



HAL
open science

Cognition, Affects et Interaction

Gérard Bailly, Sylvie Pesty

► **To cite this version:**

Gérard Bailly, Sylvie Pesty (Dir.). Cognition, Affects et Interaction. 2017, Cognition, Affects et Interaction. <hal-01483705>

HAL Id: hal-01483705

<https://hal.science/hal-01483705v1>

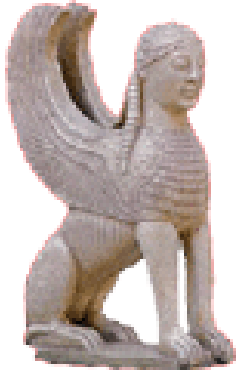
Submitted on 6 Mar 2017

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.



Copyright - All rights reserved



Master Recherche **IC2A**
**Ingénierie de la Cognition,
de la Création
et des Apprentissages**

[Master Recherche Sciences Cognitives]

**Travaux d'études et de recherche
2016-2017**

pour le cours :
Cognition, Affects et Interaction

Gérard Bailly & Sylvie Pesty

Université de Grenoble-Alpes, France

Avant-propos

Cet ouvrage rassemble les travaux d'études et de recherche effectués dans le cadre du cours « Cognition, Affects et Interaction » que nous avons animé au 1er semestre 2016-2017. Cette troisième édition de cours poursuit le principe inauguré en 2014 : aux cours magistraux donnés sur la thématique "Cognition, Interaction & Affects" qui donnent les outils méthodologiques des composantes de l'interaction socio-communicative, nous avons couplé une introduction à la robotique sociale et un apprentissage actif par travail de recherche en binômes. Le principe de ces travaux d'études et de recherche est d'effectuer une recherche bibliographique et de rédiger un article de synthèse sur un aspect de l'interaction homme-robot. Ces travaux sont revus par nos soins tant sur le fond que sur la forme, et finalement présentés devant l'ensemble des étudiants et notés. Si plusieurs sujets ont été proposés aux étudiants en début d'année, certains binômes ont choisi d'aborder l'interaction avec un angle original qui reflète souvent les trajectoires de formation variés des étudiants en sciences cognitives (ingénierie, sociologie, psychologie, etc). Le résultat dépasse nos espérances : le lecteur trouvera une compilation d'articles argumentés de manière solide, rédigés de manière claire et présentés avec soin. Ces premières « publications » reflètent les capacités singulières de réflexion de cette promotion qui explore de nouvelles problématiques de l'interaction homme-robot. Nous espérons que cette série d'ouvrages disponibles sous HAL puisse servir de point d'entrée à des étudiants ou chercheurs intéressés à explorer ce champ de recherches pluri-disciplinaire.

Gérard Bailly

Directeur de Recherches CNRS
GIPSA-Lab, Univ. Grenoble-Alpes/CNRS
email : gerard.bailly@gipsa-lab.fr

Sylvie Pesty

Professeur des Universités UPMF
LIG, Univ. Grenoble-Alpes/CNRS
email : sylvie.pesty@imag.fr

Précédents ouvrages

Cognition, Affects et Interaction *Gérard Bailly & Sylvie Pesty* 2015 [cel-01110281](#) (pdf)
Cognition, Affects et Interaction *Gérard Bailly & Sylvie Pesty* 2016 [cel-01258860](#) (pdf)

Table des matières

Articles en français	1
Le dialogue Homme-Machine : un état de l'art <i>Raphael Arcis & Nelly Sagardiluz</i>	1
La cobotique <i>Julien Mas & Romane Cecchi</i>	11
La robotique à l'épreuve des émotions. Du robot jouet au clone robotisé <i>Jane Bourgninaud & Julien Miranda</i>	16
Les comportements non-verbaux dans les interactions homme-machine <i>Roman Piekut & Bertrand Métral</i>	24
Cadres d'apprentissage pour les interactions homme-robot <i>Nicolas Roudier & Ali Saghiran</i>	29
Architectures cognitives pour l'interaction humain-robot <i>Gabriela Vives & Nolwenn Martin</i>	33
Articles en anglais	41
Socially-assistive robotics and elderly people <i>Edwige Laflaquière, Marie Palma & Annalisa Paroni</i>	41
Integration of Episodic Memory in Companion-Robots to Enhance Social Interactions <i>Fanny Broqua & Argitxu Caldichoury</i>	47
Learning-by-Teaching : a Paradigm for the Use of Robots in Education <i>Erika Godde & Martina Jurcova</i>	53
Influence of Human Personality Traits on Trust in Human-Robot Interaction <i>Hugo Pan- tecouteau & Brice Passera</i>	62
Index	70

Le dialogue Homme-Machine : un état de l'art

Raphaël Arcis

Etudiant M2 Sciences Cognitives

Grenoble INP Phelma

Grenoble, France

raphael.arcis@phelma.grenoble-inp.fr

Nelly Sagardiluz

Etudiant M2 Sciences Cognitives

Grenoble INP Phelma

Grenoble, France

nelly.sagardiluz@phelma.grenoble-inp.fr

Abstract— Que ce soit par la recherche ou la science-fiction, le dialogue homme-machine semble satisfaire tous les fantasmes d'un futur dans lequel les robots seront capables de converser au même titre que tout être humain. Soutenus par plus d'un demi-siècle de travaux dans le domaine, de tels systèmes commencent à émerger et faire leurs preuves. Néanmoins, ils demeurent relativement limités : leurs discours sont simulés jusqu'à une certaine mesure et ne tiennent pas encore exhaustivement compte de l'étendue des exigences en matière de communication. Ce papier traite des diverses composantes du dialogue, de l'implémentation de certaines d'entre elles et des technologies auxquelles elles ont conduit ainsi que de leurs limites.

