V)$. An example of decision trees is given in Figure 3.

Example 4. Let $\mathcal{T}$ = *Fireworks* with roles defined as in example 3. Then one can state a stereotypical[3] relation $msh(P, V) \in \Re$ over scale $FAM \times DOM$ located at {0,1}.
Applying decisions trees in Figure 3 to example 3 results in: $\mathbb{S}(\text{T.profit}, -, \ominus, msh) \Rightarrow$

1 Stating $\sigma_S = \text{Vag/FAM}$ FAM = 0
2 Opinion $\sigma_O = \text{Dis/FAM}$
3 Justification $\sigma_J = \text{Def/FAM}$

Intuitively, the agent will first try to express the report as vaguely as possible; then will say that it disagrees with the expected feelings of the viewer about the report; and will finally add and argument in defense of the variable profit.

3.4 Argument phase $\mathbb{A}$

In phase $\mathbb{A}$, σ operators elicited in $\mathbb{S}$ are applied in order to produce a formal structure arg$[\alpha_i]$, which will be expressed in a multimodal manner during the final phase $\mathbb{E}$.

Definition 10. Argument. The structure arg is composed of three sub structures, associated with each step in σ, where each element is in turn composed of specific sub structures:
$\text{arg} = \langle \alpha_S, \alpha_O, \alpha_J \rangle$
$\alpha_S = \langle \alpha_{now}, \alpha_{past}, \alpha_{future} \rangle$
$\alpha_O = \langle \alpha_{opinion}, \alpha_{Oemote} \rangle$
$\alpha_J = \langle \alpha_{claim}, \alpha_{Jemote} \rangle$
where any structure or sub structure α_i can be $\emptyset$. In the following, elements of $\alpha_{S,O,J}$ are illustrated through the definition of the three operators occurring in steps 1 to 3 of example 4, namely: Vag, Dis and Def.

Definition 11. Operator Vague. Operator Vag is used to control how arguments α_i in step α_S are generated. Intuitively, Vag tries not to answer to a request about the current value (now) of a topic variable by giving its value but instead it synthesizes a generalization over the trends of the variable Operator Vag is defined by the following table, mapping trend cases over:
– $\alpha_{now} = \emptyset$ for all trend cases;
– α_{past} assessment of stability of now *wrt* past trend;
– α_{future} assessment of durability of α_{past} (mind not of now) according to future trend.

3. For this example we use the cliché: the manager is neutrally familiar with the shareholders ($x = 0$) and dominates them ($y = 1$).

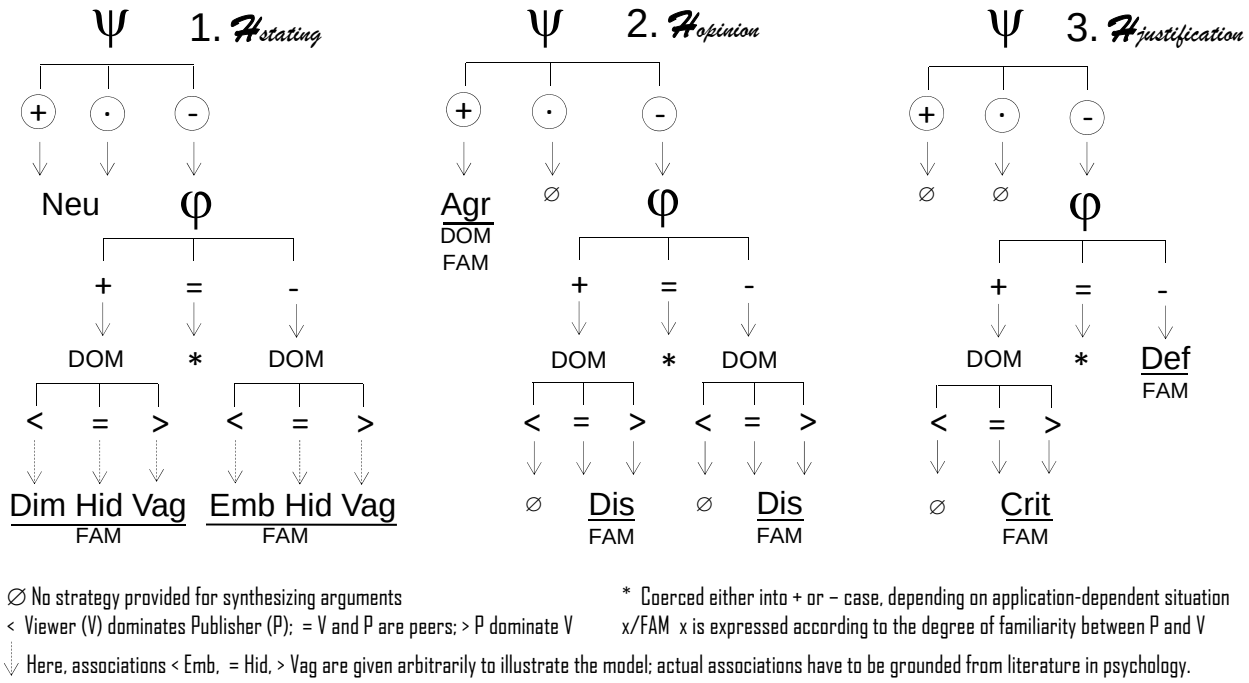


Figure 3: Decision trees eliciting operators for: $\mathcal{H}_{stating}, \mathcal{H}_{opinion}, \mathcal{H}_{justification}$

Trend cases	α_{past}	α_{future}
$past = now = future$	stable	durable
$past \simeq now \simeq future$	quite-stable	durable
$past \simeq now \neq future$	quite-stable	brief
$past \neq now \simeq future$	unstable	durable
$past \neq now \neq future$	unstable	brief

$x = y$ abridges trend(x) = trend(y); x,y $\in \mathcal{I}$ (*resp.* $\neq$);
$x \simeq y$ abridges trend(x) = $\longrightarrow \vee$ trend(y) = $\longrightarrow$;

Definition 12. Operator Disagree. Operator Dis is used to control how arguments α_O are generated. Intuitively, Dis expresses publisher's disagreement about the expected feelings of the viewer once the report expressed. Operator Dis is defined by the following table, mapping viewer's feelings over:
– $\alpha_{opinion}$ disagreement according to φ cases;
– α_{Oemote} emote expressing ψ while disagreeing.

φ	$\alpha_{opinion}$	α_{Oemote}
+	disagree-with-positive-feeling	surprised
=	disagree-with-indifference	aloof
-	disagree-with-negative-feeling	soothing

Definition 13. Operator Defend. Operator Def is used to control how arguments α_J are generated. When Def is elicited in $\mathcal{H}_{stating}$. It means that the current value of a topic variable (now) is not satisfactory for the viewer, whatever its value ($\nearrow \longrightarrow \searrow$). Hence operator Def tries to synthesize arguments in favor of the topic variable that are elicited from past and future trends. Similarly to Vag, Def is defined by the following table, mapping trend cases over:
– α_{claim} is a claim in favor of the variable;
– α_{Jemote} is an emote expressing publisher's feelings while expressing the claim.

Cases $(p, n, f \vee p, n, f)$	α_{claim}	α_{Jemote}
$\searrow\searrow\searrow \vee \nearrow\nearrow\nearrow$	always	sorry
$\searrow\searrow\overset{\uparrow}{\rightarrow} \vee \nearrow\nearrow\overset{\downarrow}{\rightarrow}$	future-better	happy
$\nearrow\searrow\searrow \vee \searrow\nearrow\nearrow$	past-ok	hoping
$\nearrow\searrow\overset{\uparrow}{\rightarrow} \vee \searrow\nearrow\overset{\downarrow}{\rightarrow}$	brief-nok	assured

3.5 Expression phase $\mathbb{E}$

During $\mathbb{E}$ the agent takes the structure arg[α_i] produced in phase $\mathbb{A}$ and expresses its arguments in a multimodal way, which is system-dependent that is according to the capabilities and modalities offered by output devices:
— Arguments are handled in sequence, according to each step: Stating, Opinion and Justification, which can be expressed in two ways that are not exclusive: a) Textual: arguments are transcribed into a natural language form; b) Illustrative: arguments are illustrated with additional information visualized on screen (pictures, plots *etc.*).
— Emotes α_{Oemote} or α_{Jemote} can be expressed in two ways that are not exclusive: a) The emote is applied to its associated textual argument by adorning the text (with extra words, stressors, emoticons *etc.*) a) The emote is played by a virtual character displayed on screen.

Example 5. Considering strategies produced in example 4 and arguments produced by operators Vag, Dis, Def we have:

Stating	σ_S = Vag/0	unstable	durable
Opinion	σ_O = Dis/0	disagree-with-negative-feeling	soothing
Justification	σ_J = Def/0	past-ok	hoping

With FAM = 0, all arguments will be expressed in a neutral manner [4] (*i.e.* neither cold nor warm). Considering again request in example 2:
`Viewer`: " *Tell me about FireWorks profits* "
We state the following system-dependent output capabilities: usage of textual modality only; one utterance by step; emotes expressed through added words at the end of the sentence. With these conditions, the arguments computed in example 5 are expressed by the agent as (textual part only):

`Agent`: " *Profits of FireWorks undergo some instability now, which could be durable. However you shouldn't take it too badly, set your mind at rest. Actually, profits have been quite good recently, so there's still hope.* "

To be contrasted with rational reaction in ex. 2:

`Agent`: " *Profit is now 140 M$* "

4 Discussion

Implementation An instance of the proposed architecture has been implemented with the support of two software tools dedicated to research experimentations about conversational assistant agents in the Internet: The DIVA*lite* [5] [19] and *lite*TALK [6] toolkits. The Web page of the FireWorks experiment [7] is available online.
Figure 4 displays the Web page of an experimentation where subjects can interact with agent *Jean* endorsing the role of manager of $FireWorks$ company. Jean performs a relational reaction to a viewer question about profits while using four modalities: 1) textual answer in balloon over head; 2) emote "SORRY" played (not captured in Figure 4); 3) then deictic gestural movement (arm pointing to left) ; 4) display of additional data (the plot of past trend).

4. Dimension Fam of a role ρ is not actually handled in the decision tree; rather it is used in phase $\mathbb{E}$ to add a modality upon the expression of an operator by the agent. For example, Dis/FAM means that the agent expresses a disagreement according to the position on the familiarity dimension of the social role encompassing the publisher and the viewer.
5. http://perso.limsi.fr/jps/online/divalitewebsite/divalite.site.html
6. —/online/litetalkwebsite/litetalk.site.main.html
7. —/online/divalite/demos/demo_fireworks/fireworks.main.html

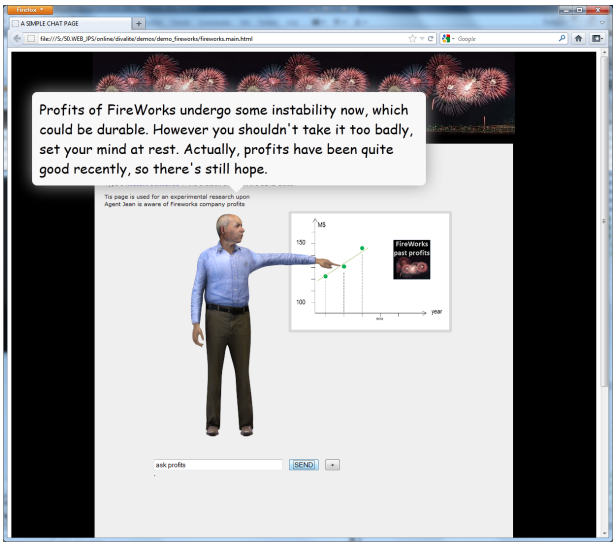


Figure 4: Web page for FireWorks experiment

Experimentation The first version of the implementation makes it possible to carry out two straightforward experiments involving subjects interacting with topic *FireWorks*:
1) users' acceptability: This factor evaluates to what extent users are ready to accept features proposed in new interfaces [14, 6]. In our context, a key issue is the additional richness of information brought by the handling in answers of the social relationships between producers and viewers. In other words, do users prefer short, straight rational answers or rich (*i.e.* multimodal), possibly vague and/or verbose, relational answers?
2) publisher's persuasiveness: This factor evaluates to what extent users are persuaded by the arguments provided by agents within an interface [5]. In our context, a key issue is the evaluation of the impact of a relational answers produced by the publisher over the viewer's feelings.
While these experiments are currently in progress, preliminary tests suggest that: a) viewers prefer *short* textual answers (they are bored with long sentences produced by relational reactions) but they like display of extra data (explicative plots); b) viewers perceive and categorize better 'negative' factors (dominance, disagreeableness, *etc.*) than positive ones.

Related works The first inspiration of this work is the notion of conversational assistant agent, first introduced by Maes [11]. It can be viewed as a mix from two main sources: a) assistant agents stem from works in in artificial intelligence, and focus on the concept of rational agent [1]; b) embodied conversational

agents [3], which are virtual characters that are given a psychological personality [10] and are supposed to be interacting with people in multimodal sessions, according to their character: social role, personality traits, affects *etc.*
The second inspiration of this work is a notion that we have termed "conversational topic" whose roots can be tracked from several sources: a) "computers as social objects"of Nass et al. [13]; entailing b) "the persuasive computer"community [7]; also *wrt* affects c) "affective computing"of Picard [15].

5 Conclusion and further work

In order to enhance enticement and acceptability factors of subjects of interest deployed in the Internet, we have proposed two forward-looking notions: conversational topics and relational reactions. Though the paper focuses on modeling, the implementation of a case-study has been achieved, enabling further evaluation.
Also some restrictive assumptions have been stated, which can be addressed in future works. For example, the topic model could be extended to handle more complex or multiple variables (*e.g.* making it possible to defend x weak trend with y strong trend); session turns could be extended to actual dialogues; relational reactions could take into account individual factors (*e.g.* publisher's personality traits).
Finally, this framework is system-independent and could be implemented in various conversational architectures and environments.

References

[1] Michael E. Bratman. What is intention? In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, Intentions in Communication, pages 15–32. The MIT Press, Cambridge, MA, 1990.

[2] J. K. Burgoon, M. L. Johnson, and P. T. Koch. The nature and measurement of interpersonal dominance. Communication Monographs, 65(4):308–335, 1998.

[3] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. Embodied Conversational Agents. MIT Press, 2000.

[4] J. Coiro. Reading Comprehension on the Internet: Expanding Our Understanding of Reading Comprehension to Encompass New Literacies. The Reading Teacher, 56(5):458–465, 2003.

[5] Chris Creed. Using computational agents to motivate diet change. In Persuasive Technology, pages 100–103. 2006.

[6] Andrew Dillon and Michael G. Morris. User acceptance of new information technology: theories and models. Annual Review of Information Science and Technology, 14(4), 1996.

[7] B. J. Fogg. Persuasive technologie. Commun. ACM, 42(5):26–29, May 1999.

[8] B. J. Fogg, Cathy Soohoo, David R. Danielson, Leslie Marable, Julianne Stanford, and Ellen R. Tauber. How do users evaluate the credibility of web sites?: a study with over 2,500 participants. In Proc. of the 2003 conference on Designing for user experiences, DUX '03, pages 1–15, New York, NY, USA, 2003. ACM.

[9] D. Galon. The Savvy Way to Successful Website Promotion: Secrets of Successful Websites, Attracting On-Line Traffic, the Most Up to Date Guide to Top Positioning on Search Engines. Trafford on Demand Pub, 1999.

[10] Oliver P. John, Richard W. Robins, and Lawrence A. Pervin, editors. Handbook of Personality: Theory and Research. The Guilford Press, 3rd edition, 2008.

[11] P. Maes. Agents that reduce work and information overload. Commun. ACM, 37(7):30–40, 1994.

[12] Luhmann N. Familiarity, confidence, trust: problems and alternatives. In D. G Gambetta, editor, Trust: Making and Breaking Cooperative Relations, pages 94–107. University of Oxford Press, 2000.

[13] C. Nass, Y. Moon, E. Kim, and B. J. Fogg. Computers are social actors: A review of current research. In B. Friedman, editor, Moral and Ethical Issues in HCI, pages 137–162. Cambridge University Press and CSLI Publications, 1997.

[14] J. Nielsen. Usability Engineering. Academic Press, Boston, 1993.

[15] R. W. Picard. Affective computing. MIT Press, 2000.

[16] Byron Reeves and Clifford Nass. The media equation: how people treat computers, television, and new media like real people and places. Cambridge university press edition, 1996.

[17] J. P. Sansonnet and F. Bouchet. Joint handling of rational and behavioral reactions in assistant conversational agents. In Helder Coehlo, Rudi Studer, and Michael Wooldridge, editors, Proc. of the 19th European Conference on Artificial Intelligence (ECAI 2010), pages 1049–1050, Lisbon, Portugal, 2010. IOS Press.

[18] J. P. Sansonnet, M. Castan, and C. Percebois. M3L: A list-directed architecture. In ACM-IEEE International Symposium on Computer Architecture ISCA-7, page 8, La Baule, France, 1980.

[19] J. P. Sansonnet, D. W. Correa, P. Jaques, A. Braffort, and C. Verrecchia. Developing web fully-integrated conversational assistant agents. In Research in Applied Computation Symposium, RACS2012, San Antonio, USA, 2012. ACM-SIGAPP.

[20] Yu-Lung Wu, Yu-Hui Tao, and Pei-Chi Yang. The discussion on influence of website usability towards user acceptability. In International Conf. on Management and Service Science (MASS'09)., pages 1–4, 2009.

[21] M. Xuetao, F. Bouchet, and J. P. Sansonnet. Impact of agent's answers variability on its believability and human-likeness and consequent chatbot improvements. In Proc. of AISB 2009, pages 31–36, Edinburgh, Scotland, 2009.

Honte ou culpabilité ? (*That is the question*)

C. Adam♣ carole.adam@imag.fr

D. Longin♠ Dominique.Longin@irit.fr

♣Laboratoire d'Informatique de Grenoble
Université Joseph Fourier, équipe MAGMA
Maison J. Kuntzmann, 110 avenue de la Chimie, 38400 Saint-Martin-d'Hères

♠Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier, équipe LILaC
118 route de Narbonne, 31062 Toulouse cedex 9

Résumé :
Sous ce titre quelque peu humoristique se cache une vraie question : qu'est-ce qui différencie la honte de la culpabilité ? Cette question est d'autant plus importante que de nombreuses études en psychologie montrent qu'on a bien souvent tendance à les assimiler et à nommer l'une pour l'autre. Mais l'intérêt d'une telle étude réside également dans le fait qu'en caractérisant une notion par ce qu'elle est, on caractérise l'autre par ce qu'elle n'est pas. À terme, notre objectif est de fournir une typologie des émotions où chacune se définirait selon ces deux types de (non)propriétés, donnant une cohérence nouvelle à des classifications bien souvent très subjectives. On se propose de formaliser ces deux émotions dans un même langage de type logique modale des états mentaux.

Mots-clés : émotions, honte, culpabilité, logique modale.

1 Introduction

De nombreuses études auprès de sujets humains montrent que honte et culpabilité sont souvent confondues. (Nous montrerons que ce n'est pas sans raison.) Il est intéressant d'étudier ces émotions pour elles-mêmes mais également d'un point de vue méthodologique.

Pour elles-mêmes, car ce sont des émotions morales, en relation avec les normes qu'un individu se fixe. Elster [9, p. 145] souligne combien les normes sociales ont une influence immensément puissante sur le comportement (« *an immensely powerful influence on behavior* »). En particulier, la honte nous touche dans ce que nous avons de plus intime, de plus personnel. Comme le notent Tangney et Ronda [25], c'est une émotion qui a une influence certaine sur l'image que nous avons de nous-mêmes et sur la manière dont on pense être socialement perçu. C'est donc une émotion clé de notre comportement, notamment en situation de prise de décision, qui constitue pour Elster *le* support des normes sociales. [1]

D'un point de vue méthodologique ensuite, car il est fréquent de constater dans la littérature l'existence de deux définitions différentes pour une même émotion (c'est le cas par exemple de l'espoir dans [21] et [14]) ou l'inverse (c'est le cas de la honte et de la culpabilité). Étudier une émotion en mettant en relief ce qu'elle a en commun avec une autre, et ce qui l'en différencie permet en définitive de mieux cerner chacune d'elles.

Enfin, ces deux émotions jouent aussi un rôle prépondérant en situation de décision stratégique : que va penser l'autre de moi si je fais telle ou telle chose ? Et l'autre justement, quel comportement doit-il adopter ? Que faire si un tuteur intelligent détecte que son élève a honte de parler anglais car il pense mal parler ? Ne devrait-il pas mettre en place quelque stratégie visant à le rassurer ?

Dans ce qui suit, nous présentons brièvement ce que nous appelons une (structure cognitive d')émotion (Section 2). Après une analyse de la littérature (Section 3) nous présentons le cadre formel (Section 4) propre à les caractériser (Section 5).

2 Qu'est-ce qu'une émotion ?

« *What is an emotion ?* » est un article de William James [13], l'un des fondateurs de la psychologie expérimentale, dans lequel l'auteur tente de présenter pour la première fois de manière moderne et rigoureuse ce qu'est une émotion. Dans la lignée de Scherer nous adoptons une vision multi-componentielle de l'émotion : *le sentiment* (le ressenti de l'émotion) ; *la réponse psychophysiologique* (accélération

1. Il prend l'exemple suivant : si un agent viole une norme sociale, nous allons refuser de traiter avec lui, ce qui va peut-être engendrer chez lui une perte matérielle quelconque, mais qui va surtout marquer notre mépris ou notre dégoût et engendrer chez lui de la honte. Et plus il nous coûtera de refuser de traiter avec lui, plus sa honte sera importante [9, p. 146].

du rythme cardiaque, de la température corporelle, *etc*.) ; *l'expression motrice* (du visage, de la voix, des gestes) ; *la tendance à l'action* (à ne pas confondre avec l'action elle-même) ; *l'évaluation cognitive* (*appraisal* en anglais). Dans les théories de l'évaluation cognitive cette dernière composante est vue comme le déclencheur des quatre autres ; elle représente le processus cognitif d'évaluation d'un certain événement qui déclenche une réponse émotionnelle différenciée, c'est-à-dire qui détermine si c'est une émotion qui est déclenchée plutôt qu'une autre, les autres composantes n'étant alors que des sortes de canaux de manifestation dans notre corps et notre esprit de l'émotion déclenchée. Cette différenciation serait rendue possible par le fait que l'on évalue (consciemment ou non) un stimulus donné par rapport à notre état mental (incluant nos préférences, buts, idéaux, et connaissances acquises au cours d'expériences passées). Ainsi, une émotion correspond alors à une *variation épisodique* de certaines de ces composantes suite à l'évaluation d'un événement donné [22].

Dans ce suit, nous nous intéressons aux émotions et non aux humeurs. Une distinction communément admise est que les premières, contrairement aux secondes, sont toujours *à propos de quelque chose* : on sera déçu *de voir son équipe préférée perdre* mais jamais *triste en général* (ce qui correspond plutôt à une humeur de tristesse). Le sentiment de l'émotion transparaît dans le fait que notre agent est introspectif (conscient) de ses émotions. Ce n'est bien sûr qu'une description partielle du phénomène : il est aussi lié à une notion d'intensité de l'émotion non traitée ici pour ne pas compliquer le formalisme (bien que des solutions techniques existent [17]). Comme dans la littérature, nous considérons l'évaluation cognitive comme la (non) congruence entre une croyance de l'agent (conséquence d'une observation ou d'un raisonnement) et ses buts/désirs ou ses idéaux (selon l'émotion considérée), ce qui est en tout point conforme avec les théories psychologiques de l'émotion. En définitive, nous formalisons dans ce qui suit ce qui correspond davantage à la structure cognitive de l'émotion, *i.e*. l'état mental nécessaire à son déclenchement.

3 La honte *versus* la culpabilité

D'après [4], des sociétés orientales comme le Japon ou la Chine ont une « culture de la honte » alors que nos sociétés occidentales ont une « culture de la culpabilité » (au sens de *se sentir coupable*). Selon cet auteur, dans les cultures de la culpabilité on restreint le comportement des individus en les rendant coupables. Dans celles de la honte, les conséquences sociales d'un acte rendu public et considéré comme honteux sont bien plus importantes et déterminantes que les sentiments individuels ; ce sont des cultures où les rangs sociaux ont une importance capitale dans l'organisation et la vie de tous les jours. L'image que dégage une personne la définit, c'est pour cela que les individus y sont particulièrement sensibles, et qu'un acte rendu public qui ternit leur image est si terrible pour eux.

La honte et la culpabilité ont été largement étudiées en psychologie [26, 24, 25, 14, 21]. Pourtant, ces émotions ont bien souvent été assimilées, ou peu différenciées l'une de l'autre (voir par exemple [25, p. 11–12] pour plus de détails). La principale raison est que l'évaluation de ces deux émotions est basée sur la violation d'une norme sociale par un comportement inapproprié par rapport à une société donnée (voir par exemple [27, 14, 9, p. 145]). Ainsi, Ortony *et al.* [21, p. 142–143] voient ces deux émotions comme le fait que l'agent qui les éprouve s'attribue la responsabilité [2] de la violation de l'un de ses propres idéaux (d'où le fait qu'ils classent ces deux émotions comme des *attribution emotions*). La seule chose qui les différencie est l'importance de l'idéal en question, dont la violation est jugée comme inexcusable dans le cas de la honte, et non dans le cas de la culpabilité (qui serait principalement composée à partir de la honte et du regret).

C'est très certainement à H. B. Lewis [15] que l'on doit d'avoir trouvé un critère discriminant (par la suite vérifié expérimentalement dans un nombre très important de travaux en psychologie) : lorsqu'un individu éprouve de la honte, c'est lui-même qu'il juge, sa propre personne dans son ensemble ; dans le cas de la culpabilité, ce sont ses actions. De même Elster [9, p. 143–144] définit la honte comme une émotion négative déclenchée par une croyance à propos de sa personne, et la culpabilité comme une émotion négative déclenchée par une croyance à propos de ses actions [3]. (Voir aussi [8, 14, 25] par exemple.)

Cette distinction explique en particulier pour-

2. Il s'agit ici d'une responsabilité causale, *i.e*. au sens où l'agent vient d'accomplir une action ayant causé l'état présent.

3. En ce sens, l'agent est causalement responsable de par ses actions de la situation présente où un de ses idéaux est violé.

quoi la honte se ressent bien plus profondément que la culpabilité, pourquoi elle est bien plus douloureuse, et pourquoi il est beaucoup plus difficile de lutter contre elle. Elle explique aussi par conséquent pourquoi la honte conduit à vouloir systématiquement chercher à ce que l'objet de notre honte ne s'ébruite pas [21], à tenter de minimiser son exposition aux autres agents. Lazarus note que dans les cas extrêmes, on se sent incapable de vivre en société selon les normes établies, d'atteindre « l'ego idéal » [14, 20] ce qui peut conduire au suicide [10, p. 274]. Plusieurs méthodes sont possibles comme nier tout lien avec la transgression ou insister sur la nature privée des évènements [20]. Dans la cas de la culpabilité, on a plutôt tendance à adopter un comportement actif et réparateur [9, 20] dans le but de minimiser ou effacer les conséquences de notre action. Un corollaire à cela est que dans le cas de la culpabilité on se sent nécessairement responsable de la situation présente (sinon on ne pourrait pas se sentir coupable) alors que dans le cas de la honte toute responsabilité, quand elle est réelle, est non assumée [20]. Elster [9, p. 150], citant en cela [27], indique que la honte peut avoir une cause indépendante de notre bonne volonté, comme avoir des parents pauvres ou devenir vieux.

Dans la culpabilité comme dans la honte, les idéaux mis en jeu ont été internalisés : se sentir coupable de s'être garé sur une place pour personnes handicapées par exemple, c'est se reconnaître dans le fait qu'il est mal de se garer sur de telles places si on n'est pas handicapé ; on considère ce principe comme devant être respecté. Si au contraire on a connaissance de ce principe mais qu'il ne nous paraît pas important de le respecter (idéal non internalisé) alors on pourra me garer sur une telle place sans pour autant se sentir coupable. (Voir [14, p. 240] par exemple.)

Il a été dit que la honte inclut nécessairement une dimension sociale, publique [9, 27, 20, 21, 8], ce qui ne serait pas le cas de la culpabilité. Elster [9, p. 149] par exemple dit que « je ressens de la honte en votre présence parce que je sais que vous me désapprouvez ». Mais des expérimentations [26] montrent que la honte ressentie en dehors de tout groupe témoin est au contraire légèrement plus fréquente que pour la culpabilité. Les auteurs citent l'exemple d'un adulte racontant que lorsqu'il était enfant, il a vu son frère se faire réprimander par leur mère pour avoir fait quelque chose d'immoral. Lui-même avait fait la même chose mais sa mère l'ignorait. Pourtant il a ressenti de la honte (dont l'objet n'est pas précisé par l'adulte). Ce qui est important n'est donc pas tant que l'objet de notre honte soit connu d'un certain groupe, mais plutôt qu'on *croie* que cela constitue une violation d'ordre moral vis-à-vis de ce groupe. Darwin abonde en ce sens lorsqu'il dit qu'un individu peut éprouver de la honte mais ne pas rougir pour autant ; pour rougir, il faut que l'objet de sa honte ait été découvert [6, p. 352]. Ainsi, le groupe face auquel on éprouve de la honte n'a ni besoin d'être participatif ou physiquement présent [27], ni même d'être au courant de la violation de la norme en question. Lazarus [14, p. 241] souligne même qu'on peut éprouver de la honte vis-à-vis d'une personne décédée. S'il y a une dimension sociale dans la honte et la culpabilité, elle se situe au niveau du groupe d'individus par rapport auquel on se projette (soit-même dans le cas de la honte, ou ses action dans le cas de la culpabilité).

En résumé. 1) Nous nous intéressons aux émotions et non aux humeurs, les notions de honte et de culpabilité que nous capturons sont donc nécessairement à propos de quelque chose.

2) La honte comme la culpabilité nécessitent la violation d'un idéal internalisé.

3) Dans la culpabilité, l'agent pense être causalement responsable de la situation présente (il se focalise sur ses actions) alors que dans la honte, il ne le pense pas (à tort ou à raison) et refuse d'assumer cette responsabilité ou cherche à la minimiser ; c'est lui-même en tant qu'individu qu'il juge dans son ensemble.

4) Ni la honte ni la culpabilité ne requièrent la présence d'une audience témoin de la violation, ni même qu'elle soit au courant des faits. Il suffit qu'on imagine qu'elle le soit.

Exemple 1 (extrait de [9]). La Princesse de Clèves se sent coupable de l'amour qu'elle éprouve pour le Duc de Nemours mais Mathilde de la Mole a honte d'être amoureuse de Julien Sorel : en trompant son mari la princesse de Clèves accomplit une action qui va à l'encontre de ses idéaux mais bien que Mathilde de la Mole puisse se sentir coupable pour les mêmes raisons, elle a surtout honte être tombée amoureuse du fils d'un charpentier ce qui lui ferait perdre la face si des personnes de son rang l'apprenaient.

Exemple 2. Une personne a faisant du shopping dans un magasin oublie involontairement un vê-

tement sur son sac et se dirige vers la sortie du magasin, étant ainsi sur le point de commettre involontairement un vol. Si la personne r responsable du magasin demande à a d'ouvrir son sac, a ne pourra pas éprouver de la culpabilité pour une action qu'il n'a pas commise volontairement. En revanche, il est probable que a éprouvera de la honte face à cette situation car sa réputation est en jeu.

Exemple 3. Une personne perd son pantalon dans la rue. Il s'agit là de la transgression involontaire d'une norme sociale ou culturelle (on ne se promène pas en sous-vêtements dans la rue). L'individu éprouvera donc probablement de la honte d'avoir perdu son pantalon (à condition bien sûr que cette norme ait été internalisée par l'agent : dans le cas contraire c'est qu'il ne la reconnaît pas comme une norme à suivre et il ne peut alors pas éprouver de la honte). À l'inverse, si l'on suppose qu'il l'a fait volontairement (pour provoquer par exemple) il ne devrait normalement pas non plus éprouver de honte. [4]

4 Cadre formel

Notre cadre formel correspond à une extension de celui développé dans [12]. Nous limitons sa présentation à ce qui est nécessaire à la compréhension de la suite. En particulier, nous présentons les différents opérateurs utilisés, mais nous ne présenterons ni la sémantique, ni même l'axiomatique complète associée. (Pour cela, se reporter à [16].)

4.1 Langage de base et attitudes mentales

Soit AGT l'ensemble fini des agents, ATM l'ensemble des formules atomiques et $ACT = \{a_1, a_2, \ldots a_n\}$ l'ensemble fini non vide des actions atomiques. Le langage de base est défini comme suit : $\varphi :: p \mid \bot \mid \neg\varphi \mid \varphi \vee \varphi \mid Bel_i\,\varphi \mid Ideal_i\,\varphi \mid SIdeal_i\,\varphi \mid \Diamond\varphi$ où p appartient à ATM, a à ACT et i à AGT. Les autres connecteurs classiques ($\wedge$, $\rightarrow$, $\leftrightarrow$ et $\bot$) sont définis de manière usuelle.

$Bel_i\,\varphi$ se lit : « l'agent i croit que φ est vrai ». La notion de croyance est celle d'un savoir subjectif, au sens où l'agent ne doute pas que φ soit vrai mais pense au contraire que φ est vrai dans le monde réel.

$Ideal_i\,\varphi$ se lit : « φ est un état de chose idéal pour l'agent i ». Les opérateurs $Ideal_i$ sont utilisés pour représenter les attitudes morales de l'agent i. Plus généralement, le fait que $Ideal_i\,\varphi$ soit vrai signifie que i se commande (s'ordonne) à lui-même de faire en sorte que φ soit vrai (quand φ est faux) ou de faire en sorte qu'il continue de l'être (quand φ est déjà vrai) [5]. En ce sens, il est moralement responsable de la réalisation de φ. $SIdeal_i\,\varphi$ (*strong ideal*) représente le fait que φ est un idéal particulièrement important pour i et dont la violation est susceptible de lui faire perdre la face (au sens de [21, p. 142–143]). Nous imposons juste que $SIdeal_i\,\varphi \rightarrow Ideal_i\,\varphi$. Nous avons ainsi deux notions d'idéal dont l'une est plus forte que l'autre en gardant un langage simple (pas besoin d'introduire des degrés qui ne seraient qu'un raffinement supplémentaire).

$\Diamond\varphi$ se lit : « φ est vrai dans au moins un état alternatif », ou plus simplement « il est possible que φ soit vrai ». L'opérateur $\Diamond$ représente la possibilité historique, c'est-à-dire qu'il représente l'existence d'au moins un état alternatif à l'état présent si la succession d'actions qui ont été accomplies jusqu'à présent n'avait pas été celle qu'elle a été (et qui a conduit à l'état présent). Autrement dit, chaque monde accessible par la relation de possibilité historique représente un présent alternatif appartenant à une histoire parallèle, c'est-à-dire un déroulement des événements différent de celui qu'on considère comme étant l'histoire réelle. Cette construction permet ainsi de représenter un futur arborescent, où différents états peuvent être atteints selon l'action accomplie car d'un point de vue sémantique, les hypothèses sur les actions font que dans chaque état (ou monde), une et une seule action est accomplie. L'opérateur dual de nécessité historique

$$\Box\varphi \stackrel{déf}{=} \neg\Diamond\neg\varphi$$

se lit : « φ est nécessairement vrai (quel que soit l'état alternatif considéré) ». Autrement dit, φ est nécessairement vrai (quoi que les agents aient fait). Par définition, nous avons $\Box\varphi \rightarrow \varphi$.

4.2 Opérateurs temporels

En étendant le langage précédent avec deux opérateurs dynamiques (voir [16]), on peut définir $X\,\varphi$ qui se lit « next φ » et qui signifie que φ sera

4. Il suffit qu'il n'ait pas internalisé l'idéal qu'il viole. Bien sûr, dans certains cas, on peut être prêt à choquer même si on doit ensuite éprouver de la honte pour ce que l'on a fait : cela signifie juste que dans ce cas, on a attribué plus d'importance à son but de choquer qu'à son idéal (qui est alors violé).

vrai l'instant juste après (quelle que soit l'action accomplie par chacun des agents) ; et $X^{-1}\varphi$ qui se lit « next-moins-un φ » et qui signifie que φ était vrai l'instant juste avant (quelle que soit l'action accomplie par chacun des agents).

On impose que $X\Box\varphi \leftrightarrow \Box X\varphi$ ce qui entraîne que $X^{-1}\Box X\varphi \leftrightarrow \Box\varphi$ (ce qui est nécessairement vrai persiste à l'être dans le futur).

4.3 Opérateurs d'agentitude

Les opérateurs d'agentitude (*agency*) [3] servent à capturer le fait qu'un état a été causé par un agent. On trouve initialement cette notion chez Davidson qui tente de définir ce qu'est une action [7, Essay 3]. Formellement (cf. [12, 16]) $[C]\varphi$ se lit : « il existe des actions accomplies conjointement par les agents du groupe C pour lesquelles, quelles que soient les actions accomplies conjointement par les agents ne faisant pas partie de ce groupe C, φ est vrai ». Plus simplement, cette formule peut se lire : « le groupe C fait en sorte que φ soit vrai ». D'où :

$$\langle i\rangle\varphi \stackrel{déf}{=} \neg[AGT \setminus \{i\}]\varphi$$

qui se lit : « il n'est pas le cas [qu'il existe des actions accomplies conjointement par les agents autres que i pour lesquelles, quelles que soit l'action accomplie par i, φ est vrai] ». Autrement dit : « quelles que soient les actions accomplies conjointement par les agents autres que i, il existe une action accomplie par i telle que φ est faux ». En raccourci, cela signifie encore que i peut faire en sorte d'empêcher que φ soit vrai.

5 Formalisation

5.1 Formalisation de la responsabilité

Comme nous l'avons montré précédemment, la culpabilité fait intervenir une notion de responsabilité causale, alors qu'au contraire la honte la rejette. Généralement, cette responsabilité peut prendre deux formes : la responsabilité **directe** (l'agent a accompli une action qui a causé directement la situation présente comme dans « L'agent i casse une tasse ») et la responsabilité **indirecte** (l'agent aurait pu intervenir pour empêcher une situation d'arriver mais il n'a rien fait, comme dans « L'agent aurait pu empêcher la tasse de se casser (mais il n'en a rien fait) »). La première correspond trivialement à l'opérateur d'agentitude ($\varphi \wedge X^{-1}[i]X\varphi$), et nous formalisons donc dans ce qui suit la seconde (bien que ces deux formes de responsabilité puissent intervenir au niveau émotionnel). [5] Soit :

$$Resp_i\,\varphi \stackrel{déf}{=} \varphi \wedge X^{-1}\langle i\rangle X\varphi$$

qui signifie que l'agent i est responsable (indirectement) du fait que φ soit vrai si et seulement si φ est vrai et qu'à l'instant juste avant, i aurait pu empêcher que l'instant juste après φ devienne vrai. [6]

Supposons que Jean autorise François à sortir une plante et que celle-ci meure brûlée par le soleil. Au sens où nous l'avons défini, François a une responsabilité directe dans la mort de la plante et Jean une responsabilité indirecte.

5.2 Formalisation de l'inévitable

Contrairement à la culpabilité, la honte suppose plutôt qu'on nie toute responsabilité dans ce qui arrive. Formellement, cela revient à écrire que l'instant juste avant il **était** inévitable (*i.e.* indépendant des actions des agents) que l'instant juste après cette situation se produise. Soit :

$$Inevitable\varphi \stackrel{déf}{=} X^{-1}\Box X\varphi$$

qui se lit « (l'instant juste avant) il *était* inévitable que φ devienne vrai maintenant ». (Il découle des propriétés précédentes et de cette définition que $Inevitable\varphi \rightarrow \varphi$.)

5.3 Formalisation des idéaux de groupe

Culpabilité et honte peuvent être ressenties (condition non nécessaire) vis-à-vis d'un groupe. Cela signifie qu'on se projette par rapport à ce groupe dont on partage les idéaux. On peut formaliser les idéaux d'un groupe comme suit (pour tout ensemble d'agents $C \in 2^{AGT}$) :

$$Ideal_C\,\varphi \stackrel{déf}{=} \bigwedge_{i\in C} Ideal_i$$

5. En fait, la nature de la responsabilité (directe ou indirecte, intentionnelle ou non, *etc*.) joue davantage un rôle au niveau de l'intensité de l'émotion qu'au niveau sa structure cognitive. Parce que nous nous intéressons ici uniquement à cette dernière la formalisation de la responsabilité n'est en soi pas central ici.

6. Dans [19, 2] et (avec un langage différent) dans [11] la responsabilité est définie par $\varphi \wedge \langle i\rangle\varphi$, ce qui ne nous semble pas intuitif car φ devrait être vrai *après* l'accomplissement de l'action et non pendant.

Autrement dit, φ est un idéal partagé (*shared ideal*) par le groupe C si et seulement si φ est un idéal internalisé pour chacun des individus de ce groupe C. Par définition, il s'ensuit que $Ideal_{\{i\}}\,\varphi = Ideal_i\,\varphi$.

Comme la notion de responsabilité, celle de groupe fait l'objet de nombreuses études en épistémologie. C'est une notion complexe dont la formalisation entraîne nécessairement la formalisation de sa structure (ou au contraire de son absence dans le cas de groupes informels), de ses propriétés, *etc*. (voir par exemple [18] pour plus de détails) et celles-ci n'entrent pas nécessairement en ligne de compte dans la définition des émotions. Seul importe le fait qu'on considère les idéaux d'un groupe, quelle que soit la nature de ce dernier.

Enfin, de manière similaire (pour tout ensemble d'agents $C \in 2^{AGT}$) :

$$SIdeal_C\,\varphi \stackrel{déf}{=} \bigwedge_{i \in C} SIdeal_i$$

se lit : « φ est un idéal fort du groupe C ».

5.4 Formalisation de la culpabilité

Dans [19] un cadre formel complet est introduit pour la responsabilité et le regret. Dans [2], cette logique est présentée de manière didactique et d'autres émotions structurellement proches sont définies dans [11]. Par ailleurs, des travaux comme par exemple ceux d'Ortony *et al*. [21] ont déjà été formalisés en logique (voir notamment [1, 23]). Cependant, ces modèles ne définissent pas des émotions telles que la culpabilité, ou le font sans introduire cette dimension relative à ce que l'agent aurait pu faire d'autre.

