HAL
open science

Une architecture cognitive et affective orientée interaction

Damien Pellier, Carole Adam, Wafa Johal, Humbert Fiorino, Sylvie Pesty

To cite this version:

Damien Pellier, Carole Adam, Wafa Johal, Humbert Fiorino, Sylvie Pesty. Une architecture cognitive et affective orientée interaction. Workshop on Affect, Compagnon Artificiel and Interaction, Jun 2016, Brest, France. hal-01365355

HAL Id: hal-01365355

https://hal.science/hal-01365355v1

Submitted on 13 Sep 2016

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

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

Une architecture cognitive et affective orientée interaction

Damien Pellier
UGA - LIG CNRS UMR 5217
Grenoble, France
damien.pellier@imag.fr

Carole Adam
UGA - LIG CNRS UMR 5217
Grenoble, France
carole.adam@imag.fr

Wafa Johal
UGA - LIG CNRS UMR 5217
Grenoble, France
wafa.johal@imag.fr

Humbert Fiorino
UGA - LIG CNRS UMR 5217
Grenoble, France
humbert.fiorino@imag.fr

Sylvie Pesty
UGA - LIG CNRS UMR 5217
Grenoble, France
sylvie.pesty@imag.fr

ABSTRACT

Les robots trouvent de nouvelles applications dans notre vie de tous les jours et interagissent de plus en plus étroitement avec leurs utilisateurs humains. Cependant, malgré une longue tradition de recherche, les architectures cognitives existantes restent souvent trop génériques et pas assez adaptées aux besoins spécifiques de l'Interaction sociale Humain-Robot, comme la gestion des émotions, du langage, des normes sociales, etc. Dans cet article, nous présentons CAIO, une architecture Cognitive et Affective Orientée Interaction. Elle permet aux robots de raisonner sur les états mentaux (y compris les émotions) et d'agir physiquement, émotionnellement et verbalement.

Keywords

Architecture cognitive, interaction sociale homme-robot, émotions

1. INTRODUCTION

Les robots sont de plus en plus présents dans notre vie quotidienne, dans des rôles d'assistants, de robots pédagogiques, de compagnons pour les enfants ou les personnes âgées, etc. rôles qui impliquent de nombreuses interactions avec l'utilisateur. Cette proximité signifie que les robots doivent partager non seulement le même espace physique mais aussi des buts et des croyances pour accomplir une tâche commune. Dans l'idéal, ils devraient pouvoir interagir en langage naturel mais aussi non-verbalement par gestes et expressions faciales. De nombreuses applications qui relèvent du domaine de l'Interaction Humain-Robot (HRI - Human-Robot Interaction) sont en cours de développement, mais l'architecture cognitive sous-jacente du robot n'est pas toujours explicite. En effet, le développement d'une architecture cognitive permettant à un robot de gérer la complexité des interactions avec des humains est un défi en soi. De nombreuses fonctionnalités sont nécessaires : gestion de la mémoire, raisonnement sur les états mentaux, gestion du dialogue, gestion des émotions et des aspects non-verbaux de l'interaction etc.

Malgré les différentes contributions dans le domaine des architectures cognitives et celles plus spécifiques pour la robotique [18, 21], la plupart des architectures sont génériques et peu d'entre elles peuvent vraiment gérer la complexité des interactions Humain-Robot. Dans cet article nous présentons notre Architecture Cognitive et Affective Orientée Interaction (CAIO) dont les contributions se situent sur les aspects suivants : autoriser le robot à raisonner sur les états mentaux (incluant les émotions) et à expliquer ses actions en termes d'états mentaux ; gérer à la fois les actions physiques et verbales ; et fournir deux boucles de réponse distinctes (réactive et délibérative).

Cet article est structuré comme suit. La section 2 présente quelques travaux sur les architectures cognitives et robotiques. La section 3 revient brièvement sur nos précédents travaux à la base de l'architecture CAIO que nous détaillons dans la suivante (section 4). La section 5 donne ensuite un court exemple sur un robot Nao, et nous positionnerons l'architecture CAIO en nous basant sur des critères définies par Langley et al. [20]. La Section 6 donne quelques perspectives.

2. ETAT DE L'ART

Les travaux dans le domaine des architectures cognitives sont une recherche de longue date et des états de l'art peuvent être trouvés dans différents articles, par exemple [34, 10]. Trois grandes familles d'architectures cognitives sont généralement distinguées : (1) les architectures dites bio-inspirées dont l'ambition est principalement la compréhension des processus de la cognition, (2) les architectures qualifiées de "purement" Intelligence Artificielle et (3) les architectures qui trouvent leur fondement dans des travaux en philosophie (la philosophie de l'esprit essentiellement). Nous donnons ici quelques éléments sur les principales architectures.