Keywords— *dialogue homme-machine; interaction située; agent conversationnel; dialogue grounding; reference tracking; coverbal; discours incrémental; turn taking.*

I. INTRODUCTION

I-Robot, Star Wars, Wall-E : la science-fiction propose de nos jours de nombreux divertissements dans lesquels elle traite de l'intelligence artificielle (IA). Plus encore, elle cherche à anticiper la capacité qu'auraient les machines à dialoguer avec les humains sur des sujets complexes, divers et précis : négociation, préférence, émotions, prévision à court ou long terme, introspection, etc... Mais où en est-on réellement sur ce sujet-là ?

La réalité est loin d'avoir rattrapé la fiction. Les systèmes actuels ont déjà beaucoup de mal à répondre à des questions, même simples, sans parler de questions plus élaborées, convoquant concepts ou émotions. On arrive à pallier à ces difficultés lorsque l'on adapte le programme à une situation donnée, mais il reste tout de même possible pour l'utilisateur de mettre le système en défaut.

En effet, le processus de dialogue entre humains fait appel à des fonctions de haut niveau. Celui-ci commence par la réception de l'énoncé prononcé par l'interlocuteur, sa compréhension, puis l'élaboration et la production d'une réponse adéquate. De nombreuses étapes jalonnent son fonctionnement et toutes sont capitales à la bonne tenue de la discussion [1]. Listées dans leur ordre d'apparition et schématisées par la figure 1, elles sont les suivantes.

- La reconnaissance automatique de la parole : identification des sons prononcés par l'interlocuteur.
- L'analyse lexicale : identification des mots correspondants

- L'analyse prosodique : identification de l'intonation, du rythme d'énonciation et d'éventuels accents toniques s'appliquant à certains mots pour les rendre saillants
- L'analyse morphosyntaxique : identification des catégories de mots et des fonctions grammaticales
- L'analyse sémantique : détermination du sens général de la phrase à partir du sens des mots
- La résolution des références : identification des personnes et des objets mentionnés
- L'identification de l'acte de dialogue : l'énoncé est-il une question ? un ordre ?
- L'identification des émotions : sous quelle influence émotionnelle l'énoncé est-il prononcé ?
- L'interprétation en contexte : analyse de l'énoncé et de l'intention de communication compte tenu des résultats de toutes les analyses qui précèdent, analyse de l'implicite par exemple
- Le raisonnement logique et détermination du « quoi dire » : face à cet énoncé et aux connaissances de la machine, comment celle-ci peut-elle réagir ? en disant quoi ?
- La détermination du « comment le dire » : quelle construction de phrase, avec quels mots, va-t-elle s'exprimer ?
- L'expression : la réalisation physique du discours.

La question du langage chez la machine semble inhérente à la question de l'intelligence. Dans l'opinion générale, il est établi que l'intelligence se développe avec le langage tandis qu'une bonne maîtrise de la langue dénote d'un signe d'intelligence. Ce parallèle est aussi fait vis-à-vis de la machine, notamment lorsque Allan Turing propose un test d'intelligence fondé sur le langage [2]. Celui-ci consiste à poser des questions à son interlocuteur pour déterminer s'il s'avère être un humain ou une machine. Cependant, la proposition de Turing est critiquable sur le fait de savoir si les réponses données le sont parce que l'interlocuteur a compris de quoi il parlait, ou s'il s'agit simplement d'une émulation de réponses adaptées à des questions précises. C'est la différence majeure entre ce qu'on appelle l'IA faible (seulement capable de simulation d'intelligence) et l'IA forte (possédant une intelligence). A ce jour, aucune IA forte n'a été conçue, mais

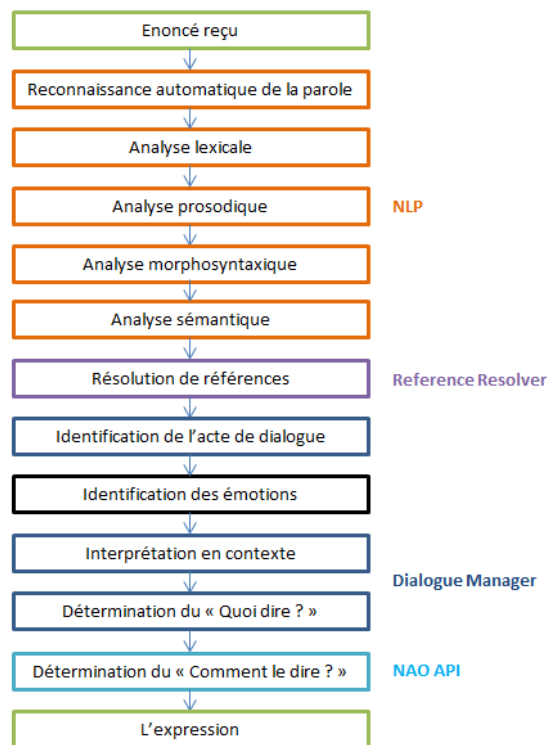


Fig. 1. Schéma général de la gestion de dialogue avec modules associés pour le robot NAO.