Ainsi, il convient de définir la culpabilité comme le fait de ne pas avoir empêché la violation de ce qu'on considère comme une norme à respecter :

$$Guilt_i\,\varphi \stackrel{déf}{=} Bel_i\,(Ideal_i\,\neg\varphi \wedge Resp_i\,\varphi)$$

Il est important de noter que comme $Resp_i\,\varphi \rightarrow \varphi$ alors $Guilt_i\,\varphi \rightarrow Bel_i\,\varphi$ (la culpabilité implique nécessairement qu'on croit que φ est vrai –même si ce n'est pas réellement le cas). Cette définition, modulo notre définition de la responsabilité qui est légèrement différente, correspond à la notion définie par ailleurs dans [2]. Il est intéressant de noter qu'à l'inverse de la honte, on peut culpabiliser même si nous n'étions pas conscient que nous pouvions empêcher ce qui est arrivé (*i.e*. il n'est pas nécessaire que $X^{-1}Bel_i\,\langle i\rangle X\varphi$ soit vrai, à comparer avec la définition de $Resp_i\,\varphi$).

Comme nous l'avons souligné précédemment, même si c'est légèrement plus fréquent pour la honte, il arrive qu'on éprouve un sentiment de culpabilité par rapport à un groupe d'agents (non nécessairement présent ni même au courant de la violation de l'idéal). Nous pouvons donc généraliser notre définition de la manière suivante (pour tout agent $i \in C$) :

$$Guilt_i\,(\varphi, C) \stackrel{déf}{=} Bel_i\,(Ideal_C\,\neg\varphi \wedge Resp_i\,\varphi)$$

5.5 Formalisation de la honte

Là encore, la honte peut être vis-à-vis de soi-même ou d'un groupe auquel on s'identifie et dont on a internalisé les idéaux (voir l'opérateur $Ideal_C$ ci-dessus).

Nous avons également montré qu'un agent qui éprouve de la honte a tendance à nier sa responsabilité (qui peut être réelle ou non). En d'autres termes, il pense qu'au moment où la situation honteuse s'est produite, elle était inévitable.

Nous obtenons la définition suivante ($i \in C$) :

$$Shame_i\,(\varphi, C) \stackrel{déf}{=} Bel_i\,(SIdeal_C\,\neg\varphi \wedge Inevitable\varphi)$$

Ainsi, l'agent i a honte du fait que φ soit vrai vis-à-vis du groupe C (qui peut se réduire à lui-même) si et seulement si il croit que :

1. φ devrait idéalement être faux pour le groupe C et que la nature de cet idéal est telle que sa violation peut lui faire perdre la face ;
2. il était inévitable au moment où la violation de l'idéal était sur le point de se produire que φ devienne vrai ;
3. φ est actuellement vrai (puisque $Inevitable\varphi \rightarrow \varphi$).

5.6 Propriétés intéressantes

Dans le reste de cette section, nous présentons quelques propriétés intéressantes dont la démonstration peut être trouvée dans [16]. Une première propriété concerne la relation entre ce

qui est nécessairement vrai (donc, indépendant de ce que font les agents) et les actions des agents. Ainsi, selon [12], on peut montrer que pour tout groupe d'agents C donné

$$\Box\varphi \rightarrow [C]\varphi \quad (1)$$

qui se lit : « si φ est inévitable, alors n'importe quelle coalition C fait en sorte que φ soit vrai, et ce quelles que soient les actions des agents extérieurs à cette coalition ».

Théorème 1. *Pour tout agent $i \in AGT$:*

$$\langle i\rangle\varphi \rightarrow \neg\Box\varphi$$

Autrement dit, si un agent i donné peut faire en sorte d'empêcher que φ soit vrai, alors nécessairement φ n'est pas inévitable.

Théorème 2. *Pour tout agent $i \in AGT$ et coalition $C \in 2^{AGT}$:*

$$Resp_i\,\varphi \rightarrow \neg Inevitable\varphi \quad \text{(a)}$$
$$Shame_i\,(\varphi, C) \rightarrow Bel_i\,\neg Resp_i\,\varphi \quad \text{(b)}$$

Le théorème (2a) signifie que pour tout agent i donné, s'il est responsable du fait que φ soit vrai alors c'est qu'il n'était pas inévitable que φ soit vrai. (Par contraposition, nous avons également que si φ était inévitable alors l'agent i n'est pas responsable du fait que φ soit vrai.)

Enfin, le théorème (2b) signifie que, pour tout agent i donné et tout groupe C d'agents, si l'agent i éprouve de la honte vis-à-vis de C par rapport à φ alors i croit qu'il n'est pas responsable du fait que φ soit maintenant vrai. On retrouve là une propriété importante de la honte et qui a été discutée plus haut, à savoir que lorsqu'une personne éprouve de la honte elle croit qu'il n'est pas le cas que l'instant d'avant elle pouvait faire en sorte d'empêcher que φ soit vrai l'instant suivant (c'est-à-dire maintenant).

Émotions miroir. Il n'est pas possible d'aborder en profondeur cet aspect pour des raisons de place, mais il intéressant de souligner que, en tant qu'émotion sociale, il existe des *émotions miroir* qui répondent en quelque sorte à celle (honte ou culpabilité dans notre cas) éprouvée par l'agent responsable de la situation. Nous illustrons cette notion par l'exemple suivant. Supposons que $Ideal_C\,\varphi$ est vrai et que tout agent $j \in C$ croit que φ est faux (*i.e.* $\bigwedge_{j\in C} Bel_j\,\neg\varphi$). Supposons en outre que tous les agents du groupe C croient que la responsabilité du fait que φ soit faux est à imputer à un agent i quelconque. Cette incongruence entre attitude morale et attitudes épistémiques correspond traditionnellement au reproche ou à la désapprobation morale à propos de $\neg\varphi$ (voir [11, p. 6] par exemple). Supposons maintenant que $i \in C$. Il en découle que i éprouve de la culpabilité à propos de $\neg\varphi$ (qui peut être vue comme de la réprobation morale envers lui-même). Autrement dit, ces deux émotions (désapprobation et culpabilité) se différencient par le fait que nous sommes d'un côté du miroir ou de l'autre (on est responsable, ou c'est quelqu'un d'autre qui l'est), mais que dans tous les cas on regarde la même chose (la valeur de φ par rapport à ses idéaux).

6 Conclusion

Nous avons montré et formalisé les différences essentielles entre la honte et la culpabilité. Certaines notions (comme la responsabilité ou l'idéal d'un groupe) ont été formalisées de manière volontairement simple car c'est le concept qui est important, plus que sa forme particulière dans la situation considérée.

Nous avons rappelé combien ces émotions étaient importantes au niveau social : exprimer du regret lorsqu'on a fait quelque chose qui a violé une norme, ou faire attention à ne pas mettre quelqu'un dans une situation où il éprouverait de la honte sont des comportements qui trouveront, selon nous, une place naturelle au sein des agents conversationnels en ayant un rôle central sur leurs décisions en situation d'action.

Cette étude préliminaire mérite bien sûr d'être affinée, notamment en prenant en compte les tendances à l'action (ce qu'un individu est tenté de faire lorsqu'il éprouve une telle émotion). Cela permettrait ainsi de prendre en compte une autre composante (au sens de Sherer, cf. Section 2) de l'émotion. Il convient également de caractériser plus finement les différents idéaux mis en jeu.

Remerciements

Ce travail a été soutenu par le contrat de recherche CECIL (Complex Emotions in Communication, Interaction and Language) No. ANR-08-CORD-005 obtenu auprès de l'ANR suite à l'appel à projet ContInt 2008. Site web du projet : `www.irit.fr/CECIL/`.

Références

[1] Carole Adam. *Emotions : from psychological theories to logical formalization and implementation in a BDI agent*. PhD thesis, INP Toulouse, France, July 2007.

[2] Carole Adam, Benoit Gaudou, Dominique Longin, and Emiliano Lorini. Logical modeling of emotions for Ambient Intelligence. In Fulvio Mastrogio and Nak-Young Chong, editors, *Handbook of Research on Ambient Intelligence and Smart Environments : Trends and Perspectives*. IGI Global, 2011.

[3] N. Belnap, M. Perloff, and M. Xu. *Facing the future : agents and choices in our indeterminist world*. Oxford University Press, New York, 2001.

[4] R. Benedict. *The chrysanthemum and the sword*. Mariner Books, 1946.

[5] H. N. Castaneda. *Thinking and Doing*. D. Reidel, Dordrecht, 1975.

[6] Charles R. Darwin. *The expression of emotions in man and animals*. Murray, London, 1872.

[7] D. Davidson. *Essay on Actions and Events*. Oxford University Press, Oxford, 2nd edition, 2001.

[8] Ramon Martinez de Pison. *Death by Despair : Shame And Suicide*. Peter Lang Pub Inc, 2006.

[9] Jon Elster. *Alchemies of the Mind : Rationality and the Emotions*. Cambridge University Press, Cambridge, 1999.

[10] N. H. Frijda. *The Emotions*. Cambridge University Press, 1986.

[11] Nadine Guiraud, Dominique Longin, Emiliano Lorini, Sylvie Pesty, and Jérémy Rivière. The face of emotions : a logical formalization of expressive speech acts (regular paper). In *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1031–1038. ACM, 2011.

[12] A. Herzig and E. Lorini. A dynamic logic of agency I : STIT, capabilities, and powers. *Journal of Logic, Language, and Information*, 19 :89–121, 2009.

[13] W. James. What is an emotion ? *Mind*, 9 :188–205, 1884.

[14] Richard S. Lazarus. *Emotion and Adaptation*. Oxford University Press, 1991.

[15] Helen Block Lewis. *Shame and guilt in neurosis*. International Universities Press, New-York, 1971.

[16] Dominique Longin. La honte versus la culpabilité : analyse conceptuelle et formelle en logique modale. Rapport de recherche IRIT/RR–2012-12–FR, IRIT, Université Paul Sabatier, Toulouse, mai 2012. 27 pages.

[17] Emiliano Lorini. A Dynamic Logic of Knowledge, Graded Beliefs and Graded Goals and Its Application to Emotion Modelling. In H. van Ditmarsch, J. Lang, and S. Ju, editors, *Proceedings of the LORI-III*, volume 6953 of *LNAI*, pages 165–178. Springer-Verlag, 2011.

[18] Emiliano Lorini, Dominique Longin, Benoit Gaudou, and Andreas Herzig. The logic of acceptance : grounding institutions on agents' attitudes. *Journal of Logic and Computation*, 19(6) :901–940, 2009.

[19] Emiliano Lorini and François Schwarzentruber. A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, 175(3-4) :814–847, 2011.

[20] M. Miceli and C. Castelfranchi. How to silence one's conscience : Cognitive defenses against the feeling of guilt. *Journal for the Theory of Social Behaviour*, 28 :287–318, 1998.

[21] Andrew Ortony, G.L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge University Press, 1988.

[22] D. Sander and K. Scherer, editors. *Traité de psychologie des émotions*. Cognitive. Dunod, 2009.

[23] B.R. Steunebrink, M. Dastani, and J.-J. Meyer. The OCC model revisited. In D. Reichardt, editor, *Proc. of the 4th Workshop on Emotion and Computing*, 2009.

[24] June Price Tangney. The self-conscious emotions : shame, guilt, embarrassment and pride. In *Handbook of Cognition and Emotion*. John Wiley & Sons, 1999.

[25] J. P. Tangney and R. L. Dearin. *Shame and Guilt*. The Guilford Press, 2002.

[26] J. P. Tangney, R. S. Miller, L. Flicker, and D. H. Barlow. Are shame, guilt, and embarrassment distinct emotions ? *Journal of Personality and Social Psychology*, 70(6) :1256–1269, 1996.

[27] Gabrielle Taylor. *Pride, Shame, and Guilt : Emotions of Self-Assessment*. Oxford University Press, New-York, 1985.

A logic of emotions: from appraisal to coping*

Mehdi Dastani
Utrecht University, The Netherlands
mehdi@cs.uu.nl

Emiliano Lorini
IRIT-CNRS, Toulouse, France
lorini@irit.fr

ABSTRACT

Emotion is a cognitive mechanism that directs an agent's thoughts and attention to what is relevant, important, and significant. Such a mechanism is crucial for the design of resource-bounded agents that must operate in highly-dynamic, semi-predictable environments and which need mechanisms for allocating their computational resources efficiently. The aim of this work is to propose a logical analysis of emotions and their influences on an agent's behaviour. We focus on four emotion types (*viz*., hope, fear, joy, and distress) and provide their logical characterizations in a modal logic framework. As the intensity of emotion is essential for its influence on an agent's behaviour, the logic is devised to represent and reason about graded beliefs, graded goals and intentions. The belief strength and the goal strength determine the intensity of emotions. Emotions trigger different types of coping strategy which are aimed at dealing with emotions either by forming or revising an intention to act in the world, or by changing the agent's interpretation of the situation (by changing beliefs or goals).

1. INTRODUCTION

Autonomous software agents are assumed to have different (possibly conflicting) objectives, able to sense their environments, update their states accordingly, and decide which actions to perform at any moment in time. The behaviour of such software agents can be effective and practical only if they are able to continuously and adequately assess their (sensed) situation and update their states with relevant information and crucial objectives. For example, a robot with a plan to transport a container to its target position may perceive it has low battery charge. The robot may assess the state of its battery charge as being relevant for the objective of having the container at its target position, and update its state by suspending the current battery-demanding transport plan. Such assessment and update may cause the agent to decide to charge its battery right away or to focus on a less battery-demanding task.

Emotion is a (cognitive) mechanism that directs one's thoughts and attention to what is relevant, important, and significant in order to ensure effective behaviour. The aim of this work is to propose a logical analysis of the relationships between emotion and cognition. An understanding of these relationships is particularly important in the perspective of the design of resource-bounded agents that must survive in highly-dynamic, semi-predictable environments and which need mechanisms for allocating their computational resources efficiently. Indeed, as it has been stressed by several authors in psychology and economics, emotions provide heuristics for preventing excessive evaluation and deliberation by pruning of search spaces [5] and for interrupting normal cognition when unattended goals require servicing [20], and signals for belief revision [15].

Our approach is inspired by the appraisal and coping models of human emotions [16, 11, 8]. According to these models, an agent continuously appraises its situation (*e.g*., low battery charge endangers the objective of having a container at its target position) after which emotions can be triggered (*e.g*., fear of failing to place the container at its target position). The triggered emotions can affect the agent's behaviour depending on their intensities. There are often a set of strategies that can be used to cope with a specific emotion, for example, by updating the agent's mental state (*e.g*., being fearful that the transportation plan will not place the container at its target position leads the agent to reconsider its plan).

We first propose, in Section 2, a dynamic logic with special operators which allow to represent the intentions of a cognitive agent as well as its beliefs and goals with their corresponding strengths. Then, in Section 3, we provide a logical analysis of the intensity of different emotions such as hope, fear, joy and distress. In Section 4, we extend the logic with special operators to formally characterize different kinds of coping strategies which are aimed at dealing with emotions either by forming or revising an intention to act in the world, or by changing the agent's interpretation of the situation (by changing belief strength or goal strength). A complete axiomatization and a decidability result for the logic are given in Section 5. Related works are discussed in Section 6.

2. LOGICAL FRAMEWORK

This section presents the syntax and the semantics of the logic DL-GA (*Dynamic Logic of Graded Attitudes*). This logic is designed to represent beliefs, goals, and intentions, where beliefs and goals have degree of plausibility and desirability, respectively.

2.1 Syntax

Assume a finite set of atomic propositions describing facts $Atm = \{p, q, \ldots\}$, a finite set of physical actions (*i.e*., actions modifying the physical world) $PAct = \{a, b, \ldots\}$, a finite set of natural numbers $Num = \{x \in \mathbb{N} : 0 \leq x \leq \mathsf{max}\}$, with $\mathsf{max} \in \mathbb{N} \setminus \{0\}$. We note $Num^- = \{-x : x \in Num \setminus \{0\}\}$ the corresponding set of negative integers. We note Lit the set of literals and $l, l', \ldots$ the elements of Lit. Finally, we define the set of propositional formulas $Prop$ as the set of all Boolean combinations of atomic propositions.

The language $\mathcal{L}$ of DL-GA is defined by the following grammar in Backus-Naur Form (BNF):

$$\begin{array}{lcll} Act & : & \alpha & ::= a \mid *\varphi \\ Fml & : & \varphi & ::= p \mid \mathsf{exc}_{\mathsf{h}} \mid \mathsf{Des}^{\mathsf{k}} l \mid \mathsf{Int}_a \mid \\ & & & \quad \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{K}\varphi \mid [\alpha]\varphi \end{array}$$

where p ranges over Atm, h ranges over Num, l ranges over Lit, k ranges over $Num \cup Num^-$, and a ranges over $PAct$. The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined in the standard way.

*This paper has been published in the proceedings of the AAMAS 2012 conference.

The set of actions *Act* includes both physical actions and sensing actions of the form $*\varphi$. A sensing action is an action which consists in modifying the agent's beliefs in the light of a new incoming evidence. In particular, $*\varphi$ is the mental action (or process) of learning that φ is true. As we will show in Section 2.3, technically this amounts to an operation of belief conditioning in Spohn's sense [21].

The set of formulas *Fml* contains special constructions exc_h, $\mathsf{Des}^\mathsf{k} l$ and Int_a which are used to represent the agent's mental state. Formulas exc_h are used to identify the degree of *plausibility* of a given world for the agent. We here use the notion of plausibility first introduced by Spohn [21]. Following Spohn's theory, the worlds that are assigned the smallest numbers are the most plausible, according to the beliefs of the individual. That is, the number h assigned to a given world rather captures the degree of *exceptionality* of this world, where the exceptionality degree of a world is nothing but the opposite of its plausibility degree (*i.e.*, the exceptionality degree of a world decreases when its plausibility degree increases). Therefore, formula exc_h can be read alternatively as "the current world has a degree of exceptionality h" or "the current world has a degree of plausibility $\mathsf{max} - \mathsf{h}$".

Formula $\mathsf{Des}^\mathsf{k} l$ has to be read "the state of affairs l has a degree of desirability k for the agent". Degree of desirability can be positive, negative or equal to zero.[1] Suppose $\mathsf{k} > 0$. Then $\mathsf{Des}^\mathsf{k} l$ means that "the agent wishes to achieve l with strength k", whereas $\mathsf{Des}^{-\mathsf{k}} l$ means that "the agent wishes to avoid l with strength k". $\mathsf{Des}^0 l$ means that "the agent is indifferent about l" (*i.e.*, the agent does not care whether l is true or false). For notational convenience, in what follows we will use the following abbreviations:

$$\mathsf{AchG}^\mathsf{k} l \stackrel{\mathtt{def}}{=} \mathsf{Des}^\mathsf{k} l \text{ for } \mathsf{k} > 0$$
$$\mathsf{AvdG}^\mathsf{k} l \stackrel{\mathtt{def}}{=} \mathsf{Des}^{-\mathsf{k}} l \text{ for } \mathsf{k} > 0$$

where AchG andAvdG respectively stand for *achievement goal* and *avoidance goal*.

Formulas Int_a capture the agent's intentions. We assume that the agent's intentions are only about physical actions and not about sensing actions. The formula Int_a has to be read "the agent has the intention to perform the physical action a" or "the agent is committed to perform the physical action a".

The logic $\mathsf{DL\text{-}GA}$ has also epistemic operators and modal operators that are used to describe the effects of a given action α. The formula $[\alpha]\varphi$ has to be read "after the occurrence of the action α, φ will be true". $\mathsf{K}\varphi$ has to be read "the agent knows that φ is true". This concept of knowledge is the standard S5-notion, partition-based and fully introspective, that is commonly used in computer science and economics [7]. The operator $\widehat{\mathsf{K}}$ is the dual of K, that is, $\widehat{\mathsf{K}}\varphi \stackrel{\mathtt{def}}{=} \neg\mathsf{K}\neg\varphi$. As we will show in the Section 2.5, the operator K captures a form of 'absolutely unrevisable belief', that is, a form of belief which is stable under belief revision with any new *evidence*.

2.2 Physical action description

Similarly to Situation Calculus [18], in our framework physical actions are described in terms of their executability preconditions and of their positive and negative effect preconditions. In particular, we define a function

$$Pre : PAct \longrightarrow Prop$$

to map physical actions to their executability preconditions. Using the notion of executability precondition, we define special dynamic operators for physical actions of the form $\langle\langle a\rangle\rangle$, where $\langle\langle a\rangle\rangle\varphi$ has to be read " the physical action a is executable and, φ will be true afterwards":

$$\langle\langle a\rangle\rangle\varphi \stackrel{\mathtt{def}}{=} Pre(a) \wedge [a]\varphi$$

Moreover, we introduce two functions

$$\gamma^+ : PAct \times Atm \longrightarrow Prop$$
$$\gamma^- : PAct \times Atm \longrightarrow Prop$$

mapping physical actions and atomic propositions to propositional formulas. The formula $\gamma^+(a, p)$ describes the *positive effect preconditions* of action a with respect to p, whereas $\gamma^-(a, p)$ describes the *negative effect preconditions* of action a with respect to p. The former represent the necessary and sufficient conditions for ensuring that p will be true after the occurrence of the physical action a, while the latter represent the necessary and sufficient conditions for ensuring that p will be false after the occurrence of the physical action a. We make the following *coherence assumption*:

(COH_γ) for every $a \in PAct$ and $p \in Atm$, $\gamma^+(a, p)$ and $\gamma^-(a, p)$ must be logically inconsistent.

It ensures that actions do not have contradictory effects.

2.3 Models and truth conditions

The semantics of the logic $\mathsf{DL\text{-}GA}$ is a possible world semantics with special functions for exceptionality and desirability.

Definition 1 (Model). *DL-GA models are tuples* $M = \langle W, \sim, \kappa_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}, \mathcal{V}\rangle$ *where:*

- W *is a nonempty set of possible worlds or states;*
- $\sim$ *is an equivalence relation between worlds in* W*;*
- $\kappa_{\mathsf{exc}} : W \longrightarrow Num$ *is a total function from the set of possible worlds to the set of natural numbers* Num*;*
- $\mathcal{D} : W \times Lit \longrightarrow Num \cup Num^-$ *is a total function from the set of possible worlds to the set of integers* $Num \cup Num^-$*;*
- $\mathcal{I} : W \longrightarrow 2^{PAct}$ *is a total function called* commitment function*, mapping worlds to sets of physical actions;*
- $\mathcal{V} : W \longrightarrow 2^{Atm}$ *is a valuation function.*

As usual, $p \in \mathcal{V}(w)$ means that proposition p is true at world w. The equivalence relation $\sim$, which is used to interpret the epistemic operator K, can be viewed as a function from W to 2^W. Therefore, we can write $\sim(w) = \{v \in W : w \sim v\}$. The set $\sim(w)$ is the agent's *information state* at world w: the set of worlds that the agent imagines at world w. As $\sim$ is an equivalence relation, if $w \sim v$ then the agent has the same information state at w and v (*i.e.*, the agent imagines the same worlds at w and v).

The function κ_{exc} represents a plausibility grading of the possible worlds and is used to interpret the atomic formulas exc_h. $\kappa_{\mathsf{exc}}(w) = \mathsf{h}$ means that, according to the agent the world w has a degree of exceptionality h or, alternatively, according to the agent the world w has a degree of plausibility $\mathsf{max} - \mathsf{h}$. (Remember that the degree of plausibility of a world is the opposite of its exceptionality degree). The function κ_{exc} allows to model the notion of belief: among the worlds the agent cannot distinguish from a given world w (*i.e.*, the agent's information state at w), there are worlds that the agent considers more plausible than others. For example, suppose that $\sim(w) = \{w, v, u\}$, $\kappa_{\mathsf{exc}}(w) = 2$, $\kappa_{\mathsf{exc}}(u) = 1$ and $\kappa_{\mathsf{exc}}(v) = 0$. This means that $\{w, v, u\}$ is the set of worlds that the agent imagines at world w. Moreover, according to the agent, the world v is strictly more plausible than the world u and the world u is strictly more plausible than the world w (as $\mathsf{max} - 0 > \mathsf{max} - 1 > \mathsf{max} - 2$).

$\mathsf{DL\text{-}GA}$ models are supposed to satisfy the following *normality* constraint for the plausibility grading to ensure that an agent's beliefs are consistent:

[1] However, we assume that exceptionality/plausibility and positive desirability are measured on the same scale *Num*.

($Norm_{\kappa_{\mathsf{exc}}}$) for every $w \in W$, there is v such that $w \sim v$ and $\kappa_{\mathsf{exc}}(v) = 0$.

The function $\mathcal{D}$ is used to interpret the atomic formulas $\mathsf{Des}^{\mathsf{k}}l$. Suppose $\mathsf{k} > 0$. Then, $\mathcal{D}(w, l) = \mathsf{k}$ means that, at world w, l has a degree of desirability k; whereas $\mathcal{D}(w, l) = -\mathsf{k}$ means that, at world w, l has a degree of desirability $-\mathsf{k}$ — or equivalently, l has a degree of undesirability k —. $\mathcal{D}(w, l) = 0$ means that the agent is indifferent about l.

The function $\mathcal{I}$ is used to interpret the atomic formulas Int_a. For every world $w \in W$, if $\mathcal{I}(w)$ is defined then $\mathcal{I}(w)$ identifies the set of physical actions that the agent intends to perform. $\mathcal{I}(w) = \emptyset$ means that the agent has no intention.

Note that in our dynamic setting an agent may be committed to perform an action even though it believes that this is a *suboptimal* choice, *i.e.*, we do not require agents to have intentions because of their desirable consequences. An agent may have an intention without desiring its consequence because, for example, its beliefs and desires may change due to a sensing action. In our running example, the robot may have the intention to transport a container to a given target position, while it believes that this is a suboptimal choice as it has just learnt that it does not have sufficient battery power to accomplish the task.

Definition 2 (Truth conditions). *Given a DL-GA model M, a world w and a formula φ, $M, w \models \varphi$ means that φ is true at world w in M. The rules defining the truth conditions of formulas are:*

- $M, w \models p$ *iff* $p \in \mathcal{V}(w)$
- $M, w \models \mathsf{exc}_{\mathsf{h}}$ *iff* $\kappa_{\mathsf{exc}}(w) = \mathsf{h}$
- $M, w \models \mathsf{Des}^{\mathsf{k}}l$ *iff* $\mathcal{D}(w, l) = \mathsf{k}$
- $M, w \models \mathsf{Int}_a$ *iff* $a \in \mathcal{I}(w)$
- $M, w \models \neg\varphi$ *iff not* $M, w \models \varphi$
- $M, w \models \varphi \wedge \psi$ *iff* $M, w \models \varphi$ *and* $M, w \models \psi$
- $M, w \models \mathsf{K}\varphi$ *iff* $M, v \models \varphi$ *for all* v *such that* $w \sim v$
- $M, w \models [\alpha]\psi$ *iff* $M^{\alpha}, w \models \psi$

where model M^{α} is defined according to Definitions 3 and 5 below.

Definition 3 (Update via physical action). *Given a DL-GA model $M = \langle W, \sim, \kappa_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}, \mathcal{V}\rangle$, The update of M by a is defined as $M^a = \langle W, \sim, \kappa_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}^a, \mathcal{V}^a\rangle$ where for all $w \in W$:*

$$\begin{array}{rcl} \mathcal{I}^a(w) & = & \mathcal{I}(w) \setminus \{a\} \\ \mathcal{V}^a(w) & = & (\mathcal{V}(w) \cup \{p : M, w \models \gamma^{+}(a, p)\}) \setminus \\ & & \{p : M, w \models \gamma^{-}(a, p)\} \end{array}$$

The performance of a physical action a makes the commitment function $\mathcal{I}$ to remove a from the set of intentions. That is, if an agent intends to perform the physical action a, then after the performance of a the agent does not intend to perform a anymore. Of course, the agent may adopt intention a again by, for example, performing an intention update operation (see Section 4.1). Physical actions modify the physical facts via the positive effect preconditions and the negative effect preconditions, defined in Section 2.2. In particular, if the positive effect preconditions of action a with respect to p holds, then p will be true after the occurrence of a; if the negative effect preconditions of action a with respect to p holds, then p will be false after the occurrence of a.[2]

[2]Note that the order of the set theoretic operations in the definition seems to privilege negative effect preconditions; however, due to the *coherence* assumption COH_{γ} made in Section 2.2 the effects of a physical action will never be inconsistent.

A sensing action updates the agent's information state by modifying the exceptionality degree of the worlds that the agent can imagine. Before defining such a model update, we follow [21] and lift the exceptionality of a possible world to the exceptionality of a formula viewed as a set of worlds.

Definition 4 (Exceptionality degree of a formula). *Let $\|\varphi\|_w = \{v \in W : M, v \models \varphi$ and $w \sim v\}$. The exceptionality degree of a formula φ at world w, noted $\kappa^w_{\mathsf{exc}}(\varphi)$, is defined as follows:*

$$\kappa^w_{\mathsf{exc}}(\varphi) \quad = \quad \begin{cases} \min\limits_{v \in \|\varphi\|_w} \kappa_{\mathsf{exc}}(v) & \text{if } \|\varphi\|_w \neq \emptyset \\ \mathsf{max} & \text{if } \|\varphi\|_w = \emptyset \end{cases}$$

As expected, the *plausibility* degree of a formula φ, noted $\kappa^w_{\mathsf{plaus}}(\varphi)$, is defined as $\mathsf{max} - \kappa^w_{\mathsf{exc}}(\varphi)$.

Definition 5 (Update via sensing action). *Given a DL-GA model $M = \langle W, \sim, \kappa_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}, \mathcal{V}\rangle$. The update of M by the sensing action $*\varphi$ is defined as $M^{*\varphi} = \langle W, \sim, \kappa^{*\varphi}_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}, \mathcal{V}\rangle$ such that for all w:*

$$\kappa^{*\varphi}_{\mathsf{exc}}(w) \quad = \quad \begin{cases} \kappa_{\mathsf{exc}}(w) - \kappa^w_{\mathsf{exc}}(\varphi) & \text{if } M, w \models \varphi \\ Cut_{\mathsf{B}}(\kappa_{\mathsf{exc}}(w) + \delta) & \text{if } M, w \models \neg\varphi \wedge \widehat{\mathsf{K}}\varphi \\ \kappa_{\mathsf{exc}}(w) & \text{if } M, w \models \mathsf{K}\neg\varphi \end{cases}$$

where $\delta \in Num \setminus \{0\}$ and

$$Cut_{\mathsf{B}}(x) \quad = \quad \begin{cases} x & \text{if } 0 \leq x \leq \mathsf{max} \\ \mathsf{max} & \text{if } x > \mathsf{max} \\ 0 & \text{if } x < 0 \end{cases}$$

The action of sensing that φ is true modifies the agent's beliefs as follows.

1. For every world w in which φ is true, the degree of exceptionality of w decreases from $\kappa_{\mathsf{exc}}(w)$ to $\kappa_{\mathsf{exc}}(w) - \kappa^w_{\mathsf{exc}}(\varphi)$, which is the same thing as saying that, degree of plausibility of w increases from $\mathsf{max} - \kappa_{\mathsf{exc}}(w)$ to $\mathsf{max} - (\kappa_{\mathsf{exc}}(w) - \kappa^w_{\mathsf{exc}}(\varphi))$.
2. For every world w in which φ is false:
   (a) if at w the agent can imagine a world in which φ is true, *i.e.* $\widehat{\mathsf{K}}\varphi$, then the degree of exceptionality of w increases from $\kappa_{\mathsf{exc}}(w)$ to $Cut_{\mathsf{max}}(\kappa_{\mathsf{exc}}(w) + \delta)$, which is the same thing as saying that, the degree of plausibility of w decreases from $\mathsf{max} - \kappa_{\mathsf{exc}}(w)$ to $\mathsf{max} - Cut_{\mathsf{max}}(\kappa_{\mathsf{exc}}(w) + \delta)$;
   (b) if at w the agent cannot imagine a world in which φ is true, *i.e.* $\mathsf{K}\neg\varphi$, then the degree of exceptionality of w does not change.

The condition 2(b) ensures that the agent's plausibility ordering over worlds does not change, if the agent learns something that he cannot imagine.[3] Cut_{B} is a minor technical device, taken from [3], which ensures that the new plausibility assignment fits into the finite set of natural numbers *Num*. The parameter δ is a *conservativeness* index which captures the agent's disposition (or personality trait) to radically change its beliefs in the light of a new evidence. More precisely, the higher is the index δ, and the higher is the agent's disposition to decrease the plausibility degree of those worlds in which the learnt fact φ is false. (When $\delta = \mathsf{max}$, the agent is minimally conservative). We assume that δ is different from 0 in

[3]Note that the tree conditions 1, 2(a) and 2(b) cover all cases. Indeed, the third condition $\mathsf{K}\neg\varphi$ is equivalent to $\neg\varphi \wedge \mathsf{K}\neg\varphi$, because $\mathsf{K}\neg\varphi \rightarrow \neg\varphi$ is valid.

order to ensure that, after learning that p is true, the agent will believe p for every proposition $p \in Atm$ (see validity (5) in Section 2.5 below).

In the sequel we write $\models_{\mathsf{DL\text{-}GA}} \varphi$ to mean that φ is *valid* in DL-GA (φ is true in all DL-GA models).

2.4 Definition of graded belief

Following [21], we define the concept of belief as a formula which is true in all worlds that are maximally plausible (or minimally exceptional).

DEFINITION 6 (BELIEF, $\mathsf{B}\varphi$). *In model M at world w the agent believes that φ is true, i.e., $M, w \models \mathsf{B}\varphi$, if and only if, for every v such that $w \sim v$, if $\kappa_{\mathsf{exc}}(v) = 0$ then $M, v \models \varphi$.*

The following concept of graded belief is taken from [10]: the strength of the belief that φ is equal to the exceptionality degree of $\neg\varphi$.

DEFINITION 7 (GRADED BELIEF, $\mathsf{B}^{\geq \mathsf{h}}\varphi$). *For all $\mathsf{h} \geq 1$, in model M at world w the agent believes that φ with strength at least h, i.e. $M, w \models \mathsf{B}^{\geq \mathsf{h}}\varphi$, if and only if, $\kappa^{w}_{\mathsf{exc}}(\neg\varphi) \geq \mathsf{h}$.*

An agent has the strong belief that φ if and only if, it believes that φ is true with maximal strength max.

DEFINITION 8 (STRONG BELIEF, $\mathsf{SB}\varphi$). *In model M at world w the agent strongly believes that φ (or at w the agent is certain that φ is true), i.e., $M, w \models \mathsf{SB}\varphi$, if and only if $\kappa^{w}_{\mathsf{exc}}(\neg\varphi) = \mathsf{max}$.*

As the following proposition highlights, the concepts of belief, graded belief and strong belief semantically defined in Definitions 6-8 are all syntactically expressible in the logic DL-GA.

PROPOSITION 1. *For every DL-GA model M, world w and $\mathsf{h} \in Num$ such that $\mathsf{h} \geq 1$:*

1. *$M, w \models \mathsf{B}\varphi$ iff $M, w \models \mathsf{K}(\mathsf{exc}_0 \rightarrow \varphi)$*
2. *$M, w \models \mathsf{B}^{\geq \mathsf{h}}\varphi$ iff $M, w \models \mathsf{K}(\mathsf{exc}_{\leq \mathsf{h}-1} \rightarrow \varphi)$*
3. *$M, w \models \mathsf{SB}\varphi$ iff $M, w \models \mathsf{K}(\mathsf{exc}_{\leq \mathsf{max}-1} \rightarrow \varphi)$*

where $\mathsf{exc}_{\leq \mathsf{k}} \stackrel{\mathtt{def}}{=} \bigvee_{0 \leq \mathsf{l} \leq \mathsf{k}} \mathsf{exc}_{\mathsf{l}}$ for all $\mathsf{k} \in Num$.

We define the dual operators in the usual way: $\widehat{\mathsf{B}}\varphi \stackrel{\mathtt{def}}{=} \neg\mathsf{B}\neg\varphi$, $\widehat{\mathsf{B}}^{\geq \mathsf{h}}\varphi \stackrel{\mathtt{def}}{=} \neg\mathsf{B}^{\geq \mathsf{h}}\neg\varphi$ and $\widehat{\mathsf{SB}}\varphi \stackrel{\mathtt{def}}{=} \neg\mathsf{SB}\neg\varphi$.

We assume that "the agent believes that φ *exactly* with strength h", *i.e.* $\mathsf{B}^{\mathsf{h}}\varphi$, if and only if the agent believes that φ with strength at least h and it is not the case that the agent believes that φ with strength at least h+1. Formally, for every $\mathsf{h} \in Num$ such that $1 \leq \mathsf{h} < \mathsf{max}$, we define:

$$\mathsf{B}^{\mathsf{h}}\varphi \stackrel{\mathtt{def}}{=} \mathsf{B}^{\geq \mathsf{h}}\varphi \wedge \neg\mathsf{B}^{\geq \mathsf{h}+1}\varphi \text{ and } \mathsf{B}^{\mathsf{max}}\varphi \stackrel{\mathtt{def}}{=} \mathsf{B}^{\geq \mathsf{max}}\varphi$$

2.5 Some properties of epistemic attitudes

The following validities highlight some interesting properties of beliefs. For every $\mathsf{h}, \mathsf{k} \in Num$ such that $\mathsf{h} \geq 1$ and $\mathsf{k} \geq 1$ we have:

$$\models_{\mathsf{DL\text{-}GA}} \mathsf{K}\varphi \rightarrow \mathsf{B}^{\geq \mathsf{h}}\varphi \quad (1)$$

$$\models_{\mathsf{DL\text{-}GA}} \mathsf{B}\varphi \leftrightarrow \mathsf{B}^{\geq 1}\varphi \quad (2)$$

$$\models_{\mathsf{DL\text{-}GA}} \mathsf{SB}\varphi \leftrightarrow \mathsf{B}^{\geq \mathsf{max}}\varphi \quad (3)$$

$$\models_{\mathsf{DL\text{-}GA}} \neg(\mathsf{B}\varphi \wedge \mathsf{B}\neg\varphi) \quad (4)$$

$$\models_{\mathsf{DL\text{-}GA}} \widehat{\mathsf{K}}\varphi \rightarrow [*\varphi]\mathsf{B}\varphi \text{ if } \varphi \in Prop \quad (5)$$

$$\models_{\mathsf{DL\text{-}GA}} (\mathsf{B}^{\geq \mathsf{h}}\varphi \wedge \mathsf{B}^{\geq \mathsf{k}}\psi) \rightarrow \mathsf{B}^{\geq \min\{\mathsf{h},\mathsf{k}\}}(\varphi \wedge \psi) \quad (6)$$

$$\models_{\mathsf{DL\text{-}GA}} (\mathsf{B}^{\geq \mathsf{h}}\varphi \wedge \mathsf{B}^{\geq \mathsf{k}}\psi) \rightarrow \mathsf{B}^{\geq \max\{\mathsf{h},\mathsf{k}\}}(\varphi \vee \psi) \quad (7)$$

According to the validity (1), knowing that φ implies believing that φ with strength at least h. According to the validity (2), belief is graded belief with strength at least 1. According to the validity (3), the agent has the strong belief that φ if and only if, it believes that φ with maximal strength max. According to the validity (4) (which follows from the normality constraint $NORM_{\kappa_{\mathsf{exc}}}$ in Section 2.3), an agent cannot have inconsistent beliefs. The validity (5) highlights a basic property of belief revision in the sense of AGM theory [2]: if φ is an objective fact and the agent can imagine a world in which φ is true then, after learning that φ is true, the agent believes that φ.[4] According to the validity (6), if the agent believes that φ with strength at least h and believes that ψ with strength at least k, then the strength of the belief that $\varphi \wedge \psi$ is at least $\min\{\mathsf{h}, \mathsf{k}\}$. According to the validity (7), if the agent believes that φ with strength at least h and believes that ψ with strength at least k, then it believes $\varphi \vee \psi$ with strength at least $\max\{\mathsf{h}, \mathsf{k}\}$. Similar properties for graded belief are given in possibility theory [6].

3. EMOTIONS AND THEIR INTENSITY

We use the modal operators of graded belief and graded goal of the logic DL-GA to provide a logical analysis of emotions such as hope, fear, joy and distress with their intensities.

According to some psychological models [17, 11, 16] and computational models [9, 4] of emotions, the intensity of hope with respect to a given event is a monotonically increasing function of the degree to which the event is desirable and the likelihood of the event. That is, the higher is the desirability of the event, and the higher is the intensity of the agent's hope that this event will occur; the higher is the likelihood of the event, and the higher is the intensity of the agent's hope that this event will occur. Analogously, the intensity of fear with respect to a given event is a monotonically increasing function of the degree to which the event is undesirable and the likelihood of the event. There are several possible merging functions which satisfy these properties. For example, we could define the merging function *merge* as an average function, according to which the intensity of hope about a certain event is the average of the strength of the belief that the event will occur and the strength of the goal that it will occur. Another possibility is to define *merge* as a product function (also used in [9, 17]), according to which the intensity of hope about a certain event is the product of the strength of the belief that the event will occur and the strength of the goal that it will occur. Here we do not choose a specific merging function, as this would much depend on the domain of application in which the formal model has to be used. The emotion intensity scale is defined by the following set:

$$EmoInt = \{y : \text{ there are } x_1, x_2 \in Num \text{ such that } merge(x_1, x_2) = y\}$$

As Num is finite, $EmoInt$ is finite too.

Let us define the notions of hope and fear with their corresponding intensities. We say that the agent hopes with intensity i that its current intention to perform the action a will lead to the desirable consequence l if and only if, there are $\mathsf{h}, \mathsf{k} \in Num \setminus \{0\}$ such that $merge(\mathsf{h},\mathsf{k}) = \mathsf{i}$ and $\mathsf{h} < \mathsf{max}$ and: (1) the agent believes with strength h that the physical action a is executable and l will be true afterwards, (2) the agent wishes to achieve l with strength k, (3) the agent intends to perform the physical action a. Formally:

$$\mathsf{Hope}^{\mathsf{i}}(a, l) \stackrel{\mathtt{def}}{=} \bigvee_{\mathsf{h},\mathsf{k} \in Num \setminus \{0\}: \mathsf{h} < \mathsf{max} \text{ and } merge(\mathsf{h},\mathsf{k}) = \mathsf{i}} (\mathsf{B}^{\mathsf{h}}\langle\langle a \rangle\rangle l \wedge \mathsf{AchG}^{\mathsf{k}} l \wedge \mathsf{Int}_a)$$

[4] The only difference with AGM theory is the condition $\widehat{\mathsf{K}}\varphi$. AGM assumes that new information φ must incorporated in the belief base (the so-called success postulate), whereas we here assume that φ must incorporated in the belief base *only if* the agent can imagine a world in which φ is true.