2.1 Les Architectures bio-inspirées

ACT-R (Adaptive Control of Thought-Rational) est une des architectures les plus connues pour modéliser les mécanismes de la cognition humaine [5] [6]. Elle repose principalement sur la séparation entre deux types de connaissance : la connaissance déclarative (chunks) et la connaissance procédurale (rules). L'architecture ACT-R est composée de 4 principaux modules, chacun ayant un équivalent chez l'humain (visual module - perception, goal module, declarative memory, and manual module - actuators), dont la coordination est assurée par un système central.

L'architecture CLARION (Connectionist Learning with Adaptive Rule Induction ON-line) [33] est également très connue ; c'est une architecture hybride qui combine des représentations connexionnistes et des représentations symboliques, et combine également des processus psychologiques implicites et explicites. L'ambition des recherches autour

de CLARION est principalement de comprendre les mécanismes de la cognition humaine (prise de décision, apprentissage raisonnement, etc.) tout en développant également des agents artificiels.

ASMO (Attentive Self-MOdifying) [22] a été développée plus récemment à partir d'une théorie biologique de l'attention, pour résoudre les compétitions entre buts éventuellement incompatibles du robot. Concrètement, le niveau d'attention détermine les priorités relatives des buts, les plus critiques étant traités comme des réflexes. ASMO a été implémentée dans un robot social en interaction avec des humains, et dans un robot Nao pour des compétitions de football.

2.2 Architectures pour la résolution de problèmes d'Intelligence Artificielle

SOAR (State, Operator And Result) [17] est une architecture purement "IA symbolique", qui met l'accent sur l'apprentissage et la résolution de problèmes. Elle dispose d'une mémoire à court-terme (ou mémoire de travail), et une mémoire à long-terme qui se décompose en mémoire procédurale, mémoire sémantique et mémoire épisodique. L'apprentissage par renforcement est déclenché quand les connaissances ne permettent pas de prendre une décision. SOAR a été étendue avec des émotions qui interviennent au niveau de l'apprentissage [15].

L'approche choisie pour l'architecture ICARUS [19] consiste à unifier les techniques réactives et délibératives de résolution de problèmes, ainsi que le raisonnement numérique et symbolique. Un but non satisfait de plus haute priorité reçoit l'attention : les capacités connues permettant de le réaliser sont amenées de la mémoire à long-terme vers la mémoire à court-terme, ou bien ce but est décomposé en sous-buts pour apprendre de nouvelles capacités.

2.3 Architectures basées sur des théories philosophiques

La philosophie de l'esprit s'intéresse à l'ensemble des problèmes qui se posent dans la vie mentale et en particulier pour l'action, la perception, le raisonnement, l'intentionnalité. En philosophie de l'esprit, il y a 3 états mentaux de base : croire et désirer et l'intention d'agir. La philosophie de l'action est une branche de la philosophie de l'esprit et s'intéresse plus particulièrement à l'*action*. La théorie de Bratman de l'action planifiée (Human Practical reasoning) [9] a ainsi servi aux travaux sur les logiques et les agents BDI (*Belief, Desire, Intention*) [27, 11]). Dans cette théorie, les croyances et les désirs sont la cause de l'intention d'action ; L'architecture PRS pour les agents rationnels (*e.g.* PRS, *Procedural Reasoning System* [38]) est une des architectures très connues. Comme nous l'expliquerons plus loin, l'architecture CAIO s'inscrit dans cette lignée, mais avec l'addition de nouveaux états mentaux (ou attitudes propositionnelles), en particulier les *émotions* qui sont essentielles pour un robot expressif social, à cause du rôle majeur qu'elles jouent dans l'interaction et le raisonnement.

Actuellement, plusieurs recherches se tournent vers la Théorie de l'esprit (Theory of Mind - ToM) [30, 16, 37] qui est également une branche de la philosophie de l'esprit. Cette théorie s'occupe de la compréhension d'autrui et de soi. IL s'agit de comprendre ce que les autres vont faire, d'être capable de le prédire afin d'anticiper les conséquences de ce comportement.

3. PRÉCÉDENTS TRAVAUX

3.1 La logique BIGRE pour représenter les états mentaux et les émotions

Guiraud et al. [14] ont proposé une logique modale dans la lignée des logiques BDI (Belief, Desire, Intention), pour représenter formellement les états mentaux qu'un agent artificiel (un robot par exemple) va pouvoir exprimer lors de son interaction avec un humain.

Cette logique, dite logique BIGRE (Belief, Ideal, Goal, Responsibility, Emotion), définit 5 états mentaux [1] :

- *Belief* (B) $Bel_i\varphi$: le robot i croit φ vrai ;
- *Ideal* (I) $Ideal_i\varphi$: idéalement pour le robot i, φ devrait être vrai (normes morales et sociales du robot [2]) ;
- *Goal* (G) $Goal_i\varphi$: le robot i a pour but que φ soit vrai ;
- *Responsibility* (R) $Resp_i\varphi$: le robot i est responsable de φ ;
- *Complex emotion* (E) (*e.g.* gratitude, admiration, reproche, etc.) qui résulte d'un raisonnement à propos de *Responsibility*.