de nombreuses IA faibles sont élaborées pour la recherche fondamentale ou des buts commerciaux. Les programmes et robots conçus par l'homme ne proposent donc qu'un simulacre de conversation puisque nous ne sommes capables de réaliser que des adaptations à des contextes donnés.

Le document présent porte donc sur les différents défis entourant la conception d'un dialogue performant à tout point de vue et présentera des technologies phares dans le contexte du dialogue homme-machine.

II. LE DIALOGUE

La perception et la production de parole nécessitent une fine coordination entre divers traitements, bouclés sur la compréhension du message verbal. Ceux-ci visent sommairement à capter le signal acoustique, comprendre sa teneur linguistique, y formuler une réponse adaptée et enfin la communiquer à son partenaire. Chez l'humain, cette logique suit le modèle psycholinguistique spécifié par Levelt [3] : de la conceptualisation d'un message préverbal (l'intention communicative se mue en forme linguistique) à la formulation d'un discours interne planifié phonétiquement, par suite prononcé en un discours externe. Les modalités propres à la robotique sont évidemment réduites, en ce sens que les machines ne conceptualisent pas d'abstractions et se limitent à délivrer une synthèse vocale ou écrite sans articulation.

Historiquement, les chercheurs travaillent dessus depuis les années 50 et pensaient aboutir très vite, or il a fallu attendre les

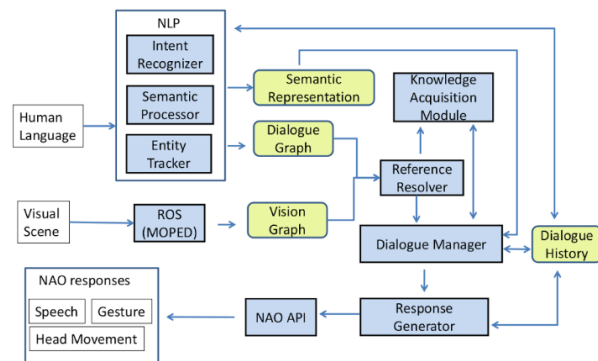


Fig. 2. Schéma d'architecture de communication du système communicant.

70 pour obtenir de véritables résultats, et, même de nos jours, la marge d'amélioration est considérable. La véritable progression a été liée à l'augmentation effrénée de la puissance de calcul en informatique, et par conséquent, à l'augmentation des bases de données disponibles pour ces programmes.

Pour détailler le fonctionnement théorique de la gestion de dialogue par une machine, nous allons nous servir du robot NAO dont le schéma d'architecture de communication est présenté figure 2. L'analogie est évidente avec le schéma figure 1 présenté en introduction, puisque les modules qui y sont détaillés traitent précisément des différentes étapes du dialogue. Seule la gestion des émotions n'est pas abordée. Notre cadre d'étude se borne également au stade central de régulation d'une interaction dyadique, c'est-à-dire n'impliquant qu'un seul humain face à un seul programme ou robot, bien que certains résultats soient généralisables.

Le discours de la machine se résume en trois actes : comprendre, formuler puis générer.

A. Le dialogue simple

1) La compréhension littérale

Tout d'abord, avant même que le système ne puisse faire la moindre analyse, il faut que celui-ci comprenne ce qui lui est énoncé. On commence donc par décoder le signal acoustique grâce au système de reconnaissance automatique de parole, dit ASR. Pour une parole continue basée sur un large vocabulaire, les unités de décodage les plus adaptées sont le phonème ou la syllabe. Après entraînement sur des échantillons cibles, ce système est prêt à retranscrire séquentiellement l'entrée acoustique au format texte. De bonnes performances sont acquises par l'application d'outils de formalisation probabilistes tels que les modèles cachés de Markov [4].