We say that the agent fears with intensity i that its current intention to perform the action a will lead to the undesirable consequence l if and only if, there are $\mathsf{h},\mathsf{k} \in Num \setminus \{0\}$ such that $merge(\mathsf{h},\mathsf{k}) = \mathsf{i}$ and $\mathsf{h} < \mathsf{max}$ and: (1) the agent believes with strength h that the action a is executable and l will be true afterwards, (2) the agent wishes to avoid l with strength k, (3) the agent intends to perform the action a. Formally:

$$\mathsf{Fear}^{\mathsf{i}}(a,l) \stackrel{\mathsf{def}}{=} \bigvee_{\mathsf{h},\mathsf{k}\in Num\setminus\{0\}:\mathsf{h}<\mathsf{max} \text{ and } merge(\mathsf{h},\mathsf{k})=\mathsf{i}} (\mathsf{B}^{\mathsf{h}}\langle\langle a\rangle\rangle l \wedge \mathsf{AvdG}^{\mathsf{k}} l \wedge \mathsf{Int}_a)$$

In the preceding definitions of hope and fear, the strength of the belief is supposed to be less than max in order to distinguish hope and fear, which imply some form of uncertainty, from happiness and distress which are based on certainty. Indeed, we have that:

$$\models_{\text{DL-GA}} \mathsf{Hope}^{\mathsf{i}}(a,l) \rightarrow \neg\mathsf{SB}\langle\langle a\rangle\rangle l \quad (8)$$

$$\models_{\text{DL-GA}} \mathsf{Fear}^{\mathsf{i}}(a,l) \rightarrow \neg\mathsf{SB}\langle\langle a\rangle\rangle l \quad (9)$$

This means that if an agent hopes/fears that its intention to perform the action a will lead to the desirable/undesirable result l, then it is not certain about that. For example, if our robot hopes to place a container at a given target position by its transport plan, then the robot is not certain that the container will be at the target position after performing the transport plan. On the contrary, to be joyful/distressed that its current intention to perform the action a will lead to the desirable/undesirable consequence l, the agent should be *certain* that its intention to perform the action a will lead to the desirable/undesirable consequence l. This is consistent with OCC psychological model of emotions [16] according to which, while joy and distress are triggered by *actual consequences*, hope and fear are triggered by *prospective consequences* (or *prospects*). Like [9], we here interpret the term 'prospect' as synonymous of 'uncertain consequence' (in contrast with 'actual consequence' as synonymous of 'certain consequence'). The following are our definitions of joy and distress about actions:

$$\mathsf{Joy}^{\mathsf{i}}(a,l) \stackrel{\mathsf{def}}{=} \bigvee_{\mathsf{k}\in Num\setminus\{0\}:merge(\mathsf{max},\mathsf{k})=\mathsf{i}} (\mathsf{SB}\langle\langle a\rangle\rangle l \wedge \mathsf{AchG}^{\mathsf{k}} l \wedge \mathsf{Int}_a)$$

$$\mathsf{Distress}^{\mathsf{i}}(a,l) \stackrel{\mathsf{def}}{=} \bigvee_{\mathsf{k}\in Num\setminus\{0\}:merge(\mathsf{max},\mathsf{k})=\mathsf{i}} (\mathsf{SB}\langle\langle a\rangle\rangle l \wedge \mathsf{AvdG}^{\mathsf{k}} l \wedge \mathsf{Int}_a)$$

where $\mathsf{Joy}^{\mathsf{i}}(a,l)$ and $\mathsf{Distress}^{\mathsf{i}}(a,l)$ respectively mean that "the agent is joyful that its current intention to perform the action a will lead to the desirable consequence l" and "the agent is distressed that its current intention to perform the action a will lead to the undesirable consequence l". Note that, when computing the intensity of joy and distress, the belief parameter in the merging function *merge* is set to max because strong belief is equivalent to graded belief with maximal strength (validity (3) in Section 2.5).

We here distinguish distress from sadness by adding a condition to the definition of distress: the appraisal variable called *controllability* or *control potential* [19]. That is, to be sad that its current intention to perform the action a will lead to the undesirable result l, the agent should be certain that it has no control over the undesirable result l, in the sense that the agent cannot prevent l to be true — which is the same thing as saying that l will be true after every executable action of the agent —.

$$\mathsf{Sadness}^{\mathsf{i}}(a,l) \stackrel{\mathsf{def}}{=} \mathsf{Distress}^{\mathsf{i}}(a,l) \wedge \mathsf{SB}\neg\mathsf{Control}\ l$$

with

$$\mathsf{Control}\ \varphi \stackrel{\mathsf{def}}{=} \bigvee_{b\in PAct} \langle\langle b\rangle\rangle\neg\varphi$$

where $\mathsf{Control}\ \varphi$ means "the agent has control over φ" (or "the agent can prevent φ to be true"). Our definition is consistent with some psychological theories [19, 11] according to which, undesirable states of affairs that not be controlled makes one to be sad.

Example 1. *Consider again our robot which can decide to transport either container number 1 or container number 2 to a given target position. The former task is more demanding than the latter task, as container number 1 is much heavier than container number 2. In particular, the former task requires at least a full battery charge, whereas the latter requires at least a half battery charge. This means that the action of transporting container number 1 ($transport_1$) and the action of transporting container number 2 ($transport_2$) have the following positive and negative effect preconditions with respect to the objective of placing a container at the target position (pos):*
$\gamma^+(transport_1, pos) = fullCharge$,
$\gamma^-(transport_1, pos) = \neg fullCharge \wedge \neg pos$,
$\gamma^+(transport_2, pos) = fullCharge \vee halfCharge$,
$\gamma^-(transport_2, pos) = \neg halfCharge \wedge \neg fullCharge \wedge \neg pos$.
Let us assume that the two actions are always executable:
$Pre(tranport_1) = Pre(tranport_2) = \top$.

Suppose that at the state w the robot intends to transport container number 1 to the target position and considers undesirable with degree k *not to have a container at the target position,* i.e.*,*

$$M, w \models \mathsf{AvdG}^{\mathsf{k}}\neg pos \wedge \mathsf{Int}_{transport_1}$$

Moreover, suppose that the robot is certain that in the current situation there is no container at the target position, i.e.*,*

$$M, w \models \mathsf{SB}\neg pos$$

Finally, suppose that the robot is minimally conservative in revising its beliefs, that is, $\delta = \mathsf{max}$.

The robot observes its battery load and realizes that it does not have a full battery charge but only a half battery charge. After the observation, the robot will strongly believe that its current place will not place any container at the target position, i.e.*,*

$$M^{*halfCharge\wedge\neg fullCharge}, w \models \mathsf{SB}\langle\langle transport_1\rangle\rangle\neg pos$$

It should be noted that the new model $M^{*halfCharge\wedge\neg fullCharge}$ is exactly the same as M except for the plausibility value κ_{exc}. This implies that we have,

$$M^{*halfCharge\wedge\neg fullCharge}, w \models \mathsf{AvdG}^{\mathsf{k}}\neg pos \wedge \mathsf{Int}_{transport_1} \wedge \mathsf{SB}\langle\langle transport_1\rangle\rangle\neg pos$$

and therefore for $merge(\mathsf{max},\mathsf{k}) = \mathsf{i}$ we have

$$M^{*halfCharge\wedge\neg fullCharge}, w \models \mathsf{Distress}^{\mathsf{i}}(transport_1, \neg pos)$$

This means that, after having observed that it only has a half battery charge, the robot is distressed with intensity i *that, if it follows its current intention, then it will not succeed in placing a container at the target position.*

4. FROM APPRAISAL TO COPING

In the previous section, we have characterized emotions in terms of beliefs, (achievement and avoidance) goals, and intentions, and formalized their intensities in terms of belief strength and goal strength. Emotions with high intensity influence the agent's behaviour in order to cope with relevant and significant events. In general, coping can be seen as a cognitive mechanism whose aim is to discharge a certain emotion by modifying one or more of the mental attitudes (*e.g.*, beliefs, goals, intentions) that triggered the emotion [11]. For example, our robot can cope with its distress that if it follows its current intention then it will not succeed in placing a container at the target position, by reconsidering its current intention. The coping mechanism determines various types of responses, also called coping strategies. We here consider three types of coping strategies: coping strategies affecting intentions, coping strategies affecting beliefs and coping strategies affecting goals. More precisely, we consider coping strategies which deal with emotion either by forming or revising an intention to act in the world, or by changing the agent's interpretation of the situation (by changing belief strength or goal strength).

4.1 Coping strategies: syntax and semantics

We extend the logic DL-GA with three different kinds of coping strategies: (1) coping strategies affecting beliefs of the form $\varphi\uparrow^{\mathsf{B}}$ and $\varphi\downarrow^{\mathsf{B}}$, (2) coping strategies affecting goals of the form $l\uparrow^{\mathsf{D}}$ and $l\downarrow^{\mathsf{D}}$, and (3) coping strategies affecting intentions of the form $-a$ and $+a$. We call DL-GA^{+} the resulting logic. $\varphi\uparrow^{\mathsf{B}}$ consists in increasing the strength of the belief that φ is true, while $\varphi\downarrow^{\mathsf{B}}$ consists in reducing the strength of the belief that φ is true. $l\uparrow^{\mathsf{D}}$ consists in increasing the desirability of l, while $l\downarrow^{\mathsf{D}}$ consists in reducing the desirability of l. Finally, $-a$ consists in removing the intention Int_a, while $+a$ consists in generating the intention Int_a.

The set of coping strategies is defined by the following grammar:

$$CStr \quad : \quad \beta \quad ::= \quad \varphi\uparrow^{\mathsf{B}} \mid \varphi\downarrow^{\mathsf{B}} \mid l\uparrow^{\mathsf{D}} \mid l\downarrow^{\mathsf{D}} \mid -a \mid +a$$

where φ ranges over *Fml*, l ranges over *Lit*, and a ranges over *PAct*. For every coping strategy β we introduce a corresponding dynamic operator $[\beta]$, where $[\beta]\psi$ has to be read "after the occurrence of β, ψ will be true".

As expected, the truth conditions of the new operators are given in terms of model transformation. For every $\beta \in CStr$ we define:

$$M, w \models [\beta]\psi \text{ iff } M^{\beta}, w \models \psi$$

The updated model M^{β} is defined according to the following Definitions 9-11.

Coping strategies affecting the strength of the belief that φ either increase or decrease the exceptionality of the worlds in which φ is false with ω unit, only if the agent does not believe that φ is false, *i.e.* $\widehat{\mathsf{B}}\varphi$. If the agent believes that φ is false, *i.e.* $\mathsf{B}\neg\varphi$, they do not have any effect on the agent's mental state. ω is a parameter which captures the agent's disposition (or personality trait) to radically change its mental state when coping with emotions (the higher is ω, and the higher is the agent's disposition to change its mental state when coping with emotions).

Definition 9 (Update via coping strategy on beliefs). *Given a DL-GA model M and $\beta \in \{\varphi\uparrow^{\mathsf{B}}, \varphi\downarrow^{\mathsf{B}}\}$, the updated model M^{β} is defined as $M^{\beta} = \langle W, \sim, \kappa^{\beta}_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}, \mathcal{V}\rangle$ where for all w:*

$$\kappa^{\beta}_{\mathsf{exc}}(w) \quad = \quad \begin{cases} \kappa_{\mathsf{exc}}(w) & \text{if } M, w \models \varphi \\ Cut_{\mathsf{B}}(\kappa_{\mathsf{exc}}(w) + \omega) & \text{if } M, w \models \neg\varphi \wedge \widehat{\mathsf{B}}\varphi \text{ and } \beta = \varphi\uparrow^{\mathsf{B}} \\ \kappa_{\mathsf{exc}}(w) & \text{if } M, w \models \neg\varphi \wedge \mathsf{B}\neg\varphi \text{ and } \beta = \varphi\uparrow^{\mathsf{B}} \\ Cut_{\mathsf{B}}(\kappa_{\mathsf{exc}}(w) - \omega) & \text{if } M, w \models \neg\varphi \wedge \widehat{\mathsf{B}}\varphi \text{ and } \beta = \varphi\downarrow^{\mathsf{B}} \\ \kappa_{\mathsf{exc}}(w) & \text{if } M, w \models \neg\varphi \wedge \mathsf{B}\neg\varphi \text{ and } \beta = \varphi\downarrow^{\mathsf{B}} \end{cases}$$

$\omega \in Num \setminus \{0\}$ and Cut_{B} has been defined in Definition 5.

Coping strategies affecting desirability of l either increase or decrease the desirability of l with ω unit.

Definition 10 (Update via coping strategy on goals). *Given a DL-GA model M and $\beta \in \{l\uparrow^{\mathsf{D}}, l\downarrow^{\mathsf{D}}\}$, the updated model M^{β} is defined as $M^{\beta} = \langle W, \sim, \kappa_{\mathsf{exc}}, \mathcal{D}^{\beta}, \mathcal{I}, \mathcal{V}\rangle$ where for all w:*

$$\mathcal{D}^{\beta}(w, l') \quad = \quad \begin{cases} Cut_{\mathsf{D}}(\mathcal{D}(w, l') + \omega) & \text{if } \beta = l\uparrow^{\mathsf{D}} \text{ and } l' = l \\ Cut_{\mathsf{D}}(\mathcal{D}(w, l') - \omega) & \text{if } \beta = l\downarrow^{\mathsf{D}} \text{ and } l' = l \\ \mathcal{D}(w, l') & \text{if } l' \neq l \end{cases}$$

$\omega \in Num \setminus \{0\}$ and:

$$Cut_{\mathsf{D}}(y) \quad = \quad \begin{cases} y & \text{if } -\mathsf{max} \leq y \leq \mathsf{max} \\ \mathsf{max} & \text{if } y > \mathsf{max} \\ -\mathsf{max} & \text{if } y < -\mathsf{max} \end{cases}$$

Cut_{D} ensures that the new desirability degree of a literal fits into the finite set of integers $Num \cup Num^{-}$.

Finally, coping strategies affecting intentions change the commitment function by either adding or removing an intention.

Definition 11 (Update via coping strategy on intentions). *Given a DL-GA model M and $\beta \in \{-a, +a\}$, the updated model M^{β} is defined as $M^{\beta} = \langle W, \sim, \kappa_{\mathsf{exc}}, \mathcal{D}, \mathcal{I}^{\beta}, \mathcal{C}, \mathcal{V}\rangle$ where for all w:*

$$\mathcal{I}^{\beta}(w) \quad = \quad \begin{cases} \mathcal{I}(w) \setminus \{a\} & \text{if } \beta = -a \\ \mathcal{I}(w) \cup \{a\} & \text{if } \beta = +a \end{cases}$$

The following validities capture some expected properties of coping strategies affecting beliefs and goals. If $\mathsf{h} \geq 1$ then:

$$\models_{\text{DL-GA}^{+}} \mathsf{B}^{\geq \mathsf{h}}\varphi \rightarrow [\varphi\uparrow^{\mathsf{B}}]\mathsf{B}^{\geq Cut_{\mathsf{B}}(\mathsf{h}+\omega)}\varphi \tag{10}$$

$$\models_{\text{DL-GA}^{+}} (\widehat{\mathsf{B}}\varphi \wedge \widehat{\mathsf{B}}\neg\varphi) \rightarrow [\varphi\uparrow^{\mathsf{B}}]\mathsf{B}^{\geq Cut_{\mathsf{B}}(\omega)}\varphi \tag{11}$$

$$\models_{\text{DL-GA}^{+}} \mathsf{B}^{\geq \mathsf{h}}\varphi \rightarrow [\varphi\downarrow^{\mathsf{B}}]\mathsf{B}^{\geq Cut_{\mathsf{B}}(\mathsf{h}-\omega)}\varphi \text{ if } Cut_{\mathsf{B}}(\mathsf{h}-\omega) > 0 \tag{12}$$

$$\models_{\text{DL-GA}^{+}} \mathsf{B}^{\geq \mathsf{h}}\varphi \rightarrow [\varphi\downarrow^{\mathsf{B}}]\neg\mathsf{B}\varphi \text{ if } Cut_{\mathsf{B}}(\mathsf{h}-\omega) = 0 \tag{13}$$

$$\models_{\text{DL-GA}^{+}} \mathsf{Des}^{\mathsf{h}}l \rightarrow [\varphi\uparrow^{\mathsf{D}}]\mathsf{Des}^{Cut_{\mathsf{D}}(\mathsf{h}+\omega)}l \tag{14}$$

$$\models_{\text{DL-GA}^{+}} \mathsf{Des}^{\mathsf{h}}l \rightarrow [\varphi\downarrow^{\mathsf{D}}]\mathsf{Des}^{Cut_{\mathsf{D}}(\mathsf{h}-\omega)}l \tag{15}$$

4.2 Triggering conditions of coping strategies

In our model coping strategies have triggering conditions which are captured by the function

$$Trg : CStr \longrightarrow Fml$$

mapping coping strategies to DL-GA-formulas. For every coping strategy β, $Trg(\beta)$ captures the conditions under which the coping strategy β is *possibly* triggered. Following current psychological and computational models of emotions [11, 16, 9], we here assume that coping strategies are triggered by the agent's positively valenced emotions (*e.g.*, hope and joy) and negatively valenced emotions (*e.g.*, fear and sadness). In what follows we only discuss coping strategies triggered by negatively valenced emotions.

We assume that an agent that is fearful or distressed because its intention a will realize the undesirable effect l will possibly reconsider its intention. Such an intention reconsideration strategy can be formulated as follows:

$$Trg(-a) = \bigvee_{l \in Lit, \mathsf{i} \in EmoInt : \mathsf{i} \geq \theta} ((\mathsf{Fear}^{\mathsf{i}}(a, l) \vee \mathsf{Distress}^{\mathsf{i}}(a, l)) \wedge \mathsf{B}\ \mathsf{Control}\ l)$$

This means that the coping strategy of reconsidering the intention to perform the action a is triggered if and only if (1) the agent is either fearful or distressed with intensity at least θ that its intention to perform the action a will lead to an undesirable result, (2) the agent believes that he has control over l, in the sense that he can prevent the undesirable result l to be true by performing a different action. θ is a threshold which captures the agent's sensitivity to negative emotions (the lower is θ, and the higher is the agent's disposition to discharge a negative emotion by coping with it). $\mathsf{Control}\ l$ captures the appraisal variable called *controllability* we have discussed in Section 3.

Furthermore, we assume that an agent that is fearful or distressed because it believes that its intended action a will realize the undesirable consequence l on which it has no control (1) will possibly decrease its belief that action a will lead to the undesirable consequence l or, (2) will increase the desirability of l. The former kind of coping strategy captures *wishful thinking* while the latter captures *mental disengagement*.

$$Trg(\langle\langle a\rangle\rangle l\downarrow^{\mathsf{B}}) = \bigvee_{l \in Lit, \mathsf{i} \in EmoInt : \mathsf{i} \geq \theta} ((\mathsf{Fear}^{\mathsf{i}}(a, l) \vee \mathsf{Distress}^{\mathsf{i}}(a, l)) \wedge \neg\mathsf{B}\ \mathsf{Control}\ l)$$

$$Trg(l\uparrow^{\mathsf{D}}) = \bigvee_{l \in Lit, \mathsf{i} \in EmoInt : \mathsf{i} \geq \theta} ((\mathsf{Fear}^{\mathsf{i}}(a, l) \vee \mathsf{Distress}^{\mathsf{i}}(a, l)) \wedge \neg\mathsf{B}\ \mathsf{Control}\ l)$$

Note that differently from intention-related coping, wishful thinking and mental disengagement are triggered if the agent appraises that it has no controllability of the undesirable consequence l, in the sense that it cannot prevent l to be true (on this see also [14]).

Example 2. *Let us continue the example of Section 3. We have,* $M^{*halfCharge \wedge \neg\neg fullCharge}, w \models \mathsf{Distress}^{\mathsf{i}}(transport_1, \neg pos)$

Suppose $\mathsf{i} \geq \theta$. *Given the assumption that* $Pre(tranport_2) = \top$ *and the positive effect preconditions of* $transport_2$ *with respect to pos, the robot believes that the action* $transport_2$ *will place the second container at the target position,* i.e.,

$$M^{*halfCharge \wedge \neg fullCharge}, w \models \mathsf{B}\langle\langle transport_2 \rangle\rangle pos$$

and therefore,

$$M^{*halfCharge \wedge \neg fullCharge}, w \models \mathsf{B}\ \mathsf{Control}\ \neg pos$$

Following the specification of the triggering condition for intention-related coping, the robot can now reconsider its intention $\mathsf{Int}_{transport_1}$ *and possibly form the new intention* $\mathsf{Int}_{transport_2}$, i.e.,

$$M^{*halfCharge \wedge \neg fullCharge}, w \models Trg(-transport_1) \wedge Trg(+transport_2)$$

5. AXIOMATIZATION AND DECIDABILITY

The logic DL-GA of Section 2 is axiomatized as an extension of the normal modal logic **S5** for the epistemic operator K with (1) a theory describing the constraints imposed on the agent's mental state, (2) the reduction axioms of the dynamic operators $[\alpha]$, and (3) an inference rule of replacement of equivalents.

Theory of the agent's mental state.

$$\bigvee_{\mathsf{h} \in Num} \mathsf{exc}_\mathsf{h}$$
$$\bigvee_{\mathsf{k} \in Num \cup Num^-} \mathsf{Des}^\mathsf{k} l$$
$$\mathsf{exc}_\mathsf{h} \rightarrow \neg \mathsf{exc}_\mathsf{l} \text{ if } \mathsf{h} \neq \mathsf{l}$$
$$\mathsf{Des}^\mathsf{k} l \rightarrow \neg \mathsf{Des}^\mathsf{m} l \text{ if } \mathsf{k} \neq \mathsf{m}$$
$$\widehat{\mathsf{K}}\ \mathsf{exc}_0$$

Reduction axioms for the dynamic operators $[\alpha]$.

$$[\alpha]p \leftrightarrow \begin{cases} (\gamma^+(a,p) \wedge \neg\gamma^-(a,p)) \vee (p \wedge \neg\gamma^-(a,p)) \text{ if } \alpha = a \\ p \text{ if } \alpha = *\varphi \end{cases}$$

$$[\alpha]\mathsf{Int}_a \leftrightarrow \begin{cases} \bot \text{ if } \alpha = a \\ \mathsf{Int}_a \text{ otherwise} \end{cases}$$

$$[\alpha]\mathsf{exc}_\mathsf{h} \leftrightarrow \begin{cases} \mathsf{exc}_\mathsf{h} \text{ if } \alpha = a \\ ((\varphi \wedge \bigvee_{\mathsf{l},\mathsf{m} \in Num \setminus \{0\} : \mathsf{l}-\mathsf{m}=\mathsf{h}} (\mathsf{B}^\mathsf{m} \neg\varphi \wedge \mathsf{exc}_\mathsf{l})) \vee \\ (\varphi \wedge (\widehat{\mathsf{B}}\varphi \wedge \mathsf{exc}_\mathsf{h})) \vee \\ (\neg\varphi \wedge \widehat{\mathsf{K}}\varphi \wedge \bigvee_{\mathsf{l}:Cut_\mathsf{B}(\mathsf{l}+\delta)=\mathsf{h}} \mathsf{exc}_\mathsf{l}) \vee \\ (\mathsf{K}\neg\varphi \wedge \mathsf{exc}_\mathsf{h})) \text{ if } \alpha = *\varphi \end{cases}$$

$$[\alpha]\mathsf{Des}^\mathsf{k} l \leftrightarrow \mathsf{Des}^\mathsf{k} l$$
$$[\alpha]\neg\psi \leftrightarrow \neg[\alpha]\psi$$
$$[\alpha](\psi_1 \wedge \psi_2) \leftrightarrow ([\alpha]\psi_1 \wedge [\alpha]\psi_2)$$
$$[\alpha]\mathsf{K}\psi \leftrightarrow \mathsf{K}[\alpha]\psi$$

Rule of replacement of equivalents.

$$\text{From } \psi_1 \leftrightarrow \psi_2 \text{ infer } \varphi \leftrightarrow \varphi[\psi_1/\psi_2]$$

Given a formula φ, let $red(\varphi)$ be the formula obtained by iterating the application of the reduction axioms from the left to the right, starting from one of the innermost dynamic operators $[\alpha]$. *red* pushes the dynamic operators inside the formula, and finally eliminates them when facing an atomic proposition. Obviously, $red(\varphi)$ does not contain dynamic operators $[\alpha]$. The following proposition is proved using the reduction axioms above and the rule of replacement of equivalents.

PROPOSITION 2. *Let* φ *be a formula in the language of DL-GA. Then,* $red(\varphi) \leftrightarrow \varphi$ *is DL-GA valid.*

THEOREM 1. *Satisfiability in DL-GA is decidable.*

SKETCH OF PROOF. Let L-GA be the fragment of the logic DL-GA without dynamic operators. The problem of satisfiability in L-GA is reducible to the problem of *global* logical consequence in **S5**, where the set of global axioms Γ is the theory of the agent's mental state given above. That is, we have $\models_{\text{L-GA}} \varphi$ if and only if $\Gamma \models_{\mathbf{S5}} \varphi$. Observe that Γ is finite. It is well-known that the problem of global logical consequence in **S5** with a finite number of global axioms is reducible to the problem of satisfiability in **S5**. The problem of satisfiability checking in **S5** is decidable [7]. It follows that the problem of satisfiability checking in the logic L-GA is decidable too. Proposition 2 and the fact that L-GA is a conservative extension of DL-GA ensure that *red* provides an effective procedure for reducing a DL-GA formula φ into an equivalent L-GA formula $red(\varphi)$. As L-GA is decidable, DL-GA is decidable too. ■

The logic DL-GA$^+$ of Section 4 is axiomatized by the axioms and the rules of inference of the logic DL-GA *plus* the following reduction axioms for the dynamic operators $[\beta]$.

Reduction axioms for the dynamic operators $[\beta]$.

$$[\beta]p \leftrightarrow p$$

$$[\beta]\mathsf{Int}_a \leftrightarrow \begin{cases} \top \text{ if } \beta = +a \\ \bot \text{ if } \beta = -a \\ \mathsf{Int}_a \text{ otherwise} \end{cases}$$

$$[\beta]\mathsf{exc}_\mathsf{h} \leftrightarrow \begin{cases} ((\varphi \vee (\neg\varphi \wedge \mathsf{B}\neg\varphi)) \wedge \mathsf{exc}_\mathsf{h}) \vee \\ (\neg\varphi \wedge \widehat{\mathsf{B}}\varphi \wedge \bigvee_{\mathsf{l}:Cut_\mathsf{B}(\mathsf{l}+\omega)=\mathsf{h}} \mathsf{exc}_\mathsf{l}) \text{ if } \beta = \varphi\uparrow^\mathsf{B} \\ ((\varphi \vee (\neg\varphi \wedge \mathsf{B}\neg\varphi)) \wedge \mathsf{exc}_\mathsf{h}) \vee \\ (\neg\varphi \wedge \widehat{\mathsf{B}}\varphi \wedge \bigvee_{\mathsf{l}:Cut_\mathsf{B}(\mathsf{l}-\omega)=\mathsf{h}} \mathsf{exc}_\mathsf{l}) \text{ if } \beta = \varphi\downarrow^\mathsf{B} \\ \mathsf{exc}_\mathsf{h} \text{ if } \beta = l\uparrow^\mathsf{D} \text{ or } \beta = l\downarrow^\mathsf{D} \end{cases}$$

$$[\beta]\mathsf{Des}^\mathsf{k} l \leftrightarrow \begin{cases} \bigvee_{\mathsf{m}:Cut_\mathsf{D}(\mathsf{m}+\omega)=\mathsf{k}} \mathsf{Des}^\mathsf{m} l \text{ if } \beta = l\uparrow^\mathsf{D} \\ \bigvee_{\mathsf{m}:Cut_\mathsf{D}(\mathsf{m}-\omega)=\mathsf{k}} \mathsf{Des}^\mathsf{m} l \text{ if } \beta = l\downarrow^\mathsf{D} \\ \mathsf{Des}^\mathsf{k} l \text{ if } \beta = \varphi\uparrow^\mathsf{B} \text{ or } \beta = \varphi\downarrow^\mathsf{B} \end{cases}$$

$$[\beta]\neg\psi \leftrightarrow \neg[\beta]\psi$$
$$[\beta](\psi_1 \wedge \psi_2) \leftrightarrow ([\beta]\psi_1 \wedge [\beta]\psi_2)$$
$$[\beta]\mathsf{K}\psi \leftrightarrow \mathsf{K}[\beta]\psi$$

The following Theorem 2 is proved in the same way as Theorem 1.

THEOREM 2. *Satisfiability in DL-GA*$^+$ *is decidable.*

6. RELATED WORK

Although psychological models of emotion emphasize the role of emotion intensity and its role in the coping mechanism, most existing works on logical modeling of emotions have ignored either the intensity of emotions or the coping strategies.

Adam [1] has proposed a logical formalization of the OCC model, while Lorini & Schwarzentruber [13] have formalized counterfactual emotions such as regret and disappointment. Both approaches ignore the quantitative aspect of emotions. In a previous work [12] we formalized emotion intensity by using a similar logic, but we did not consider the coping strategies.

The logical approach to emotion proposed by Steunebrink et al. [22] has both characteristics of our approach: it provides a formal model of emotions extended with their intensities and coping strategies. In this model, an intensity function is assigned to each appraised emotion to determine its intensity at each state of the model. The coping mechanism introduced in this model is inspired by Frijda's theory of action tendencies [8]. According to this theory, specific emotions give agents the tendency to perform particular actions. In the proposed model, coping strategies are developed for negative emotions and their aim is to reduce the intensity of negative emotions. However, unlike the present approach, Steunebrink et al.'s approach takes emotion intensity as a primitive without explaining how it depends on more primitive cognitive ingredients such as belief strength and goal strength. The other important difference between the present work and Steunebrink et al.'s work is

that we here provide a decidable logic of emotion with a complete axiomatization, whereas Steunebrink et al. do not provide any decidability result or complete axiomatization for their logic of emotion.

In the computational model proposed by Gratch and Marsella [9, 14], the eliciting conditions of emotions are defined in terms of quantitative measures such as desirability and likelihood of events. The model is based on several thresholds that determine when emotions are elicited and how emotions are coped with. The implementation of the proposed model is called EMA and is applied to generate predictions about human emotions and their coping strategies. Since the model is quantitative and the authors do no provide any details about its underlying logic, it is hard to compare this model with other logical approaches. One can only conclude that the model proposed by Gratch and Marsella considers both emotion intensities and coping strategies, although it does not provide a logical characterization of the emotions, their intensities, or the corresponding coping strategies.

7. DISCUSSION

In this work we provided a logical characterization of emotions enriched with intensities and coping strategies. Emotions are defined in terms of graded beliefs, graded (achievement and avoidance) goals, and intentions. The intensity of emotions, which are determined in terms of belief and goal strengths, are used to trigger specific coping strategies. We considered only a few coping strategies triggered by negative emotions. In future work, we intend to extend our analysis to coping strategies triggered by positive emotions. For example, hope or joy with respect to a particular literal and plan can trigger coping strategies that suspend other plans in order to create a focus on the literal and plan for which the agent is hopeful or joyful. Moreover, the emotions discussed in this paper are defined with respect to an agent's action. We would like to extend our model in order to characterize emotions in terms of events that are independent from the agent's actions and intentions. Finally, since in the present work emotions are characterized in terms of the agent's actions and intentions, they model the so-called prospective emotions, rather than actual emotions. We believe that our model can be easily extended to characterize actual emotions such as being joyful to have already placed a container at its target position or being hopeful that the current state of the battery charge is not empty.

8. REFERENCES

[1] C. Adam. *The emotions: from psychological theories to logical formalization and implementation in a BDI agent*. PhD Thesis, IRIT, Toulouse, 2007.

[2] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *J. of Symbolic Logic*, 50(2):510–530, 1985.

[3] G. Aucher. A combined system for update logic and belief revision. In *Proc. of PRIMA 2004*, volume 3371 of *LNAI*, pages 1–18. Springer-Verlag, 2005.

[4] T. Bosse and E. Zwanenburg. There's always hope: Enhancing agent believability through expectation-based emotions. In *Proc. of ACII 2009*, pages 111–118. IEEE Computer Society Press, 2009.

[5] A. R. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam Pub Group, 1994.

[6] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty*. Plenum Press, 1988.

[7] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

[8] N. Frijda. *The Emotions*. Cambridge University Press, 1987.

[9] J. Gratch and S. Marsella. A domain independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.

[10] N. Laverny and J. Lang. From knowledge-based programs to graded belief-based programs, part II: off-line reasoning. In *Proc. of IJCAI'05*, pages 497–502, 2005.

[11] R. S. Lazarus. *Emotion and adaptation*. Oxford University Press, 1991.

[12] E. Lorini. A dynamic logic of knowledge, graded beliefs and graded goals and its application to emotion modelling. In *Proc. of LORI-III*, LNCS, pages 165–178. Springer, 2011.

[13] E. Lorini and F. Schwarzentruber. A logic for reasoning about counterfactual emotions. *Artificial Intelligence*, 175(3-4):814–847, 2011.

[14] S. Marsella and J. Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10:70–90, 2009.

[15] W. U. Meyer, R. Reisenzein, and A. Schützwohl. Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21:251–274, 1997.

[16] A. Ortony, G. L. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge University Press, 1988.

[17] R. Reisenzein. Emotions as metarepresentational states of mind: naturalizing the belief-desire theory of emotion. *Cognitive Systems Research*, 10:6–20, 2009.

[18] R. Reiter. *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. MIT Press, 2001.

[19] I. J. Roseman, A. A. Antoniou, and P. E. Jose. Appraisal determinants of emotions: constructing a more accurate and comprehensive theory. *Cognition and Emotion*, 10:241–277, 1996.

[20] H. A. Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74:29–39, 1967.

[21] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In *Causation in decision, belief change and statistics*, pages 105–134. Kluwer, 1998.

[22] B. R. Steunebrink, M. Dastani, and J.-J. Ch. Meyer. A formal model of emotion-based action tendency for intelligent agents. In *Proc. of EPIA'09*, LNAI. Springer-Verlag, 2009.

Un ACA sincère comme compagnon artificiel

J. Rivière Jeremy.Riviere@imag.fr
S. Pesty Sylvie.Pesty@imag.fr
C. Adam Carole.Adam@imag.fr

Equipe MAGMA
Laboratoire d'Informatique de Grenoble (LIG)

Résumé :
Nous proposons dans cet article de rendre **sincère** un Agent Conversationnel Animé (ACA) pour, d'une part, améliorer sa crédibilité du point de vue de l'humain, et d'autre part contribuer à le rendre acceptable dans une relation privilégiée compagnon artificiel - humain. Pour cela, nous avons introduit dans des travaux précédents le modèle logique BIGRE, représentant les états mentaux de l'agent, et le Langage de Conversation Multimodal (LCM) qui permet d'exprimer ces états mentaux. Dans le but de permettre à l'agent de raisonner dans le dialogue, c'est-à-dire mettre à jour ses états mentaux et ses émotions et sélectionner son intention communicative, un moteur de raisonnement est ici présenté. Ce moteur de raisonnement est basé sur le cycle de comportement BDI - Perception, Décision, Action -, les opérateurs logiques du modèle BIGRE et le LCM, utilisé dans le moteur pour atteindre l'intention communicative de l'ACA. Le moteur de raisonnement a été implémenté dans l'agent MARC, dont les expressions multimodales sont construites à partir de l'évaluation des checks de Scherer dans le contexte du dialogue. Une évaluation du moteur de raisonnement montre que les états mentaux déduits par le moteur sont appropriés à la situation, et que leur expression (l'expression de la sincérité de l'agent) est également appropriée.

Mots-clés : Moteur de raisonnement, Dialogue, Compagnon artificiel, Emotions

1 Introduction

Les progrès en informatique graphique et l'augmentation des capacités graphiques des ordinateurs ont participé, ces dernières années, à rendre les personnages virtuels de plus en plus **réalistes**, en termes de rendu de leur apparence et d'animation. Ce réalisme engendre des attentes importantes chez les utilisateurs : ils prêtent à ces personnages de grandes capacités de raisonnement (intentions, états mentaux...) et les extrapolent pour prévoir leur **comportement** [15]. Lorsque ces attentes ne correspondent pas au comportement réel du personnage (ce qu'on peut qualifier de **dissonance**), il se produit chez les utilisateurs un phénomène de rejet ou de répulsion à l'égard de ces personnages : ce phénomène perceptif est connu sous le nom d'*uncanny valley* [17]. Un utilisateur ne pourra donc pas construire de relation affective durable avec le compagnon artificiel tant qu'il y aura cette dissonance entre ses attentes, issues du réalisme de l'agent, et le comportement de cet agent, qui doit être « acceptable et confortable pour l'humain » [9].

Dans le but d'améliorer la crédibilité des personnages virtuels, et donc de réduire la distance entre leur comportement et les attentes des utilisateurs, la majorité des travaux sur les personnages virtuels se sont intéressés principalement au rôle de l'émotion dans le comportement, du fait de son importance dans la communication et la cognition [20] : de ce fait, un grand nombre de personnages virtuels sont dotés de modèles cognitifs des émotions leur permettant de prendre en compte des variables de leur environnement pour en déduire leurs propres émotions et ensuite les exprimer [8, 18, 14]. Les interactions entre l'humain et le personnage virtuel s'en trouvent améliorées, tant en termes de crédibilité que de performance, c'est-à-dire de succès de la communication : rendre **explicite** les émotions des personnages virtuels (de manière réaliste, i.e. avec les expressions appropriées) améliore donc leur crédibilité, en cohérence avec d'autres travaux [9, 6].

Cependant, les émotions ne sont pas les seuls états mentaux qu'un individu exprime lorsqu'il communique avec d'autres individus [30] : on peut vouloir communiquer ses croyances, ses valeurs, ses buts, son intention de faire quelque chose etc. Dans le but d'augmenter la crédibilité des ACA, et ainsi réduire encore la dissonance entre attentes et comportement, il leur faut donc rendre **explicite** l'ensemble de leurs états mentaux, et pas seulement leurs émotions. La proposition que nous faisons consiste à rendre **sincère** l'ACA, et à ce qu'il soit crédible dans l'expression de sa sincérité : un ACA **expressif, affectif et sincère**, capable de raisonner dans le dialogue, exprime donc l'ensemble de ses états mentaux (dont ses émotions) de manière crédible.

Nous faisons l'hypothèse que l'apport de la sincérité dans le raisonnement et dans l'expression des ACA est double :

1. L'expression de l'ensemble des états mentaux de l'ACA sincère, de manière crédible,

rend son comportement explicite et transparent pour l'humain, et permet de réduire ainsi l'effet de dissonance entre son comportement et les attentes de l'humain ;

2. La sincérité étant un élément important dans le jugement affectif [11, 25], un ACA sincère est plus crédible dans un rôle de compagnon artificiel qu'un ACA insincère, et plus à même de mettre en place une relation affective et à long terme.

Rendre un ACA sincère nous amène à poursuivre deux objectifs principaux. Premièrement, une représentation des états mentaux de l'agent et un moyen d'exprimer ces états mentaux sont nécessaires. Pour cela, nous avons introduit d'une part le modèle BIGRE [23], basé sur une logique de type BDI (Belief, Desire, Intention, [21]), représentant les états mentaux de l'agent dont les émotions que nous appelons *complexes* ; d'autre part, nous avons mis en place un Langage de Conversation Multimodal (LCM, [22, 24]) qui exprime ces états mentaux de manière crédible. Le modèle BIGRE et le LCM sont présentés brièvement dans la section suivante (section 2).

Deuxièmement, il faut permettre à l'ACA de raisonner dans le dialogue, mettre à jour ses états mentaux (**dont ses émotions**) et décider de son *intention communicative*, i.e. les états mentaux qu'il va vouloir exprimer dans le contexte. Pour cela, nous présentons en section 3 un **moteur de raisonnement**, basé sur les croyances et les émotions de l'agent, qui déduit son intention communicative en utilisant le Langage de Conversation Multimodal. Une dernière section détaille l'implémentation du moteur de raisonnement dans l'agent MARC [8] et les résultats d'une évaluation de ce moteur menée dans le cadre d'un scénario de réconciliation entre MARC et l'humain.

2 Le modèle BIGRE

La formalisation des états mentaux de l'agent proposée dans [12] se base sur les opérateurs de la logique **MLC** (*Modal Logic of Communication*) de type BDI : cette logique permet de représenter d'une part les états mentaux de l'agent et d'autre part l'aspect explicite et social de la communication. Elle est constituée des opérateurs suivant, qui forment ce que l'on appelle le **modèle BIGRE** :

– (B) $Bel_i\varphi$: l'agent *i* croit φ ;
– (I) $Ideal_i\varphi$: idéalement pour l'agent *i*, φ devrait être vrai ; cet opérateur correspond aux normes de l'agent, tant morales que sociales ;
– (G) $Goal_i\varphi$: l'agent *i* a pour but que φ soit vrai ;
– (R) $Resp_i\varphi$ l'agent *i* est responsable de φ.

La prise en compte de la notion de responsabilité, issue du raisonnement contrefactuel sur ses propres actions et celles d'autrui, permet de formaliser les émotions que nous appelons *complexes* (E), par combinaison des opérateurs B,I,G et R (BIGR $\rightarrow$ E) : le regret, la déception, la culpabilité, le reproche, la satisfaction morale, l'admiration, la réjouissance et la gratitude. Par exemple, l'émotion de culpabilité correspond à un regard porté par l'agent sur ce qu'il a fait et ce qu'il aurait pu faire à un instant donné : l'agent *i* a un idéal ($Ideal_i\varphi$) et croit qu'il est **responsable** de sa violation ($Bel_i Resp_i \neg\varphi$). De même, le reproche est dû à la comparaison par l'agent des capacités de l'interlocuteur et de ce qu'il a fait : l'agent *i* a un idéal ($Ideal_i\varphi$) et croit que l'*interlocuteur j* est **responsable** de sa violation ($Bel_i Resp_j \neg\varphi$).