Les émotions complexes dont il est question ici jouent un rôle central dans l'interaction dialogique et sont exprimées principalement par le langage. Ce sont des émotions qu'il faut distinguer des émotions de base (voir ci-dessous) que l'on peut représenter à partir des *Belief*, des *Goal* et des *Ideal* qui sont quant à elles associées généralement à des expressions faciales prototypiques.

Les opérateurs B, I, G et R (BIGR $\rightarrow$ E) ont permis de représenter 12 émotions : 4 émotions de base (*joie*, *tristesse*, *approbation*, *désapprobation*) et 8 émotions complexes (*regret*, *déception*, *honte*, *reproche*, *satisfaction morale*, *admiration*, *réjouissance* et *gratitude*, voir tableau 1).

La logique BIGRE a de plus permis de définir formellement des actes de conversation expressifs (regretter, reprocher, complimenter, etc.), actes qui expriment des émotions complexes (le regret, le reproche, l'approbation morale, etc.). (voir section suivante).

3.2 Le Langage de Conversation Multimodal

Pour qu'un agent artificiel, type personnage virtuel, puisse s'exprimer de manière crédible, Riviere *et al.* [29] ont défini un langage de conversation pour agent basé sur la Théorie des Actes de Discours de Searle [32]. Ce langage est appelé *Langage de Conversation Multimodal (LCM)* car sa particularité est de relier étroitement les aspects verbaux (les énoncés) aux aspects non-verbaux (les expressions faciales/gestes) afin d'augmenter l'expressivité de l'agent. Ce langage est dans la lignée des langages de communication pour agent tels que FIPA [26].

Ce langage LCM regroupe 38 *Actes de Conversation Multimodaux* (ACM) répartis en 4 classes :

- actes assertifs : *dire*, *affirmer*, *nier*, etc.
- actes directifs : *demander*, *suggérer*, *réclamer*, etc.
- actes engageants : *promettre*, *accepter*, *refuser*, etc.
- actes expressifs : *remercier*, *reprocher*, *féliciter*, etc.

Chaque acte ACM a été formalisé en logique BIGRE :

1. Le lecteur est invité à se référer à [14] pour plus de détails sur la sémantique et l'axiomatique de cette logique.
2. Par exemple, une obligation morale à aider une personne en danger ou une règle sociale à payer les taxes, etc.

$\wedge$	$Bel_{i\varphi}$	$Bel_i Resp_{i\varphi}$	$Bel_i Resp_{j\varphi}$
$Goal_{i\varphi}$	$Joy_{i\varphi}$	$Rejoicing_{i\varphi}$	$Gratitude_{i,j\varphi}$
$Goal_{i\neg\varphi}$	$Sadness_{i\varphi}$	$Regret_{i\varphi}$	$Desappointment_{i,j\varphi}$
$Ideal_{i\varphi}$	$Approval_{i\varphi}$	$MoralSatisfaction_{i\varphi}$	$Admiration_{i,j\varphi}$
$Ideal_{i\neg\varphi}$	$Disapproval_{i\varphi}$	$Guilt_{i\varphi}$	$Reproach_{i,j\varphi}$

Table 1: Formalisation des 4 émotions de base et des 8 émotions complexes.

Rejoice	$Exp_{a,h,H}(Rejoicing_a\varphi)$ $\equiv Exp_{a,h,H}(Goal_a\varphi \wedge Bel_a Resp_a\varphi)$
Preconditions	$Goal_a\varphi \wedge Bel_a Resp_a\varphi \stackrel{déf}{=} Rejoicing_a\varphi$ Agent a "feels" rejoicing ; it believes it is responsible for having achieved its goal φ.
Sending effects	$Bel_a Bel_h Rejoicing_a\varphi$ Agent a believes that human h believes that a is rejoicing about φ.
Reception effects	$Bel_a Goal_h\varphi \wedge Bel_a Bel_h Resp_h\varphi$ Agent a believes that human h expressed his rejoicing about φ. Therefore, a believes that h has achieved its goal φ and believes himself to be responsible for this.

Table 2: Exemple : l'acte ACM *Se réjouir - Rejoice* du point de vue de l'agent *a* en conversation avec un humain *h*.

— préconditions de l'acte, que l'agent doit satisfaire avant de pouvoir émettre cet acte, ceci pour garantir la *sincérité* de l'agent au sens de la théorie de Searle (sincerity conditions) ;
— effets en émission de l'acte, pour l'agent artificiel qui s'exprime ;
— effets en réception de l'acte, pour l'agent artificiel qui reçoit l'acte d'un interlocuteur.

Par exemple, le tableau 2 donne la formalisation de l'acte ACM *Se réjouir - Rejoice*.

Cette représentation formelle permet à l'agent de mettre à jour ses états mentaux, en particulier ses croyances mais aussi ses émotions, les états mentaux de son interlocuteur, et de savoir s'il lui est possible de s'exprimer.