Ensuite, les modules TAL (pour Traitement Automatique des Langues), encore appelé NLP (pour Natural Language Processing), visent à construire une représentation sémantique de tout propos tenu au cours du dialogue. Le processeur sémantique, principal d'entre eux, réalise les analyses morphologique, syntaxique et sémantique [4]. Elles consistent respectivement en l'identification des termes et l'assignation de

ceux-ci à des catégories lexicales ; la transformation de la séquence de mots en structure hiérarchique de mots et l'association de l'entité linguistique à l'élément physique auquel elle correspond. Enfin, on applique une analyse de grammaire : pour une phrase donnée, on essaie de retourner la meilleure structure grammaticale. Suivant la qualité de l'énoncé traité, diverses méthodes peuvent entrer en jeu pour faciliter cette étape. Une grammaire catégorielle combinatoire (CCG) peut se montrer assez efficace vis-à-vis de discours fragmentés, informels et pourvus d'un lexique apparenté au domaine, ce qui est souvent le cas. Quoi qu'il en soit, une grande flexibilité est tout de même attendue des concepteurs pour éviter certaines impasses communicatives. Par exemple lors d'un simple manque de vocabulaire, on pallie au problème à l'aide d'un système de Large Vocabulary Speech Recognition de la parole non contrainte ou d'apprentissage sur le terrain après mise en ligne [5]. Concrètement, les fonctions sus-décrites ne suivent pas une chronologie stricte et peuvent se combiner dans une certaine mesure pour remédier aux erreurs de transcription de l'ASR et aux multiples ambiguïtés d'ordre phonologique, lexical ou syntaxique.

Pour approfondir la qualité de la compréhension, on peut procéder à une analyse pragmatique d'identification de l'intention communicative de l'humain. Cette fonction réalisée par l'Intent Recognizer s'opère pour clarifier le sens en contexte de la phrase et apporter une réponse pertinente. Un exemple classique est celui de la question "As-tu l'heure ?", dont on espère obtenir autre chose qu'un simple oui/non.

2) La résolution de référence

Ici, la problématique est de taille : il s'agit de faire en sorte que notre machine soit capable de reconnaître un élément quand on le désigne, même s'il est désigné de différentes façons. Par exemple "Felix" ou "le chat" ou "il" peuvent être trois appellations différentes du même objet. En fonction de leur utilisation, certaines machines n'utilisent pas ce concept. Cependant son utilisation n'est pas à négocier si le concepteur cherche à établir une gestion de dialogue de qualité. Pour ce faire, différents modules nous aident à atteindre notre objectif.

La vision du monde du robot est en partie gouvernée par un graphe : le graphe de dialogue. Il correspond à la représentation de l'environnement qui découle de la conversation entretenue avec l'utilisateur humain. Il est composé de nœuds correspondant aux différents éléments nommés au fur et à mesure de la conversation, munis de leurs caractéristiques. L'entity tracker vise à déterminer le caractère nouveau ou préétabli de chaque entité linguistique référée dans le discours. Pour cela, il analyse des données linguistiques, de récence et de compatibilité sémantique entre entités. Si la coréférence est avérée, c'est-à-dire qu'un même objet du monde physique est désigné plusieurs fois, alors il y a réorganisation du graphe de dialogue, et les entités coréférentes fusionnent en un nœud unique, dont les attributs capturent les propriétés sémantiques de l'entité linguistique en jeu. La mémorisation des informations issues des modules NLP et de l'analyse sémantique est donc nécessaire à son fonctionnement ; elle est assurée par l'historique de dialogue et le module d'acquisition de connaissances. Ces derniers ont en effet pour rôle de conserver les données sémantiques des différents constituants du monde que l'on décrit au robot et qu'il perçoit.

Enfin vient la résolution de références. Celle-ci nécessite un graphe de vision à comparer au graphe de dialogue par l'intermédiaire d'algorithmes pertinents. Le graphe de vision capture la représentation interne que se fait le robot de l'espace partagé via un système de reconnaissance d'objets en temps-réel. Chaque nœud qui le compose correspond à un objet perçu du voisinage et est détaillé en attributs recensant ses propriétés visuelles. Le graphe de dialogue contient pour sa part la description fournie par l'humain. En cherchant à fonder des correspondances entre les deux, la machine infère quels référents sont mentionnés. Mais pour s'assurer d'enrichir ses représentations internes d'une connaissance correcte, c'est-à-dire approuvée par son interlocuteur donc partie d'un common ground (cf. section correspondante pour plus de détails sur ce sujet), le robot utilise d'abord le gestionnaire de dialogue pour exiger confirmation.

3) La formulation

La sélection du "quoi dire" se fait au sein du gestionnaire de dialogue. Une stratégie de dialogue y précise quelle action effectuer ou quelle phrase formuler en réponse à un état du dialogue. Il ne s'agit finalement que de ne choisir la meilleure réponse adaptée aux stimuli d'entrée, ce qui peut être implémenté sous la forme d'un algorithme simple.

4) La génération

Enfin, l'action sélectionnée ouvre la voie à la production d'une réponse dans une ou plusieurs modalités, dont les voies orale, écrite, ou encore via des manifestations coverbales. Ces dernières peuvent prendre la forme de backchannels ("uh-uh", "ok"), de hochements de la tête ou d'expressions faciales. Le dialogue manager associe au moment opportun des combinaisons prédéfinies, à activer par consultation de cette stratégie comportementale.