Le Langage de Conversation Multimodal, basé sur la théorie des Actes de Discours [29], est composé de 38 Actes de Conversation Multimodaux (ACM) qui expriment les états mentaux BIGRE. Ces ACM sont répartis en 4 catégories : les assertifs (`Affirmer`, `Nier`...), les directifs (`Demander`, `Suggerer`...), les engageants (`Promettre`, `Accepter`...) et les expressifs (`S'excuser`, `Se satisfaire`...) qui expriment en particulier les émotions *complexes*. Nous identifions pour ces actes leurs préconditions et leurs effets (différents suivant qu'ils soient émis ou reçus par l'agent). Par exemple, la définition de l'ACM `Se réjouir` du point de vue de l'agent est montrée par le tableau 1.

Les préconditions ainsi définies sont les états mentaux que doit avoir l'agent pour faire l'acte de manière sincère, au sens des conditions de sincérité de la théorie des Actes de Discours [29] ; en ce sens, elles garantissent donc la sincérité de l'agent. La définition des préconditions et effets de chaque ACM permet par la suite, comme présenté ci-dessous, de mettre à jour les états mentaux de l'agent lors de la réception (ou de l'émission) d'un acte et de construire des plans composés d'ACM.

Préconditions	$Goal_a\varphi \wedge Bel_a Resp_a\varphi \stackrel{déf}{=} Rejouissance_a\varphi$ L'agent *a* "ressent" l'émotion de réjouissance ; il est responsable d'avoir atteint son but φ.
Effets d'émission	$Bel_a Bel_h Rejouissance_a\varphi$ L'agent *a* croit que l'humain *h* croit que *a* se réjouit de φ.
Effets de réception	$Bel_a Goal_h\varphi \wedge Bel_a Bel_h Resp_h\varphi$ L'humain *h* s'est réjoui de φ à l'agent *a*, alors *a* croit que *h* a atteint son but φ et que *h* se croit responsable de cela.

TAB. 1 – Préconditions et effets de l'ACM `Se réjouir` pour l'agent *a* dans un dialogue avec un humain *h*.

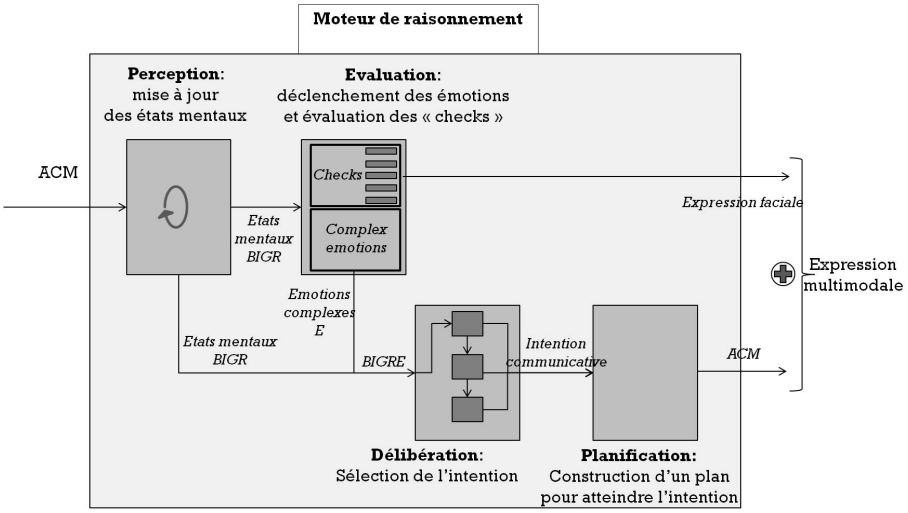


FIG. 1 – Architecture du moteur de raisonnement.

3 Le moteur de raisonnement

Le moteur de raisonnement permet à l'ACA, dans le dialogue, de répondre à un acte de discours (ACM) par un autre ACM. La réception des ACM de l'humain, reconnus dans l'énoncé en langage naturel [1], entraîne :

1. **la perception** de l'ACM et la mise à jour des états mentaux en fonction de ses effets ;
2. **l'évaluation** qui déclenche les émotions complexes (E) de l'agent à partir de ses états mentaux (BIGR), suivant leur formalisation logique, et qui construit une expression faciale dynamique résultant de l'évaluation de l'ACM reçu ;
3. **la délibération** qui décide des différentes intentions de l'agent à partir de ses états mentaux BIGRE et sélectionne, parmi ces intentions, la plus appropriée ;
4. **la planification** qui construit un plan composé d'une séquence d'ACM pour atteindre l'intention sélectionnée, et qui exécute le premier ACM du plan ainsi construit.

Chacune de ces phases est implémentée dans un sous-module distinct de l'architecture du moteur de raisonnement (cf. figure 1).

3.1 Le sous-module de perception

La perception d'un ACM par ce premier sous-module déclenche la mise à jour des états mentaux de l'agent, qui s'effectue *via* des règles d'inférence : les effets de l'ACM perçu sont ajoutés dans la base de connaissances de l'agent. Celle-ci est divisée en deux parties : une partie statique qui représente ses connaissances initiales (bibliothèque de plans du domaine, ontologie et connaissances du domaine, *e.g.* sur la politique), et une partie dynamique qui contient l'ensemble de ses connaissances construites au cours du dialogue (les états mentaux de l'agent et ses croyances sur ceux de l'humain).

Lorsque l'agent reçoit un ACM émis par l'hu-

[1] La traduction de l'énoncé en ACM ne fait pas l'objet de ce travail. La traduction de l'énoncé a été réalisée de façon simple et *ad hoc via* la reconnaissance d'un certain nombre de mots-clés, sélectionnés à partir des différents scénarios.

main, il déduit un certain nombre de croyances sur les états mentaux de l'humain (dont les émotions dans le cas des ACM expressifs) à partir des effets de l'ACM. Ce type "d'évaluation inverse" (appelée *reverse appraisal* en anglais) a été discuté en psychologie [13], et abordé dans le domaine des agents virtuels [16, 10].

Le sous-module de perception perçoit également les ACM émis par l'agent lui-même et effectue la mise à jour en conséquence.

3.2 Le sous-module d'évaluation

Le sous-module d'évaluation reçoit les états mentaux BIGR de l'agent, dont les croyances sur ceux de l'humain, et a deux fonctions différentes. Premièrement, il déclenche les émotions *complexes* à partir de leur définition logique en termes d'états mentaux BIGR. Par exemple, l'émotion de gratitude est déclenchée lorsque les BIGR de l'agent contiennent $Goal_i\varphi \wedge Bel_i Resp_j\varphi$ [12] ; l'intensité de l'émotion est fonction du degré du but ou de l'idéal contenu dans sa définition. Le déclenchement des émotions dépend d'un des facteurs de la personnalité de l'agent : son émotivité [2].

Deuxièmement, le sous-module d'évaluation construit une expression faciale dynamique pour accompagner l'ACM sélectionné par le moteur de raisonnement en réponse à l'humain. Pour cela, il évalue chaque ACM perçu selon cinq variables d'évaluation ("*checks*"), introduites par la théorie d'évaluation cognitive de Scherer [27, 26], que nous avons adaptés à la théorie des Actes de Discours :

1. la nouveauté de l'acte (était-il attendu dans le schéma de dialogue ?) ;
2. le plaisir intrinsèque (défini suivant le type de l'acte, *e.g.* `Refuser` *vs.* `Accepter`, et le contenu propositionnel) ;
3. la congruence avec les buts de l'agent et l'attribution de la responsabilité ;
4. le potentiel d'adaptation (est-ce que l'agent peut s'adapter aux conséquences de l'ACM ?) ;
5. la compatibilité avec les idéaux de l'agent.

Ce processus d'évaluation de l'acte consiste en une séquence de *checks*, évalués l'un après l'autre. Scherer a montré [28] que chaque *check* pouvait être lié à une expression faciale définie à l'aide d'Unités d'Action (*Action Units* en anglais - AU), c'est-à-dire de muscle du visage. Une expression facaile dynamique peut donc être construite à partir de l'évaluation des *checks* : pour chaque acte reçu, l'expression faciale de chaque *check* est calculée, combinée avec la précédente puis est exprimée par l'ACA. L'expression globale ainsi construite est exprimée par l'agent quand il répond à l'humain (pour un exemple concret, voir la figure 2).

Dans la théorie de Scherer, nous pouvons identifier des *checks* communs avec notre définition des émotions *complexes* : la congruence aux buts et la compatibilité avec les idéaux sont représentées par les opérateurs $Goal_i\varphi$ et $Ideal_i\varphi$, et l'attribution de la responsabilité par l'opérateur $Resp_i\varphi$. Ces similarités montrent que les émotions *complexes* forment un sous-ensemble des émotions déclenchées par le processus d'évaluation des *checks*.

3.3 Le sous-module de déliberation

L'intention communicative de l'ACA est sélectionnée à partir de ses états mentaux (BIGR+E) *via* un raisonnement pratique [5]. Nous définissons trois types d'intentions communicatives : l'intention "émotionnelle" et l'intention "d'obligation", utiles à la régulation locale du dialogue [3], et l'intention "globale" qui définit la direction générale du dialogue.

L'intention "émotionnelle" est l'intention d'exprimer une émotion. Après chaque perception et mise à jour des états mentaux, les émotions de l'agent sont déclenchées à partir de ses BIGR dans le sous-module d'évaluation. Dans l'hypothèse d'un agent sincère, émotionnel et expressif, lorsque cet agent "ressent" une émotion, il va adopter l'intention de l'exprimer, de la communiquer à l'humain. Ce type d'intention est modulé par un second paramètre de la personnalité de l'agent, son expressivité [3]. L'intention *émotionnelle* participe à la régulation du dialogue au niveau "local", en permettant notamment une interaction plus naturelle entre l'agent et l'humain [4]. Cette intention a donc la priorité sur les autres intentions : l'agent va d'abord essayer de satisfaire cette intention (i.e. d'exprimer son émotion) avant de considérer ses autres intentions.

[2] L'émotivité représente sa capacité à "ressentir" les émotions : un agent très émotif ressent plus facilement les émotions, et *vice-versa*

[3] Un agent très expressif aura l'intention d'exprimer toutes ses émotions tandis qu'un agent peu expressif aura l'intention d'exprimer seulement ses émotions les plus intenses

L'intention "d'obligations" est l'intention de satisfaire une règle *d'obligations du discours* ; ces règles sont des normes sociales, définies par Traum et Allen [32], qui guident le comportement de l'agent et le rendent réactif au niveau du discours. Le tableau 2 montre quelques règles d'obligations du discours : à chaque énoncé reçu ou émis par l'agent correspond un certain nombre d'obligations du discours (représentant ce que l'agent devrait faire en réponse).

Source de l'obligation	Obligation correspondante
S1 accepte ou promet A	S1 réalise A
S1 demande (requête) A	S2 accepte ou refuse A
S1 demande (question) P	S2 informe S1 de P
Enoncé incompris ou incorrect	"Réparer" l'énoncé

TAB. 2 – Quelques règles d'obligations du discours entre un locuteur (S1) et un interlocuteur (S2) définies par [32].

Dans le moteur de raisonnement, à chaque acte reçu par l'agent, les règles correspondant à l'acte s'appliquent : lorsque les conditions d'une règle se vérifient, l'agent adopte l'intention communicative d'accomplir l'acte provenant de l'obligation. Par exemple, lorsque l'humain `Demande` ou `Offre` quelque chose, l'agent doit `Accepter` ou `Refuser`, suivant ses buts et ses connaissances (tout en vérifiant les préconditions de ces actes, décrites précédemment). L'intention "d'obligations" a une priorité inférieure à l'intention "émotionnelle" (i.e. l'agent tente de satisfaire son intention "émotionnelle", puis adopte l'intention "d'obligations"), mais a une priorité supérieure à l'intention *globale*.

L'intention "globale" correspond au niveau global du dialogue : c'est cette intention qui va donner la direction du dialogue et définir son type ([33], *p.ex.* délibération, persuasion). L'agent adopte une intention "globale" de poursuivre un de ses buts lorsqu'il s'engage à atteindre ce but. Un tel engagement peut être *public*, par l'intermédiaire d'ACM engageants tels que `Promettre, Accepter`, ou *privé* : l'agent peut s'engager sur un de ses buts à partir de l'ensemble de ses connaissances et par un raisonnement pratique (en lien avec les plans qu'il connait). Ces types d'engagements sont cohérents avec la définition donnée par Cohen et Levesque [7], qui parlent de "*social commitment*" pour les engagements publics et d'"*internal commitment*" pour les engagements privés. Un exemple d'engagement privé peut être que l'agent a le but de se réconcilier avec l'utilisateur, et qu'il connait un plan pour y parvenir ; si le contexte s'y prête (i.e. si l'agent n'a pas déjà d'autres intentions "globales" ou d'autres engagements incompatibles), l'agent adopte l'intention "globale" de se réconcilier avec l'utilisateur.

3.4 Le sous-module de planification

Le plan pour atteindre l'intention communicative sélectionnée par le sous-module de délibération est construit par le sous-module de planification de l'intention, selon une approche par plan du dialogue [19, 2] : les ACM qui composent notre Langage de Conversation Multimodal constituent les opérateurs (actions) de ces plans. La formalisation des préconditions et des effets des ACM permet une planification par chaînage arrière : l'agent cherche l'ensemble des actions qui satisfont son intention ; il cherche ensuite l'ensemble des préconditions de ces actions qui sont fausses ; les sous-actions satisfaisant ces préconditions sont ajoutées au plan, et ainsi de suite.

Dans le cas des intentions "émotionnelles" et "d'obligations", le plan construit inclut généralement un seul ACM. Par exemple, si l'intention "émotionnelle" de l'agent est d'exprimer sa gratitude, le plan contiendra l'ACM `Remercier` ou `Féliciter`, selon l'intensité de l'émotion. Dans le cas des intentions "globales", les plans du domaine contenus dans la base de connaissances de l'agent peuvent être nécessaires : les actions qui composent ces plans sont également décrites en termes de préconditions et d'effets. Par exemple, si l'agent a l'intention de réserver un billet de train pour l'utilisateur, il a besoin de savoir que pour réserver un billet de train, il lui faut connaître la date et le lieu de départ, ainsi que le lieu d'arrivée ; à lui ensuite de décider des ACM appropriés (*p.ex.* `Demander`), grâce aux mêmes mécanismes de planification, pour obtenir ces informations.

Lorsque le plan est déjà connu, il est mis à jour à partir des dernières actions effectuées (connues par l'agent) et des actions impossibles à réaliser (*p.ex.* si l'utilisateur refuse de donner une information, l'agent doit trouver un autre plan pour obtenir cette information). Si le dernier

ACM accompli par l'agent n'a pas été compris par l'humain (*p.ex.* si celui-ci ne lui répond pas ou répond par une autre question), l'ACM reste dans le plan pour être de nouveau tenté ultérieurement.

La première action du plan ainsi construit (ou mis à jour) est exécutée en sortie du moteur de raisonnement par l'agent, conjointement à l'expression faciale dynamique provenant du sous-module d'évaluation (cf. figure 1).

4 Implementation et évaluation

MARC [8] est un ACA capable de s'exprimer *via* plusieurs modalités (dont des expressions faciales encodées par des Unités d'Action). Nous avons implémenté le moteur de raisonnement en Prolog (en nous basant sur les travaux précédant de [1]) et nous l'avons déployé dans MARC pour conduire une évaluation suivant un certain scénario. Ce scénario fait intervenir, dans un dialogue, un ACA compagnon (joué par MARC) et un utilisateur. La scène a lieu après une dispute (imaginaire) entre MARC et l'utilisateur, que vont incarner les sujets humains. MARC `regrette` tout d'abord la dispute, car il s'en croit responsable et c'est contre un de ses buts (l'ACM `Regretter` est le moyen d'atteindre son intention "émotionnelle"). Dans un deuxième temps, il `demande` à l'utilisateur de lui pardonner : cet acte fait partie du plan construit pour atteindre son intention "globale", se réconcilier avec l'humain. Les sujets ont alors le choix entre `Accepter` de pardonner MARC, `Refuser` (cf. figure 2) et `Demander` un peu de temps pour réfléchir.

Il est alors demandé aux sujets d'observer la réaction de l'agent suivant leur choix. Dans le cas où le sujet `Demande` un peu de temps pour réfléchir, quatre intentions communicatives sont sélectionnées en séquence par le moteur de raisonnement :

1. intention "émotionnelle" : MARC exprime sa déception (`Se plaindre`) car l'utilisateur ne veut pas répondre à sa demande ;
2. intention "d'obligations" : MARC doit `Accepter` ou `Refuser` la demande de l'utilisateur ; comme son but est d'avoir une réponse à sa propre requête, MARC choisit de `refuser` d'attendre ;
3. intention "émotionnelle" : MARC `s'excuse` car il croit, d'une part, que l'utilisateur avait le but qu'il `accepte`, et d'autre part, qu'il est responsable de n'avoir pas `accepté` la requête ;
4. intention "globale" : MARC continue de suivre son intention "globale" de se réconcilier avec l'utilisateur, et `demande` de nouveau que l'utilisateur lui pardonne.

L'évaluation a été menée auprès de 17 sujets entre 12 et 30 ans (la moyenne d'âge est de 16.9 ans), lors de l'évènement RoboFesta[4]. Les résultats de cette évaluation montrent que les états mentaux déduits par le moteur sont jugés crédibles par les sujets et que leurs expressions (i.e. l'expression de la sincérité de l'agent) sont également jugées crédibles. De plus, les résultats tendent à montrer que les sujets prêtent des états mentaux à l'ACA que celui-ci n'exprime pas obligatoirement ; ce résultat est cohérent avec les résultats précédents montrés par [8] et à la base des attentes des utilisateurs générées par l'apparence des personnages virtuels [15], mais doit encore être approfondi. Un autre résultat intéressant montre que les sujets reconnaissent d'autres émotions que celles proposées dans le questionnaire, telles que la colère, la tristesse ou la surprise ; certaines de ces émotions sont cohérentes avec notre formalisation logique et les expressions faciales de MARC, tandis que la reconnaissance de la colère lors de l'expression du reproche (13 sujets sur 17 - soit 76%) semble montrer une relation forte entre ces deux émotions. Enfin, cette évaluation n'a pas permis de montrer une amélioration de la relation affective entre l'ACA et l'humain par notre moteur de raisonnement. Une possible influence négative du scénario (une tentative de réconciliation après une dispute), ou le fait que MARC `refuse` de laisser du temps au sujet lorsque celui-ci en demande (réaction jugée peu crédible) peuvent expliquer ce résultat.

5 Conclusion et perspectives

Le moteur de raisonnement que nous avons présenté ici déduit, pour un ACA expressif, affectif et sincère, les états mentaux à exprimer (son intention communicative) dans un dialogue avec l'humain : les émotions déduites du modèle BIGRE assurent son côté affectif, les préconditions des actes du LCM assurent sa sincérité, et son expressivité est maintenue par la définition de trois types d'intentions communicatives.

Ce moteur de raisonnement a été implémenté dans l'agent MARC. L'expression multimodale

[4] Rencontres Académiques de Robotique collèges et lycées, http ://crdp2.ac-rennes.fr/robofesta/

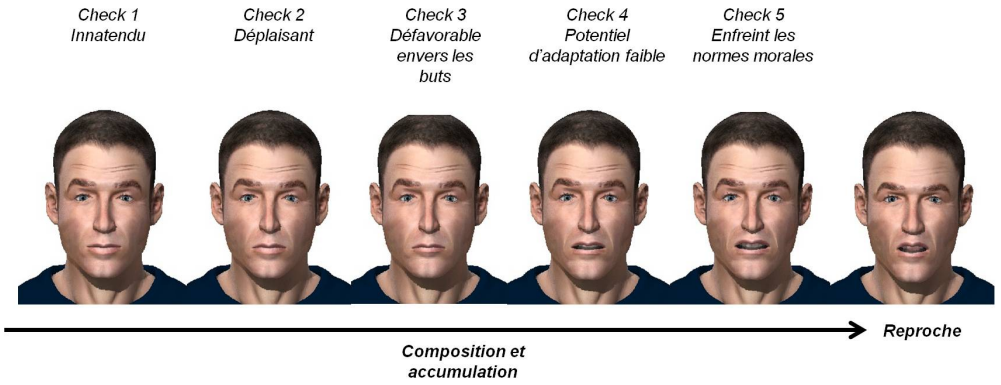


FIG. 2 – L'utilisateur vient de `refuser` de pardonner à MARC. De gauche à droite, la figure montre l'expression de l'évaluation de chaque *check* et leur accumulation. L'expression globale ainsi construite est exprimée par MARC lorsqu'il `reproche` à l'utilsateur son refus de lui pardonner.

de l'agent lors de l'acte est celle déduite de l'évaluation d'une partie des *checks* de Scherer, que nous avons adaptés au dialogue : à chaque acte reçus ou accompli par l'agent, les *checks* sont évalués en séquence et mènent à une expression faciale dynamique, associée à l'ACM déduit par le moteur de raisonnement. Une évaluation du moteur de raisonnement avec l'agent MARC a montré que les états mentaux déduits par le moteur sont appropriés à la situation, et que leur expression (l'expression de la sincérité de l'agent) est également appropriée. Les résultats de cette évaluation ne permettent cependant pas de conclure sur l'effet de la sincérité de l'agent sur sa relation affective avec l'humain.

Une des perspectives à court terme consiste en de nouvelles évaluations plus complètes de l'effet de la sincérité sur la relation affective entre un agent compagnon et un humain : un exemple d'évaluation intéressante pourrait impliquer, sur quelques semaines, un ensemble de sujets et le compagnon artificiel, doté de capacités et de connaissances propres à plusieurs domaines. Ce compagnon devrait de plus être à même de reconnaitre les ACM de l'humain dans la langue naturelle, qui reste la modalité privilégiée des utilisateurs et augmente le naturel de l'interaction et l'implication de l'humain [31].

Enfin, il nous semble que la notion d'évaluation inverse (*reverse appraisal* en anglais) est importante dans le développement des compagnons artificiels. Le fait de conserver et de mettre à jour une image des états mentaux de l'humain pendant l'interaction permet à l'agent de déduire par exemple, lors d'un acte émis par l'humain, son émotion ressentie et la cause de cette émotion. Lorsque l'humain `culpabilise`, l'agent serait capable de savoir, d'après les effets formalisés des ACM, que l'humain avait un certain idéal, que celui-ci a été violé et qu'il se croit responsable. Une analyse de l'énoncé de l'humain permettrait de savoir quel était l'idéal de l'humain, et le compagnon pourrait alors suivre plusieurs stratégies, comme tenter de réconforter l'humain par exemple (« tu culpabilises parce que tu te sens responsable de ... ? »).

Remerciements

Ce travail a été réalisé dans le cadre du projet CECIL (*Complex Emotions in Communication, Interaction, Language*), et bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-08-CORD-005.

Références

[1] Carole Adam. *Emotions : from psychological theories to logical formalization and implementation in a BDI agent*. Thèse de doctorat, INP, Toulouse, France, juillet 2007.

[2] J Allen and C R Perrault. Analyzing intention in utterances. In *Readings in natural language processing*, pages 441–458. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1986.

[3] Michael J. Baker. A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence and Education*, 5(2) :199–254, 1994.

[4] Joseph Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7) :122–125, 1994.

[5] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, November 1987.

[6] C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. *Intelligent Robots and Systems*, 2 :858–863, 1999.

[7] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3) :213–261, 1990.

[8] Matthieu Courgeon, Jean-Claude Martin, and Christian Jacquemin. Marc : a multimodal affective and reactive character. In *Proc. of The 1st workshop on Affective Interaction in Natural Environments (AFFINE)*, Chania, Crete, 2008.

[9] Kerstin Dautenhahn. Ants don't have friends - thoughts on socially intelligent agents. Technical report, In Socially Intelligent Agents, 1997.

[10] Celso M. de Melo, J. Gratch, and P. Carnevale. Reverse appraisal : Inferring from emotion displays who is the cooperator and the competitor in a social dilemma. In *The 33rd Annual Meeting of the Cognitive Science Society (CogSci) 2011*, Boston, MA, 2011.

[11] S. Fiske, A. Cuddy, and P. Glick. Universal dimensions of social cognition : warmth and competence. *Trends in Cognitive Sciences*, 11(2) :77–83, 2007.

[12] Nadine Guiraud, Dominique Longin, Emiliano Lorini, Sylvie Pesty, and Jérémy Rivière. The face of emotions : a logical formalization of expressive speech acts. In *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, pages 1031–1038, Taipei, Taiwan, May 2011. International Foundation for AAMAS.

[13] Shlomo Hareli and Ursula Hess. What emotional reactions can tell us about the nature of others : An appraisal perspective on person perception. *Cognition and Emotion*, 24(1) :128–140, 2010.

[14] S. Kopp, B. Jung, N. Lessmann, and I. Wachsmuth. Max - a multimodal assistant in virtual reality construction. *KI - Künstliche Intelligenz*, 4/03 :11–17, 2003.

[15] Sau-lai Lee, Ivy Y. Lau, S. Kiesler, and Chi-Yue Chiu. Human Mental Models of Humanoid Robots. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005.*, pages 2767–2772. IEEE Computer Society, 2005.

[16] Celso M. Melo, Liang Zheng, and Jonathan Gratch. Expression of moral emotions in cooperating agents. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA'09)*, pages 301–307, Berlin, Heidelberg, 2009. Springer-Verlag.

[17] M. Mori. The uncanny valley. In *Energy*, pages 33–35, 1970.

[18] Ana Paiva, Ruth Aylett, and Joao Dias. An affectively driven planner for synthetic characters. In *International Conference on Automated Planning and Scheduling (ICAPS)*, pages 2–10. AAAI press, 2006.

[19] C. Raymond Perrault and James F. Allen. A plan-based analysis of indirect speech acts. *Comput. Linguist.*, 6(3-4) :167–182, July 1980.

[20] J. Piaget. Les émotions. In B. Rimé and K. Scherer, editors, *Les relations entre l'intelligence et l'affectivité dans le developpement de l'enfant*, pages 75–95. Delachaux et Niestlé, Neuchâtel-Paris, 1989.

[21] Anand S. Rao and Michael P. Georgeff. Modeling Rational Agents within a BDI-Architecture. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484. Morgan Kaufmann publishers Inc. : San Mateo, CA, USA, 1991.

[22] Jérémy Rivière, Carole Adam, and Sylvie Pesty. Langage de Conversation Multimodal pour Agent Conversationnel Animé. *TSI : Revue des sciences et technologies de l'information*, 31, 2012.

[23] Jérémy Rivière, Carole Adam, and Sylvie Pesty. A reasoning module to select eca's communicative intention. In *Proc. of the 12th Int. Conf. on Intelligent Virtual Agents (IVA'12)*, volume 7502 of *Lecture Notes in Computer Science*, pages 447–454, Heidelberg, 2012. Springer Berlin / Heidelberg.

[24] Jérémy Rivière, Carole Adam, Sylvie Pesty, Catherine Pelachaud, Nadine Guiraud, Dominique Longin, and Emiliano Lorini. Expressive multimodal conversational acts for saiba agents. In *Proc. of the 11th Int. Conf. on Intelligent Virtual Agents (IVA'11)*, volume 6895 of *Lecture Notes in Computer Science*, pages 316–323, Heidelberg, 2011. Springer.

[25] Seymour Rosenberg, Carnot Nelson, and P. S. Vivekananthan. A multidimensional approach to the structure of personality impressions. In *Journal of Personality and Social Psychology*, volume 9, pages 283–294, 1968.

[26] K. R. Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion : Theory, methods, research*, pages 92–120, 2001.

[27] K. R. Scherer and P. Ekman. *Approaches to emotion*. Erlbaum, Hillsdale, NJ, 1984.

[28] K. R. Scherer and H. Ellgring. Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal ? *Emotion*, 7 :113–130, 2007.

[29] John R. Searle and Daniel. Vanderveken. *Foundations of illocutionary logic*. Cambridge University Press, New York, USA, 1985.

[30] J.R. Searle. *Intentionality, an Essay in the Philosophy of Mind*. Cambridge Paperback Library. Cambridge University Press, 1983.

[31] William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, Chad Lane, Jacquelyn Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu, and Kyle White. Ada and grace : toward realistic and engaging virtual museum guides. In *Proceedings of the 10th international conference on Intelligent virtual agents*, IVA'10, pages 286–300, Berlin, Heidelberg, 2010. Springer-Verlag.

[32] David R. Traum and James F. Allen. Discourse obligations in dialogue processing. In *Proc. of the 32th annual meeting of the Association for Computational Linguistics (ACL)*, pages 1–8, 1994.

[33] Douglas Walton and Eric Krabbe. *Commitment in Dialogue : Basic concept of interpersonal reasoning*. State University of New York Press, 1995.

Synchronie interpersonnelle : un panorama des méthodes d'évaluation

E. Delaherche⋆ M. Chetouani⋆
delaherche@isir.upmc.fr chetouani@isir.upmc.fr

⋆Institut des Systèmes Intelligents et de Robotique - UMR 7222
Université Pierre et Marie Curie
75005 Paris, France

Résumé :
La synchronie désigne la coordination temporelle des individus au cours des interactions sociales. L'analyse de ce phénomène est complexe, nécessitant la perception et l'intégration de signaux de communication multimodaux. L'évaluation de la synchronie a reçu une attention multidisciplinaire en raison de son rôle dans le développement précoce, l'apprentissage des langues et l'établissement des liens sociaux. Cet article soulève les questions posées par l'évaluation de la synchronie et les méthodes de l'état-de-l'art dédiées à cette évaluation. Nous présentons tout d'abord les définitions et les fonctions de la synchronie dans l'enfance et à l'âge adulte. Ensuite, nous passons en revue les approches manuelles et computationnelles pour l'annotation, l'évaluation et la modélisation de la synchronie interactionnelle. Enfin, les limitations actuelles et les orientations futures de la recherche dans les domaines de la robotique développementale, la robotique sociale et les études cliniques sont abordées.

Mots-clés : Synchronie, comportements non verbaux, traitement des signaux sociaux

1 Introduction

La synchronie est un phénomène complexe qui nécessite la perception et la compréhension de signaux sociaux et de communication ainsi qu'une adaptation constante à son partenaire. Une approche multidisciplinaire, à l'interface des traitements des signaux sociaux, des neurosciences computationnelles, de la psychologie développementale et de la pédopsychiatrie semble indispensable pour évaluer la synchronie [16]. La mise en oeuvre d'algorithmes interactifs dans des interfaces homme-machine nécessite une meilleure compréhension des stratégies de régulation de l'interaction, en particulier de la synchronie [64]. Le lien étroit entre la synchronie et la qualité de l'interaction porte d'ailleurs des perspectives prometteuses dans le domaine des interfaces sociales, robots ou agents conversationnels [49, 9]. Par ailleurs, le manque d'outils automatiques pour l'évaluation de la synchronie limite l'étude des affections psychiatriques qui touchent les aptitudes sociales, que ce soit de manière permanente (e.g., l'autisme) ou temporaire (e.g., la dépression). Récemment, Meltzoff et al. ont souligné l'intérêt d'une recherche interdisciplinaire psychologie / modèles computationnels du développement [36]. En particulier, les mécanismes de l'apprentissage social intéressent les chercheurs de la robotique développementale, pour lesquels l'objectif à long terme est la création de robots qui, comme les enfants, apprennent par l'observation, l'imitation et les échanges synchrones. La synchronie est délicate à circonscrire. De nombreux termes ont été utilisés dans la littérature pour décrire l'interdépendance des comportements dyadiques (mimétisme, résonance sociale, coordination, synchronie, effet caméléon, etc.). Par ailleurs, plusieurs concepts sont étroitement liés à la synchronie ou sont des conditions préalables à la synchronie, comme les tours de parole et l'attention conjointe. Le premier objectif de cette étude est de clarifier le concept de synchronie et ses fonctions à la fois dans l'enfance et à l'âge adulte. Nous donnons par la suite un aperçu des méthodes d'évaluation de la synchronie, manuelles puis automatiques, dans les interactions humain-humain. Enfin, nous présentons les limites actuelles de ces méthodes et les perspectives de recherche pour l'évaluation de la synchronie.

2 Définitions et fonctions

2.1 Définitions et concepts liés

L'étude de la synchronie est inextricablement liée à l'étude de la communication et du langage. Selon les théories du dialogue, la conversation est une activité "jointe" qui nécessite une coordination à deux niveaux : le contenu et les processus [11].

Coordination du contenu. Au niveau du contenu, les partenaires conversationnels doivent coordonner ce qui est dit et parvenir à une compréhension commune. Cette compréhension commune est obtenue par un "alignement interactif" de leurs représentations des situations (informations sur l'espace, le temps, la causalité, l'intentionnalité et les personnes concernées) [48]. Les partenaires alignent leurs représentations à différents niveaux linguistiques au même

moment. Concrètement, un partenaire aura tendance à utiliser des formes pragmatiques et lexicales identiques à celles qui ont été utilisées précédemment dans le dialogue [19]. Cet alignement sert l'efficacité de la communication : les partenaires conversationnels ont tendance à formuler leurs déclarations de manière à minimiser le temps et les efforts nécessaires à la compréhension mutuelle. Cet alignement interactif a été observé au niveau lexical [19], syntaxique [6], de l'accent ou du débit de parole [20]. La Communication Accomodation Theory (CAT) traite de cette tendance à minimiser ou accentuer inconsciemment les différences avec son interlocuteur dans le contenu du discours, les caractéristiques vocales ou les gestes [20]. Le "behavior matching" [3] ; mimétisme [33] ; la congruence et l'effet caméléon [10] concernent l'alignement des comportements non verbaux, tels que les postures, les manières ou les mimiques.

Coordination des processus. Au niveau des processus, les partenaires conversationnels sont capables de prédire avec précision les débuts et les fins des phases, grâce à la syntaxe et à l'intonation [11]. En prédisant avec précision la fin du tour de l'orateur, l'auditeur peut commencer son tour de parole au moment opportun, permettant ainsi aux partenaires d'atteindre la synchronie. Bernieri et al. définissent la synchronie comme "le caractère non-aléatoire, répétitif ou synchronisé des comportements interactifs que ce soit au niveau de la forme ou du timing" [3]. Le terme "synchronie" fait référence à la coordination temporelle entre les individus. La synchronie est liée à l'adaptation d'un individu au rythme et aux mouvements du partenaire d'interaction [13] et au degré de congruence entre les cycles d'engagement et de désengagement des deux personnes. En opposition au "behavior matching", la synchronie est un phénomène dynamique [44]. Par ailleurs, la synchronie peut exister entre différentes modalités sensorielles, comme c'est le cas par exemple, lorsque le nourrisson bouge à l'unisson du comportement vocal de sa mère [14].

2.2 Fonctions

Fonctions dans l'enfance. Il reste probablement beaucoup à comprendre sur le rôle de la synchronie au cours du développement précoce. Certaines fonctions importantes ont d'ores et déjà été mises en évidence. Tout d'abord, la synchronie semble être impliquée dans la corégulation des états affectifs [18], un "processus par lequel la mère et l'enfant font correspondre leurs états affectifs en modérant le niveau d'excitation positive". Les mères ont tendance à utiliser ce mécanisme pour maintenir et réguler les échanges avec leur bébé pendant le face-à-face. La synchronie semble aider l'enfant à faire l'expérience des liens sociaux [24]. Lorsque les cycles d'interaction sont interrompus, puis rétablis, "le sentiment de confiance de l'enfant dans sa capacité à s'auto-réguler et à engager les autres efficacement" est améliorée [24]. La synchronie devrait également faciliter un attachement sécurisant. Le nombre croissant d'enfants privés d'attachement sécurisant ou montrant des troubles du comportement après l'exposition pendant l'enfance à des comportements maternels perturbateurs [55], la dépression [28], ou la privation sociale [53] justifie l'importance de la synchronie pour un développement émotionnel adéquat de l'enfant. Enfin, la synchronie joue également un rôle lors de l'acquisition du langage. Dans la production et la compréhension de la parole, l'interaction avec un adulte influe fortement sur l'apprentissage [21]. Enfin, il convient de mentionner l'imitation [47, 38], qui a été largement étudiée en psychologie du développement. Dans un premier temps, l'imitation est un moyen d'apprendre par l'observation et la reproduction (apprentissage par observation) et contribue également à la construction d'un code social pour l'enfant afin de reproduire ce qu'il a observé dans des situations adéquates. Ensuite, l'imitation est un moyen de communiquer tant que l'enfant n'a pas accès au langage.

Fonctions à l'âge adulte. A l'âge adulte, la synchronie interactionnelle agit comme un facilitateur des interactions sociales, pour atteindre "une coordination des attentes chez les partenaires" [27]. La synchronie non verbale joue également un rôle dans la construction des relations interpersonnelles [31, 60]. Dans leur étude sur l'effet caméléon, Chartrand et al. ont établi un lien entre le degré de mimétisme, la perception de fluidité dans l'interaction et la sympathie entre les partenaires d'interaction [10]. Un lien a également été établi entre le degré de synchronie et la façon dont sont perçus les partenaires dyadiques [41]. Par exemple, Lakens et al. ont manipulé le rythme de couples de figurines animées et ont demandé à des juges d'évaluer l'entitativité perçue (i.e. l'unité, l'émergence d'une unité sociale). Il ont trouvé une relation linéaire entre la différence de rythmes des couples de figurines et la perception d'entitati-

vité [32]. L'exécution d'une tâche en synchronie semble promouvoir la coopération entre les individus [65] et améliorer la mémoire des paroles de son partenaire et l'apparence de son visage [35]. Ramseyer et Tschacher ont étudié la synchronie non verbale entre un patient et un thérapeute au cours de séances de psychothérapie et ont montré que le degré de synchronie non verbale était corrélé avec les résultats de la thérapie [50]. Bouhuys et al. ont constaté que le manque de coordination dans les comportements non-verbaux constitue un facteur de risque de récidive de la dépression [5].

3 Evaluation manuelle

Plusieurs méthodes ont été proposées pour évaluer manuellement la synchronie, allant de la micro-analyse des comportements à la perception globale de la synchronie. Les méthodes de codage du comportement proposent d'évaluer le comportement de chaque partenaire de l'interaction à une échelle très fine. Ces méthodes nécessitent l'utilisation de logiciels d'annotation (e.g., Observer ou Anvil [29]) et des observateurs qualifiés. Cappella [8] a proposé d'analyser des micro-unités de comportement (segments de parole et direction du mouvement des différentes parties du corps). Des grilles permettent d'analyser directement les comportements interactifs (sourires, regards, gestes) ou les états fonctionnels (vigilance, orientation entre les partenaires, émotion) [63, 26]. En règle générale, une mesure de synchronie est déduite de la covariation des comportements annotés. Cappella [8] et Bernieri et al. [2] ont proposé une alternative à la micro-analyse des comportements, souvent fastidieuse : la méthode par jugement global. Dans leurs études, ils ont proposé d'évaluer globalement des vidéos de nourrissons qui interagissent avec leur mère. Les observateurs évaluent les mouvements simultanés, la similitude du rythme et la coordination et la fluidité à l'échelle de l'interaction entière sur une échelle de Likert. Cappella a montré que des juges non formés étaient cohérents entre eux et fiables pour juger de la synchronie entre les partenaires [8].

4 Mesures automatiques

Des techniques automatiques ont été proposées pour capturer les signaux sociaux pertinents et évaluer la synchronie de mouvement dans les interactions humain-humain. Ces études mesurent le degré de similitude entre les comportements non verbaux des partenaires. Ces études permettent soit de comparer le degré de synchronie dans différentes conditions (e.g., avec ou sans contact visuel entre les partenaires) soit d'étudier la corrélation entre le degré de synchronie et une variable (e.g., l'amitié). Par conséquent, ces méthodes sont la plupart du temps non supervisées : la mesure de synchronie n'est pas validée en tant que tel, c'est la capacité de la mesure à prédire la variable ou à discriminer les différentes conditions qui importe. Nous décrivons ci-après les étapes classiques d'un modèle d'évaluation de la synchronie.

4.1 Caractéristiques

La première étape consiste à extraire les éléments pertinents du mouvement de la dyade. A l'exception de Delaherche et Chetouani [15], les études existantes s'intéressent à une modalité unique. On peut distinguer les études centrées sur le mouvement d'une partie du corps (mouvement du doigt [45], mouvement des yeux [51], mouvement de la main [57], mouvement des jambes [56], posture des participants [58], mouvement de la tête [1]) de celles qui capturent le mouvement d'ensemble de la dyade [4, 15, 50, 60]. Plusieurs techniques d'acquisition sont mises en avant dans la littérature : dispositifs de capture de mouvement, techniques de traitement d'image (algorithmes de suivi, de différence d'images) et capteurs physiologiques.

4.2 Mesures

Corrélation. La corrélation est certainement la méthode la plus couramment utilisée pour évaluer la synchronie. Après extraction des séries temporelles caractérisant le mouvement des partenaires, un calcul de corrélation croisée avec retard est appliqué entre les deux séries temporelles sur de courtes fenêtres d'interaction. Les cartes de corrélation sont le moyen le plus courant pour représenter les coefficients de corrélation croisée [1, 4, 60]. Le temps est représenté sur un axe, et les différents retards temporels sont représentés sur l'autre axe. Le degré de corrélation est représenté par des nuances de couleur différentes. Les cartes de corrélation ont l'avantage de montrer un aperçu global de l'interaction. Les séquences de forte synchronie sont faciles à identifier, et la différence entre deux dyades peut être perçue immédiatement.

La représentation par dendrogrammes, s'appuyant sur un clustering hiérarchique des caractéristiques extraites sur les partenaires (prosodie, geste...) a également été proposée [15]. Pour comparer quantitativement des dyades ou étudier la relation entre la synchronie et une variable de sortie (e.g., la fluidité de l'interaction), il est courant d'agréger la mesure de synchronie en méta-paramètres comme le degré de synchronie, l'orientation de la synchronie et le délai entre partenaires [1].

Analyse des récurrences. L'analyse des récurrences ("recurrence analysis") a été inspirée par la théorie des systèmes dynamiques couplés, afin de fournir des représentations graphiques de ces systèmes. L'analyse des récurrences évalue les points dans le temps où les deux systèmes transitent par des états similaires, appelés "points de récurrence". Une matrice de récurrence est créée en comparant deux à deux des vecteurs extraits des deux séries temporelles à l'aide d'une mesure de distance (e.g., euclidienne). Le carte de récurrence est la représentation en deux dimensions de la matrice de récurrence. Plusieurs paramètres permettent d'illustrer la structure de la coordination entre les deux systèmes comme le pourcentage de "points de récurrence" ou la longueur des diagonales.