3.3 Le moteur de raisonnement PLEIAD

PLEIAD (ProLog Emotional Intelligent Agent Designer) [1] est à l'origine un moteur de raisonnement pour des agents de type BDI, développé en SWI-Prolog. Il confère aux agents des capacités génériques de raisonnement. Il permet également l'implémentation de différents modèles d'émotions, comme le modèle OCC [2], ou de théories sur des émotions particulières comme la honte [3]. PLEIAD a été plus récemment complété avec des stratégies de *coping* [4] et des traits de personnalité.

Le moteur de raisonnement PLEIAD a été utilisé pour implémenter la logique BIGRE et constitue le coeur de l'architecture CAIO, principalement pour le module *Deliberation* et pour la partie cognitive du module *Evaluation des émotions* (voir section 4.2).

4. L'ARCHITECTURE CAIO

4.1 Vue globale de l'architecture

Une des spécificités de l'architecture CAIO (cf. Figure 1) est d'avoir 2 boucles de raisonnement : *une boucle délibérative* qui permet de raisonner sur les états mentaux BIGRE et de produire des plans d'actions au sens de la planification automatique et *une boucle sensori-motrice* plus rapide qui permet d'exprimer de manière réactive les émotions. Chaque boucle prend en entrée les perceptions multimodales en provenance de l'environnement.

Au cours de la boucle délibérative, la partie cognitive du module d'évaluation des émotions déduit les émotions complexes à partir des états mentaux. Le module de délibération déduit alors les *intentions* du robot à partir de ses états mentaux et choisit parmi celles-ci la plus appropriée. A partir de cette intention, le module de planification produit un plan, i.e., un ensemble ordonné d'actions communicatives ou physiques, pour atteindre l'intention retenue et choisir la prochaine action à exécuter. Finalement, le moteur de rendu multimodal exécute l'action précédemment choisie. Les modules de l'architecture fournissent un retour aux autres modules sur le résultat de leur traitement. Par exemple, le module de planification informe le module de délibération lorsque l'intention choisie n'est pas atteignable ou encore le moteur de rendu multimodal informe le planificateur du succès ou de l'échec de l'exécution d'une action.

Simultanément, au cours de la boucle sensori-motrice, le module d'évaluation des émotions évalue les données provenant du module de perception multimodale selon les critères de Scherer (critères SEC - Stimulus Evaluation Checks).Ce processus d'évaluation détermine le déclenchement des émotions et envoie au module de rendu multimodal les actions faciales et gestuelles à exécuter pour exprimer l'émotion.

4.2 Principaux modules de l'architecture

4.2.1 Le module de perception multimodale

Le rôle de ce module est de fusionner les informations en provenance des capteurs du robot et de les exprimer sous la forme d'états mentaux du locuteur. Pour l'instant, nous nous sommes focalisés essentiellement sur les aspects associés au langage. Néanmoins, nous envisageons à court terme d'intégrer d'autres données en entrée telles que les expressions faciales ou encore les signaux para-linguistiques.

Concrètement, ce module effectue dans un premier temps un traitement de la parole en utilisant l'interface de Google Speech. A partir de l'énoncé textuel, le module, dans un

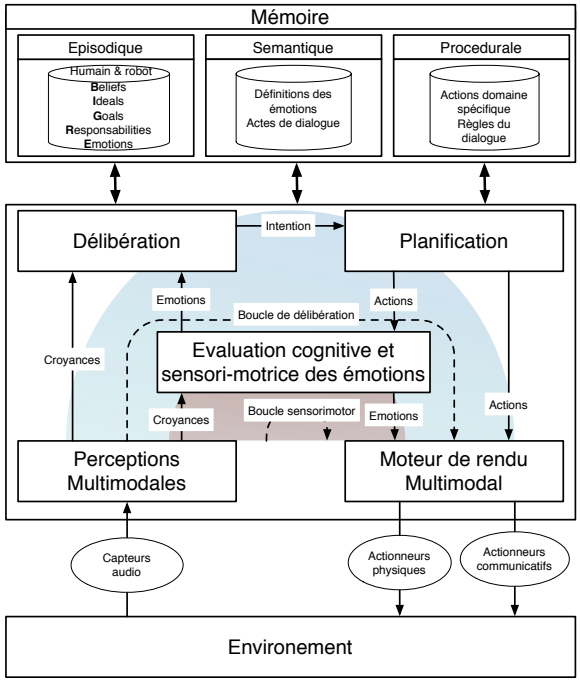

Figure 1: Schéma conceptuel de l'architecture CAIO.

second temps, extrait l'acte de conversation ACM. [3].

4.2.2 La mémoire

La mémoire au sein de l'architecture CAIO est divisée classiquement en 3 parties. La mémoire *épisodique* contient d'une part les états mentaux BIGRE propre au robot et d'autre part ceux de l'humain. La mémoire *sémantique* contient la définition des émotions et des actes de conversation. La mémoire *procédurale* gère la description des actions physiques que peut réaliser le robot et les règles du discours (cf. Tableau 3).