Pour ancrer l'action du robot dans son environnement et appuyer sa compréhension de la situation, il est classique pour ses modules décisionnels d'initier la réalisation de gestes déictiques (pointage). Cela permet à son interlocuteur de mieux identifier l'objet ciblé (soit inféré par le robot des stimuli reçus, soit précisément décrit par lui-même), et procure une expérience plus interactive et naturelle [6]. Pour un bon fonctionnement, les sous-systèmes doivent se coordonner temporellement et invoquer cette modalité en complémentarité, non en redondance (par exemple, désambiguïsation dans le bruit), pour parfois dissoudre une certaine confusion. Une autre modalité à prendre en considération est le regard. Sa direction et les fixations opérées lors d'une saccade vers l'objet désigné pourraient théoriquement permettre de meilleurs résultats. Cependant, en tant que feedback intermédiaire en interaction située, cette donnée semble malheureusement comporter par nature trop de bruit pour être exploitée avec fiabilité [6].

5) Exemple de dialogue avec un robot NAO

Dans les travaux de Chai et al. [7], les auteurs ont effectué différents tests avec le robot NAO incluant des tâches de dénomination. On observe figure 3 un court échantillon de dialogue. En sortie de l'ASR et des modules NLP, les phases de dialogue sont stockées dans l'historique de dialogue, de sorte que les informations mutuellement transmises sont tour à tour paramétrées sous (H1, R1), puis (H2, R2), etc, où H désigne les propositions de l'humain et R celles du robot,

H1:	The green cup is called Bill.
R1:	Ok, the green cup is Bill. (point to the inferred object)
H2:	Do you see a car?
R2:	Yes, I see a car.
H3:	Do you see a blue can behind it?

H1: Human Act: DescribeObjectName [Type: Statement]; Focus: x1, Refs: [x1] Sem: Color(x1,green)^ Type(x1,cup)^ Name(x1,bill)	H2: Human Act: RequestIdObject [Type: Question(y/n)]; Focus: x2 Refs: [x2] Sem: Type(x2,car)
R1: Robot Act: Accept [Type: Explicit]; Act: DescribeObjectName [Type: Statement] Focus: y1 Refs: [y1 y1->x1] Sem: Color(y1,green)^ Type(y1,cup)^ Name(y1,bill)	R2: Robot Act: Accept [Type: Explicit(yes)]; Act: DescribeObjectProperty [Type: Statement] Focus: y2 Refs: [y2 y2->x2] Sem: Type(y2,car)

Fig. 3. Court échantillon de dialogue: en haut, sa retranscription écrite, et en bas la mémoire sémantique constituée.

rangées dans l'ordre d'apparition, comme montré figure 3. Au cours de l'analyse sémantique, sont déterminés : l'objet qui focalise l'attention (représenté par "Focus"), l'entité linguistique correspondante (dans "Refs") et ses propriétés ("Sem" regroupe couleur, forme, relations spatiales entre objets). La ligne Act correspond à l'intention communicative de l'humain perçue par l'intent recognizer. Par exemple, lorsque l'humain souhaite déclarer une information, l'action est représentée par "Act: Statement". Enfin le robot doit faire une résolution de référence concernant la voiture : celle-ci est appelée une première fois "car" puis une seconde fois "it". Pour que la machine comprenne de quoi on parle, elle va devoir associer ces deux entités linguistiques.

Dans le cas du robot NAO, la représentation de l'état du dialogue prend appui sur le résultat de la résolution de références et de l'historique du dialogue. L'espace des actions

du robot peut quant à lui être plus ou moins étendu (rejette, accepte, décrit, etc.). Celui-ci peut également utiliser une modalité coverbale en réalisant des gestes déictiques pour faciliter la compréhension mutuelle avec son homologue humain. On en voit l'exemple dans le cas de R1.

L'architecture expliquée ci-dessus est employée en HRI située pour notamment satisfaire aux enjeux collaboratifs favorisés par dialogue grounding, au cours d'une tâche de dénomination d'objets effectuée par le robot NAO. Les résultats principaux propres à l'incorporation de ce processus sont discutés plus en avant en section II. Néanmoins, les briques opérationnelles de cet ensemble sont pertinentes dans un contexte plus général.

B. Le discours incrémental

La plupart des systèmes de dialogue utilisent un seuil de silence pour déterminer l'instant où leur interlocuteur a fini de parler. L'énoncé en entrée est ensuite traité par les différents modules, un à la fois. Après analyse, le robot répondra de son côté par un synthétiseur vocal. Si les processus prennent du temps, par exemple suite à une tentative d'accès à une ressource externe, cela peut résulter en un délai de réponse excessif.

Un système incrémental peut être la solution pour ce cas de figure. En effet, ce dernier construit de façon continue une réponse prête à être énoncée à l'utilisateur. Au fur et à mesure que celui-ci parle, la réponse du robot évolue et s'adapte avant d'être fournie. Lorsque l'humain a fini de parler, la machine commence immédiatement à communiquer tout en poursuivant ses dernières étapes de calcul et modifiera la phrase si nécessaire.

Il y a de nombreuses raisons pour lesquelles le système pourrait requérir un temps additionnel afin de terminer les processus de détermination du "quoi dire". Par exemple, pour éviter de longs délais de réponse, il a été établi que les modules d'ASR devaient fonctionner en temps réel. Ce qui ralentit évidemment les autres fonctionnalités. Pour un besoin de performance et pour rendre la discussion beaucoup plus naturelle, il serait donc plus qu'utile que le système occupe son tour de parole dès que son interlocuteur a fini de parler.