Méthodes spectrales. Les méthodes spectrales constituent une alternative intéressante aux méthodes temporelles lorsqu'il s'agit de tâches rythmiques. Les méthodes spectrales mesurent l'évolution de la phase relative entre les deux partenaires comme une indication d'un décalage stable entre eux. Par exemple, Oullier et al. [45] et Richardson et al. [52] proposent de tracer l'histogramme des phases relatives pour l'interaction entière. La stabilité de la coordination interpersonnelle entre les partenaires dyadiques est indiquée par l'uniformité de la distribution de la phase. Les méthodes spectrales mesurent également la similitude des fréquences de mouvement des partenaires, appelée "cross-spectral coherence" [52, 15] ou "power spectrum overlap" [45]. Ce paramètre est mesuré par l'aire de la zone d'intersection des spectres normalisés des participants et indique la force de l'entraînement entre les deux partenaires.

4.3 Test de significativité : bootstrap

Une question cruciale lorsque l'on tente de détecter les relations de dépendance entre des caractéristiques est se savoir où placer le seuil de décision entre les degrés de synchronie significatifs et non significatifs. Une mesure de référence est nécessaire pour comparer les scores et déterminer la significativité de la mesure. Pour cela, Bernieri et al. ont proposé de synthétiser des interactions de substitution (pseudo-interactions) : les images vidéo des partenaires dyadiques sont isolés et recombinés dans un ordre aléatoire [3]. Puis des observateurs évaluent la synchronie dans les interactions originales et les pseudo-interactions. Les scores dans les pseudo-interactions constituent une mesure de référence pour juger des scores obtenus dans l'interaction originale. L'idée de produire des données de substitution et de comparer les scores de synchronie sur les ensembles de données réelles et de substitution a été étendue aux méthodes automatiques d'évaluation de la synchronie. Tout d'abord, les caractéristiques sont extraites pour chaque partenaire dyadique. La structure temporelle de la série temporelle du premier partenaire est détruite et associée avec la série originale du second partenaire. Les scores de synchronie sont évalués en utilisant les bases de données originales et de substitution. Les scores de synchronie sur l'ensemble des données de substitution constituent une base de référence pour juger de la coordination de la dyade [1, 60].

5 Modélisation dynamique de la communication

Une pratique alternative adoptée par les chercheurs du domaine consiste à modéliser et prédire l'apparition et la fréquence d'événements de haut niveau tels que des sourires, des gestes, des mouvements de tête ou des changements de locuteur. Ces événements comportementaux peuvent être soit extraits d'une base de données annotée ou prédits à partir de signaux de bas niveau extraits automatiquement à partir des données. Ces méthodes résultent d'un intérêt particulier pour comprendre les mécanismes de la dynamique de l'interaction.

5.1 Modélisation des interactions sociales comme séquences de comportements

Les méthodes d'apprentissage automatique offrent un cadre intéressant pour l'exploration de comportements interactifs. Un défi majeur est de proposer des modèles du contenu et de la structure temporelle des interactions dyadiques. Différents modèles d'apprentissage séquentiels,

tels que les modèles de Markov cachés (HMM) ou les Champs Aléatoires Conditionnels (CAC), sont généralement utilisés pour caractériser la structure temporelle des interactions sociales. Messinger et al. se concentrent sur certains signaux sociaux spécifiques (e.g., les sourires) et proposent des approches statistiques pour la caractérisation de ce signal sur la réponse du partenaire [39]. Dans [54], une approche intégrative est proposée pour caractériser explicitement la synchronie des comportements dans le cadre du diagnostic différentiel de pathologies du développement.

5.2 Prédiction des tours de parole et back-channels

Comme évoqué dans la section 2, la synchronie est liée à l'adaptation continue des comportements entre les partenaires d'interaction. Plusieurs équipes cherchent "à développer des modèles prédictifs de la dynamique de la communication intégrant les événement précédents et courants afin d'anticiper les actions les plus probables à venir de l'un ou de tous les interlocuteurs" [46]. La prédiction des tours de parole a été largement étudiée dans la perspective de construire des systèmes de dialogue plus fluides. L'objectif est de prédire avec précision le temps entre les transitions de locuteur et le type d'énoncé à venir (locuteur gardant la parole, changement de locuteur) tels qu'ils apparaissent dans les interactions humain-humain. Par exemple, Neiberg and Gustafson [43] ont proposé de prédire si un changement de locuteur aura lieu ou non à la fin d'un énoncé à l'aide de caractéristiques prosodiques, spectrales et de la durée du segment de parole précédent. Huang et al. [25] ont proposé de combiner des caractéristiques de plusieurs modalités pour prédire la fin d'un tour de parole. Une autre option pour améliorer les performances des prédicteurs de tours de parole est de regarder le comportement des deux partenaires, au lieu de se concentrer sur le comportement du locuteur seul [34]. Les back-channels sont intrinsèquement liés aux tour de rôle et comprennent entre autre des mots d'acquiescement ("hum", "aha") ou des gestes de régulation (hochements de tête, rires). Les back-channels assurent au locuteur que l'auditeur est attentif et au même niveau dans la conversation [62]. Plusieurs équipes ont étudié la façon dont le comportement du locuteur déclenche les back-channels de l'auditeur. Gravano et al. ont étudié la façon dont l'intonation, l'intensité, la fréquence fondamentale, la qualité de la voix et de la durée inter-pauses accompagnaient les mots d'acquiescement [23]. Morency et al. ont proposé d'étudier quelle caractéristique du locuteur (prosodie, pause, mots, regard) est importante pour prédire l'apparition et le timing des hochements de tête de l'auditeur [42].

6 Questions ouvertes et perspectives

6.1 Questions ouvertes

Plusieurs questions relatives à la dimension et à la perception de la synchronie restent à explorer. Ces questions sont fondamentales pour le développement d'un modèle automatique pour évaluer la synchronie. Une première question se rapporte à l'échelle de temps de la synchronie : seconde, minute, toute l'interaction. Est-il approprié d'étudier de courts fragments de comportement ? Est-il possible de comptabiliser des occurrences de synchronie ? Une deuxième question concerne la dimension de la synchronie : la synchronie est-elle une notion discrète ou continue [24] ? Diverses sources indiquent que la synchronie varie au cours de l'interaction, étant plus forte au début et à la fin d'un échange [27] ou à des moments d'engagement particulier [7]. La question du corpus est également cruciale. Jusqu'à la contribution récente de Sun et al. [59] et leur base de données de mimétisme, aucun corpus annoté accessible au public n'a été consacré à la détection de la synchronie.

6.2 Perspectives en robotique développementale

Dans la dernière décennie, les chercheurs dans le domaine de la robotique, du traitement du signal et de l'intelligence artificielle ont développé un intérêt croissant pour les phénomènes du développement, comme la synchronie parent-enfant, l'acquisition du langage et l'attention conjointe [37]. L'objet de la robotique développementale est de permettre aux robots et autres systèmes artificiels de développer de façon autonome des compétences utilisables dans n'importe quelle situation plutôt que de les programmer à résoudre des objectifs particuliers dans un environnement spécifique. Prepin et Gaussier [49] ont ainsi proposé une architecture robotique (ADRIANA) capable de mesurer le degré de synchronie avec un humain et d'adapter son comportement en conséquence.

6.3 Perspectives pour les robots sociaux et agents conversationnels animés

Gratch et al. ont évalué l'importance de la contingence, une condition sine qua non de la synchronie, dans différentes situations d'interactions humain-humain et humain-agent virtuel [22]. Ils ont comparé la sympathie des participants pour un agent réactif et un agent non contingent. Les chercheurs ont trouvé que la contingence des réactions de l'agent influençait le comportement du participant et était impliquée dans l'établissement d'une bonne entente avec l'agent. Michalowski et al. ont conçu le robot Keepon, pour initier des interactions synchrones avec des enfants [40]. Keepon est programmé pour danser en effectuant des mouvements périodiques qui évoluent en douceur au rythme perçu par le robot. Le rythme peut être extrait de différents capteurs en fonction des conditions de test (vision, audio, capteurs de pression, des accéléromètres, etc.). Les auteurs ont étudié l'effet de mouvements synchronisés sur l'engagement dans différentes conditions où le robot suit le rythme de la musique ou s'adapte au mouvement de l'enfant. Ils ont également effectué une étude longitudinale auprès d'enfants autistes interagissant avec Keepon [30].

6.4 Perspectives dans les études développementales et cliniques

Dans le domaine de la psychiatrie de l'enfant, de nombreux avantages potentiels à l'utilisation de robots interactifs dans les milieux cliniques avec des individus atteints de Troubles du Spectre Autistique (TSA) ont été envisagés. Ces avantages comprennent l'attrait intrinsèque de la technologie pour des individus du spectre autistique, la capacité des robots à produire des comportements sociaux simples, isolées et répétitifs, et le fait qu'ils peuvent être adaptés pour fournir un traitement individualisé (pour une revue voir [17]). Cependant, malgré l'intérêt des médias, la recherche dans ce domaine a été seulement exploratoire visant à évaluer la préférence pour la machine ou à utiliser un robot pour susciter des comportements, travailler une compétence ou fournir un feedback. Tartaro et al. ont proposé de concevoir des compagnons virtuels pour aider les enfants TSA à acquérir des compétences de communication [61]. Les auteurs ont observé que, par rapport à une interaction avec un homologue humain, les enfants TSA ont produit plus de réponses contingentes avec le compagnon virtuel.

7 Conclusion

L'évaluation de la synchronie comporte des questions complexes à la frontière de plusieurs domaines de recherche. Les méthodes d'évaluation manuelle sont essentielles pour les ingénieurs afin d'identifier les signaux pertinents, de valider les méthodes d'apprentissage automatique qui mesurent la synchronie ou modélisent des patterns interactifs. De nouvelles interfaces socialement adaptées pourraient émerger d'une meilleure analyse de ces mécanismes sociaux. En retour, les psychologues pourraient bénéficier de méthodes d'évaluation automatique et objective de la synchronie pour l'étude de troubles psychiatriques, comme la dépression et l'autisme. Bien que peu d'études soient actuellement disponibles dans ce domaine spécifique, elles semblent déjà très prometteuses (succès de la psychothérapie [50], interaction mère-enfant [12]). Un autre potentiel réside dans la possibilité de construire des robots ou des agents virtuels avec des capacités interactives. De telles interfaces permettraient de contrôler les variables de l'interaction et de tester différents scénarios et comportements. De telles études pourraient bénéficier à la fois aux ingénieurs et aux cliniciens et conduire à une meilleure compréhension des mécanismes sous-jacents aux interactions sociales.

Remerciements

Ces travaux sont financés par les programmes UPMC Emergence 2009, European Union Seventh Framework (subvention n°288241) et Syned-Psy.

Références

[1] K. T. Ashenfelter, S. M. Boker, J. R. Waddell, and N. Vitanov. Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *J Exp Psychol Hum Percept Perform*, 35(4) :1072–91, 2009.

[2] F. Bernieri, J. Reznick, and R. Rosenthal. Synchrony, pseudo synchrony, and dissynchrony : Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology*, 54(2) :243–253, 1988.

[3] F. Bernieri and R. Rosenthal. *Interpersonal coordination : Behavior matching and interactional synchrony. Fundamentals of nonverbal behavior*. Cambridge University Press, 1991.

[4] S. M. Boker and J. L. Rotondo. *Symmetry building and symmetry breaking in synchronized movement*, volume 42, pages 163–171. Amsterdam, Netherlands : John Benjamins Publishing Company, 2002.

[5] A. L. Bouhuys and M. M. Sam. Lack of coordination of nonverbal behaviour between patients and in-

terviewers as a potential risk factor to depression recurrence : vulnerability accumulation in depression. *Journal of Affective Disorders*, 57(1-3) :189 – 200, 2000.

[6] H. P. Branigan, M. J. Pickering, and A. A. Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75(2) :B13–B25, May 2000.

[7] N. Campbell. An audio-visual approach to measuring discourse synchrony in multimodal conversation data. In *Interspeech*, pages 2159–2162, 2009.

[8] J. Cappella. Behavioral and judged coordination in adult informal social interactions : vocal and kinesic indicators. *Pers. Soc. Psychol.*, 72 :119–131, 1997.

[9] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Conversation as a system framework : Designing embodied conversational agents. In *Embodied Conversational Agents*, pages 29–63. MIT Press, 2000.

[10] T. L. Chartrand and J. A. Bargh. The chameleon effect : The perception-behavior link and social interaction. *Journal of personality and social psychology*, 76(6) :893–910, June 1999.

[11] H. H. Clark. *Using Language*, volume 23. Cambridge University Press, 1996.

[12] J. F. Cohn. Advances in behavioral science using automated facial image analysis and synthesis. *IEEE Signal Processing Magazine*, 27(November) :128–133, 2010.

[13] W. Condon and W. Ogston. A segmentation of behavior. *Journal of Psychiatric Research*, 5 :221–235, 1967.

[14] W. Condon and L. Sander. Neonate movement is synchronized with adult speech : interactional participation and language acquisition. *Science*, 183 :99–101, 1974.

[15] E. Delaherche and M. Chetouani. Multimodal coordination : exploring relevant features and measures. In *Second International Workshop on Social Signal Processing, ACM Multimedia 2010*, 2010.

[16] E. Delaherche, M. Chetouani, M. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen. Interpersonal synchrony : A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 2012. to appear.

[17] J. J. Diehl, L. M. Schmitt, M. Villano, and C. R. Crowell. The clinical use of robots for individuals with autism spectrum disorders : A critical review. *Research in Autism Spectrum Disorders*, In Press :–, 2011.

[18] R. Feldman. Infant-mother and infant-father synchrony : the coregulation of positive arousal. *Infant Mental Health Journal*, 24(1) :1–23, 2003.

[19] S. Garrod and A. Anderson. Saying what you mean in dialogue : a study in conceptual and semantic coordination. *Cognition*, 27(2) :181–218, 1987.

[20] H. Giles, J. Coupland, and N. Coupland. *Accommodation theory : Communication, context, and consequence*, pages 1–68. Number 1984. Cambridge University Press, 1991.

[21] M. H. Goldstein, A. P. King, and M. J. West. Social interaction shapes babbling : Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13) :8030–8035, 2003.

[22] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In *IVA '07 : Proceedings of the 7th international conference on Intelligent Virtual Agents*, pages 125–138. Springer-Verlag, 2007.

[23] A. Gravano and J. Hirschberg. Backchannel-inviting cues in task-oriented dialogue. In *INTERSPEECH*, pages 1019–1022. ISCA, 2009.

[24] A. Harrist and R. Waugh. Dyadic synchrony : Its structure and function in children‘s development. *Developmental Review*, 22(4) :555–592, 2002.

[25] L. Huang, L.-P. Morency, and J. Gratch. A multimodal end-of-turn prediction model : learning from parasocial consensus sampling. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, AAMAS '11, pages 1289–1290, Richland, SC, 2011.

[26] C. J. *A Procedure to Measure Interactional Synchrony in the Context of Satisfied and Dissatisfied Couples' Communication*, pages 199–208. Lawrence Erlbaum, 2005.

[27] A. Kendon. Movement coordination in social interaction : some examples described. *Acta Psychologica*, 32 :100–125, 1970.

[28] J. Kim-Cohen, G. A. Light, D. L. Braff, C. L. M. Caton, R. E. Drake, D. S. Hasin, P. E. Shrout, S. Samet, W. B. Schanzer, T. E. Moffitt, A. Taylor, S. J. Pawlby, A. Caspi, J. M. Hettema, C. A. Prescott, J. M. Myers, M. C. Neale, K. S. Kendler, M. R. Liebowitz, A. J. Gelenberg, and D. Munjack. Maternal depression and children's antisocial behavior : Nature and nurture effects. *Archives of General Psychiatry*, 62 :173–181, 2005.

[29] M. Kipp. Spatiotemporal coding in anvil. In *LREC*, 2008.

[30] H. Kozima, M. Michalowski, and C. Nakagawa. Keepon. *International Journal of Social Robotics*, 1 :3–18, 2009. 10.1007/s12369-008-0009-8.

[31] M. Lafrance. *Posture Mirroring and Rapport*, pages 279–298. Human Sciences Press, New York, NY, 1982.

[32] D. Lakens. Movement synchrony and perceived entitativity. *Journal of Experimental Social Psychology*, 46(5) :701 – 708, 2010.

[33] J. L. Lakin and T. L. Chartrand. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14(4) :334–339, 2003.

[34] C.-C. Lee and S. Narayanan. Predicting interruptions in dyadic spoken interactions. In *ICASSP'10*, pages 5250–5253, 2010.

[35] C. N. Macrae, O. K. Duffy, L. K. Miles, and J. Lawrence. A case of hand waving : Action synchrony and person perception. *Cognition*, 109(1) :152 – 156, 2008.

[36] A. N. Meltzoff, R. Brooks, A. P. Shon, and R. P. N. Rao. "Social" robots are psychological agents for infants : A test of gaze following. *Neural Networks*, 23(8-9) :966–972, 2010.

[37] A. N. Meltzoff, P. K. Kuhl, J. Movellan, and T. J. Sejnowski. Foundations for a new science of learning. *Science*, 325(5938) :284–288, 2009.

[38] A. N. Meltzoff and M. K. Moore. Imitation et développement humain : les premiers temps de la vie. *Terrain*, 44 :71–90, 2005.

[39] D. M. Messinger, P. Ruvolo, N. V. Ekas, and A. Fogel. Applying machine learning to infant interaction : The development is in the details. *Neural Networks*, 23(8-9) :1004 – 1016, 2010. Social Cognition : From Babies to Robots.

[40] M. Michalowski, R. Simmons, and H. Kozima. Rhythmic attention in child-robot dance play. In *Proceedings of RO-MAN 2009*, 2009.

[41] L. K. Miles, L. K. Nind, and C. N. Macrae. The rhythm of rapport : Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, 45(3) :585 – 589, 2009.

[42] L.-P. Morency, I. Kok, and J. Gratch. Predicting listener backchannels : A probabilistic multimodal approach. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA '08, pages 176–190, Berlin, Heidelberg, 2008. Springer-Verlag.

[43] D. Neiberg and J. Gustafson. Predicting speaker changes and listener responses with and without eye-contact. In *INTERSPEECH*, pages 1565–1568, 2011.

[44] B. Newman and P. Newman. *Development through life : A psychosocial approach*, pages 171–175. Cengage/Wadsworth, 2009.

[45] O. Oullier, G. C. de Guzman, K. J. Jantzen, J. A. Scott Kelso, and J. Lagarde. Social coordination dynamics : Measuring human bonding. *Social Neuroscience*, 3(2) :178–192, 2008.

[46] D. Ozkan, K. Sagae, and L.-p. Morency. Latent mixture of discriminative experts for multimodal prediction modeling. *Computational Linguistics*, (August) :860–868, 2010.

[47] J. Piaget. *La formation du symbole chez l'enfant : imitation, jeu et rêve, image et représentation*. Actualités pédagogiques et psychologiques. Delachaux et Niestlé, 1976.

[48] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02) :169–190, 2004.

[49] K. Prepin and P. Gaussier. How an agent can detect and use synchrony parameter of its own interaction with a human ? In A. Esposito et al., editors, *Development of Multimodal Interfaces : Active Listening and Synchrony*, volume 5967, pages 50–65. Springer Berlin / Heidelberg, 2010.

[50] F. Ramseyer and W. Tschacher. Nonverbal synchrony in psychotherapy : Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, 79(3) :284 – 295, 2011.

[51] D. C. Richardson and R. Dale. Looking to understand : The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6) :1045–1060, 2005.

[52] M. J. Richardson, K. L. Marsh, R. W. Isenhower, J. R. Goodman, and R. Schmidt. Rocking together : Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26(6) :867 – 891, 2007.

[53] M. Rutter and T. G. O'Connor. Are there biological programming effects for psychological development ? Findings from a study of Romanian adoptees. *Developmental Psychology*, 40(1) :81–94, 2004.

[54] C. Saint-Georges, A. Mahdhaoui, M. Chetouani, M. Laznik, F. Apicella, P. Muratori, S. Maestro, F. Muratori, and D. Cohen. Do parents recognize autistic deviant behavior long before diagnosis ? taking into account interaction using computational methods. *PLOS ONE*, 6(7) :e22393, 2011.

[55] D. S. Schechter and E. Willheim. Disturbances of attachment and parental psychopathology in early childhood. *Child and Adolescent Psychiatric Clinics of North America*, 18(3) :665 – 686, 2009.

[56] R. Schmidt, C. Carello, and M. Turvey. Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology : Human Perception and Performance*, 16(2) :227 – 247, 1990.

[57] R. C. Schmidt, B. O'Brien, and R. Sysko. Self-organization of between-persons cooperative tasks and possible applications to sport. *Int J Sport Psychol.*, 30 :558–579, 1999.

[58] K. D. Shockley, A. A. Baker, M. J. Richardson, and C. A. Fowler. Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology : Human Perception and Performance*, (33) :201–208, 2007.

[59] X. Sun, J. Lichtenauer, M. Valstar, A. Nijholt, and M. Pantic. A multimodal database for mimicry analysis. In S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, editors, *Affective Computing and Intelligent Interaction, Part I*, volume 6974 of *Lecture Notes in Computer Science*, pages 367–376, Berlin, Germany, October 2011. Springer Verlag.

[60] X. Sun, K. P. Truong, M. Pantic, and A. Nijholt. Towards visual and vocal mimicry recognition in human-human interactions. In E. Tunstel, S. Nahavandi, and A. Stoica, editors, *IEEE International Conference on Systems, Man, and Cybernetics, SMC 2011 : Special Session on Social Signal Processing*, pages 367–373, USA, November 2011. IEEE Computer Society.

[61] A. Tartaro and J. Cassell. Playing with virtual peers : bootstrapping contingent discourse in children with autism. In *Proceedings of the 8th international conference on International conference for the learning sciences - Volume 2*, ICLS'08, pages 382–389. International Society of the Learning Sciences, 2008.

[62] K. R. Thorisson. *Natural Turn-Taking Needs No Manual : Computational Theory And Model, From Perception to Action*, pages 173–207. Kluwer Acedemic Publishers, Dordrecht, The Netherlands, 2002.

[63] C. Trevarthen and S. Daniel. Disorganized rhythm and synchrony : Early signs of autism and Rett syndrome. *Brain and Development*, 27(1) :S25–S34, 2005.

[64] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing : Survey of an emerging domain. *Image and Vision Computing*, 27(12) :1743–1759, November 2009.

[65] S. S. Wiltermuth and C. Heath. Synchrony and cooperation. *Psychological Science*, 20(1) :1 – 5, 2009.

Modélisation de dialogues narratifs pour la conception d'un ACA narrateur

Alexandre Pauchet ⋆ François Rioult † Émilie Chanoni ‡ Zacharie Alès ⋆ Ovidiu Şerban ⋆

⋆ INSA Rouen - LITIS, {prenom.nom}@insa-rouen.fr

† Université de Caen - Greyc, francois.rioult@unicaen.fr

‡ Université de Rouen - Psy-NCA, emilie.chanoni@univ-rouen.fr

Résumé :
Dans l'optique de la conception d'un Agent Conversationnel Animé (ACA) narratif et affectif, cet article démontre l'importance de l'interaction dans le processus de narration à des enfants. Un corpus de 30 dialogues de narration entre des parents et leur enfant de 3 à 5 ans, a été annoté à l'aide d'une grille dite "mentaliste". Deux méthodes d'extraction de connaissances ont été appliquées aux annotations des dialogues afin de les modéliser. Celles-ci révèlent des régularités dans les explications données par les parents dans la description des émotions des personnages. Ces travaux fournissent une ligne directrice pour la conception du modèle dialogique d'un ACA narrateur affectif.

Mots-clés : Modélisation du dialogue ; Extraction de connaissances ; Agent Conversationnel Animé

1 Introduction

La conception d'un modèle de dialogue est une tâche difficile et souvent pluridisciplinaire. Qu'il soit dédié à la narration interactive ou non, il implique de nombreux mécanismes : traitement de signaux multimodaux (parole, gestes, regards, etc.), reconnaissance et génération de langage naturel, gestion du dialogue, modélisation des émotions, prosodie et comportement non verbal. En particulier, la gestion de la multimodalité et des émotions dans le dialogue reste à ce jour insuffisante au sein des Agents Conversationnels Animés (ACA) [Cassell et al., 2000], bien que ces aspects soient essentiels pour des interactions efficaces [Swartout et al., 2006].

Avec l'émergence des environnements numériques, et plus particulièrement de systèmes de narration participative, les situations d'interactions enfant-agents humanoïdes sont de plus en plus fréquentes. Le comportement d'un ACA dédié à la narration interactive, devrait correspondre aux standards d'interaction adulte-enfant afin de faciliter la compréhension de ce dernier. En particulier, le modèle dialogique de l'agent devrait être adapté aux compétences socio-cognitives et langagières de l'enfant.

Le but de cet article est de proposer une méthode de modélisation du dialogue appliquée à des dialogues de narration entre parents et enfants. Nous souhaitons fournir des outils permettant de faciliter l'extraction de régularités à partir de dialogues interactifs. Pour ce faire, nous proposons, d'une part d'extraire des motifs dialogiques à partir d'un corpus, et d'autre part une méthode prédictive permettant de guider une session de narration interactive, comme modèle de dialogue pour ACA narrateurs.

Un état de l'art sur le dialogue homme-machine est présenté section 2. La méthode ainsi que le corpus de dialogues sont décrits section 3. Les détails techniques des deux procédures sont décrits section 4 et le modèle extrait est analysé et expliqué section 5. L'article se termine par une courte conclusion ainsi qu'un exposé de nos futurs travaux.

2 Modèles du dialogue pour ACA narrateur : un (court) état de l'art

Les ACA sont des interfaces autonomes et anthropomorphiques, incarnés par des personnages animés aux compétences multi-modales : langage naturel, expressions du visage, regards, attitudes et gestes [Cassell et al., 2000]. Les ACA peuvent être catégorisés selon leur expressivité : systèmes de type présentateurs télé, interactions face à face (un ACA avec un humain) et conversations multipartites (plusieurs ACA et utilisateurs) [André and Pelachaud, 2010]. Les projets de recherche récents se focalisent sur l'interactivité des ACA en perfectionnant les expressions faciales et le comportement non verbal afin d'améliorer la qualité générale de l'agent [Cassell et al., 2000, Pelachaud, 2009]. Greta [Pelachaud, 2009], MARC [Courgeon et al., 2009] et le projet européen SEMAINE [Schröder, 2010] sont de bons exemples des capacités actuelles des ACA.

Dans le domaine de la narration, les agents virtuels intelligents, qu'ils soient incarnés ou non, peuvent être utilisés en tant que personnages

expressifs (ex : [Seif El-Nasr and Wei, 2008]) ou en tant que narrateurs. Le projet GV-LEx (Gesture and voice for an expressive reading), par exemple, a pour but de fournir aux robots Nao [Gouaillier et al., 2009] et à l'ACA Greta [Pelachaud, 2009] la capacité de lire du texte sans ennuyer l'auditeur [Gelin et al., 2010]. Ce projet propose d'utiliser une intonation expressive ainsi que des gestes tout en parlant afin de produire des narrations crédibles.

En ce qui concerne les systèmes et modèles du dialogue pouvant être intégrés dans les ACA, plusieurs approches existent.
L'approche *à états finis* (voir par exemple [McTear, 2004]) qui représente la structure du dialogue par un automate à états finis dans lequel chaque énoncé conduit à un nouvel état. En pratique, cette approche est limitée aux systèmes de dialogue directifs.
L'approche *par formulaire* représente le dialogue comme un processus de remplissage de formulaire contenant des entrées prédéfinies (voir par exemple [Aust et al., 1995]). Les contributions possibles sont fixées à l'avance.
L'approche *par planification* (exemple : [Allen and Perrault, 1980]) combine la reconnaissance de plans et la théorie des Actes de Langage [Searle, 1969]. Cette approche est complexe du point de vue calculatoire et requiert des composants très avancés de TAL afin d'inférer les intentions du locuteur.
Le framework ISU (*Information State Update*) [Larsson and Traum, 2000] utilise une représentation formelle du terrain commun, l'*état d'information*, ainsi qu'une structure gérant le raisonnement de l'agent.
L'approche *logique* représente le dialogue et son contexte par un formalisme logique et utilise des mécanismes tels que l'inférence et les jeux de dialogue (voir par exemple [Hulstijn, 2000]). La plupart des travaux concernant les approches logiques ne sont actuellement qu'au stade de la théorie.
Les approches *par apprentissage* proposent des techniques telles que l'apprentissage par renforcement [Frampton and Lemon, 2009] afin de modéliser le dialogue via des processus de Markov. Ces approches requièrent un travail d'annotation considérable.

En raison de la complexité des systèmes de dialogue complet, la plupart des ACA existants n'intègrent que des processus basiques de gestion du dialogue, telles que la détection de mots-clés au sein d'une approche de type états finis ou frame-based (ex : le projet SEMAINE [Schröder, 2010]). La gestion du dialogue reste donc inefficace dans les ACA actuels [Swartout et al., 2006]. Les approches précédentes utilisent comme représentation des structures de données régulières (ex : l'automate [McTear, 2004]), extraites manuellement ou apprises automatiquement à partir d'un corpus de dialogues, de traces ou de fichiers log. Ces structures de données ne permettent de représenter que des motifs d'interactions linéaires. Cependant la gestion du dialogue implique plusieurs dimensions et non une seule [Bunt, 2011]. Le modèle d'interaction d'un ACA nécessite la gestion de tous les aspects associés aux interactions humaines (gestion de tâches individuelles et collectives, feedbacks, aspects affectifs, obligations sociales, etc.), exprimés suivant différentes modalités (sémantique, prosodie, gestes, expressions, etc.).

En comparaison, le modèle de dialogue proposé dans la suite de cet article combine planification au service de la résolution de la tâche - prédiction et planification des interventions de l'enfant - et une gestion plus réactive par jeux de dialogue pour les conventions/motifs dialogiques. La représentation matricielle encodant le dialogue permet en plus de tenir compte du caractère multidimensionnel du dialogue.

3 Modélisation de dialogues

La méthode de modélisation du dialogue proposée est présentée Figure 1 :

1. *collecte et numérisation* d'un corpus de dialogues au format audio ou vidéo. Le corpus que nous considérons dans cet article est composé d'histoires enfantines racontées par des parents à leur enfant ;
2. l'étape *transcription et codage* consiste à produire des données brutes à divers niveaux de détails (tours de parole, énoncés, onomatopées, pauses, etc.) selon les caractéristiques que l'on souhaite exhiber ;
3. une phase d'*extraction de connaissances*, suivant un schéma de codage spécifique, est ensuite appliquée aux énoncés encodés afin d'obtenir une description précise des comportements dialogiques. Les dialogues sont alors considérés comme annotés.
4. une phase d'*extraction de régularités* (modélisation) est appliquée aux annotations.
5. le modèle peut alors être *exploité* [1].

1. Cette étape n'est pas présentée dans cet article.

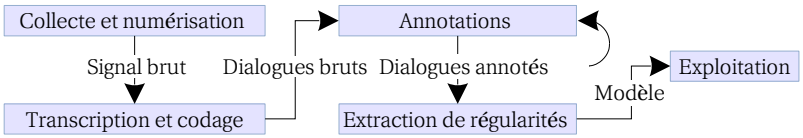


FIGURE 1 – Analyse du dialogue

Ligne	Locuteur	Énoncé	Annotations				
25	P	T'inquiète pas	A	P	E	-	-
26	P	Donc là ils se cachent	A	P	B	-	-
27	P	Ils cherchent	A	-	F	-	-
28	P	qui pourrait avoir pris la couronne.	Q	-	F	-	-
29	E	Elle est dedans, elle est dedans la couronne.	A	-	F	-	-
30	P	Donc là ils suspectent plein de monde, Cornélius, Céleste, la vieille dame...	A	P	Y	C	J
31	P	Qui a bien pu prendre la couronne ?	Q	-	F	-	-
32	E	La couronne elle est dedans.	A	-	F	-	-
33	P	Tu crois ? !	Q	H	K	-	-
34	E	Oui.	A	-	F	-	-
35	P	Mais Babar il ne sait pas qu'elle est dedans.	A	P	N	O	J
36	P	Donc il se dit que c'est une bombe, la couronne	A	P	N	C	J
37	P	ou je ne sais quoi.	A	R	N	-	-

TABLE 1 – Représentation matricielle des annotation d'un dialogue narratif parent-enfant

Dans les sous-sections suivantes, nous présentons l'expérimentation réalisée pour collecter le corpus de dialogues de narration d'histoires enfantines (étape 1) ainsi que le schéma de codage utilisé pour obtenir une représentation matricielle des dialogues (étape 2).

3.1 Corpus de dialogues narratifs

La narration d'histoires d'un parent à son enfant est une situation classique participant au développement de l'enfant. Les contextes sociaux et langagiers apportés par l'adulte, sont nécessaires à l'enfant dans son processus d'apprentissage des compétences socio-communicatives, cognitives et morales. Les enfants développent une *théorie de l'esprit* [Astington and Baird, 2005] durant leurs premières années et deviennent ainsi capables d'assimiler le fait qu'une personne est déterminée par ses propres intentions, émotions et états mentaux. Ce développement n'est possible qu'au travers des situations sociales de dialogue. Le discours des adultes concernant les états mentaux, en particulier, se révèle être un médiateur d'apprentissage du concept de cognition sociale - grâce à une participation active au dialogue et à des interactions dynamiques.

Dans cette étude, nous utilisons un corpus de 90 dialogues entre parents et enfants âgés de 3, 4 et 5 ans, filmés en situation de récit d'histoires enfantines (10 enfants par tranche d'âge x 3 histoires différentes). Ces enregistrements sont retranscrits et annotés suivant une *grille mentaliste* [Chanoni, 2009] afin de faire ressortir les informations relatives aux états mentaux (croyances, volition, émotions, etc.) contenues dans les énoncés. La longueur moyenne des dialogues est de 89,3 énoncés.

3.2 Représentation des dialogues

Comme le souligne Bunt, la gestion du dialogue est multi-niveaux [Bunt, 2011]. Afin de concevoir un modèle du dialogue multi-dimensionnel, les annotations sont représentées matriciellement. Chaque énoncé est caractérisé par un vecteur d'annotations dont les composantes correspondent aux différentes dimensions de codage : une ligne par énoncé et une colonne par espace/dimension de codage.

Le Tableau 1 présente un exemple de dialogue provenant du corpus collecté. Chaque énoncé est caractérisé par un numéro de ligne, un locuteur (P : parent, E : enfant), une transcription et des annotations encodées suivant 5 dimensions :

– la première colonne caractérise la nature de l'énoncé : une (A)ffirmation, une (Q)uestion, une demande d'attention - générale (G) ou concernant l'histoire (D) ;
– la seconde colonne définit la référence de l'énoncé. Il peut se référer à un personnage (P), à l'auditeur (H) ou au narrateur (R) ;
– la troisième colonne est dédiée aux états mentaux. Les interlocuteurs peuvent exprimer une

(E)motion, une (V)olition, une cognition observable (B) ou non (N), une déclaration épistémique (K), une hypothèse (Y) ou une (S)urprise. La surprise se distingue des autres émotions de par son lien avec les croyances ;
– les deux dernières colonnes représentent les explications par (C)ause/conséquence, (O)pposition ou empathie (M), qui peuvent être utilisées, soit pour expliquer l'histoire (J), soit pour préciser une situation par l'évocation d'un contexte personnel (F).

Par exemple, la ligne 35 est encodée ainsi : l'énoncé est une affirmation (A) portant sur un état mental se référant à un personnage - "*Babar*" - (P) ; l'état mental correspondant - "*sait*" - se réfère à une cognition non observable (N) ; "*Mais*" dénote une justification par opposition (O) ; enfin, l'énoncé se réfère à l'histoire (J).

La construction de cette représentation matricielle nécessite un processus d'annotation manuel, semi-automatique et/ou automatique - un pour chaque dimension/colonne. Les matrices obtenues peuvent également être vues comme des séquences de vecteurs d'annotations. Une fois les dialogues annotés, l'étape d'extraction de régularités est appliquée sur les matrices.

4 Extraction de régularités pour la modélisation du dialogue

Nous proposons deux approches d'extraction de régularités : un calcul de similarité par programmation dynamique permettant de collecter des motifs dialogiques et une méthode de prédiction d'événements se concentrant sur la caractérisation des interactions de l'enfant.

4.1 Extraction de motifs dialogiques

Avec notre représentation matricielle, un *motif dialogique* est défini comme un ensemble d'annotations dont la disposition apparaît - de manière exacte ou approchée - dans plusieurs dialogues. Un motif peut contenir des annotations non adjacentes en ligne ou en colonne (i.e. il peut avoir des trous), et deux instances d'un même motif peuvent contenir des insertions, des suppressions ou des substitutions de caractères. Deux motifs sont donc considérés comme similaires si leur distance d'édition est faible.

La figure 2 présente la méthode utilisée pour extraire un ensemble de motifs dialogiques pertinents. Elle est composée d'une extraction de régularités basée sur un alignement de matrices par programmation dynamique, permettant de collecter un ensemble de paires de motifs similaires, suivi d'une étape de clustering afin de regrouper les motifs dialogiques récurrents.

La méthode d'extraction de motifs en deux dimensions s'apparente à l'alignement de matrices. Il s'agit d'une généralisation de la distance d'édition locale entre deux vecteurs de caractères. La distance d'édition *ed* (ou distance de Levenshtein) entre deux vecteurs de caractères s_1 et s_2 correspond au coût minimal des trois opérations d'édition élémentaires (insertion et suppression de caractères, ainsi que la substitution d'un caractère par un autre) permettant de transformer s_1 en s_2. Un alignement local de deux matrices de caractères s_1 et s_2, de tailles respectives $m_1 \times n_1$ et $m_2 \times n_2$, consiste à chercher les portions de s_1 et s_2 qui sont les plus similaires (parmi toutes les portions de s_1 et s_2). Pour ce faire, une table à 4 dimensions T de taille $(m_1+1)\times(n_1+1)\times(m_2+1)\times(n_2+1)$ est calculée, de telle sorte que $T[i][j][k][l]$ soit égal à la distance d'édition locale entre $S_1[0..i-1][0..j-1]$ et $S_2[0..k-1][0..l-1]$, $\forall i \in [\![1, m_1-1]\!]$, $j \in [\![1, n_1-1]\!]$, $k \in [\![1, m_2-1]\!]$ et $l \in [\![1, n_2-1]\!]$. Dans notre méthode, le calcul de T est obtenu par minimisation d'une formule de récurrence. Une fois T calculée, le meilleur alignement local est obtenu en effectuant un algorithme de tracé arrière à partir de la position où T atteint sa valeur maximale. Ce tracé arrière permet d'inférer les caractères faisant partie de l'alignement. La figure 3, commentée en section 5, présente un exemple d'alignement issu de notre corpus. Pour de plus amples informations concernant l'extraction de motifs en deux dimensions, se référer à [Lecroq et al., 2012].

L'alignement de matrices permet d'extraire les motifs par paires. Nous les regroupons à l'aide d'algorithmes heuristiques de clustering (voir tableau 2). L'idée sous-jacente est que les clusters les plus conséquents représentent les comportements les plus communs, tandis que les petits clusters reflètent des comportements plus marginaux. Une matrice de similarité entre les différents motifs est calculée grâce à une distance d'édition globale appliquée aux paires de motifs détectés. Cette matrice est utilisée comme entrée des algorithmes de clustering.

Cette méthode a été testée sur le corpus de dialogues de narration. Durant la phase d'extraction, 1740 motifs dialogiques ont été collectés,

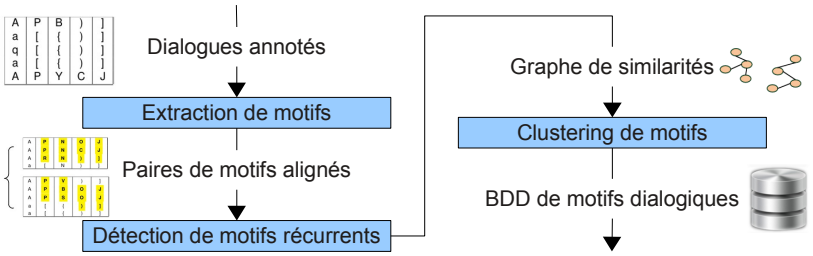


FIGURE 2 – Extraction de motifs dialogiques.

Méthode	Référence	Nombre de clusters trouvés					
		5	20	50	80	116	150
Single-Link	[Florek et al., 1951]	41	97	183	270	320	360
CHAMELEON	[Karypis et al., 1999]	458	605	**628**	-	-	-
ROCK	[Guha et al., 2000]	520	600	621	626	629	630
Spectral clustering non normalisé	[Von Luxburg, 2007]	277	**658**	563	155	194	226
Spectral clustering selon Shi and Malik	[Von Luxburg, 2007]	524	615	**628**	**631**	631	**632**
Spectral clustering selon Jordan and Weiss	[Von Luxburg, 2007]	**555**	616	**628**	630	631	**632**
Propagation d'affinité	[Frey and Dueck, 2007]	-	-	-	-	**632**	-

TABLE 2 – Indice de Dunn en fonction du nombre de clusters pour les heuristiques implémentées. Le caractère '-' est utilisé lorsqu'une solution n'est pas produite pour un nombre de clusters donné. Les valeurs surlignées correspondent aux meilleur(s) résultat(s) pour chaque colonne.

de taille variant entre 10 et 124 énoncés pour une moyenne de 28,9.

Le nombre de solutions des méthodes de clustering étant trop élevé pour une comparaison au cas par cas, l'indice de Dunn [Dunn, 1973] a été utilisé afin d'évaluer les méthodes. Si s_{ij} mesure la similarité entre deux motifs i et j, et $c(i)$ est le nombre de clusters contenant i pour une solution donnée, l'indice de Dunn est égal à

$$\frac{\min_{c(i)\neq c(j)} s_{ij}}{\max_{c(k)=c(l)} s_{kl}}$$

Ainsi, les solutions comportant un indice de Dunn élevé sont susceptibles d'être pertinentes, car composées de clusters compacts et clairement séparés. Le nombre de clusters étant lui-même inconnu, les méthodes ont été testées sur un grand nombre de valeurs. Le tableau 2 présente une partie représentative des résultats de l'indice de Dunn. Les meilleures méthodes semblent être la propagation d'affinité et les méthodes de type spectral clustering.