La mémoire est dynamiquement mise à jour en trois temps. Tout d'abord, les nouvelles croyances déduites à partir des perceptions du monde ou résultantes des interactions, i.e., les effets des actes ACM du locuteur et les effets produits par les propres actes ACM du robot, sont ajoutées à la mémoire. Puis les états mentaux sont mis à jour par inférence.

4.2.3 Le module d'évaluation des émotions

Le module d'évaluation est constitué de deux sous-modules : l'évaluation cognitive et l'évaluation sensori-motrice :

— l'évaluation cognitive prend en entrée les perceptions du robot ainsi que ses états mentaux et calcule l'émotion correspondante à partir de leur définition logique (cf. tableau 1). Par exemple, une émotion de gratitude est calculée, lorsque le robot a le but φ et croit que l'humain est responsable de ce but, i.e., lorsque le robot i à l'état mental $Goal_i\varphi \wedge Bel_i Resp_j \varphi$. Une intensité pour l'émotion est dérivée à partir des priorités du but ou de l'idéal précisé dans sa définition. Cette émotion est stockée dans la mémoire épisodique.

— L'évaluation sensori-motrice évalue tous les actes ACM reçus du locuteur ou ceux émis par le robot lui-même selon des critères dévaluation définis par Scherer (*SEC-Stimulus Evaluation Checks*) [31] que nous avons adaptés dans le cadre de nos travaux :
  — la *nouveauté* de l'acte de conversation (est-ce que l'acte était prévu dans le schéma général du dialogue ?) ;
  — le *plaisir intrinsèque* (en fonction du type d'acte, e.g., *Refuse vs. Accept*, et de son contenu propositionnel) ;
  — la *congruence* rapport aux buts
  — le *potentiel d'adaptation* (est-ce que le robot peut influencer les conséquences des actes ?) ;
  — la *compatibilité* avec les idéaux du robot et les normes sociales.

  Le résultat de cette évaluation sensori-motrice est propagé au module de rendu afin d'être traduit en expressions faciales et corporelles. (cf. tableau 4).

3. le traitement de la langue est un champ de recherche complexe ; nous n'abordons pas cet aspect dans ce travail, l'extraction des actes ACM à partir des énoncés est faite de façon ad-hoc.

Origine de l'obligation	**Action à réaliser**
S_1 Accept ou Promise A	S_1 achieves A
S_1 Request A	S_2 address Request : accept A **ou** reject A
S_1 Question fermée est-ce que P	S_2 Answer-If P
S_1 Question ouverte $P(x)$	S_2 Inform-ref x

Table 3: Exemples de règles du discours définies par Traum et Allen entre deux locuteurs (S1) et (S2) portant sur une proposition (P).

4.2.4 Le module de délibération

La délibération est le processus qui permet au robot de décider de sa prochaine *Intention* (ou but). Nous définissons trois types d'*Intention*.

Les *Intentions émotionnelles* sont les intentions qui vont se traduire en général par un acte ACM expressif (par exemple regretter, remercier, reprocher, se réjouir,...). Afin que le robot soit sincère, affectif et expressif, nous avons décidé que toute émotion "ressentie" par le robot devait être exprimée, ce qui aide à une meilleure régulation locale du dialogue. [8]. Notons que les intentions émotionnelles ont la priorité la plus haute.

Les *Intentions fondées sur les obligations* contribuent aussi à la régulation du dialogue et conduisent à des interactions humain-robot plus naturelles [7]. Les intentions de ce type sont adoptées à partir des *règles du discours* définies par Traum et Allen [35] pour représenter les normes sociales et guider le comportement du robot. Ces intentions jouent un rôle important dans la réaction du robot au niveau du dialogue (cf. tableau 3). Concrètement le robot adopte toujours l'intention d'atteindre ses obligations (déduites à partir des règles du discours) mais contrairement aux intentions émotionnelles, il leur affecte une priorité moyenne.

Finalement, l'*Intention globale* exprime la direction générale du dialogue et définit son type [36], e.g., délibération, persuasion, etc. Ce type d'intention est adoptée lorsque le robot s'engage à l'atteindre, soit explicitement, e.g., en l'ex-

plicitant par un acte ACM de type *Promettre* ou *Accepter*, ou en interne en raisonnant sur ses croyances. L'intention globale a toujours la priorité la plus faible.

4.2.5 Le module de planification

Le module de planification est en charge de trouver un moyen d'atteindre l'Intention décidée par le module de délibération. Ce module repose sur l'idée proposée à l'origine par [25] que l'ordonnancement des actes de dialogue peut être réalisé par la planification. En termes d'implémentation, le module repose sur la librairie PDDL4J [23] et les travaux de [24]. Il est important de noter que les plans produits peuvent contenir à la fois des actes de conversation ACM et des actions physiques, qui peuvent tous les deux être exécutés par le robot. Les pré-conditions et les effets des actions sont formalisés dans le langage PDDL (*Planning Domain Description Language*) utilisé dans la communauté planification rendant ainsi l'architecture CAIO compatible avec la plupart des planificateurs.