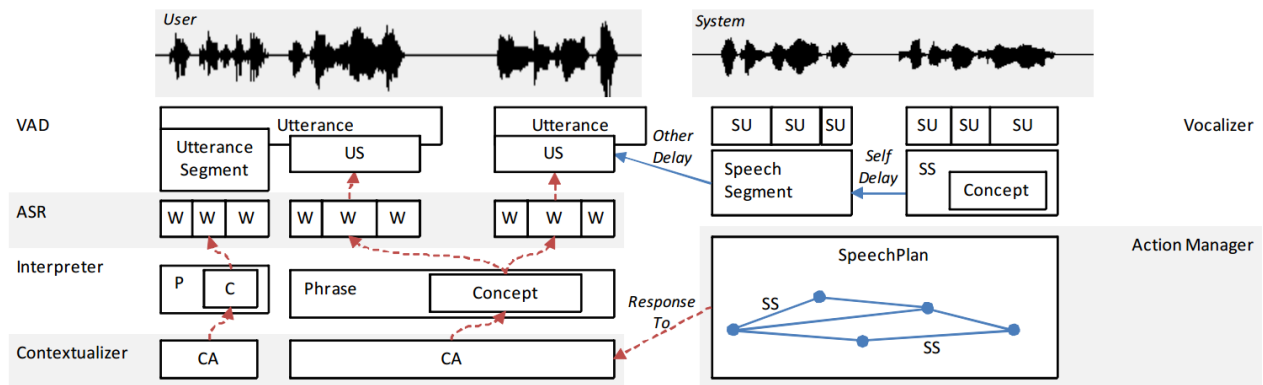


Fig. 4. Utilisation du VAD pour segmenter les énoncés. W=Mot ; SS=Segment de parole ; US=Unité de parole ; CA=Acte de dialogue

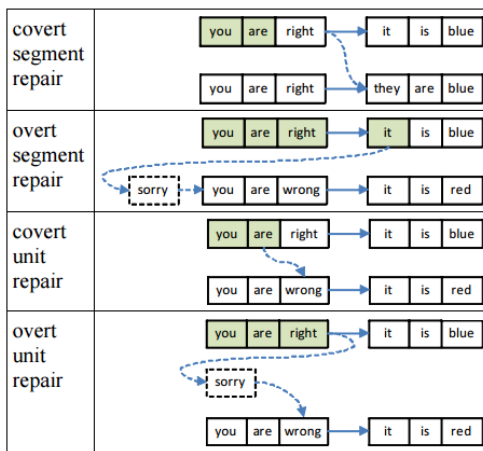


Fig. 5. Différentes façon d'adapter le discours en fonction des informations reçues.

L'idée du discours incrémental est que le locuteur choisit et modifie ses énoncés au fur et à mesure du discours. Lorsqu'il y a besoin de dévier du plan initial, deux portes s'ouvrent à lui : faire un overt repair ou un covert repair. L'overt repair consiste à récupérer une erreur sur un élément déjà prononcé, par exemple à l'aide de formules telles que "pardon". Le covert repair, quant à lui, s'applique justement quand l'élément n'a pas été prononcé et permet de modifier le plan de discours. Une application classique est de se laisser du temps pour réorganiser nos propos à l'aide de formules telles que "euh..." ou "je veux dire...".

Dans leurs travaux [8], Skantze et al ont implémenté le fonctionnement d'un discours incrémental sur une plateforme Jindigo élaborée sous JAVA open source. Ils y ont démontré une amélioration certaine du dialogue en termes de temps de réponse et de confort pour l'utilisateur humain. Le schéma de l'architecture Jindigo est donné figure 4. En plus des modules nécessaires à la gestion de dialogue, les auteurs ont rajouté un module VAD qui permet de séparer les énoncés en segments plus courts afin de les traiter les uns après les autres.

La machine va donc par la suite pouvoir générer un plan de discours (ici SpeechPlan) qu'elle adaptera. Un exemple type de modification est donné figure 5. On y différencie les segments et les unités par le fait qu'une unité est un mot tandis qu'un segment est une phrase.

L'un des logiciels les plus en vogue dans ce domaine est le programme InproTK conçu par Baumann et al [9]. Leurs travaux, basés sur ceux de Skantze et al., ont permis d'élaborer un synthétiseur de discours incrémental.

C. Une application du dialogue simple : les chatbots

De nos jours, l'agent conversationnel, ou chatbot, est omniprésent dans la vie digitale. De nombreux sites internet ou logiciels proposent à leurs utilisateurs des assistants virtuels pour les aider. Qui, par exemple, ne se souvient pas du trombone de Microsoft, Clippy, qui nous proposait de façon intrusive et inopinée son aide ? Ou alors tout simplement de Cortana, l'assistant virtuel de Google, et de Siri, son



Fig. 6. Clippy.

homologue d'Apple, qui proposent des réponses à toute requête - humoristiques lorsqu'elles relèvent de domaines trop conceptuels. En témoigne sa conception du sens de la vie, lorsque la question lui est adressée... La réponse est sans appel: "42".