4.2 Prédiction des interventions de l'enfant

Cette section est consacrée à la définition d'un modèle du dialogue permettant de stimuler l'interaction. Dans cette optique, les interventions de l'enfant doivent être finement modélisées en se concentrant sur la *prédiction d'événement*. Nous recherchons des séquences d'événements dialogiques entraînant une interaction particulière afin de s'en servir comme plan pour générer les interventions d'un ACA narrateur.

Les contributions majeures sur l'extraction de connaissances à partir de séquences sont principalement consacrées aux épisodes et à la classification de séquences. La prédiction d'événement sur des données discrètes n'y est que très peu abordée [Antunes and Oliveira, 2001]. Nous proposons de découper les données en *tours de parole*, caractérisés par un ensemble d'énoncés successifs provenant d'une seule personne (ici le parent ou l'enfant). Le problème revient à prévoir la fin du tour. Dans ce but, nous considérons des séquences de séries de vecteurs d'annotations (une série de vecteurs d'annotations par tour de parole) se terminant par une intervention de l'enfant. Les séquences du tableau 1 sont : $< (APE)(APB)(AF)(QF) >$, $< (APYCJ)(QF) >, < (QHK) >, ...$

Pour extraire les régularités menant à la fin des séquences, les épisodes sont explorés par projections récursives grâce à un algorithme glouton. Dans l'exemple ci-dessus, l'algorithme débute avec l'épisode $< (Q) >$, commun à toutes les fins de séquences. L'algorithme est ensuite appelé une nouvelle fois sur les séquences projetées $< (APE)(APB)(AF) >, < (APYCJ) >$. (A) est ajouté à l'épisode qui devient $< (Q)(A) >$, qui est lui même projeté une nouvelle fois : les

séquences résultantes sont $< (APE)(APB) >$, ... L'explosion combinatoire est limitée par deux contraintes anti-monotones : la fréquence d'apparition et la longueur moyenne des séquences et la distance moyenne en nombre d'énoncés à la fin de la séquence.

Au cours du traitement - dans lequel la séquence est parcourue de la fin vers le début - les épisodes obtenus ne sont pas nécessairement tous appropriés à la prédiction de la fin du tour de parole. Supposons, par exemple, que chaque séquence commence et termine par une (Q)uestion, l'algorithme décrit précédemment donnera comme prédicteur de fin $< (Q) >$, bien qu'il soit aussi un bon prédicteur de début. Pour éviter ce type de résultats défavorables, la distance moyenne de chaque épisode au début de la séquence doit être prise en compte. Si cette dernière est trop faible, l'épisode n'est pas conservé. Ce processus assure que les régularités extraites sont pertinentes.

Le processus d'extraction fournit un très grand nombre d'épisodes. Afin que l'expert puisse manuellement les évaluer, il est nécessaire d'en limiter le nombre. Dans cette optique, une approche par clustering de trajectoire [Lee et al., 2007] a été adoptée, en considérant que les épisodes sont des séquences de déplacements entre deux ensembles de vecteurs d'annotations. Les déplacements sont classés et un représentant est obtenu pour chaque classe. Ce regroupement permet de passer de plusieurs centaines d'épisodes à seulement quelques dizaines de représentants.

5 Analyse des modèles obtenus

L'évaluation des modèles calculés montre qu'un agent narrateur doit être interactif avec l'enfant (au travers de questions et de demandes d'attention) et ce d'autant plus avec les enfants en bas âge. Il apparaît essentiel de solliciter les enfants afin qu'ils interagissent. De plus, la compréhension des émotions et états mentaux des personnages peut être améliorée par une explication du comportement entraîné par l'état mental.

Dans la suite de cette section, l'évaluation des modèles réalisée par une psychologue spécialiste des interactions parents-enfant est détaillée. Ces modèles devraient permettre d'expliquer les comportements relatifs aux états mentaux observés et, à terme, d'être intégrés dans un ACA narrateur afin d'en guider le comportement.

5.1 Motifs dialogiques

L'extraction de motifs dialogiques a permis de collecter un ensemble de motifs et de les regrouper selon leur score de similarité calculé par programmation dynamique. La figure 3 présente un exemple d'alignement de motifs. Ce motif montre que les parents parlent, tout d'abord, des causes ou des conséquences du comportement du personnage (P, C, J) sans référence à l'état mental. Après quelques affirmations ou questions, les parents insistent sur la justification du comportement du personnage (ligne 6), puis le mettent en relation directe avec l'état mental du personnage (ligne 7). Enfin, le parent vérifie que l'enfant a compris en posant des questions ou en demandant son attention (ligne 8).

Ligne	Dialogue B3 (4 ans)					Dialogue C8 (5 ans)				
0	A	-	-	-	-	A	-	-	-	-
1	A	**P**	E	**C**	**J**	A	**P**	-	**C**	**J**
2	Q	**-**	-	**-**	**-**	A	**-**	-	**-**	**-**
3	A	**-**	-	**-**	**-**	A	**-**	-	**-**	**-**
4	A	**-**	-	**-**	**-**	A	**-**	-	**-**	**-**
5	A	**-**	-	**-**	**-**	A	**-**	-	**-**	**-**
6	A	**P**	-	**C**	**J**	A	**P**	-	**C**	**J**
7	A	**P**	**E**	**-**	**-**	Q	**P**	**E**	**-**	**-**
8	**Q**	**-**	**-**	**-**	**-**	**D**	**-**	**-**	**-**	**-**
9	A	-	-	-	-	A	P	E	-	-

FIGURE 3 – Alignement de motif dialogique

Ce motif démontre parfaitement qu'il n'est pas suffisant de nommer un état mental pour l'expliquer. En effet, le développement narratif implique une démonstration pratique de la théorie de l'esprit. Le motif décrit le lien entre le comportement du personnage et l'état mental, le second expliquant le premier.

5.2 Prédiction d'interaction

Nous décrivons ici les conditions nécessaires à une amélioration significative de la narration interactive, en fonction de l'âge de l'enfant (3, 4 ou 5 ans). Le tableau 4 résume les modèles des interactions de l'enfant (voir section 3.2). Pour chaque âge, les modèles sont caractérisés par :

– leur *longueur moyenne*, qui correspond au nombre moyen d'énoncés entre le modèle et l'interaction de l'enfant. Plus une séquence est courte, moins il y a d'énoncés entre elle et l'intervention de l'enfant ;
– le *modèle*, qui décrit une séquence d'annotations. Par exemple la séquence E-Q symbolise une annotation E suivi, plus ou moins tard, d'une annotation Q. Les annotations peuvent ne pas être dans la même dimension ;

3 ans			4 ans			5 ans		
longueur	**modèle**	**fréquence**	**longueur**	**modèle**	**fréquence**	**longueur**	**modèle**	**fréquence**
3,2	**E-Q**	**10,4%**	2,1	D-Q	14,9%	1,9	Q	35.4%
3,4	D-Q	16,8%	**2.2**	**E-Q**	**7,5%**	**2,2**	**E-E**	**9,1%**
3,5	J-Q	9,6%	2,2	Q-Q	12,7%	2,6	J-D	8,1%
3,5	D-Q-Q	9,6%	2,6	D-E	10,4%	2,7	E-D	6,1%
3,5	E-J	8,8%	2,8	D-D	11,2%	3,1	J-E	8,1%
4,3	D-E	12,8%	3,5	J	14,9%	3,4	V	13,1%
4,3	D-J	8,0%	3,8	B	7,5%	3,7	D-E	7,1%
5,4	B	10,4%	4,0	E-E	7,5%	3,8	J-J	6,1%
5,6	V	13,6%	4,1	E-D	6,7%	4,1	E-J	7,1%
			4,3	V	6,7%			

FIGURE 4 – Longueurs moyennes et fréquences de séquences en fonction de l'âge.

– la *fréquence*, qui est le pourcentage de fois où le modèle apparaît.

Les demandes d'attention (codées D, par exemple "regarde !" ou "tu as vu ?"), essentielles pour la narration interactive, sont présentes pour tous les âges. Plus l'enfant est âgé, plus sa réaction à la demande d'attention est rapide. Plus l'enfant est jeune, plus les demandes d'attention doivent être répétées ou ponctuées de questions. Ceci peut s'observer dans des séquences telles que D-D ou D-Q ou D-Q-Q.

110 séquences contenant une demande d'attention ont été recensées dans les dialogues des enfants de 3 ans, 58 pour les enfants de 4 ans. Les enfants de 3 ans interagissent en effet après un nombre d'énoncés moyen compris en entre 3,5 et 4,6. Les enfants de quatre ans réagissent plus rapidement (entre 2,2 et 2,8 énoncés) : l'efficacité du modèle s'améliore avec l'âge. Par contre, les parents se comportent différemment avec les enfants de 5 ans se comportent : les demandes d'attention sont moins fréquentes et soit associées à des états mentaux (D-E ou E-D) soit à des justifications (D-C ou C-D). Nous n'avons dénombré que 21 séquences comportant des demandes d'attention, rapidement suivies d'une interaction de l'enfant (entre 2,6 et 2,7 énoncés). Les séquences comprenant des justifications (codées J, par exemple "puis, Leo casse le château !") sont essentielles au processus d'interaction émotionnelle narrative.

En conclusion, nous souhaitons mettre l'accent sur certains points notables :
– quel que soit l'âge, les séquences contenant des justifications sont fréquemment associées à divers indices (émotion, demande d'attention ou question). Dans ce contexte, l'interaction de l'enfant ne survient qu'entre 3,1 et 4,3 énoncés après le modèle ;
– la longueur des interactions décroît avec l'âge, de 3,2 à 1,9 énoncés ;
– le nombre d'énoncés auxquels sont associées des émotions est quasiment équivalent pour tous les âges. Néanmoins, plus l'enfant est âgé, plus les séquences d'émotions sont variées. Les séquences complexes (émotions et justifications : J-E ou E-J) n'apparaissent qu'avec les enfants les plus âgés ;
– à l'exception des demandes d'attention, les modèles les plus efficaces (en rouge et gras dans le tableau 4) contiennent toujours des émotions (E-Q ou E-E).

6 Conclusion

Nous avons montré, dans cet article, les raisons pour lesquelles la narration nécessite de nombreuses interactions et émotions. Nous avons présenté une méthodologie et des outils permettant d'améliorer la modélisation du dialogue. Ces modèles sont dédiés à la narration interactive et peuvent être intégrés dans des ACA narrateurs. La méthodologie proposée consiste, d'une part, à extraire des motifs dialogiques et à les classifier afin d'encoder les conventions dialogiques et, d'autre part, à une prédiction d'événements afin d'encourager l'interaction avec l'auditeur. Une représentation matricielle de l'interaction est utilisée afin d'encoder les aspects multidimensionnels du dialogue. Nos algorithmes ont été appliqués à un corpus de dialogues de narration parents-enfants.

Les interactions enfant-agent peuvent différer des interactions parent-enfant, non seulement par les capacités dialogiques de l'agent, mais aussi en raison de la représentation numérique de l'agent. Afin d'évaluer l'impact de l'incarnation, une seconde expérience a été menée. Des dialogues enfant-adulte ainsi que des dialogues enfant-avatar - durant une expérience de type Wizard of Oz (WOz) - ont été enregistrés via un système de vidéoconférence. Ce corpus de dialogues est actuellement en cours d'an-

notation. Nous prévoyons d'appliquer les algorithmes présentés précédemment afin de comparer les modèles de dialogue obtenus par visioconférence (interaction enfant-adulte) et ceux du WOz (interaction enfant-avatar).

Enfin, nos travaux futurs seront dédiés à l'intégration effective des modèles de dialogue obtenus au sein d'un ACA et à son évaluation en situation de narration interactive. Pour ce faire, nous proposons d'utiliser, comme Hulstijn [Hulstijn, 2000], des jeux de dialogues afin de gérer les conventions représentées par les motifs dialogiques.

Remerciements

Ce travail a bénéficié du soutien du projet CNRS PEPS INS2I-INSHS « ACAMODIA ».

Références

[Allen and Perrault, 1980] Allen, J. and Perrault, C. (1980). Analyzing intention in utterances. *Artificial Intelligence*, 15(3) :143–178.

[André and Pelachaud, 2010] André, E. and Pelachaud, C. (2010). Interacting with embodied conversational agents. *Speech technology*, pages 123–149.

[Antunes and Oliveira, 2001] Antunes, C. M. and Oliveira, A. L. (2001). Temporal data mining : An overview.

[Astington and Baird, 2005] Astington, J. W. and Baird, J. (2005). *Why language matters for theory of mind*. Oxford University Press, New York.

[Aust et al., 1995] Aust, H., Oerder, M., Seide, F., and Steinbiss, V. (1995). The philips automatic train timetable information system. *Speech Communication*, 17(3-4) :249–262.

[Bunt, 2011] Bunt, H. (2011). Multifunctionality in dialogue. *Computer Speech and Language*, 25(2) :222–245.

[Cassell et al., 2000] Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., and Yan, H. (2000). Embodied conversational agents. chapter Human conversation as a system framework : designing embodied conversational agents, pages 29–63. MIT Press.

[Chanoni, 2009] Chanoni, E. (2009). Comment les mères racontent une histoire de fausses croyances à leur enfant de 3 à 5 ans ? *Enfance*, (2) :181–189.

[Courgeon et al., 2009] Courgeon, M., Clavel, C., and Martin, J.-C. (2009). Appraising emotional events during a real-time interactive game. In *AFFINE'09*, pages 7 :1–7 :5, New York, NY, USA. ACM.

[Dunn, 1973] Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3) :32–57.

[Florek et al., 1951] Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H., and Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. In *Colloquium Mathematicum*, volume 2, pages 282–285.

[Frampton and Lemon, 2009] Frampton, M. and Lemon, O. (2009). Recent research advances in reinforcement learning in spoken dialogue systems. *Knowledge Engineering Review*, 24(04) :375–408.

[Frey and Dueck, 2007] Frey, B. and Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315(5814) :972.

[Gelin et al., 2010] Gelin, R., d'Alessandro, C., Le, Q. A., Deroo, O., Doukhan, D., Martin, J.-C., Pelachaud, C., Rilliard, A., and Rosset, S. (2010). Towards a storytelling humanoid robot. In *AAAI Fall Symposium Series*.

[Gouaillier et al., 2009] Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., and Maisonnier, B. (2009). Mechatronic design of nao humanoid. *Proc. of the Int. Conf. on Robotics and Automation*, pages 769–774.

[Guha et al., 2000] Guha, S., Rastogi, R., and Shim, K. (2000). Rock : A robust clustering algorithm for categorical attributes. *Information Systems*, 25(5) :345–366.

[Hulstijn, 2000] Hulstijn, J. (2000). Dialogue games are recipes for joint action. In *Proc. of Gotalog'00*.

[Karypis et al., 1999] Karypis, G., Han, E., and Kumar, V. (1999). Chameleon : Hierarchical clustering using dynamic modeling. *Computer*, 32(8) :68–75.

[Larsson and Traum, 2000] Larsson, S. and Traum, D. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, 6(3&4) :323–340.

[Lecroq et al., 2012] Lecroq, T., Pauchet, A., Chanoni, E., and Solano, G. A. (2012). Pattern discovery in annotated dialogues using dynamic programming. *Int. J. of Intelligent Information and Database Systems*, 6(6) :603–618.

[Lee et al., 2007] Lee, J.-G., Han, J., and Whang, K.-Y. (2007). Trajectory clustering : a partition-and-group framework. In *Int. conf. on Management of data*, pages 593–604. ACM.

[McTear, 2004] McTear, M. (2004). *Spoken dialogue technology : toward the conversational user interface*. Springer-Verlag New York Inc.

[Pelachaud, 2009] Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Trans. of the Royal Society B : Biological Sciences*, 364(1535).

[Schröder, 2010] Schröder, M. (2010). The SEMAINE API : towards a standards-based framework for building emotion-oriented systems. *Advances in HCI*, 2010 :2–2.

[Searle, 1969] Searle, J. (1969). *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University.

[Seif El-Nasr and Wei, 2008] Seif El-Nasr, M. and Wei, H. (2008). Exploring non-verbal behavior models for believable characters. In *Interactive Storytelling*, volume 5334 of *Lecture Notes in Computer Science*, pages 71–82.

[Swartout et al., 2006] Swartout, W. R., Gratch, J., Jr., R. W. H., Hovy, E. H., Marsella, S., Rickel, J., and Traum, D. R. (2006). Toward virtual humans. *AI Magazine*, 27(2) :96–108.

[Von Luxburg, 2007] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4) :395–416.

Impact of the Social Behaviours of the Robot on the User's Emotions: Importance of the Task and the Subject's Age

A. Delaborde[1,2] agdelabo@limsi.fr
L. Devillers[1,3] devil@limsi.fr

[1]LIMSI-CNRS
Orsay – FRANCE

[2]Université Paris-Sud XI

[3]Université Paris-Sorbonne IV

Résumé :
Cette étude est menée dans le cadre du projet français ROMEO, qui vise à concevoir un robot social humanoïde capable d'aider les personnes en perte d'autonomie dans leurs activités domestiques quotidiennes, et également de jouer à des jeux avec les enfants. Nous mettons au point un système dans lequel les indices émotionnels audio extraits du signal de parole permettent la création d'un profil émotionnel et interactionnel de l'utilisateur. Ce profil détermine la sélection du comportement du robot. Il est alors obligatoire d'analyser l'impact des comportements du robot possibles sur les émotions de l'utilisateur. Nous avons mené deux expériences mettant en scène des enfants jouant à des jeux avec le robot, et des personnes déficientes visuelles interagissant avec le robot dans un contexte d'assistance. Nous remarquons que lorsque le robot exprime des comportements socialement non désirables, les personnes déficientes visuelles réagissent différemment que les enfants.
Mots-clés : Traitement du signal social, Interaction Humain-Robot, Émotions

Abstract:
This study is carried out in the framework of the French ROMEO project, which aims to design a social humanoid robot to assist elderly and/or disabled people at home in everyday life activities, but which will also be able to play games with children. We design a system in which the emotional audio cues extracted from speech are used to build an emotional and interactional profile of the user. This profile will determine the robot's behaviour selection. It is therefore compulsory to analyse the impact of the various robot behaviours on the user's expression of emotions. We designed two experiments featuring children playing with the robot, and visually-impaired people interacting with the robot in a context of assistance. We notice that, when facing undesirable behaviours from the robot, the visually-impaired react differently from the children.
Keywords: Social signal processing, Human-Robot Interactions, Emotions

1 Introduction

Social interaction is characterised by a continuous and dynamic exchange of communication and information signals. Producing and understanding these signals allow the human to interact with his kind. Human beings interact with each other by several means: voice, gesture, facial expressions, etc. Among these means, vocal expression (verbal and nonverbal components) conveys a great amount of information, for both carrying meaning and emotion. The interpretation and the production of non-verbal signals (tone of the voice, facial expressions and gestures) have been identified as major obstacles for the development of robotic systems endowed with social and affective intelligence.

This study is carried out in the framework of the French ROMEO project1. This project aims to design a social humanoid robot which will be able to assist elderly and/or disabled persons at home in everyday life activities, but also which will be able to play games with children (for example with the grand-children of the user). So as to interact as naturally as possible with the user, the robot will be endowed with a multi-level processing of non-verbal audio cues [1] (see Figure 1). Low level cues can be computed from the speech signal [2]: duration of speaker turns, F0, energy, and other acoustic coefficients. Mid-level markers can be derived from machine learning techniques such as support vector machines trained on various

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

statistical functionals and transformations applied to these cues and provide a system with emotional information such as positive/negative emotion, activation/non activation behaviour, emotion labels (Joy, Sadness, Fear, Anger), speech delivery, rhythm and duration. On a higher level of analysis, these data can be processed through fuzzy logic rules so as to get cues about the emotional and interactional tendencies of the speaker: we can obtain emotional and interactional markers such as ill-at-ease, talkative, shy, or dominant. A speaker identification system would also bring sociological metadata such as the age bracket of the speaker, the gender, and to be able to recognise a specific user and then keep an automatic track of his or her emotional and interactional profile.

The robot behaviour selection is based on this profile. It seems compulsory to determine, in the ROMEO project, the impact of the robot's behaviour on the user. The system built (see Figure 1) uses Voxler libraries for the emotion detection system and the Spirops AI for the user modeling module.

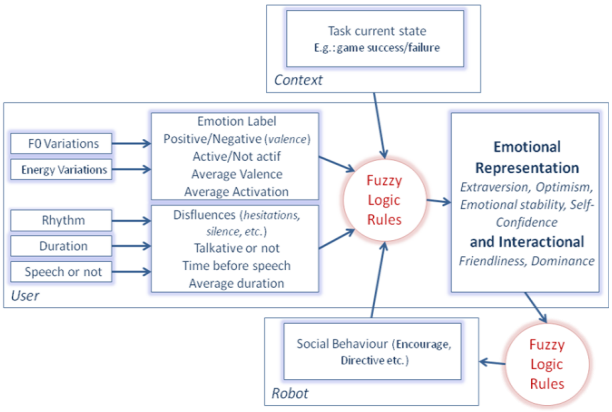


FIG. 1 – Multi-level processing of the non-verbal audio cues extracted from speech, in order to build an emotional and interactional representation of the user, and for the selection of the robot's social behaviour.

This article presents, in section 1, a global overview of studies dealing with the impact of social virtual agents' or robots' behaviours on the user. Section 2 introduces two experiments we carried on children and visually impaired adult people interacting with a robot, here the humanoid robot Nao (Aldebaran French Company). Speakers express several affective states, and the robot presents different behaviours. Section 3 gives an analysis of this study: what is the impact of the robot's behaviour on the speaker's expression of emotion? Are their differences between what is undesirable for a senior user and for a child, in specific tasks? Section 4 gives our conclusions.

2 Impact of the System's Behaviour on the User: an Overview

2.1 Considering the user and the task

The behaviours which the interface (robots, knowbots such as conversational agent, etc.) should present are closely linked to the task and the user's characteristics. However, it is in the first place compulsory to make sure that the behaviours are correctly perceived, by analyzing the feedback from the user [3].

As for persons suffering from loss of autonomy, it is required to give the system a certain degree of autonomy in the fulfilment of the task [4][5]. For the interaction not to be prohibitive (as it could be for elderly people dealing with new technologies), one ought to analyse the preferred communication modalities between the interface and the user [6].

2.2 Importance of the behaviour selection and expression

A robot endowed with personality, as described in Kiesler and Goetz's study [7], impacts the user's mental representation of the robot, as well as their involvement in the interaction. Carefully choosing the way the desires or intentions of the robot are designed (e.g. via emotion expression as described in Breazel's works [8]) can lead the user to better interact with the robot, and even fulfil the robot's needs. On the contrary, a robot which disobeys the user's orders will have an impact on the emotional expression of the user [9].

In Human-Robot Communication, beyond the functional expectation of the user, meaning that the robot carries out the task which is devolved to it and that its use is intuitive and ergonomic, a social robot should share the human's interpersonal communication codes [10][11], so

as to efficiently respond to the user's messages. In this way, the robot will be more likely to establish and maintain a natural and socially acceptable relationship.

In our present study, the interaction modalities are carefully selected. We take into account the fact that we deal with two different types of speakers (children and visually-impaired adults) and two different types of tasks (gaming and assistance). Although they are based on the same coding, the robot's behaviours are then expressed slightly differently.

3 Experiments

We carried two series of recordings, based on the interaction with the robot Nao presenting alternatively positive and negative behaviours. We name "positive" the behaviours which are deemed desirable in a Human-Robot Interaction, "negative" when the behaviour does not seem to fit a socially acceptable interaction.

The behaviour selection is based on Ray's study on users' expectation in robotics [12]. Besides, when engaging a first interaction with an interlocutor, Isbister [13] highlights two prevailing questions which allow the speaker to define his social position towards the interlocutor: is he a friend or a foe? Is he socially more powerful than me or not? Isbister offers an adapted version of the Interpersonal Circumplex based on the dimensions Dominance and Friendliness. In the context of an interaction between a virtual or robotic entity, the latter should allow the human to socially position himself, and thus should behave coherently with these axes. In an assistive context for example, the user should feel that the robot is sensitive to his needs.

The Figure 2 presents the localisation of the desirable behaviours which can be expected from the robot in the course of the IDV-HR experiment: encouragement, empathy and friendly. We also selected undesirable behaviours as counter-examples: for example, in an assistive context, we deem that a doubtful or a too directive robot will not satisfy the user. As far as game is concerned, if the robot disregards the rules or is not cooperative for example, the player will not accept it.

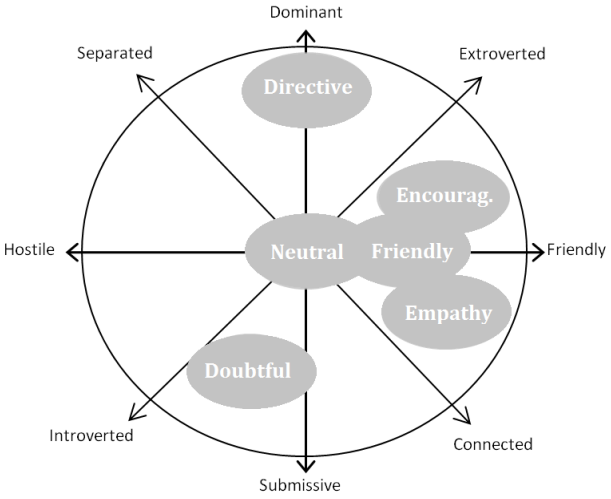


FIG. 2 – Representation of the robot behaviours on the Dominance and Friendliness axes (example for IDV-HR experiment). Adapted from [13]

3.1 NAO-Children Data Collection

This recorded experiment features twelve children from eight to thirteen playing Question/Answers games in pairs with Nao, supervised by a human game master (experimenter). Figure 3 shows an example of two boys aged 8 and 12 playing gaming with Nao. Each recording session lasts approximatively thirty minutes.

In the course of the games, the robot displays positive and negative social attitudes, expressed by oral and gestural means. Among positive behaviours, the robot could congratulate the child or offer its help ("Do you need some hints?"). When behaving negatively, the robot presented different types of behaviours. First, it is completely disconnected from the imperatives of the game: for example, it gives away an answer or counts a correct answer whereas a wrong answer was given. It could also persist in making mistakes: e.g. giving wrong answers repeatedly. Lastly, Nao could be not conciliatory, by refusing to give hints or by being excessively proud ("I'm the best!"). The utterances and the sequence of positive/negative behaviours was fixed in advance.

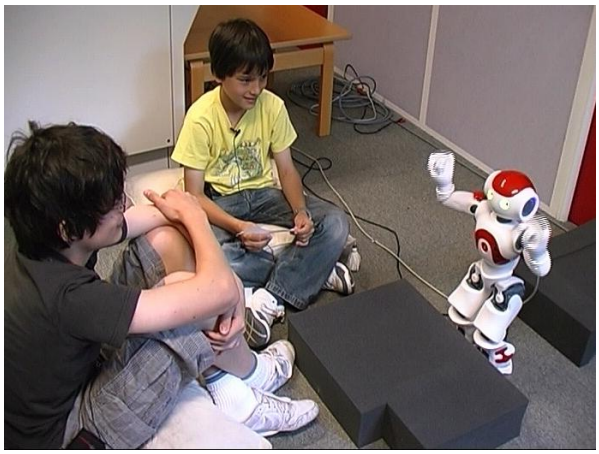


FIG. 3 – Two boys aged 8 and 12 playing games with Nao (NAO-Children data collection)

The robot is remotely controlled by a Wizard-of-Oz experimenter, using the programming software Choregraphe (developed by Aldebaran Robotics). Automata allow selecting the robot's next attitude depending on the answer of the children (manually entered: for example, has the child given a wrong answer or a correct one?). Figure 4 presents the sequence for the first question of the game: depending on the child's answer, manual input will trigger different behaviours. This protocol is detailed in [14]. For each recording session, the robot behaved 69% of the times in a positive way, 28% negatively and 3% were neutral utterances.

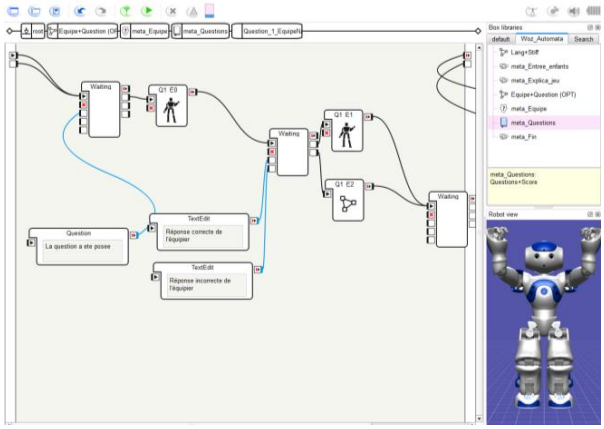


FIG. 4 – Sequence of behaviours of the robot for the first question of the Question/Answer game. *Choregraphe GUI (by Aldebaran Robotics)*

We make a distinction between the moments when the children spoke to the robot at their own initiative or answered the robot's questions, and the moments when they talked together. In the first case, the emotions expressed are deemed induced by the robot's reactions, whereas in the other case they chatted out of the protocol (this is what we will subsequently call "free discussion").

3.2 IDV-HR Data Collection

Fourteen visually impaired speakers (median age fifty-nine) are asked to play three sessions of five scenarios in which they picture themselves in a situation of waking up in the morning (see Figure 5). Each experiment lasts approximately forty-five minutes. The robot comes to them to chat about their health and set the program of the day. In each scenario the speaker is asked to act a different general affective state, either positive (in peak form, glad for something to come) or negative (depressed, emergency, sick). The robot behaviour in the course of each of the three sessions is pre-selected by the Wizard. Each speaker is presented one desirable robot behaviour and one (or even two) undesirable behaviour(s). The robot could be either "Dominant", or "Doubtful", "Friendly", "Machine-like", "Empathetic" or "Encouraging" (see Figure 2).

FIG. 5 – A visually-impaired 39 year-old woman speaking to Nao. Nao is sitted in front of her on the coffee table, out of shot of the camera. (IDV-HR data collection)

Since the speakers are not able to correctly see the robot, its behaviours are solely lexically coded, based on Charaudeau's works on classification of French language. This classification is realised according to the speaker's consideration (or absence of

consideration) of the interlocutor [15]. For example, Encouragement presupposes that the speaker implicates him/herself and the interlocutor in his/her speech, for he/she tries to convince that his/her advice has to be followed, so as to reach a goal deemed positive. In this context, the robot will say sentences such as: "Everything'll be fine as soon as you have breakfast." The robot then asks a question, to initiate the discussion: "Do you want me to dial your doctor's number for you?"

We wanted the speakers to be fully aware of the fact they were dealing with a robot, and not with simple loudspeakers. For this reason we asked them to touch the top of the robot's head at the beginning of the session. Touching the tactile sensors there would trigger the start of the experiment (see Figure 6).

As another example, in the "in peak form" scenario, the robot will say, depending of its behaviour: "I think you'll be able to do a lot of things today !" (Encouraging); "I'm so glad you're in peak form!" (Empathetic); "I notice you're in peak form." (Friendly); "Detected emotion: dynamism" (Neutral); "Since you're in peak form, I want you to do a lot of things today." (Directive).

The speakers are also recorded when they speak with the experimenter at the end of the session (what we refer to as "Free discussion" subsequently).

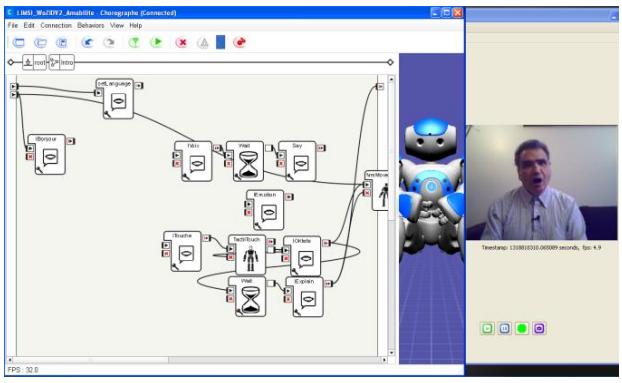


FIG. 6 – Nao's introductory speech. As soon as the speaker touches the robot's head tactile sensor, the experiment starts. Right window allows the Wizard to see the speaker through the robot's camera. *Choregraphe GUI (on the left) and Telepathe (on the right) (by Aldebaran Robotics)*

4 Impact of the robot's behaviour on the speaker's expression of emotions

4.1 Emotional data annotation and results

On each child's track, two expert labellers defined segment boundaries. A segment is emotionally homogenous, i.e. the emotion is considered as being the same and of a constant intensity along the segment [16][17]. The labellers performed an emotional annotation on each segment:

- An Emotion macro-class, among "Joy" (satisfaction, joy, amused), "Anger" (hot and cold anger, annoyed, impatience), "Fear" (fear, anxiety), "Sadness" (sadness, boredom) and "Neutral".
- An Activation label, among -2 (very low) to 2 (very high), representing the strength of the emotion.

For the purpose of this study, the macro-classes "Anger", "Fear" and "Sadness" are merged into the valence "Negative"; "Joy" is considered as a "Positive" valence. Table 1 presents the annotation results for the two corpora. The following emotional attitudes analysis is based on the segments on which the labellers agree in terms of Valence (for the valence analysis) and Activation.

TAB. 1 – Annotation Results

Corpus	# segments	Kappa Values	
		Valence	*Activation*
Nao-Children	1287	0.7	0.58
IDV-HR	3933	0.82	0.57

4.2 Emotional attitudes analysis

The speakers we recorded are not professional actors: in the case of the children playing with the robot, they are only reacting spontaneously to coded behaviours of the robot. As for the visually-impaired adult people, they have to play a role, but they do not have a tough control on their emotion expressions, as a professional actor would. Thus we picture that, even if they are given rules about the emotion they are expected to act, the behaviour of the robot and its type of responses will have a noticeable

impact on their expression of emotions.

We now evaluate the impact of the positive and negative social behaviours of the robot on the speakers in terms of valence and activation, according to the task (game or assistance).

4.2.1 Children's reactions

In the course of the game the children played with the robot, the latter behaved in either a desirable or an undesirable way.

TAB. 2 – Valence average according to the Behaviour of the Robot (for children)

		Robot Behaviour	
		Positive	***Negative***
Speaker's reactions	**In interaction w/ the robot**	0.56[a]	0.68
	Free discussion	0.7	

a. Valence average between 0 (negative) and 1 (positive)

TAB. 3 – Activation average according to the Behaviour of the Robot (for children)

		Robot Behaviour	
		Positive	***Negative***
Speaker's reactions	**In interaction w/ the robot**	0.61[a]	0.63
	Free discussion	0.56	

a. Activation average between 0 (low) and 1 (high)

Table 2 presents the overall distribution of emotional reactions which were expressed by the children in reaction to the robot's behaviours ("Positive" and "Negative"), and during the free discussion moments. We selected only the 901 segments where labellers agreed on the valence. We can see that, on the whole, the negative strategies applied through the robot did not trigger an increase in negative reactions from the children. Although we knew that they could laugh at the robot, we expected the children to express a soft or average irritation, or a soft stress. On the contrary, children were 12% more positive when the robot behaved undesirably. These actions were considered as funny or at least did not lead to any frustrating or angry expressions of emotions.

We also analysed the impact of the behaviours on the activation of the expressed emotions. We used the 746 segments presenting an agreement on the activation between the labellers. We notice in Table 3 that the negative behaviours of the robot triggered a slight increase of the intensity of the emotions expressed by the children. We can link that to the change in valence, the emotions being more positive and linked to laughter.

4.2.2 Visually-impaired adults' reactions

The scenarios are categorised in two types: positive situations ("happy for something to come" and "peak form") and negative situations ("depressed", "emergency" and "sick"). We expected the speakers to express predominantly negative emotions during negative scenarios and vice-versa.

Table 4 presents the average valence expressed by the speakers during the recording. We analyse the 3756 segments which present an agreement between the labellers on the valence. We notice a difference according to the attitude of the robot: when the robot displays an undesirable behaviour, the speakers tend to express more neutral emotions, or more negative. During the positive scenarios, valence falls by 9% and by 5% during negative scenarios.

TAB. 4 – Valence Average according to the Behaviour of the Robot and to the Scenario (for adults)

		Robot Behaviour	
		Positive	***Negative***
Scenario	Positive	0.73[a]	0.64
	Negative	0.12	0.07
Free discussion		0.45	

a. Valence average between 0 (negative) and 1 (positive)

We now analyse the differences in the expression of activation. Our analysis is based on the 2313 segments presenting an agreement on the activation between the labellers. We notice (cf. Table 5) that the negative behaviours of the robot lead to a drop of the average activation by 4.5%.

TAB. 5 – Activation Average according to the Behaviour of the Robot and to the Scenario (for adults)

		Robot Behaviour	
		Positive	***Negative***
Scenario	Positive	0.47[a]	0.5
	Negative	0.58	0.64
Free discussion		0.15	

a. Activation average between 0 (low) and 1 (high)

We notice that the undesirable behaviours played by the robot in this context have an impact on the speakers' expression of emotions: valence tends to be negative, and activation increases. The intervals, however, are minor (6% on average).

We expected that the speakers would express a modification in their emotional behaviour, when the robot did not behave in desirable way. Nonetheless, it seems obvious that the in vitro context (no real task to perform and experimental conditions) smoothes the speakers' reactions.

Conclusion

When one wants to design a social robot, it is compulsory to make sure that the behaviours of the robot will be deemed socially acceptable by the user. We studied the impact of different behaviours on the expression of emotions of two types of publics in different tasks: children at play with the robot, and visually impaired adult people in a context of assistance in interaction with the robot.

In order to study this impact, we designed two experiments. The first one featured children playing question/answer games with Nao, where the robot either encouraged and motivated the children, or made mistakes and was not conciliatory. In the second experiment, visually-impaired adult persons played scenarios in interaction with Nao: they had to pretend to be in various emotional states, and the robot was more or less friendly and helpful with them.

The analysis of our recordings allows us to see that the undesirable behaviours of the robot, when displayed in an assistance context, have a negative impact on the expression of emotions of the adults: their emotions tend to be more negative, even if they are in a scenario when they are expected to express positive emotions. As far as children are concerned, in a context of gaming, what we expected to be undesirable behaviours did not trigger the expression of negative emotions, but were rather perceived as funny by the speakers. Although the impact is moderate in experimental conditions, it reveals most certainly the emotional behaviour tendencies of a user in vivo.

The higher activation in both contexts could be interpreted as an increase in the engagement of the speaker toward the robot. But this kind of engagement is not what one should expect in a human-robot interaction: either the speaker laughed at the robot, or tried to make him-herself clearer so as to be better understood by the robot, or even expressed irritation.

This experiment underlines the importance of the task and the age of the subject when defining the area of "desirable" and "undesirable" behaviours of the robot: the user will not expect the same performance in the achievement of a gaming task, nor the same social attitude from the robot, than when he/she is expecting a practical realisation and understanding from the robot, which is the case in an assistance context.

Acknowledgment

This work is financed by national funds FUI6 under the French ROMEO project labelled by CAP DIGITAL competitive centre (Paris Region). We also thank the SME Voxler and Spirops, our partners in the ROMEO project.

References

[1] A. Delaborde and L. Devillers, "Use of Nonverbal Speech Cues in Social Interaction between Human and Robot: Emotional and Interactional markers". In proc. 3rd Int. Workshop on Affective Interaction in Natural Environments, ACM Multimedia. Firenze, Italy, 2010.

[2] L. Devillers, L. Vidrascu and L. Lamel, “Challenges in real-life emotion annotation and machine learning based detection”. Journal of Neural Networks, 18:407 – 422, 2005.

[3] M. El-Nasr, J. Yen and T.R. Ioerger. FLAME. Fuzzy Logic Adaptive Model of Emotions, Autonomous Agents and Multi-Agent Systems, 3, 219.257, 2000

[4] K. Dautenhahn and I. Werry, “A quantitative technique for analysing robot–human interactions”. In proc. of the IEEE/RSJ, International Conference on Intelligent Robots and Systems (pp. 1132–1138). Lausanne, Switzerland, 2002.

[5] A. Tapus and M. Matarić, “User personality matching with hands-off robot for post-stroke rehabilitation therapy”. In proc. Int. Symp. on Experimental Robotics (ISER). Rio de Janeiro, Brazil, 2006.

[6] C. Granata, M. Chetouani, A. Tapus, P. Bidaud and V. Dupourque, “Voice and Graphical based Interfaces for Interaction with a Robot Dedicated to Elderly and People with Cognitive Disorders”. 19th IEEE Int. Symp. in Robot and Human Interactive Communication, 2010.

[7] S. Kiesler and J. Goetz, “Mental Models and Cooperation with Robotic Assistants”. In Human Factors in Computing Systems, pp. 576-577. IEEE Press, New Yors (2002).

[8] C. Breazeal. “Robot in Society: Friend or Appliance?” In Agents 99: Workshop on Emotion-based Agent Architectures (1999), pp. 18-26.

[9] A. Batliner, C. Hacker, S. Steidl, E. Nöth, S. D'Arcy, M. Russel and M. Wong, “’You stupid tin box’ - children interacting with the aibo robot: a cross-linguistic emotional speech corpus”. In proc. of the 4th Int. Conf. of Language Resources and Evaluation. pp. 171-174, 2004

[10] D. Duhaut and S. Pesty. “Acceptability in Interaction: From Robots to Embodied Conversational Agents”. In Computer graphics theory and applications, Algarve, Portugal, 2011.

[11] D. Feil-Seifer, K. Skinner and M.J. Matarić, “Benchmarks for evaluating socially assistive robotics”. In Interaction Studies: Psychological Benchmarks of Human-Robot Interaction, 8(3), 423-429 Oct 2007.

[12] C. Ray, F. Mondada, and R. Siegwart, “What do people expect from robots?” IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. Nice, France. IROS 2008. pp.3816 – 3821, 2008.

[13] K. Isbister, “Better Game Characters by Design: A Psychological Approach”. The Morgan Kaufmann Series in Interactive 3D Technology. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2006. p. 26.

[14] A. Delaborde, M. Tahon, C. Barras and L. Devillers, “Wizard-of-Oz game for collecting emotional audio data in a children-robot interaction”, in International Workshop on Affective-aware Virtual Agents and Social Robots, ICMI-MLMI. 2009: Boston, USA.