Dans le cas des Intentions émotionnelles et de celles fondées sur les obligations, le plan généré est souvent constitué d'un seul acte de conversation ACM. Par exemple, l'intention émotionnelle pour exprimer la gratitude peut être atteinte avec un acte de conversation ACM *Remercier* ou *Féliciter* en fonction de l'intensité de l'émotion.

Dans le cas des autres intentions, il est nécessaire de mettre en œuvre des actions spécifiques au domaine dont les préconditions et les effets sont décrits dans la mémoire procédurale du robot. Par exemple, pour réserver des billets de train, il est nécessaire de connaître l'heure et la date de départ ainsi que le lieu de départ et la destination. Le planificateur peut donc être amené dans ce cas à produire un plan contenant à la fois des actes de conversation ACM, e.g., demander les informations pertinentes à l'utilisateur, et des actions spécifiques au domaine, e.g., réserver des billets de train.

Si aucun plan ne peut être généré pour atteindre l'intention déterminée par le module de délibération, l'intention est abandonnée et le module de délibération est une nouvelle fois sollicité pour calculer une nouvelle intention.

4.2.6 Le moteur de rendu multimodal

Le moteur de rendu multimodal prend en entrée la première action orchestrée par le module de planification et l'émotion produite par le module d'évaluation des émotions. Puis, il agit sur les actionneurs du robot pour offrir un rendu physique de l'action intégrant dynamiquement l'expression faciale et les postures correspondant à l'émotion. Comme mentionné précédemment, la génération des postures s'appuie sur les valeurs des critères d'évaluation SEC [31, 13, 12] calculées par le sous-module sensori-moteur du module d'évaluation des émotions, par exemple le tableau 4 montre les postures produites pour exprimer le reproche.

5. EXEMPLE ET ÉVALUATION

5.1 Exemple pour le robot Nao

Une version préliminaire de l'architecture CAIO a été implémentée pour un personnage virtuel afin d'évaluer la crédibilité des comportements produits par l'architecture. Elle a notamment été testée dans le cadre d'un scénario de dispute entre un personnage virtuel compagnon et son utilisateur [28]. Cette première version était conçue pour un personnage virtuel et n'était pas modulable. Une nouvelle version de l'architecture CAIO a été développée avec ROS (Robot Operating System). ROS est utilisé et largement répandu dans la communauté robotique. Les modules sont principalement implémentés en python et interfacés pour certains d'entre eux : avec SWI-Prolog pour le module de délibération et avec Java pour le module de planification PDDL4J.

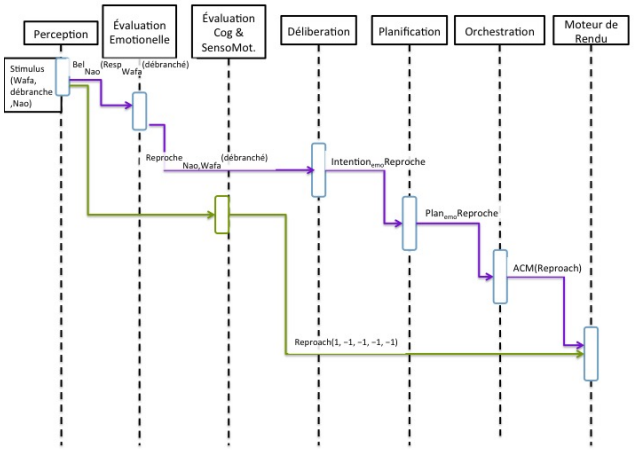

Figure 2: Diagramme de séquence UML

Nous présentons ci-dessous un court scénario illustrant l'encapsulation des différents modules de l'architecture CAIO dans des nœuds ROS (cf. le diagramme de séquence UML de la figure 2)

Ce scénario implique deux acteurs : un utilisateur, *Wafa*, et le *robot Nao*. Nao a une batterie en fin de vie qui nécessite d'être branchée en permanence. Dans sa mémoire épisodique, Nao possède l'Idéal d'être branché exprimé de la manière suivante : $Ideal_{Nao}(\neg unplugged)$. De son côté Wafa doit sortir ce soir et doit se sécher les cheveux, mais Nao est branché à la seule prise disponible près du miroir. Par conséquent, Wafa annonce à Nao : " Nao, je vais te débrancher, j'ai besoin de me sécher les cheveux ".