Ces chatbots prennent en entrée le langage naturel et répondent de même grâce au TAL. Ils reposent en général sur l'utilisation d'une base de données. On parle alors de don cognitif puisque ce sont des humains qui les composent : ces programmes se servent de la réflexion d'humains pour offrir la meilleure réponse possible à une sollicitation. Il existe deux types de chatbots, ceux à conversation libre et ceux à conversation finalisée. Les seconds entretiennent une discussion avec un humain dans un but précis, comme par exemple réserver un avion, et auront donc un dialogue manager construit dans le but d'atteindre cet objectif. Les premiers, pour leur part, servent à entretenir la conversation et nous allons développer leur fonctionnement plus en détail par la suite.

La recherche sur ces interfaces homme-machine est influencée par le prix Loebner [10] qui récompense chaque année la machine ayant le mieux répondu au test de Turing, c'est-à-dire qui a le mieux créé l'illusion que le programme pense par un dialogue sensé. Les chatbots ne s'essaient pas à comprendre la discussion qu'ils entretiennent, mais se limitent généralement à repérer et réagir à des mots dits déclencheurs (méthode des mots clés) afin de provoquer un schéma de réponse adapté. Si l'interface ne sait pas réagir, elle va tenter de répondre de façon générique et rebondir sur une autre partie de la conversation. Tout ceci est réalisé dans le but de poursuivre la discussion et de tromper l'utilisateur le plus longtemps possible.

Le véritable problème dans la conception des agents virtuels vient de la base de données sur laquelle ils s'appuient. Pour commencer, la méthode des "mots-clés" a comme défaut majeur d'exiger une base de données importante pour donner des résultats, comptant entre 5000 et 10000 entrées. En effet, les mots-clés reconnus sont souvent des sous-expressions couramment employées telles que "j'aime..." ou "je déteste...", utilisées par le programme pour faire écho aux propos de l'utilisateur ("Depuis quand aimez-vous... ?", "Quelqu'un d'autre dans votre famille aimait-il... ?"). À elle seule, cette méthode se montre souvent insuffisante et ne fonctionne que peu de temps (typiquement durant deux ou trois réponses et rarement plus de dix) avant que le discours de la machine soit perçu comme artificiel. Afin d'y remédier, les concepteurs ajoutent souvent des expressions complètes telles que "j'aime le chocolat", et cela tend à encore rapidement accroître la taille de la base de données. La méthode ci-dessus peut ainsi difficilement se prêter aux logiciels possédant un trop faible espace de stockage de données, d'où une faible persistance des

I am the psychotherapist. Please, describe your problems.
 Each time you are finished talking, type RET twice.

I have a terrible headache.

Maybe you should consult a doctor of medicine, I am a psychiatrist.

Okay, I'll do that then.

Can you elaborate on that?

What are you talking about? No I cannot elaborate. See you later.

Why do you say that?

Now you're just talking nonsense!

What makes you believe now I am just talking nonsense?

You are not making any sense.

Perhaps your plans have something to do with this.

Fig. 7. Exemple de discussion avec ELIZA.

chatbots à garder en mémoire des informations sur les utilisateurs avec lesquels ils communiquent.

Pour les autres systèmes disposant d'un plus large espace de stockage, les développeurs utilisent une meilleure méthode de communication opérant éventuellement en deux étapes. Elle commence par l'analyse linguistique de la phrase entrée par l'utilisateur, pour tenter d'en extraire les informations permettant de répondre de façon appropriée selon les données contenues en mémoire. Si ce processus échoue, intervient alors la méthode précédente classique des "mots-clés". Cette dernière méthode n'améliore certes pas la taille de la base de données, mais elle permet au chatbot de traiter de nouveaux cas de langue puisqu'elle est apte à déterminer la proximité sémantique d'énoncés inconnus à ceux enregistrés. Par exemple, avec la méthode des mots-clés, si la base de données contient une entrée "who is that" et que l'utilisateur demande "whos that", le robot a toutes les chances de ne pas répondre juste, tandis qu'avec une analyse de l'entrée utilisateur, il est susceptible de répondre à la question.

Parmi les chatbots les plus connus, on peut citer le cas d'ELIZA [11]. Ce programme informatique, conçu par Joseph Weizenbaum en 1966, simule un psychologue rogière en reformulant la plupart des affirmations du patient en question, et en les lui posant. Ce qui est intéressant à noter est le fait qu'elle n'apporte aucune réponse « utile » mais se contente de faire parler son interlocuteur. Typiquement, si le patient prononce le mot « mère », ELIZA lui répondra « parlez-moi de votre mère ». Elle est également à la source de l'effet du même nom qui provoque chez les patients un attachement émotionnel alors même qu'ils savent pertinemment qu'ils ont affaire à une machine. Un historique de dialogue est disponible figure 7. On y voit un fonctionnement classique de ce programme avec une mise en défaut évidente. Eliza est la preuve depuis longtemps que certaines machines peuvent tromper leurs interlocuteurs et sembler passer le test de Turing, alors qu'elles ne constituent de façon évidente que des IA faibles.