[15] P. Charaudeau, “Grammaire du sens et de l’expression.” Hachette, pp. 576-629. 1992.

[16] L. Devillers, R. Cowie, J.-C. Martin, E. Douglas-Cowie, S. Abrilian and M. McRorie, “Real life emotions in French and English tv video clips: an integrated annotation protocol combining continuous and discrete approaches”. In proc. 5th Int. Conf. on Language Resources and Evaluation (LREC06), Genoa, Italy, 2006.

[17] L. Devillers and J.-C. Martin, “Coding Emotional Events in Audiovisual Corpora”. In proc. 6th Int. Conf. on Language Resources and Evaluation, Marrakech, Morocco, 2008..

Collecte de données pour la détection du stress dans les interactions sociales

M. Soury [1,2] soury@limsi.fr
L. Devillers [1,3] devil@limsi.fr

[1] Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur - CNRS
[2] Université Paris Sud Orsay - Orsay – France
[3] Université Paris-Sorbonne PIV - France

Résumé :
Ce papier présente une collecte de données multimodales relative au stress dans la prise de parole lors d'interactions sociales. Cette expérimentation a été faite afin de constituer un corpus de stress pour la détection automatique dans le traitement des phobies sociales, et afin de valider une application d'animations dynamiques pour la collecte de données. Ces travaux sont réalisés dans le cadre du projet FEDER E-Therapie.
Mots-clés : phobie sociale; stress; logiciel de remédiation; corpus multimodal;

Abstract :
This paper presents the collection of multimodal data relative to stress in the context of speaking in social interactions. This experiment was conducted in order to build a stress corpus for automated detection in the treatment of social phobia, and in order to validate the use of dynamic animations for such data collection. This work is conducted in the context of the FEDER E-Therapie project.
Keywords: social phobia; stress; remediation software; multimodal corpus;

1 Introduction

Les techniques de réalité virtuelle permettent de projeter l'utilisateur dans un environnement interactif. La thérapie cognitive comportementale a su tirer parti de cet outil dans le traitement des troubles cognitifs, notamment des phobies [5][10][12].

Il a été établi que les tâches de prise de parole en public sont génératrices de stress [8]. Plusieurs expériences de recherche ont utilisé cette tâche pour éliciter du stress chez les participants afin d'observer leurs réactions [1][9][11][14].

La prise de parole dans le cadre d'interactions sociales est une tâche difficile pour les phobiques sociaux, et les exercices de remédiation utilisés en thérapie cognitive comportementale pour le traitement de cette phobie sont centrés sur l'entrainement à la prise de parole dans des situations sociales anxiogènes.

Le but de ce projet est de créer un outil de réalité virtuel utilisant des systèmes de détection automatique. Nous proposons ici l'étude d'un logiciel de remédiation cognitive, avec un double objectif : d'une part évaluer la pertinence des animations de formes géométriques pour la mise en scène de situations anxiogènes contrôlées à la voix; et d'autre part collecter des données émotionnelles et attentionnelles inhérentes à la tâche de prise de parole dans le cadre d'interactions sociales, afin de concevoir des modèles de détection automatique adaptés.

Ce papier présente la constitution d'un corpus multimodal audio, vidéo RGB et vidéo profondeur de données spontanées relatives à l'expression du stress. Les études existantes sur le stress utilisent la prise de parole en public pour éliciter l'émotion. Nous proposons la constitution d'un corpus de stress dans les interactions sociales, il n'existe à notre connaissance aucun corpus de ce type accessible dans la communauté aujourd'hui.

Nous présentons d'abord le protocole de collecte de données, en particulier l'application interactive utilisée pour éliciter les réactions des utilisateurs. Dans la section suivante, nous proposons une première analyse des données collectées lors de la campagne initiale, réalisée auprès de non phobiques. Enfin nous énonçons quelques perspectives envisagées à la suite de

cette expérimentation.

2 Protocole d'acquisition de données

L'expérimentation teste trois hypothèses :

- la projection d'un contexte réaliste sur une animation abstraite réagissant au comportement de l'utilisateur peut figurer des interactions sociales [4].

- le stress de l'utilisateur peut être perçu dans sa voix.

- la posture de l'utilisateur du système, spécifiquement sa distance à l'écran et la façon dont elle varie, peut être utilisé comme un marqueur d'attention que celui ci porte à la tâche.

Cette étude a été réalisée dans le cadre du projet FEDER E-Therapie, dont l'objectif est la conception d'un outil de remédiation à destination des phobiques sociaux.

3 scénarii mettant en scène des situations sociales anxiogènes ont été conçus en partenariat avec les thérapeutes du Centre Emotion de l'hôpital de la Pitié-Salpêtrière à Paris.

Pour jouer ces scénarii, un logiciel d'animations a été développé au LIMSI, où des formes géométriques en mouvement figurent les scènes anxiogènes. L'utilisateur est représenté dans ces scènes par un rond réagissant (taille et couleur) à ses comportements.

2.1 Scénarii

Le participant dispose d'une minute de parole dans chaque scénario pour atteindre un objectif émotionnel. L'émotion exprimée est symbolisée par la couleur du rond représentant l'utilisateur à l'écran, les codes varient selon les scénarii; L'interaction sociale avec le ou les interlocuteurs est symbolisée par la distance entre les formes.
Dans le premier scénario (FIG 1 gauche), le participant doit parler à un public de quelques interlocuteurs (N ∈ [6-12]) avec une voix neutre (i.e. sans intonation émotionnelle particulière).
Les phobiques sociaux s'imposent des standards de performance exagérés, ce qui rend ce type de présentation difficile car chaque maladresse (tremblement, rougissement, hésitation...) les déstabilise [2].

Le vert indique une voix neutre, la couleur vire au bleu dans le cas d'une émotion négative. Plus les interlocuteurs sont convaincus et plus ils sont nombreux; à l'inverse si le participant ne parle pas ou peu, il ne parvient pas à retenir l'attention du public : les interlocuteurs s'éloignent et disparaissent au fur et à mesure.

L'objectif est donc de maintenir une couleur verte et d'attirer un maximum de triangles en une minute.
Dans le deuxième scénario (FIG 1 milieu), le participant doit convaincre un proche avec une voix enjouée de l'accompagner, par une invitation au restaurant ou en vacances.
Les phobiques sociaux ont une image dévalorisée d'eux même, et ont des difficultés à présenter des arguments positifs pour supporter leurs idées [2].

La couleur rouge symbolise l'expression d'une émotion positive; au début de la scène les symboles ont des couleurs différentes pour illustrer le désaccord. Plus l'interlocuteur est convaincu par le participant et plus il se rapproche et plus sa couleur tend vers celle de l'utilisateur.
L'objectif est donc de rendre le $2^{ème}$ rond rouge et de réduire la distance entre les deux ronds en moins d'une minute.
Dans le troisième scénario (FIG 1 droite), le participant doit repousser les avances d'un vendeur, en exprimant vocalement sa colère.
Les phobiques sociaux se persuadent que leurs actions et opinions impactent de façon démesurée l'image que les autres se font d'eux, et ont des difficultés à exprimer leur désaccord [2].
La couleur rouge symbolise ici l'expression d'une émotion négative. Plus l'interlocuteur perçoit l'énervement du participant, et plus il s'éloigne de lui, jusqu'à disparaitre.
L'objectif est donc de rendre le rond rouge et de

faire disparaitre le triangle en moins d'une minute.

2.2 Application

Un outil d'animations dynamiques a été développé afin d'implémenter les scénarii. Les animations sont réalisées à l'aide de la librairie Microsoft MFC. L'application intègre un module de détection automatique des émotions dans la voix et un module de capture de distance entre l'utilisateur et le système.

Dans cette première expérimentation, un système de détection de 3 classes émotionnelles (positif, négatif, neutre) a été utilisé. La détection des émotions utilise des modèles SVM construits à partir de données prototypiques collectées dans le cadre d'un jeu sur environ 60 locuteurs [16].

La détection de distance utilise des modèles d'êtres humains se tenant debout à une distance de quelques mètres, mais les sujets étaient assis dans cette tâche.

Afin d'éliminer les erreurs de détection automatique lors de l'expérimentation, celle ci a été conduite en Wizard of Oz (WOz) : les émotions et la posture étaient envoyées au système par un humain observant la scène sans que les participants n'en soient conscients.

L'un des objectifs de l'expérimentation était la collecte de données relatives au stress et à l'attention des participants, spécifiquement dans la voix et la posture.

La voix a été enregistrée à 16KHz à l'aide du logiciel Audacity, les participants ont été filmés à l'aide d'une Kinect et de l'API OpenNi. L'application enregistre également certain évènements représentatifs de l'interaction (émotion exprimée, changement de posture, début et fin de scénario).

Chaque stimulus collecté a été évalué a posteriori avec les systèmes automatiques.

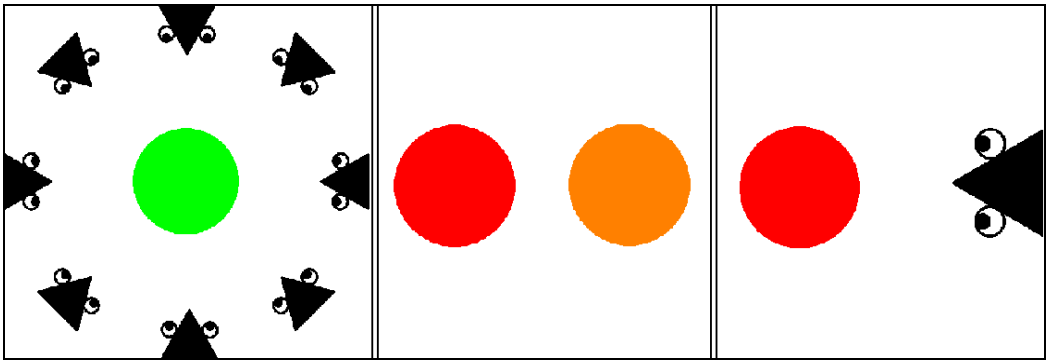


FIG 1– Animations des 3 scénarii anxiogènes à partir de formes abstraites. Les inconnus sont des triangles avec un regard orienté vers l'utilisateur. Les proches ont la même forme que l'utilisateur, un rond.

2.3 Expérimentation

Afin de tester les scénarii et de collecter des données nous avons organisé une campagne sur 2 jours. Les participants ont été recrutés parmi la population du laboratoire sur la base du volontariat.

Les volontaires étaient accueillis par 2 expérimentateurs. L'un présentait les différents scénarii et assistait les participants, l'autre se chargeait des enregistrements et contrôlait l'application en WOz à l'insu des utilisateurs. Le volontaire était équipé d'un micro-cravate AK40 (équivalent à celui utilisé pour la création du modèle de détection automatique [16]) et installé à un bureau face à un écran d'ordinateur surmonté d'une caméra Kinect, sans accès au clavier ou à la souris. Chaque session durait environ 30 minutes, afin de laisser le temps aux volontaires de réaliser tous les scénarii et de

remplir plusieurs questionnaires : le BFI comme évaluation des traits de personnalité [3], le L-SAS comme évaluation de la tendance à la phobie sociale [7], et un questionnaire destiné à évaluer l'application.

FIG 2 – Distance entre l'utilisateur et le capteur de profondeur

TAB 1 – Résultats des volontaires au BFI et L-SAS. (Echelle L-SAS : < 55 pas de phobie; 55-65 phobie modérée; 65-80 phobie marquée; 80-95 phobie sévère; > 95 phobie très sévère)

Sujet	1	2	3	4	5	6	7	8	9	10	11
Sexe	F	F	F	H	H	H	H	H	H	H	F
Age	60	59	25	26	28	42	29	23	23	25	24
Test BFI											
Extraver-sion	10	8	3	11	17	7	8	3	5	9	8
Amabilité	16	9	8	13	14	5	1	0	9	4	16
Contrôle	17	10	12	1	19	4	-4	3	-1	-5	17
Névrotrisme	4	1	0	-3	-4	4	9	8	0	3	-1
Ouverture	26	22	23	20	30	25	29	23	25	29	34
Test L-SAS											
Résultat	?	48	37	31	7	17	6	29	29	**63**	**60**

Lors de cette expérimentation, nous avons reçu 11 volontaires (4 femmes, 7 hommes) âgées de 20 à 60 ans (μ=33.9, σ=14.09), dont 2 se sont avérés phobiques sociaux modérés après étude de leurs réponses au L-SAS (sujets 10 et 11). Les résultats sont visibles dans TAB 1.

3 Analyse des données collectées

3.1 Analyse automatique

Les données collectées ont été analysées par les mêmes systèmes de détection que ceux intégrés à l'application en mode automatique.

Les données audio ont été segmentées par tâche (un segment par scénario, d'une minute ou moins si l'objectif était atteint plus rapidement) afin d'éliminer les commentaires des expérimentateurs, et les remarques hors-tâche des volontaires. Les données nettoyées ont été soumises à un modèle de détection de la valence à trois niveaux (négative, neutre, positive). Ce modèle a été conçu sur des données prototypiques et a priori sans présence spécifique de stress. Il a été validé lors d'expérimentations en laboratoire dans le cadre d'interactions hommes-machines, notamment avec le robot Nao [15]. Les données de l'un des participants n'ont pas été correctement détectées par le système, et n'ont pas pu être analysées. Les résultats détaillés de l'analyse sont visibles dans TAB 2.

TAB 2 – Analyse automatique de la valence par sujet

sujet	global sur l'expérimentation		
	négative	neutre	positive
1	100,00%	0,00%	0,00%
2	94,12%	0,00%	5,88%
3	100,00%	0,00%	0,00%
4	93,75%	4,78%	0,00%
5	79,41%	7,54%	11,76%
6	92,31%	5,02%	0,00%
7	50,00%	32,56%	0,00%
8	93,75%	4,01%	0,00%
9			
10	100,00%	0,00%	0,00%
11	52,38%	30,75%	0,00%
global	84,00%	13,14%	2,86%

Au niveau global de l'expérimentation, les voix

des participants ont été perçues comme ayant une valence négative dans 84% des cas. Une validation perceptive a confirmé que seul un participant sur 11 a atteint l'objectif de voix enjouée pour le scénario 2. Dans la majeure partie des cas, la perception était un mélange de voix neutre et stressée, l'aspect positif du scénario étant plutôt exprimé lexicalement.

En recoupant ces informations avec la difficulté perçue des scénarii (le scénario 3 a été jugé comme le plus facile, le scénario 2 comme le plus difficile, voir TAB 3), on peut émettre l'hypothèse que la valence négative dans la voix des utilisateurs est un indicateur de leur gêne.

De nouvelles expériences seront menées avec des modèles de détection automatique adaptés à cette tâche, et avec un plus grand nombre de participants, ainsi que des patients phobiques.

TAB 3 – Difficulté perçue des scénarii par sujet

Sujet	1	2	3	4	5	6	7	8	9	10	11	Moyenne
Scénario le plus facile	1	3	2	3	1	3	3	2	1	1	3	3
Scénario le plus difficile	3	2	1	2	2	2	2	1	2	2	1	2

Les données vidéo 3D ont également été segmentées par tâche. Les informations d'arrière-plan et de premier plan ont été éliminées afin de ne garder que les informations collectées entre 50cm et 100cm du capteur, ce qui correspond à la position générale du participant dans notre expérimentation (voir FIG 2).

Nous avons observé la variation de la profondeur du point central du rectangle englobant le blob de pixels contigus représentant l'utilisateur (voir FIG 3 droite). Nous souhaitions explorer en quoi cette information de posture (i.e. est-ce que l'utilisateur est plus ou moins penché vers l'écran) pouvait être significative de l'attention portée à l'animation.

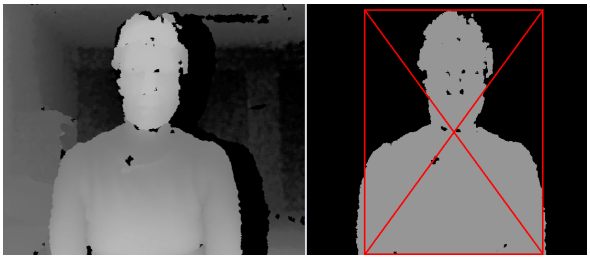


FIG 3 – Image profondeur du participant (à gauche les données brutes; à droite les données nettoyées et le rectangle englobant le blob)

Les résultats se sont avérés décevants : le capteur utilisé a une précision d'environ 10cm, et les mouvements des participants n'ont pas été suffisants pour être pertinents. Une explication possible de ces résultats est que les participants étaient assis lors de l'expérimentation, ce qui a contraint leur posture.

D'autres mesures attentionnelles mentionnées en conclusion de cet article, seront étudiées dans les expérimentations suivantes.

3.2 Analyse manuelle

Les données audio ont été annotées manuellement pour la détection du stress.

TAB 4 – Stress annoté manuellement (0 : pas de stress ; 0,5 : stressé selon au moins un annotateur ; 1 : stressé selon tous les annotateurs)

Sujet	Scénario 1	Scénario 2	Scénario 3	Moyenne	Stress ressenti
1	0	0,5	1	50,00%	n
2	0	0	1	33,33%	n
3	1	1	1	100,00%	n
4	0	0	0	0,00%	o
5	0,5	1	1	83,33%	o
6	0	1	1	66,67%	n
7	0	0,5	0,5	33,33%	n
8	0	0	1	33,33%	n
9	0	0	0	0,00%	n
10	0	1	1	66,67%	o
11	1	1	1	100,00%	o
Moyenne	22,73%	54,55%	77,27%	51,52%	36,36%

L'annotation a été réalisée par 2 annotateurs experts, sur les mêmes segments audio que ceux soumis au système de détection automatique de la valence.

Une note unique de 0 (pas de stress audible) ou 1 (présence de stress) a été attribuée à chaque segment (durée ∈ [14s-60s]) pour tous les scénarii (Ns=3) et tous les participants (Np=11). Le score d'accord inter-annotateurs sur l'ensemble des segments (N=33) est de κ=0.76.

Ces annotations sont comparées au ressenti des participants, qui ont indiqué s'ils étaient stressés suite à l'expérimentation dans le questionnaire d'évaluation. Les résultats sont visibles dans TAB 4.

Les participants ont été jugés comme stressés à 51,52%, sur l'ensemble de l'expérimentation. Les participants se sont sentis stressés à 36,36%. En détaillant ces résultats par scénario, il apparait que le scénario 3 a été jugé par les annotateurs comme le plus stressant (77,27%), suivis du scénario 2 (54,55%) et du scénario 1 (22,73%). La colère exprimée dans le scénario 3 est une colère stressée. Les participants ont évalué le scénario 3 comme le plus facile, le scénario 2 comme le plus difficile (voir TAB 3).
Ces données indiquent un décalage entre le stress perçu par un observateur extérieur, et le stress ressenti par les participants. La matrice de confusion entre stress perçu et stress ressenti et avoué présentée en TAB 5 montre une différence pour plus de la moitié des sujets.

TAB 5 – Matrice de confusion entre stress perçu et stress ressenti et avoué

	Pas de stress perçu	Stress perçu
Pas de stress avoué	9,09%	54,55%
Stress ressenti	9,09%	27,27%

Le public ciblé par l'application étant les phobiques sociaux, nous avons distingué les sujets ayant obtenu un score supérieur à 55 au L-SAS (voir TAB 1).
En observant les mêmes données, les sujets phobiques ont été jugés comme stressés par les annotateurs à 83,33% (44,44% pour les non phobiques). Ils ont reconnu leur stress à 100% (22,22% pour les non phobiques) et les observateurs l'ont perçu à 100% également.
Pour les 2 sujets les plus phobiques, les scénarii 3 et 2 ont été jugés stressant à 100%, et le scénario 1 a été stressant à 50%.
Tous les sujets sont d'accord pour estimer le scénario 3 comme le plus facile et le 2 comme le plus difficile.

Etant donné le petit nombre de participants et leur tendance peu phobique dans l'ensemble, ces chiffres sont à considérer comme des indicateurs de pistes à explorer.

4 Conclusion et perspectives

Les résultats de cette expérimentation sont encourageants. Les volontaires ont bien réagi à l'application : suite à l'expérimentation ils en ont évalué différents aspects sur une échelle de Likert de 1 à 5. Ils ont considéré l'utilisation d'animations abstraites peu gênante pour se projeter dans la situation des scénarii (2.97/5) ; ils ont évalué l'outil comme pertinent pour l'entrainement à la prise de parole (3.52/5).

Ces résultats semblent indiquer que les formes géométriques abstraites, lorsqu'elles réagissent à la voix de l'utilisateur et lorsque celui ci est placé dans un contexte spécifique, peuvent effectivement simuler une certaine interaction sociale.

Le petit nombre de candidats reçu n'est pas suffisant pour affirmer l'acceptation d'un tel outil. Une autre campagne de collecte est prévue avec les thérapeutes de Centre Emotion auprès de phobiques, et des volontaires supplémentaires seront reçus au laboratoire pour enrichir le corpus constitué et valider cet outil.

La difficulté perçue des scénarii est à pondérer par l'assimilation des participants du fonctionnement du système, malgré un premier exercice d'entrainement en début de session. Pour limiter ce biais lors de futures expérimentations, on pourra tester 3 scénarii

choisis au hasard dans un ensemble plus grand, et dans un ordre aléatoire.

Concernant les données de stress collectées, l'analyse automatique de la valence semble indiquer que le stress ressenti par les participants impacte l'expression de leurs émotions. Cette piste sera explorée par l'étude d'indices audio plus spécifiques au stress [6][13].

L'analyse automatique de la posture des candidats à partir d'une mesure unique de distance n'a pas donné de résultats comme mesure attentionnelle. Une étude combinant cette mesure avec d'autres indices attentionnels comme l'orientation du regard et la qualité des mouvements sera menée sur ces données.

L'annotation manuelle du stress dans les données audio a révélé un décalage entre le stress ressenti par les participants et le stress perçu par un observateur extérieur. Cette hypothèse pourrait être renforcée par un plus grand nombre d'annotateurs [6]. Une étude incluant les relevés d'un cardio-fréquencemètre sera menée dans une prochaine expérimentation dont les résultats seront corrélés aux mesures perceptives et à l'auto-évaluation du sujet.

Les données annotées vont permettre la constitution de modèles SVM de reconnaissance automatique spécifique au stress. Un modèle spécifique pour les phobiques sociaux sera peut-être nécessaire, cette hypothèse sera à vérifier lors de la collecte avec le Centre Emotion. De même, l'application présentée ici sera à valider lors de cette expérimentation.

Remerciements

Nous remercions Fan Yang, dont le travail réalisé lors de son stage de M2R a contribué à la réalisation de cette étude. Nous remercions également les chercheurs du Centre Emotion de la Pitié-Salpêtrière, le professeur Antoine Pelissolo et le docteur Albert Moukheiber, pour leur collaboration lors de la rédaction des scénarii.

Références

[1] Carrillo et al., "Gender differences in cardiovascular and electrodermal responses to public speaking task: the role of anxiety and mood states", *International Journal of Psychophysiology*, pp 253-264, **Vol. 42**, Num. 3, 2001

[2] D. M. Clark "A Cognitive Perspective on Social Phobia", *International Handbook of Social Anxiety: Concepts, Research and Interventions Relating to the Self and Shyness*, John Wiley & Sons Ltd., 2001

[3] O. P. John, E. M. Donahue & R. L. Kentle, "The Big Five Inventory--Versions 4a and 54." Berkeley: University of California, Berkeley, Institute of Personality and Social Research, 1991

[4] F. Heider, M. Simmel, "An experimental study of apparent behaviour" *American Journal of Psychology*, **Vol. 13**, 1944

[5] E. Klinger, "Apports de la réalité virtuelle à la prise en charge de troubles cognitifs et comportementaux." Thèse Doctorat Informatique ENST, 2006

[6] P. Laukka et al., "In a Nervous Voice: Acoustic Analysis and Perception of Anxiety in Social Phobics' Speech", *Journal of Nonverbal Behavior*, pp 195–214, **Vol. 32**, 2008

[7] M. R. Liebowitz, "Social phobia" *Modern Problems of Pharmacopsychiatry*, pp 141-173, **Vol. 22**, 1987

[8] S. M. Palma, F. S. Guimarães, A.W Zuardi, "Anxiety induced by simulated public speaking and stroop color word test in healthy subjects: effects of different trait-anxiety levels.", *Brazilian journal of medical and biological research*, pp 2895-2902, **Vol. 27**, Num. 12 ,1994

[9] D. P. Pertaub, M. Slater, C. Barker, "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience", *Presence: Teleoperators & Virtual Environments*, pp 68-78, **Vol. 11**, Num. 1, 2002

[10] G. Riva, "Virtual reality in psychotherapy: review", *Cyberpsychology & Behavior*, **Vol. 8**, Num. 3, pp 220-230, 2005

[11] Rodebaugh, "Self-efficacy and social behavior", *Behaviour research and therapy*, pp 1831-1838, Vol. 44, Num. 12, 2006

[12] B. O. Rothbaum, A. Garcia-Palacios, A. O. Rothbaum "Treating anxiety disorders with virtual reality exposure therapy", *Revista de Psiquiatria y Salud Menta*, pp67-70, **Vol. 5**, Num. 2, 2012

[13] R. Ruiz "Analyse acoustique de la voix pour la détection de perturbations psychophysiologiques – Application au contexte aéronautique", Habilitation à Diriger des Recherches, Laboratoire de Recherche en Audiovisuel (LA.R.A), 2012

[14] S. R. Sumter et al., "Age and puberty differences in stress response during a public speaking task: do adolescent grow more sensitive to social evaluation?", *Journal of Psychoneuroendocrinology*, pp 1510-1516, **Vol. 35**, Num. 10, 2010

[15] M. Tahon, A. Delaborde, L. Devillers "Real-Life Emotion Detection from Speech in Human-Robot Interaction: Experiments Across Diverse Corpora with Child and Adult Voices." *proc. of the 12th Annual Conference of the International Speech Communication Association*, pp 3121-3124, 2011

[16] M. Tahon, L. Devillers, "Acoustic measures characterizing anger across corpora collected in artificial or natural context", *proc. of the 5th International Conference Speech Prosody*, 2010

AFFIMO: Toward an open-source system to detect AFFinities and eMOtions in user's sentences

Magalie Ochs[1], Jeremy Ollivier[2], Brieuc Coic[2],Thomas Brien[2] and Fabien Majeric[2]

[1]CNRS LTCI Télécom ParisTech, France
magalie.ochs@telecom-paristech.fr,

[2]Ecole Centrale Marseille, France
firstname.lastname@centrale-marseille.fr

Résumé :
Dans cet article, nous présentons AFFIMO (AFFInités et eMOtions), une première étape vers le développement d'un système open-source de détection des émotions et des affinités de l'utilisateur à partir de phrases en langue naturelle. AFFIMO intègre une analyse lexicale à partir du dictionnaire *SentiWordNet*, une analyse syntaxique permettant de considérer des intensifieurs ainsi que les négations et une analyse sémantique pour calculer la connotation émotionnelle d'une phrase à la lumière des affinités de l'utilisateur détectées. AFFIMO a été implémenté et testé sur différentes phrases.

Mots-clés : Emotions, sentiments, traitement automatique de la langue (TAL)

Abstract:
In this paper, we present *AFFIMO* (AFFInities and eMOtions), a first attempt to develop an open-source system to detect user's emotions and affinities from sentences in natural language. AFFIMO integrates a lexical analysis based on *SentiWordNet* dictionary, a syntactic analysis to consider intensifiers and negation, and a semantic analysis to compute the valence of a sentence (positive or negative) in the light of the user's detected affinities. AFFIMO has been implemented and tested on different sentences.

Keywords: Emotions, sentiments, natural language processing (NLP)

1 Introduction

During an interaction between a user and an Embodied Conversational Agent (ECA), the detection of the user's emotions and sentiments may enable the ECA to adapt effectively the interaction. For instance, if the ECA detects that the user does not like another agent or object, the ECA may decide to avoid the use of the object or the communication with the agent. Moreover, the user's sentiments are key elements to understand the valence of a sentence. For instance, "I eat chocolate" may be either positive or negative depending if the user likes or dislikes the chocolate.

The objective of the presented research work is to develop system to determine the valence of a user's sentence (positive or negative) considering her affinities. The user's affinities toward the agents (human or virtual) and the objects are automatically detected and updated based on a specific analysis of the user's sentences.

Several research has been done to develop systems to detect emotions or sentiments in a sentence or a text. However, most of the proposed tools is specifically designed to a particular domain (movies for instance) and/or not freely available. The objective of the proposed tool, called AFFIMO (AFFInities and eMOtions), is to provide a multi-language open-source system not limited to a specific domain.

The paper is organized as follows. In Section 1, a state of art on the methods used to detect emotions and resulting existing systems are presented. Section 2 focuses on the proposed system AFFIMO. In Section 3, illustrative examples of the capabilities of AFFIMO are presented.

2 Emotion detection from text: State of Art

In this section, existing methods and techniques used to determine the user's emotions from a text typed or said by the user are presented.

2.1 Approaches to detect emotions from text

Different approaches have been explored to determine user's emotions from a text. In this section, we present the principle of each method, the resources needed to apply the method, the pros and cons of the method and we cite an example of a model using each method. In the next section, we detail some models based on the presented methods.

Keywords spotting.
- *Principle*: to determine user's emotions based on unambiguous affective terms (such as "happy", "distressed").
- *Resources needed*: dictionary of affective terms
- *Pros*: simplicity of the method (rule-based method)
- *Cons*:superficial recognition: no recognition of emotion in affective sentences without affective terms (for instance: "My husband just filed for divorce and he wants to take my children away from me").
- *Example*: Linguistic Inquiry and Word Count [11].

Lexical Affinity.
- *Principle*: to determine user's emotions based a set of words with a probabilistic affinity to an emotion (for instance the word "accident" might be assigned a 75% probability of being indicating a negative affect).
- *Resources needed*: dictionary of words with a probabilistic affinity to an emotion (information generally learned from linguistic corpora).
- *Pros*: method more sophisticate that the keywords spotting method
- *Cons*: poor recognition of emotion in metaphorical sentences (for instance: "I met my girlfriend by accident").
- *Example*: The Affective Semantic Similarity [15].

Statistical natural language processing.
- *Principle*: use machine learning algorithm on a large training corpus of affective annotated texts.
- *Resources needed*: large corpus of affective annotated text.
- *Pros*: good method for a sufficiently large text input (such as paragraph of text).
- *Cons*:
  - no semantic consideration: poor recognition on sentence
  - domain-dependent: recognition may be dependent on the training corpus.
- *Example*: Children stories automatic classification [1].

Appraisal-based Model.
- *Principle*: analyze the state of the intentions, beliefs and/or desires of the user in the text to try to determine his emotions based on psychological cognitive theory of emotions (appraisal theory).
- *Resources needed*: method to extract the state of the intentions, beliefs and/or desires of the user from a text.
- *Pros*: not only recognition, but also understanding of the causes of the user's emotions.
- *Cons*:
  - need to consider an interaction (not only a sentence or a text)
  - need a deep understanding of the sequence of events occurring in the interaction.
- *Example*: EDAMS [9].

This method is not described in details in this article since it's somehow out of scope (no recognition from a text but from an interaction).

Several researchers propose to combine different approaches to determine the user's emotions. In the next sections, we describe in more details different existing tools and techniques.

2.2 Implemented models of emotion detection from text

In this section, we present different models developed to recognize user's emotions from a sentence or a text [1]. For each model, we highlight the approach used, the emotion considered, the type of texts on which the model may be used (the domain), the originality and the limits of the model.

Emologus [7]
- *Approach*: keywords spotting + semantic relations rules
- *Emotions considered*: valence (positive and negative) and intensity of the emotions.
- *Domain*: stories for children
- *Method*: rule-based method using emotional value of words and the semantic relations between words. A value, representing a valence and an intensity, is assigned to each word (+ for positive word and - for negative word). The values are defined by experts. The corpus of words considered are vocabulary of children from 5 to 7 years old. To consider the relations between words to compute the valence

1. We do not present all existing models but some of them that we found especially relevant and original.

and intensity of the sentence, experts have defined the effects of adjectives and verbs on the nouns. For instance, "to break" has a negative effect: $(x, y) - > -y$ that means that if the noun has a positive value, the sentence has a negative valence ("break a jewelery" is negative) and if the noun has a negative value, the sentence is positive ("break a monster's leg" is positive). The rules defined on the adjective is used to compute the intensity. For instance, the rules on "kind": $x - > x + 1$ is used to increase the intensity of the sentence "kind object".
- *Originality*: rules considering the semantic relations between words.
- *Limits*:
  - domain-dependent
  - rules have to be defined manually for each verb and adjective.

EmoText [10]
- *Approach*: keyword spotting + grammatical rules
- *Emotions considered*: valence (positive/negative) and intensity (low/high)
- *Domain*: movies
- *Method*: rule-based method: the keyword spotting is based on a set of 4 500 affective words extracted from different dictionaries: WordNet Affect, Levin [5], and GI [12]. The authors have also defined neutral words corresponding to movies concepts that not convey emotions (for instance "Happy Gilmore" does not convey a positive emotion but corresponds to the title of a film). Eleven grammatical rules are used to consider intensifiers in a sentence. Rules are those defined in [4]. For instance, the rule on interjections: "Oh, what a beautiful present". The interjection "Oh" intensifies the emotional meaning of the emotion word beautiful. The 11 grammatical rules are described in the paper [10]. Moreover, negation words (for instance not, never) are considered to compute the valence of a sentence. The text is fragmented in sub-sentences and rules determine how to compute the emotion related to a sequence of sub-sentences. For instance, "Alexander is very sad, but everybody else is happy" is fragmented in two sub-sentences: the first is high negative and the second is low positive, based on the emotion word sad and happy, and on the intensifier very. The global sentence is then assumed as low negative.
- *Originality*: definition of neutral concepts + grammatical rules used to consider the intensifiers
- *Limits*:
  - domain-dependent
  - method does not consider the context (for instance "It was a good book" is positive, but if it is in a context of lost, it should be negative)

EMMA - Emotion Metaphor and Affect [17]
- *Approach*: keyword spotting + syntactic detection of affective metaphor
- *Emotions considered*: emotion types
- *Domain*: metaphor
- *Method*: method specifically designed to detect metaphorical affective expression in which emotions are considered as physical objects or events, for instance "joy ran through me", "my anger returns in a rush", "fear is killing me". The model first recognizes the specific syntactic structure of the metaphor: "a singular common noun subject + present-tense form + object/event". Then, the subject is analyzed through the affective dictionary WordNet-Affect [16] to determine the associated emotion.
- *Originality*: affect detection in metaphorical language
- *Limit*: only for a specific type of metaphor.

EmoHeart [8]
- *Approach*: keyword spotting + symbolic analysis + syntactic rules
- *Emotions considered*: anger, disgust, fear, guilt, interest, joy, sadness, shame, and surprise.
- *Domain*: internet chatting environment
- *Method*: rule-based method: the keyword spotting is based on a database created by the authors containing emoticons (English and Japanese), popular acronyms and abbreviations (for example, "BL" for "belly laughing", "cul8r" for "see you later"), interjections (such as "alas", "wow", "yay") modifiers (for instance "very", "extremely"), and emotional words extracting from WordNet-Affect. This database has been annotated manually with intensity tags. Rules are defined to determine the emotions of a message. In a symbolic analysis, if the message contains an emoticon or an emotional abbreviation, the emotion of the message corresponds to the emotion associated to this element. In a syntactic analysis, adjectives are analyzed to determine the intensity of the emotion. The negation and some prepositions (such as "without", "except") neutralize the emotional content. The sentences expressing attitudes (think, must, would, believe, know, sure,...) or condition

(if, when, whenever, ...) are not considered as emotional. The emotion associated to the sentence is the dominant emotion if there is a contradiction (subject and object corresponding to opposite emotions).
- *Originality*: consider informal messages (abbreviations, emoticons, etc).
- *Limits*: simplistic method with the only advantage of considering emoticons and abbreviations.

Empathy Buddy [6]
- *Approach*: keyword spotting + statistical natural language processing
- *Emotions considered*: 6 basic emotions (anger, disgust, fear, joy, sadness, surprise).
- *Domain*: not specific
- *Method*: keyword spotting rules based on a knowledge base of commonsense: Open Mind Common Sense (OMCS) (half a million sentences in English with 10% of affective sentences). *Affective sentences* of the knowledge base are extracted and annotated using keyword spotting with the affective lexicon defined in the OCC model. *Concepts* in the sentences are associated to emotions. For instance, "car accident" is associated to fear because there is a sentence in the data base annotated with "fear", such as "car accident can be scary". Emotions are also associated to modifier given the label of the sentence. For instance, "Modly"is associated to disgust given the sentence "Modly bread is disgusting". Some hand-coded rules have been defined to determine the valence of sentences. For instance, "I wrecked my car": narrator neg-verb pos-object -> neg-valence. A propagation of the affective annotation with two or three passes (with a certain factor *d* to define) are settled. For instance:
  "Something exciting is both happy and surprising":
  Pass 1: the word "exciting" is associated to joy (intensity 1) and surprise (intensity 1).
  "Rollercoasters are exciting"
  Pass 2 (with d=0.5): the word "rollercoaster" is associated to joy (intensity 0,5) and surprise (intensity 0,5).
  "Rollercoasters are typically found at amusement parks"
  Pass 3 : the word "amusement park" is associated to joy (intensity 0.25) and surprise (intensity 0.25).
  Moreover, various techniques are used to smooth the transition of emotions from one sentence to the next. For instance, a decay technique enables to decrease the emotion associated to the sentence if the next one is neutral, an interpolation method defines that a neutral sentence between two angry sentences is associated to angry with smaller intensity. Meta-emotions, emotions that are not part of the six basic emotions, can emerge. For example, the meta-emotion frustration emerges in the case of repetition of low-magnitude anger, and relief if fear is followed by happy. The authors use the dynamic of emotions to infer new emotions.
  The model has been implemented in a mailing system *Empathy Buddy* that detects emotions when a user types a message and changes a virtual face accordingly. The system has been evaluated positively (entertainment, intelligent, interactive, users will use it).
- *Originality*: used on a knowledge base of commonsense that constantly evolves + propagation rules + emergence of meta-emotions
- *Limit*: the proposed method does not consider the context of the sentence.

Sentistrenght [14]
- *Approach*: statistical natural language processing
- *Emotions considered*: valence (positive/negative) and strength (intensity-arousal).
- *Domain*: identify the sentiment expressed in a message. The sentiment may correspond to the author's hidden internal state, the intended message interpretation, or the reader's hidden internal state.
- *Method*: machine learning with a specific affective lexicon and algorithms: authors have collected *affective terms* (298 positive terms and 465 negative terms with intensity values) from the annotation of comments and words of MySpace (2 600 human-classified MySpace comments and words). The classification of affective terms are *optimized* using an algorithm. This algorithm starts with the human-allocated term and strength for the predefined list, and then, for each term, the algorithm assesses whether an increase or decrease of the strength by 1 would increase the accuracy of the classifications. Algorithms to *correct misspelling*, to consider *negation* and *emoticon* are used.
  A *booster algorithm* is defined to take into account some words such as "very" "some" and repeated letter or punctuation, that may boost or reduce the strength of emotions. Finally, the message is assigned with *both the most positive and most negative emotion identified* in it. The results of the evaluation shows 60%

of good detection of the valence of a message with appropriate intensity. The comparison of *sentistrenght* with learning algorithms reveals that *sentistrenght* offers globally better results.
- *Originality*: determine the strength of emotions and identify both positive and negative emotions in a sentence. For instance, "I have no opinion about anything at all" has positive valence with strength 1 and negative valence with strength 1 .
- *Limit*: method does not consider the context of the text and this method is not adapted for long text.

To summarize this state of art section, different approaches may be considered to detect emotions from a text. Rule-based methods using keyword spotting technique and specific rules to consider semantic relations between words or syntax are often used. The keyword spotting is generally based on an affective dictionary that may be enriched using machine learning techniques (optimization, feature extraction). The specific rules to refine the detection should at least consider negation and intensifiers (such as adjectives, punctuation, interjections, ...). Other rules may be considered for specific text. For instance, for informal messages, rules to take into account emoticons, abbreviations, and metaphors may be defined. Some methods are based on annotated corpus of sentences. Note that in the context of emotions, it may be difficult to obtain agreement between annotators, especially to annotate intensity of emotions. Finally, in existing works, the emotions are represented either by a valence (positive versus negative) or a type (joy, anger, sadness,...) and sometimes by intensity. The choice of emotion representation may be motivated by the emotional classes occurring in the affective resources (labels in the dictionary or in the annotated text). It makes sense to consider intensity only if rules on intensifiers are defined. Concerning domain-dependency, since each domain has its specificity, and in particular emotional and non-emotional terms may depend on the domain, the affective dictionary should be adapted to each domain through specific rules, for instance.

The main limit of the presented existing systems to detect emotions from text is their usability for human-ECA interactions. Indeed, they are either specifically designed for certain domains (e.g. children's stories or movies) or not freely available. The objective of the work presented in this paper is to start to develop an open-source system to detect user's emotions expressed in sentences during a dialog with an ECA.

3 AFFIMO: an open-source system to detect AFFinities and eMOtions in user's sentences

The system AFFIMO is a first attempt to develop an open-source tool to detect emotions and affinities in a written sentence.

Emotions and Affinities. The *emotion* detected in a sentence is represented by a valence (positive - negative) and an intensity. Note that our objective is not to detect the emotion felt by the user. We aim at detecting the emotion expressed by the user in a sentence. The emotion expressed may be different from the felt emotion. The *affinities* correspond to the *degree of appreciation* of the speaker toward the other objects or agents cited in the sentences. An affinity is represented by a 3-uplet $< agent1, agent\_object2, value >$ to characterize the degree of appreciation $value$ of $agent1$ toward $agent\_object2$. The value $value$ varies in the interval $[-10, 10]$. For instance, $< bob, chocolate, 10 >$ means that the agent bob highly appreciates the chocolate. The initial affinities is set up in a text file [2] They are updated depending on the affinities detected in sentences. The algorithm used is described in more details in the following.

Lexical analysis. In AFFIMO, the first step is the study of the valence of the words contained in the sentence. For this purpose, we use the affective dictionary *SentiWordNet* [3]. This dictionary has the advantage to be freely available [3]. In this dictionary, to each term is associated a positive and negative score [4]. These values correspond to the positive or negative connotation of the term. The *SentiWordNet* dictionary is particularly well-adapted to our research purpose to detect emotions and affinities (through the positive and negative connotations of words). Moreover, AFFIMO considers several smilies (such as ":-)" to compute the valence of the sentence.