1. Le nœud *perception* reçoit le premier acte ACM initialisant le dialogue suivant : $Stimulus(unplugged, wafa, nao)$;
2. Cette acte est traduit en termes d'états mentaux comme suit : $Bel_{Nao}(unplugged)$ et $Bel_{Nao}(Resp_{Wafa}(unplugged))$;
3. le module d'évaluation des émotions (sous-module cognitif) déduit l'émotion $Reproach_{Nao,Wafa}(unplugged)$ à partir des états mentaux ;
4. Le module d'évaluation des émotions (sous-module sensori-moteur) évalue l'acte selon les 5 critères d'évaluation ;
5. Le nœud *déliberation* reçoit l'émotion de *Reproach* et publie une liste d'intentions (dans cet exemple, l'intention prioritaire est l'intention émotionnelle d'exprimer le reproche) ;
6. Le nœud *planification* reçoit cette liste d'intentions et choisit la plus prioritaire, et publie une liste de plans (ici, un plan unique contenant l'acte ACM *Reproach*).
7. Le plan est reçu par le *scheduler* qui choisit la première action (ici, l'acte ACM *Reproach*) ;

Nouveauté	⇒	Plaisir intrinsèque	⇒	Effet sur But/Besoin	⇒	Potentiel d'adaptation	⇒	Norme Compatibilité
Nouveau	⇒	Déplaisant	⇒	Attentes insatisfaites	⇒	Peu de contrôle	⇒	Norme violée

Table 4: Séquence d'évaluation sensori-motrice qui exprime le reproche.

8. Finalement le *moteur de rendu* "joue" l'acte de conversation *Reproach* de façon multimodale (verbale et non-verbale) sur le robot Nao.

5.2 Évaluation conceptuelle

Nous proposons ici une courte évaluation conceptuelle de l'architecture CAIO en suivant les critères d'évaluation des architectures cognitives proposés par Langley *et al.* [20] :

— **Généralisable & Polyvalent** (facilement adaptable à différents domaines et indépendant de la tâche) : la mémoire à long-terme de CAIO contient une base de connaissances génériques et statiques, et un nouveau domaine de connaissance peut être chargé dynamiquement afin de s'adapter au contexte.

— **Rationnel & Optimal** (prise de décision logique, sélection de comportement optimal pour un contexte donné) : le raisonnement produit par l'architecture CAIO s'appuie sur 5 états mentaux et est dans la lignée des raisonnements des agents rationnels BDI, assurant que le plan choisi par le robot déduit de son intention est optimal dans le contexte actuel de sa base de connaissances.

— **Efficace & Extensible** (réponse et action en temps réel) : l'implémentation ROS de CAIO permet une interaction humain-robot en temps réel.

— **Réactive & Pérenne** (réponse à des stimuli tout en gardant des buts à long terme) : CAIO gère à la fois des intentions réactives à court-terme (intentions émotionnelles et de discours) et des intentions pérennes à long-terme (intentions globales).

— **Améliorable** (comportements améliorables dans le temps, soit par programmation soit par apprentissage automatique) : pour l'instant, CAIO est améliorable seulement par programmation (chaque module peut être remplacé ou amélioré séparément), mais l'apprentissage automatique de nouvelles règles sera possible dans des extensions futures.

— **Autonome & Capable d'opérations prolongées** (choix de ses propres buts) : CAIO permet à un robot de choisir ses propres intentions sur la base de ses perceptions, de ses états mentaux et de sa base de connaissances.

Cette rapide évaluation conceptuelle montre que l'architecture cognitive et affective CAIO possède la plupart des propriétés identifiées comme indispensables à toutes les architectures cognitives. L'architecture CAIO nécessite toutefois d'être expérimentée dans différents scénarii d'interaction Humain-Robot. Elle nécessite d'être améliorée au niveau de la représentation des connaissances (utilisation d'un langage de représentation des connaissances comme RDF ou OWL) et en se dotant de capacité d'apprentissage.

6. CONCLUSION

Dans cet article, nous avons décrit l'architecture Cognitive et Affective Orientée Interaction CAIO pour robot social. Cette architecture est en cours de validation plus poussée par des interactions en temps réel avec des enfants afin de vérifier que le robot exprime clairement ses intentions, et est perçu comme sincère. Nous étudierons les limites du système sur différents types de tâches. Des travaux sur la représentation des connaissances sont nécessaires afin d'être compatible avec les ontolgies web. Le module de perception multimodale doit également être complété pour que l'extraction de l'acte de conversation ACM de l'énoncé de l'utilisateur soit améliorée, mais aussi pour y intégrer d'autres modalités (reconnaissance faciale des émotions, reconnaissance de gestes par exemple). Un module d'apprentissage serait aussi une extension intéressante pour s'assurer que le robot s'améliore durant l'interaction, et apprenne progressivement à connaître son utilisateur, à s'y adapter et à l'engager dans l'interaction.

7. REFERENCES

[1] C. Adam. *The emotions : from psychological theories to logical formalisation and implementation in a BDI agent*. PhD thesis, INP Toulouse, July 2007.

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

[3] C. Adam and D. Longin. Shame, when reasoning and emotions are linked. In *EUMAS'13*, 2013.

[4] C. Adam and E. Lorini. A BDI emotional reasoning engine for an artificial companion. In *A-Health workshop @ PAAMS*, 2014.

[5] J. R. Anderson. Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, 29(3), 2005.

[6] J. R. Anderson and G. H. Bower. *Human associative memory*. Winston and Sons, Washington, DC, 1973.

[7] M. J. Baker. A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence and Education*, 5(2) :199–254, 1994.

[8] J. Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7) :122–125, 1994.