Syntactic analysis. To consider the negation in the sentence, the valence of a sentence in a neg-

2. The text file describing the affinities should ideally be defined for each user. Indeed, the affinity may vary from one user to another.

3. http://sentiwordnet.isti.cnr.it/

4. An objective score is also associated to each term. We do not consider this score in AFFIMO.

ative form is inversed. To characterize the effects of intensifiers such as "very" or "small", a list of intensifiers is described in a text file. This list is extracted from [2] in which 179 English intensifiers are described with associated values traducing their impacts. Finally, the syntactic analysis studies the negative form of the sentence and the intensifiers. The identification of the role of each word in the sentence (subject, object, verb, etc.), i.e. the lemmatization, is done during the semantic analysis.

Semantic analysis. The valence of a sentence depends a lot on the *point of view*. For instance, the sentence "the cat eats the dog" could be emotionally positive from the cat's point of view (if we suppose that the dog has a good taste) but negative from the dog's point of view (if we suppose that the dog liked his dog's life). Moreover, a sentence may be interpreted differently depending on the affinities. For instance, "John breaks the vase" could be positive if John does not like the vase but negative in the contrary. To consider the point of view and the affinities to compute the valence of a sentence, a semantic analysis is performed. Firstly, we use the *TreeTagger* tool [13] for the lemmatization of the sentence. This tool enables one to identify the subject, adjectives, objects, and verbs in a sentence. A simple algorithm has been developed, based on the outputs of the *TreeTagger*, to moreover detect the passive form of a sentence.

Secondly, the semantic analysis in AFFIMO is largely inspired from the model *Emologus* [7] (described in the previous section). The effects of the verbs on the nouns are defined in a text file. For instance, the verb "eat" is supposed to have a positive effect on the subject and a negative one on the object. As described previously, the initial affinities are defined in an external text file. The affinities are updated based on a analysis of the structure "adjective noun" in a sentence. The effect of an adjective on the following noun is computed based on the emotional value of this adjective in the *SentiWordNet* dictionary. For instance, "a nasty cat" is considered as "negative". Consequently, the sentence "It's a nasty cat" changes the affinity value of the speaker toward the cat as negative [5]. Then, if the speaker says "the dog eats the nasty cat", this sentence is interpreted as positive.

Finally, without any semantic information (affinities or verbs' effects), the computation of the valence of an input sentence is based only on a lexical analysis (mean of the valence of the words contained in the sentence) and on a simple syntactic analysis. The emotional detection may be fine-grained with a semantic analysis considering the effects of verbs on the subject and the objects, and considering the current affinities of the speaker.

Multi-language analysis. Most of the affective dictionaries is available only in one language (mainly English). To enable AFFIMO to detect emotions and affinities in sentences written in different languages, we have used the *Web Translator Java API* [6] to automatically translate any sentence in English. This API supports the translation for 14 languages (including French, Spanish, German, etc.). A sentence written in a language different from English is automatically translated in English before starting the analysis. The result of AFFIMO becomes dependent from the accuracy of the translation.

4 Evaluation

To evaluate AFFIMO, we have selected 10 sentences from the French *EmoLogus* database [7]. The database contains several sentences extracted from a French fairy tale "Comment le Grand Nord découvrit l'été" ("How the big North discovered the summer"). The database has the advantage that 31 annotators have attributed an emotional value to every sentences through a single scalar value including valence and intensity.

In the presented evaluation, we have compared the valence of the emotion provided by AFFIMO to the valence indicated by the annotators. To assess the multi-language capabilities of AFFIMO, the sentences entered in the system are in French, as the originals. The sentences and the comparisons between the annotators' rates and the results from AFFIMO are described Table 1.

On 10 sentences, AFFIMO misclassified 3 of them, compared to the annotators' rates. The differences between the annotators' rates and the AFFIMO results for these three sentences may be explained by the lexical analysis. Indeed, some of the words that could have a valence (such as "tiède" in the sentence 6 or "'soupiré" in the sentence 8) are not described

5. In this case, the affinity of the speaker toward the cat (and not all the cats) is modified; we suppose that there is only one cat. The model does not deal with the deictic and the reference.

6. http://webtranslator.sourceforge.net/

Table 1: Evaluated sentences rated by annotators and analyzed by AFFIMO (*neg.* - resp. *pos.* - means that the sentence has been rated with a negative - resp. positive - valence).

Id	*Sentences extracted from the EmoLogus database [7]*	*Annotators*	*AFFIMO*
1	Jadis il y a très longtemps c'était toujours l'hiver sur les terres du Grand Nord.	neg.	neg.
2	Toute l'année le soleil se levait tard et disparaissait très vite.	neg.	pos.
3	neige et glace ne fondaient jamais.	neg.	pos.
4	les animaux avaient toujours froid aux pattes.	neg.	neg.
5	Ils étaient tristes et engourdis.	neg.	neg.
6	Là-bas l'air est tiède et parfumé.	pos.	pos.
7	les oiseaux chantent tout le temps.	pos.	pos.
8	Le vent a soupiré	neg.	neutral.
9	il y a que les oiseaux qui pourraient rapporter l'été mais ils sont prisonniers.	neg.	neg.
10	Que c'était beau !	pos.	pos.

emotionally as expected ("tiède" in *SentiWordNet* has a negative connotation and "soupirer" is not present in the dictionary). Moreover, some sentence should be analyzed in the light of the context of the story. For instance, the sentence 2 may be interpreted either positive or negative depending on the context.

5 Conclusion

In conclusion, in this paper, we have presented AFFIMO a first attempt to develop a multi-language open-source system to detect the valence of a sentence based on a lexical, syntactic and semantic analysis considering the detected user's affinities.

Compared to the presented existing models (Section 2.2), AFFIMO has the advantages to be open-source, multi-language and to consider the user's affinities to compute the valence of a written sentence. However, the proposed system presents several limits. For instance, the user's affinities is inferred from the valence of adjectives preceding nouns in the sentences. Simple improvements, for instance by analyzing sentences containing "like" or "hate", could be integrated. The valence of a sentence is computed according to the user's affinities. However, the utility of the object/person should be considered to compute the valence. Indeed, one may dislike an object but the fact that this object is broken could be negative if this object is extremely useful. Grammatical rules, as proposed in the system *EmoText*, could improve the computation of the intensity. Compared to the model *EMMA*, the metaphors are not considered. The system computes only the valence positive or negative of a sentence whereas others existing models compute specific emotion types. Moreover, to evaluate the capacity of AFFIMO, the evaluation should be extended by considering a larger corpus of sentences. Other affective dictionaries, such as WordNet-Affect, could be integrated to compute the valence of the sentence or to consider specific types of emotion. An experimentation with users evaluating the AFFIMO results on the inferred affinities should enable us to improve the system.

Last but not least, a major problem of AFFIMO is that it is not real-time whereas our final purpose is to use AFFIMO to automatically detect, in real-time, the valence of a user's sentence interacting with a virtual character.

6 Acknowledgments

This research has been supported by the European Community Seventh Framework Program (FP7/2007-2013), under grant agreement no. 231287 (SSPNet).

Moreover, the authors would like to thank *François Brucker* (Ecole Centrale Marseille) for his participation to this project and *Jean-Yves Antoine* (Université François Rabelais) for the EmoLogus database that enables us to perform a first evaluation of AFFIMO.

References

[1] C. O. Alm, D. Roth, and R. Sproat. Emotions from text : machine learning for textbased emotion prediction. In *Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 2005.

[2] J. Brooke. *A Semantic Approach to Automated Text Sentiment Analysis*. PhD thesis, Simon Fraser University, 2009.

[3] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation*, pages 417–422, 2006.

[4] G. N. Leech. *A communicative grammar of English. Third edition.* Longman Publishing Group, 2003.

[5] B. Levin. *English verb classes and alternations*. The University of Chicago Press, 1993.

[6] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of the Conference on Intelligent user interfaces (IUI)*. ACM-Press, 2003.

[7] Le Tallec M., Villaneau J., Antoine J.-Y., and Duhaut D. Affective interaction with a companion robot for vulnerable children: a linguistically based model for emotion detection. In *Proc. LTC'2001, Language Technology Conference*, 2011.

[8] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Textual affect sensing for sociable and expressive online communication. In *Proceedings of the Affective Computing and Intelligent Interaction Conference (ACII)*. Springer-Verlag, 2007.

[9] M. Ochs, C. Pelachaud, and D. Sadek. An empathic virtual dialog agent to improve human-machine interaction. In *Autonomous Agent and Multi-Agent Systems (AAMAS)*, 2008.

[10] A. Osherenko. Emotext: Applying differentiated semantic analysis in lexical affect sensing. In *Proceedings of the Affective Computing and Intelligent Interaction Conference (ACII)*, 2009.

[11] J.W. Pennebaker, M.E. Francis, and R.J. Booth. *Linguistic Inquiry and Word Count: LIWC2001*. Mahwah, NJ: Erlbaum Publishers, 2001.

[12] R. Quirk and S. A Greenbaum. *University Grammar of English*. Longman Publishing Group, 1988.

[13] H. Schmid. Improvements in part-of-speech tagging with an application to german. In *In Proceedings of the ACL SIGDAT-Workshop*, pages 47–50, 1995.

[14] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment in short strength detection informal text. *Journal of the American Society for Information Science*, 61(12):2544–2558, December 2010.

[15] A. Valitutti, C. Strapparava, and O. Stock. Lexical resources and semantic similarity for affective evaluative expressions generation. In *Proceedings of the Affective Computing and Intelligent Interaction Conference (ACII)*, 2005.

[16] R. Valitutti. Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086, 2004.

[17] L. Zhang. Exploration of affect sensing from speech and metaphorical text. In *Proceedings of the 4th International Conference on E-Learning and Games: Learning by Playing. Game-based Education System Design and Development*. Springer-Verlag, 2009.

Fusion anticipée de descripteurs bas niveau pour la détection d'émotions dans les textes

F. Dzogang⋆ fabon.dzogang@lip6.fr
M.-J. Lesot⋆ marie-jeanne.lesot@lip6.fr
M. Rifqi† maria.rifqi@lip6.fr

⋆LIP6, Université Pierre et Marie Curie-Paris 6, UMR7606, Paris, France

†LIP6 & Université Panthéon-Assas, Paris, France

Résumé :
Cet article considère la tâche de classification de textes selon leur contenu émotionnel. La méthode proposée consiste à combiner, dans une approche de fusion anticipée, des descripteurs bas niveau définis comme des n-grammes de plusieurs ordres. Elle effectue de plus une extraction automatique de dictionnaires spécialisés pour chacune des émotions considérées. Elle applique un processus de décision de type « un contre tous » à deux niveaux, utilisant des classifieurs linéaires. L'approche proposée est appliquée à un corpus de textes réels dont chaque phrase est étiquetée selon 15 émotions.

Mots-clés : détection automatique d'émotions, fouille de textes, fusion anticipée

Abstract:
This paper addresses an emotion detection task, i.e. text classification according to their emotional content. The proposed method consists in combining, in an early fusion setting, low level features defined as n-grams of different orders. Furthermore it performs an automatic extraction of dictionaries specialised for each emotion. The decision process exploits a two-level "one vs all" approach based on linear classifiers. The proposed method is applied to a real-world text corpus where each sentence is labelled according to a 15 emotion model.

Keywords: emotion mining, text mining, early fusion

1 Introduction

La tâche de détection d'émotions dans les textes vise à effectuer une classification automatique de textes selon leur contenu émotionnel, par exemple pour déterminer si un texte exprime de la joie, de la colère ou de l'amour. La difficulté de cette tâche peut être expliquée par plusieurs facteurs, parmi lesquels le manque de consensus concernant les émotions à considérer, qui entraîne entre autres le problème de constitution de corpus étiquetés, ou la complexité des expressions linguistiques émotionnelles.

En effet, la définition des émotions et sentiments est complexe, et les travaux en psychologie ont donné lieu à de nombreux modèles et représentations. On peut citer en particulier les modèles dimensionnels et les modèles catégoriels, qui ont été mis en œuvre pour le traitement automatique de textes [6] : les premiers représentent les sentiments comme des vecteurs dans des espaces multi-dimensionnels continus, dont les axes varient selon les auteurs [9] ; les modèles catégoriels décrivent les émotions par un ensemble discret d'étiquettes, dont le nombre et la liste varient selon les auteurs [7, 19].

Le manque de consensus des modèles d'émotions rend la tâche d'annotation de documents complexe, et il existe peu de corpus permettant la mise en œuvre d'algorithmes d'apprentissage automatique. Les principaux corpus disponibles ont été développés dans le cadre de compétitions. Ainsi la tâche 14 de SemEval 2007 [21] a conduit à une base de dépêches annotées, mais elle est de petite taille ; la compétition I2B2 [17] a conduit à la collecte de 600 textes dont chaque phrase a été étiquetée manuellement par des experts. Dans ces deux cas, des modèles catégoriels, conduisant à des problématiques de classification, ont été utilisés.

Une autre source de difficulté de la reconnaissance d'émotions dans les textes vient de la complexité des expressions linguistiques émotionnelles : celles-ci sont naturellement subjectives, mais aussi le plus souvent imprécises, ambiguës et contextuelles [6]. Les outils classiques de classification automatique de textes peuvent alors se révéler insuffisants. En particulier, à titre d'exemple, pour des tâches de classification thématique, le rôle de la négation est faible ; au contraire distinguer si une phrase est affirmative ou négative est essentiel pour la détection d'émotions et en particulier de la détection de la polarité d'une portion de texte.

Cet article propose une méthode de classification selon le contenu émotionnel exclusivement basée sur des *descripteurs bas niveau*, qui n'exploite pas d'information sémantique. Cette caractéristique permet d'éviter le problème de la constitution de ressources de haut niveau, construites par des linguistes ou des psychologues, qui nécessitent des expertises ; elle offre aussi une indépendance par rapport à la langue

dans laquelle les textes sont écrits. La méthode proposée exploite néanmoins différents descripteurs, combinés dans une procédure de *fusion anticipée* : plusieurs représentations complémentaires sont exploitées simultanément dans l'étape de description des données. De plus, pour s'adapter à la particularité des catégories émotionnelles, la méthode construit automatiquement des *dictionnaires spécifiques* à chaque émotion : elle se place dans l'hypothèse selon laquelle les émotions sont associées à des vocabulaires propres, qui contribuent à distinguer les émotions les unes par opposition aux autres. Enfin la méthode proposée applique un processus de décision selon le principe *un contre tous, à deux niveaux* : une première étape distingue les textes neutres des textes exprimant une émotion ; une seconde étape pour les textes non neutres combine les prédictions de chaque émotion par opposition à toutes les autres.

L'article est organisé de la façon suivante : la section 2 donne un état de l'art des descripteurs utilisés pour identifier automatiquement le contenu émotionnel de textes. La section 3 décrit l'espace de représentation proposé ; la section 4 présente la méthode de classification employée et la section 5 les résultats expérimentaux obtenus sur les données réelles I2B2 [17].

2 Etat de l'art : représentation de textes pour la détection d'émotions

Bien que des approches dimensionnelles aient été proposées, le plus souvent pour la caractérisation de contenu émotionnel [5], la plupart des méthodes de détection d'émotions exploitent une représentation catégorielle des émotions : elles répondent à une tâche de classification automatique multi-classes et visent à associer à un document d une classe parmi M, où M désigne le nombre d'émotions considérées. Nous présentons dans cette section les descripteurs que ces méthodes exploitent, en les distinguant suivant leur richesse sémantique.

2.1 Descripteurs bas niveau

n-grammes Un document textuel est un ensemble de paragraphes composés de phrases, elles-mêmes composées de mots. Il est d'usage, en apprentissage sur les textes, de considérer le mot comme l'unité d'information atomique. L'ensemble des mots présents dans un corpus de documents est appelé dictionnaire. Les entrées du dictionnaire sont appelées descripteurs ou dimensions.

Le dictionnaire peut être enrichi par ajout d'autres entrées contenant plus d'information, en considérant non seulement les mots isolés, mais aussi des suites de n mots consécutifs, appelés n-grammes. On parle d'unigramme pour $n = 1$, de bigramme pour $n = 2$ et de trigramme pour $n = 3$. Les n-grammes permettent de modéliser le contexte d'apparition des mots : le n-ième mot d'un n-gramme peut être vu comme un unigramme associé à son contexte d'apparition de taille $n - 1$. Un exemple intéressant de contexte est l'emploi de la négation, qui joue un rôle important pour la détection d'expressions émotionnelles : ainsi le changement de polarité dû à l'adverbe *pas* dans l'expression *pas mauvais* est capturé par un bigramme et non par les unigrammes. Des constructions plus subtiles telles que *pas vraiment mauvais* sont représentées par des trigrammes ; des grammes d'ordre supérieur permettent de décrire des expressions plus complexes encore.

De plus, dans le cas de la détection du contenu émotionnel de textes, des descripteurs spécifiques ont été proposés pour enrichir les dictionnaires : certains auteurs tiennent compte de la ponctuation [10, 5]. De même, pour tenir compte du rôle de la négation dans l'expression des émotions, certains auteurs doublent la taille du dictionnaire afin de modéliser les négations de termes [4]. Enfin, les *émoticones* contiennent de précieuses informations et sont souvent ajoutées aux dictionnaires [10, 15, 11].

Filtrage Le dictionnaire brut contient généralement de nombreuses entrées non pertinentes pour décrire les textes, susceptibles de parasiter le processus d'analyse. Aussi il est filtré lors d'une étape de pré-traitement : une phase de lemmatisation regroupe les différentes flexions d'un même mot en une forme canonique. De plus, il est parfois restreint à certaines catégories grammaticales, comme les verbes et adjectifs dans le cas de l'analyse de sentiments [12, 2]. D'autres types de traitement éliminent les mots les plus fréquents ou les moins fréquents d'un corpus ou utilisent une liste de *stop words* pour écarter les mots communs d'une langue.

On appelle dictionnaire d'ordre n, noté $\mathcal{D}^n$, l'ensemble des n-grammes pertinents apparaissant dans un corpus,la pertinence étant définie par les critères de filtrage précédents.

Représentation d'un texte Les documents sont ensuite représentés comme des vecteurs dans l'espace composé de l'ensemble des descripteurs du dictionnaire. Les composantes du vecteur peuvent être binaires, indiquant la présence ou l'absence du descripteur dans le document considéré ou tenir compte de leur fréquence d'apparition relative (pondération fréquentielle ou *tfidf*).

2.2 Descripteurs haut niveau

Le domaine de l'analyse de textes est confronté au problème du « fossé sémantique » (*semantic gap*) : la représentation précédente n'est qu'un ensemble de symboles dénués de toute sémantique. L'utilisation de descripteurs haut niveau vise à enrichir la représentation des documents en y incorporant des connaissances induites par un corpus d'étude auxiliaire ou des connaissances sémantiques externes.

En effet l'analyse statistique d'un corpus d'étude, par analyse sémantique latente (LSA) par exemple, peut conduire à constituer automatiquement un espace de concepts, d'un niveau d'abstraction supérieur aux mots isolés. Des espaces de concepts affectifs peuvent également être constitués par des connaissances externes, par exemple fournies par des taxinomies ou des ontologies. Celles-ci peuvent être construites par des linguistes [14, 18] ou obtenues par augmentation de ressources sémantiques existantes comme *Wordnet* par des propriétés affectives [22, 1, 8]. Enfin de nombreuses ressources projetant des termes affectifs dans un espace affectif dimensionnel ont été constituées par des psychologues [20, 13, 9].

3 Représentation proposée : combinaison spécialisée de dictionnaires

La méthode proposée pour la détection d'émotions dans les textes est schématisée sur la figure 1 : elle est classiquement constituée de deux étapes, respectivement de représentation des données et de classification, successivement décrites dans cette section et la suivante.

3.1 Principe

Nous proposons d'utiliser uniquement des descripteurs bas niveau, afin d'éviter une dépendance à des expertises extérieures, linguistiques ou psychologiques, et une dépendance à la langue des textes à traiter.

Selon l'approche classique, nous utilisons pour dictionnaire un ensemble de n-grammes fréquents, et nous proposons de combiner différents ordres : nous considérons que les unigrammes ne suffisent pas à discriminer les concepts affectifs, intrinsèquement subtils et complexes. Toutefois, ils sont également nécessaires car ils définissent des descripteurs simples et génériques qui garantissent une bonne couverture. Aussi l'espace de description considéré contient à la fois les unigrammes et des descripteurs plus riches qui permettent de tenir compte du contexte d'apparition des mots isolés.

Nous proposons également de spécialiser le dictionnaire pour chaque émotion, selon l'hypothèse que chacune possède un vocabulaire propre : certains mots sont particulièrement utiles pour identifier une émotion par opposition à une autre, et leur pouvoir de discrimination dépend de l'émotion considérée.

Nous décrivons dans les sous-sections suivantes la procédure de filtrage proposée afin de spécialiser le dictionnaire selon les émotions, puis la stratégie de combinaison, effectuée selon un principe de fusion anticipée.

3.2 Spécialisation du dictionnaire selon les émotions

Nous proposons d'enrichir l'étape de filtrage des n-grammes pertinents, pour construire un dictionnaire propre à chacune des émotions. On peut ainsi filtrer les descripteurs non pertinents pour certaines des classes à identifier.

Le critère de filtrage proposé est basé sur une mesure de la quantité d'information apportée par un descripteur f issu d'un dictionnaire $\mathcal{D}^n$ pour la prédiction d'une émotion e. Des mesures basées sur la log-vraisemblance pondérée [16] ou le score du χ^2 [3] ont été exploitées, nous proposons d'utiliser la mesure d'entropie de Shannon : formellement, en notant p la proportion de documents étiquetés e parmi les documents contenant le descripteur f, le critère proposé est défini comme

$$H_e(f) = -\left[(1-p)\log_2(1-p) + p\log_2(p)\right]$$

Il est maximal si f est uniformément distribué, c'est-à-dire autant présent dans les documents étiquetés e que les autres et minimal si tous les

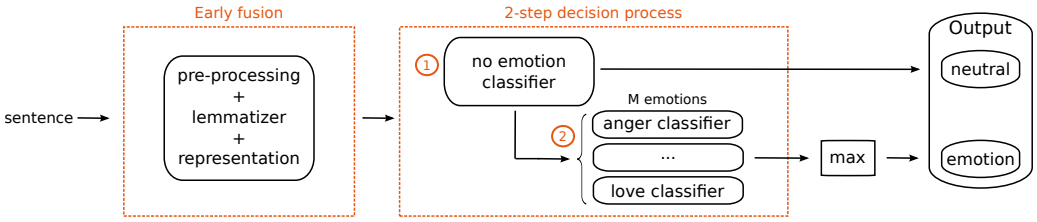


FIGURE 1 – Architecture globale de la méthodologie proposée

documents contenant f ($p = 1$) ou aucun ($p = 0$) sont étiquetés e, c'est-à-dire la présence ou l'absence de f sont totalement discriminantes.

Nous proposons donc de filtrer un dictionnaire en conservant uniquement les descripteurs associés à une entropie inférieure à un seuil :

$$\mathcal{D}^n(e) = \{f \in \mathcal{D}^n / H_e(f) < \epsilon(n, e)\} \quad (1)$$

Le seuil $\epsilon(n, e)$, défini indépendamment pour chaque ordre n et chaque émotion e, peut par exemple être estimé automatiquement, en fonction des performances des classifieurs basés sur les dictionnaires correspondants.

3.3 Combinaison des dictionnaires : fusion anticipée

Chaque émotion est alors associée à plusieurs dictionnaires filtrés, $\mathcal{D}^n(e)$ pour $n = 1..p$, qui correspondent à différents ordres de n-grammes, chacun fournissant son propre espace de description comme indiqué dans la section 2.

Trois stratégies de combinaison classiques peuvent être distinguées : la *fusion anticipée* combine les différents espaces au niveau du vecteur construit ; la *fusion intermédiaire* agit au niveau des mesures de similarité associées aux espaces de description ; la *fusion tardive* effectue la combinaison au niveau des classifieurs, appris indépendamment dans chaque espace.

Nous proposons de combiner les dictionnaires $\mathcal{D}^n(e)$, $n = 1..p$, associés à une émotion e fixée par fusion anticipée : la représentation d'un texte est obtenue par concaténation des vecteurs obtenus dans chaque espace de description. Formellement, pour un document x et une émotion e, la représentation utilisée pour prédire cette émotion est : $\vec{x}(e) = \vec{x}^1(e) \oplus \ldots \oplus \vec{x}^p(e)$.

Globalement, un document x est donc associé à M vecteurs, chacun construit dans $\mathcal{D}(e) = \cup_{n=1}^{p} \mathcal{D}^n(e)$. Ce processus est justifié par l'architecture du processus de décision proposé détaillé dans la section suivante, qui s'inscrit dans le cadre de l'apprentissage « un contre tous » : un classifieur est appris indépendamment pour chacune des émotions à prédire. Il est donc pertinent de définir un espace de description propre à chaque émotion.

4 Classification : un contre tous à 2 niveaux

4.1 Stratégie de classification

L'apprentissage « un contre tous » consiste à décomposer une tâche de classification multi-classes en multiples sous-tâches de classification binaires, chacune discriminant une des classes par opposition à toutes les autres.

La stratégie de classification proposée, schématisée sur la figure 1, s'inscrit dans ce cadre, en considérant toutefois deux niveaux : un premier classifieur vise à distinguer si un texte est porteur d'une émotion ou non, c'est-à-dire considère une classe « neutre » par opposition à toutes les « vraies » émotions. Dans un second temps, si le texte a été classé comme exprimant une émotion, il est soumis à une seconde chaîne de traitement, dans laquelle M systèmes de décision discriminent chacun une émotion e contre toutes les autres. Le classifieur prédisant une émotion avec la confiance maximale détermine la classe globalement prédite.

Le choix de cette structuration est motivé par sa capacité à distinguer d'abord les textes objectifs des textes subjectifs, puis à raffiner l'analyse de ces derniers.

Le corpus d'apprentissage soumis au classifieur associé à une classe e (neutre ou non) contient comme exemples positifs les documents étiquetés e et comme contre-exemples les autres documents. Toutefois, au second niveau de la classification, seuls les textes subjectifs sont considérés : les contre-exemples excluent les documents étiquetés comme neutres.

Enfin, pour le classifieur associé à une classe e (neutre ou non), les documents sont représentés dans l'espace $\mathcal{D}(e)$ établi comme indiqué dans la section précédente.

4.2 Classifieur mis en œuvre

Chaque classifieur est un séparateur à vaste marge (SVM), avec noyau linéaire : d'une part leur efficacité pour la classification de textes a été observée dans de multiples applications ; d'autre part, les frontières de décision produites permettent d'interpréter le rôle des différents descripteurs. La fonction de prédiction s'écrit sous la forme

$$f(\vec{x}) = \text{sign}(\langle \vec{x}, \vec{\alpha} \rangle + b) \qquad (2)$$

où le vecteur $\vec{\alpha}$ et le réel b sont les paramètres optimisés lors de la phase d'apprentissage.

Pour traiter le déséquilibre des classes, nous utilisons la version asymétrique du coût de classification, qui le définit comme inversement proportionnel à la fréquence des classes dans le corpus d'apprentissage. Sa valeur optimale est estimée par validation croisée.

5 Résultats expérimentaux

Cette section décrit les données et le protocole expérimental mis en œuvre, puis les résultats obtenus, en analysant les scores de classification des émotions ainsi que quelques caractéristiques des dictionnaires spécialisés obtenus.

5.1 Corpus de documents

Les expériences sont réalisées avec le corpus de la compétition I2B2 track 2 [17], constitué de lettres de suicide dont chaque phrase est étiquetée manuellement selon 15 classes. Celles-ci comportent 12 émotions auxquelles sont ajoutées les catégories *neutral*, *instruction* et *information* qui permettent de préciser éventuellement la classe des phrases sans contenu émotionnel. Bien qu'objectives, les deux dernières catégories sont considérées comme associées à des vocabulaires spécifiques comme les émotions, et sont traitées comme celles-ci.

Le corpus est alors constitué de 4241 documents, qui ont la particularité d'être très courts, puisque réduits à une seule phrase. Aussi, une pondération de type *tfidf* n'est pas pertinente.

Classes	Fréquence
Neutral	2460
Instruction	800
Hopelessness	455
Love	296
Information	295
Guilt	208
Blame	107
Thankfulness	94
Anger	69
Sorrow	51
Hopefulness	47
Happiness/Peacefulness	25
Fear	25
Pride	15
Abuse	9
Forgiveness	6

TABLE 1 – Distribution des étiquettes au sein du corpus d'apprentissage, ordonnées de la plus fréquente à la plus rare.

Nous adoptons donc le schéma binaire : chaque composante indique si l'entrée correspondante apparaît ou non dans le document considéré.

Le dictionnaire, avant spécialisation par émotion, est basé sur les mots apparaissant dans le corpus soumis à une étape de lemmatisation, auxquels sont ajoutés les signes de ponctuation. Il est ensuite constitué des n-grammes qui peuvent en être déduits. Un filtrage par fréquence est appliqué, pour supprimer les éléments apparaissant dans moins de 3 documents.

5.2 Performances par émotion

La performance est mesurée par la moyenne et l'écart-type du score F1, du rappel et de la précision évalués par validation croisée en 10 *folds* pour les émotions pour lesquelles plus de 60 exemples d'apprentissage sont disponibles. La figure 2 indique les valeurs obtenues pour les représentations basées sur les unigrammes, les bigrammes et la fusion de ces représentations.

Il faut noter que pour le système final qui implémente un processus de décision en deux étapes, les performances sur les étiquettes émotionnelles sont bornées par les performances du classifieur individuel sur l'étiquette *neutral*.

Résultats globaux Les résultats montrent une grande disparité dans les performances de prédiction, qui sont très élevées pour *thankfulness*,

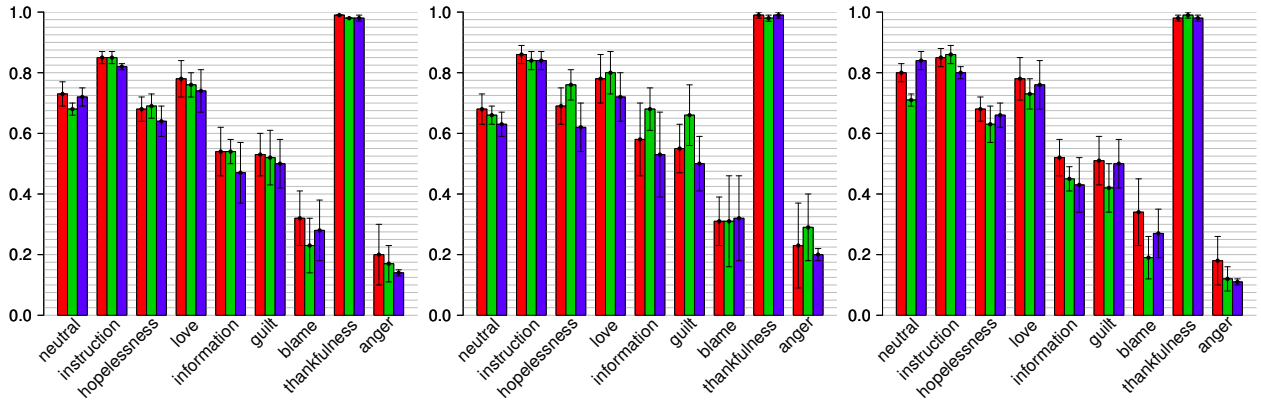


FIGURE 2 – Moyennes et écart-types du score de F1 (à gauche), de rappel (au centre) et de précision (à droite) pour les représentations basées sur les unigrammes (en vert au centre), les bigrammes (en bleu à gauche) et leur fusion (en rouge à droite), pour les émotions les plus fréquentes.

mais faibles pour *blame* ou *anger*.

Globalement une certaine corrélation des performances avec la fréquence est observée. Toutefois, certaines émotions se distinguent naturellement des autres : bien que *love* et *thankfulness* ne soient pas les plus représentées, elles sont plutôt bien prédites.

Etude selon les ordres des n-grammes On constate que les unigrammes conduisent globalement à des rappels supérieurs et à des précisions inférieures aux bigrammes : ils capturent un vocabulaire simple et générique, alors que les bigrammes capturent des constructions plus riches au détriment de la généricité des descripteurs. Le gain en précision ne compense pas le manque de couverture, et les bigrammes sont en général légèrement moins performants en termes de score F1 que les unigrammes.

Des expériences non détaillées ici montrent que cette tendance est également suivie lorsqu'on considère les trigrammes, qui ont globalement un rappel plus faible mais une précision plus élevée que les bigrammes. Ils n'apportent en général pas de gain en termes de score F1. On observe néanmoins des exceptions, en particulier pour *hopelessness* ou, de façon plus marquée encore, pour *sorrow*, pour lesquelles les trigrammes conduisent à des scores F1 particulièrement élevés, respectivement 0.80 ± 0.02 et 0.98 ± 0.01. Ceci signifie que ces émotions font appel à des constructions linguistiques plus complexes que les autres émotions.

Etude de la combinaison des unigrammes et des bigrammes Des expériences non détaillées ici ayant montré qu'une fusion exploitant les trigrammes n'apporte pas de gain significatif, nous considérons la combinaison par fusion anticipée des unigrammes et des bigrammes selon la procédure décrite dans la section 3.3.

On observe que la combinaison augmente légèrement les performances par rapport aux n-grammes considérés isolément, ou obtient des résultats similaires. Ainsi, pour *love*, la fusion améliore à la fois la précision et le rappel et donc le score F1. Pour *instruction*, les bigrammes obtiennent individuellement une précision moins bonne que les unigrammes, et la fusion n'améliore pas les performances obtenus par les unigrammes seuls.

5.3 Dictionnaires spécifiques obtenus

Cette section étudie l'effet de l'étape de spécialisation des dictionnaires, à la fois en termes de taille et de contenu.

Tailles des dictionnaires On constate une grande disparité des tailles de dictionnaires après spécialisation : le plus petit, associé à la classe *neutral*, contient 1904 entrées, le plus grand, associé à *abuse*, en contient 3511. Ceci est une conséquence des valeurs de $\epsilon(n, e)$ estimées d'après les performances des classifieurs appris sur les dictionnaires induits : nous avons constaté que pour les unigrammes ainsi que pour les émotions rares des valeurs de seuil très proches de 1 donnent de meilleurs résultats. Ces observations sont compatibles avec l'intuition que les descripteurs issus des unigrammes sont moins spécifiques que ceux issus d'ordres supérieurs. Pour les émotions rares, cet effet est dû

Émotions	Descripteurs
Thankfulness	thank / appreciate / than / nice / effort / kindness / under
	be swell / than you / you dear / appreciate it / too . / have be / for your
Instruction	cremate / call / please / sell / funeral / teach / notify
	to be / forget me / be good / to have / bury me / dispose of / care of
Love	love / wonderful / bless / watch / beloved / most / loving
	you . / do . / be wonderful / love you / god bless / your john / me on
Hopelessness	cancer / am / suffer / die / struggle / everybody / tired
	without you / go on / dear jane / can not / . my / be . / of all
Information	bldg / insurance / key / paper / owe / ticket / in
	of cincinnati / be pay / ohio . / in this / no . / and my / the key
Guilt	sorry / forgive / excuse / fail / hurt / could / burden
	have be / forgive me / please forgive / have do / understand . / not to / to help
Blame	sorry / thank / love / please / give / wish / go
	to be / cause you / of it / you . / you to / in the / to go

TABLE 2 – Classements des 7 meilleurs unigrammes et bigrammes pour les classifieurs de score F1 supérieur à 0.3, classés par valeur décroissante des performances des classifieurs.

au fait que de nombreuses entrées du dictionnaire apparaissent uniquement chez les contre-exemples de l'émotion considérée, et sont donc considérées comme discriminantes.

Il faut aussi noter que la spécialisation permet dans tous les cas de réduire significativement les dictionnaires obtenus avant spécialisation, qui ont pour taille $|\mathcal{D}^1| = 1206$ et $|\mathcal{D}^2| = 2968$, autorisant une taille maximale de dictionnaire $|\mathcal{D}| = 4174$.

Entrées discriminantes des dictionnaires Afin d'étudier les vocabulaires identifiés comme spécifiques des émotions sans considérer l'intégralité de leurs contenus, le tableau 2 donne les descripteurs les plus discriminants : pour les émotions associées à des classifieurs de score F1 supérieur à 0.3, le tableau indique les 7 unigrammes et les 7 bigrammes correspondant aux valeurs maximales du vecteur $\vec{\alpha}$ qui définit la frontière de décision (cf éq. 2). La présence de ces termes influence en effet la détection de l'émotion correspondante.

On observe alors par exemple naturellement que les descripteurs *love* et *thank* sont discriminants pour les émotions correspondantes.

On peut noter que les unigrammes discriminants peuvent être communs à plusieurs émotions, comme *please* associé à *instruction* et *blame*, ou *love* associé à *love* mais aussi à *blame*. Au contraire, les bigrammes discriminants apparaissent comme spécifiques à une émotion donnée, et non partagés.

Il est aussi intéressant d'observer que les unigrammes les plus influents n'apparaissent pas nécessairement dans les bigrammes les plus pertinents, et que, réciproquement, les bigrammes apparaissent comme plus discriminants que les unigrammes qui les constituent pris isolément.

On peut constater que dans une certaine mesure la position des mots dans la phrase importe : des bigrammes de la forme *terme* ., indiquant que le terme est le dernier de la phrase, apparaissent fréquemment comme bigrammes influents.

6 Conclusion et perspectives

Nous avons proposé une méthode de détection d'émotions dans les textes basée sur la combinaison de classifieurs dans une approche « un contre tous » spécialisés pour chacune des émotions et agrégés selon la confiance qu'ils associent à leur prédiction. Chaque classifieur utilise une représentation bas niveau des textes selon un dictionnaire spécialisé pour l'émotion à détecter. Les dictionnaires combinent, dans une approche de fusion anticipée, des unigrammes et des bigrammes. Les résultats obtenus sur un corpus réel illustrent la difficulté de la tâche : alors que l'approche proposée atteint de très bons résultats pour certaines émotions, que l'on peut donc supposer être associées à un vocabulaire spécifique en effet identifié par la méthode, d'autres restent difficiles à prédire. On constate que la combinaison des unigrammes et des bigrammes donne les meilleurs résultats, même si les différences ne sont pas toujours significatives par rapport aux n-grammes pris isolément.

Les perspectives incluent l'étude d'autres types de fusion, intermédiaire ou tardive. Une autre perspective est l'enrichissement de la représentation proposée par des connaissances sémantiques, et la comparaison des performances, en termes de score F1, mais aussi de temps de calcul et de coût global, incluant l'expertise nécessaire à la constitution des ressources externes.

Références

[1] A. Andreevskaia and S. Bergler. Mining wordnet for fuzzy sentiment : Sentiment tag extraction from wordnet glosses. In *FUZZ-IEEE*, 2006.

[2] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. Subrahmanian. Sentiment analysis : Adjectives and adverbs are better than adjective alone. In *Int. Conf. on Weblogs and Social Media*, 2007.

[3] H. Cui, V. Mittal, and M. Datar. Comparative experiments on sentiment classification for online product reviews. In *21st National Conf. on Artificial Intelligence, AAAI*, pages 1265–1270, 2006.

[4] S. Das and M. Chen. Yahoo ! for amazon : Sentiment extraction from small talk on the web. *Management Science*, 53 :1375–1388, 2007.

[5] F. Dzogang, M.-J. Lesot, M. Rifqi, and B. Bouchon-Meunier. Analysis of texts' emotional content in a multidimensional space. In *Int. Conf. on Kansei Engineering and Emotion Research, KEER*, pages 877–886, 2010.

[6] F. Dzogang, M.-J. Lesot, M. Rifqi, and B. Bouchon-Meunier. Expressions of graduality for sentiments analysis - a survey. In *FUZZ-IEEE*, 2010.

[7] P. Ekman. Basic emotions. In T. Dalgleish and M. Power, editors, *Handbook of Cognition and Emotion*, chapter 3, pages 45–60. John Wiley, 1999.

[8] A. Esuli and S. Fabrizio. Sentiwordnet : A publicly available lexical resource for opinion mining. In *LREC*, 2006.

[9] J. Fontaine, K. Scherer, E. Roesch, and P. Ellsworth. The world of emotions is not two-dimensional. *Psychological science*, 18 :1050–1057, 2007.

[10] M. Gilad. Experiments with mood classification in blog posts. In *1st Workshop on Stylistic Analysis of Text for Information Access (Style 2005)*, 2005.

[11] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. Technical report, Stanford, 2009.

[12] V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *18th Conf. on Computational linguistics*, volume 1, pages 299–305, 2000.

[13] S. Leleu. Un atlas sémantique de concepts d'émotions : normes et validation. Master's thesis, Université catholique de Louvain, 1987.

[14] Y.-Y. Mathieu. A computational semantic lexicon of french verbs of emotion. In *Computing Attitude and Affect in Text : Theory and Applications*. Springer, 2006.

[15] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Textual affect sensing for sociable and expressive online communication. In *Affective Computing and Intelligent Interaction*, volume 4738, pages 218–229. Springer, 2007.

[16] V. Ng, S. Dasgupta, and S. Niaz Arifin. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *COLING/ACL*, 2006.

[17] J. Pestian, P. Matykiewicz, M. Linn-Gust, B. South, O. Uzuner, J. Wiebe, K. Cohen, J. Hurdle, and C. Brew. Sentiment analysis of suicide notes : A shared task. *Biomedical Informatics Insights*, 5 :3–16, 2012.

[18] A. Piolat and R. Bannour. An example of text analysis software (emotaix-tropes) use : The influence of anxiety on expressive writing. *Current psychology letters*, 25, 2009.

[19] R. Plutchik. *The emotions*. University Press of America, 1990.

[20] J. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3) :273–294, 1977.

[21] C. Strapparava and R. Mihalcea. SemEval-2007 task 14 : Affective text. In *Proc. of the 4th Int. Workshop on the Semantic Evaluations (SemEval)*, 2007.

[22] C. Strapparava and A. Valitutti. Wordnet-affect : an affective extension of wordnet. In *4th Int. Conf. on Language Resources and Evaluation*, 2004.