[9] M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Nov. 1987.

[10] H.-Q. Chong, A.-H. Tan, and G.-W. Ng. Integrated cognitive architectures : a survey. *Artificial Intelligence Review*, 28(2) :103–130, Feb. 2009.

[11] P. R. Cohen and H. J. Levesque. Persistence, intention and commitment. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 33–69. MIT Press, 1990.

[12] M. Coulson. Expressing emotion through body movement : A component process approach. *Animating Expressive Characters for Social Interaction : Advances in Consciousness Research*, 74, 2008.

[13] M. S. Erden. Emotional Postures for the Humanoid-Robot Nao. *International Journal of Social Robotics*, 5(4) :441–456, July 2013.

[14] J. Guiraud, D. Longin, E. Lorini, S. Pesty, and J. Rivière. The face of emotions : a logical formalization of expressive acts. In *Int. Conf. on Autonomous Agent and Multiagent Systems*, pages 1031–1038, 2011.

[15] E. Hogewoning, J. Broekens, J. Eggermont, and E. Bovenkamp. Strategies for affect-controlled action-selection in soar-rl. In *International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC)*, volume 4528 of *LNCS*, pages 501–510. Springer, 2007.

[16] H. Jones and N. Sabouret. Intentions stratégiques basées sur un modèle affectif et une théorie de l'esprit. In *WACAI*, 2014.

[17] J. E. Laird. *The Soar cognitive architecture*. MIT Press, 2012.

[18] S. Lallée, U. Pattacini, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, K. Hamann, J. Steinwender, E. Sisbot, G. Metta, J. Guitton, R. Alami, M. Warnier, T. Pipe, F. Warneken, and P. Dominey. Towards a platform-independent cooperative human robot interaction system : III an architecture for learning and executing actions and shared plans. *IEEE T. Autonomous Mental Development*, 4(3) :239–253, 2012.

[19] P. Langley, D. Choi, and S. Rogers. Interleaving learning, problem solving, and execution in the icarus architecture. Technical report, Computational Learning Laboratory, CSLI, Stanford University, 2005.

[20] P. Langley, J. Laird, and S. Rogers. Cognitive architectures : Research issues and challenges. *Cognitive Systems Research*, 10(2) :141–160, 2009.

[21] S. Lemaignan, R. R. Espinoza, R. Alami, and M. Beetz. What are you talking about ? grounding dialogue in a perspective-aware robotic architecture. In *IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, pages 107–112, 2011.

[22] R. Novianto, B. Johnston, and M.-A. Williams. Attention in asmo cognitive architecture. In *1st International Conference on Biologically Inspired Cognitive Architectures (BICA)*, 2010.

[23] D. Pellier. PDDL4J : Planning Domain Description Language Library for Java, 2015.

[24] D. Pellier, H. Fiorino, and M. Métivier. Planning when goals change : A moving target search approach. In *Int. Conf. on Advances in Practical Applications of Heterogeneous Multi-Agent Systems*, 2014.

[25] C. R. Perrault and J. F. Allen. A plan-based analysis of indirect speech acts. *Comput. Linguist.*, 6(3-4) :167–182, July 1980.

[26] S. Poslad. Specifying protocols for multi-agent systems interaction. *ACM Trans. Auton. Adapt. Syst.*, 2(4), Nov. 2007.

[27] A. S. Rao and M. P. Georgeff. Modeling Rational Agents within a BDI-Architecture. In *Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484, 1991.

[28] J. Rivière. *Interaction affective et expressive Compagnon artificiel - humain*. PhD thesis, de Grenoble, 2012.

[29] J. Riviere, C. Adam, S. Pesty, C. Pelachaud, N. Guiraud, D. Longin, and E. Lorini. Expressive multimodal conversational acts for saiba agents. In *IVA*, pages 316–323. Springer, 2011.

[30] N. Sabouret and M. Belkaid. Un modèle logique de théorie de l'esprit pour un agent virtuel dans le contexte de simulation d'entretien d'embauche. In *WACAI*, 2014.

[31] K. R. Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion : Theory, methods, research*, pages 92–120, 2001.

[32] J. Searle and D. Vanderveken. *Foundations of illocutionary logic*. Cambridge University Press, 1985.

[33] R. Sun. *Duality of the Mind : A Bottom-up Approach Toward Cognition*. Lawrence Erlbaum Associates, Mahwah, NJ, 2002.

[34] K. Thórisson and H. Helgasson. Cognitive Architectures and Autonomy : A Comparative Review. *Journal of Artificial General Intelligence*, 3(2) :1–30, Jan. 2012.

[35] D. R. Traum and J. 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.

[36] D. Walton and E. Krabbe. *Commitment in Dialogue : Basic concept of interpersonal reasoning*. State University of New York Press, 1995.

[37] M. Warnier. *Gestion des croyances de l'homme et du robot et architecture pour la planification et le contrôle de la tâche collaborative homme-robot*. PhD thesis, Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse), Décembre 2012.

[38] M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, 2009.