Une autre illustration des avancées en la matière est le fameux logiciel Watson d'IBM [12]. Vainqueur début 2011 au terme de trois épisodes du jeu télévisé américain "Jeopardy", ce robot démontre ainsi sa capacité à comprendre l'énoncé des questions, buzzer pour prendre la main, trouver et énoncer les réponses en quelques secondes notamment grâce à une

synthèse vocale, et à choisir le thème et le montant de la prochaine question, comme l'exigent les règles du jeu.

De nos jours, l'un des programmes les plus pointus est le chatbot ALICE [13] élaboré par Richard Wallace en 1995. Celui-ci a remporté les prix Loebner de 2000, 2001 et 2004.

III. AMÉLIORATIONS PAR LE DIALOGUE SITUÉ

On ne se limite cependant pas à la conception d'agents conversationnels virtuels. L'évolution inévitable vers une contextualisation du dialogue homme-machine répond à la volonté d'insérer des machines physiques et intelligentes au sein des sociétés humaines afin d'en faciliter leur fonctionnement par automatisation. Conçues dans l'optique d'interagir avec un environnement constitué de personnes non-expertes en informatique et en robotique, elles sont dotées de capacités à communiquer de façon spontanée et fluide, et ce dans des langages accessibles à tous. D'actives recherches sont menées pour leur permettre de reproduire ce qui leur est montré par voies motrice et verbale, afin qu'elles deviennent des collaborateurs plus efficaces [14][15][16].

Vue sous un éclairage développemental, l'interaction située peut aussi être envisagée comme une étape vers des aspirations plus élevées relatives à l'abstraction. Cette analogie s'inspire de l'enrichissement linguistique progressif chez l'humain : l'enfant étant simplement capable de convoquer le moment et l'espace présents, il devient doué, à l'âge adulte, d'un spectre de capacités beaucoup plus complexes [17]. Ainsi, pour hisser les performances des systèmes artificiels à la hauteur de celles de l'être humain en situation d'interaction sociale, la technologie doit d'abord traiter de la grande variabilité des phénomènes issus du contexte ici-présent, pour plus tard évoluer vers des niveaux conceptuels. Plusieurs niveaux de conception des IA peuvent être envisagés, selon qu'on exige qu'elles pensent ou agissent, rationnellement ou comme les humains. Les différences sont notables, et vu que le cerveau humain est encore à l'aube de révéler tous ses secrets, s'en inspirer, littéralement parlant, demeure une approche approximative.

Quelle qu'en soit la motivation, l'interaction située est donc une préoccupation déterminante de la recherche en robotique actuelle. Des processus multimodaux variés et fondamentaux en sont des piliers majeurs, parmi lesquels le dialogue grounding, le turn taking, et le perspective taking. Prenant place au cours de conversations finalisées ou libres, ceux-ci visent à fluidifier l'échange en le rendant plus naturel mais également à permettre à chaque protagoniste d'appréhender la conception du monde de l'autre.

Au cadre d'étude précédemment établi s'ajoutent désormais les dimensions contextuelle et coopérative de l'IHR, manifestées par mobilisation de ressources de l'espace partagé entre agents co-présents (pas de capacités de prédiction ni d'abstraction) pour mener à bien une activité supposant un but commun.

A. Dialogue grounding

Au sein d'un contexte commun, les représentations internes diffèrent significativement d'un interlocuteur à un autre. En

effet, ils n'ont pas vécu les mêmes expériences préalables, quand bien même ils évoluent dans un environnement proche commun. Or, l'accomplissement de toute tâche collaborative requiert une communication efficace, il est alors nécessaire de faire en sorte que tout le monde se comprenne et d'établir ce que chacun perçoit de la scène en question. L'objectif est d'établir un common ground, c'est-à-dire une vision partagée, afin de partir sur de bonnes bases de dialogue. Ce processus est naturel dans les interactions entre humains. Il peut par exemple s'agir de rappeler de quel objet il est question dans le propos après l'avoir au fur et à mesure remplacé par un pronom personnel, et peut-être ainsi oublié. Que cela procède d'une requête de l'interlocuteur ou d'une initiative propre au locuteur, cette précision se voit motivée par un souci des protagonistes de se comprendre mutuellement. La coordination des états mentaux des participants en présence, en termes d'intentions et d'objectifs vis-à-vis de la tâche à traiter par dialogue grounding requiert ainsi le déploiement de certains efforts conversationnels supplémentaires. Chai et al [7] proposent une méthode pour la mise en application de différents niveaux d'efforts collaboratifs fournis par le robot. Il s'agit pour un robot NAO et un humain de construire un common ground. L'expérience se termine lorsque l'humain estime que la vision qu'il a du monde est partagée par le robot. Elle suscite les résultats suivants.

Soit le robot opère un faible effort pour établir un common ground. Par exemple, le robot prononce seulement "Compris" pour accepter une présentation de l'humain lors d'une tâche de dénomination. On ne sait pas si celui-ci a associé le bon terme au bon objet ou encore si les bonnes caractéristiques ont été attribuées au bon élément. Il peut donc arriver que l'écart perceptif entre interlocuteurs soit malencontreusement interprété par l'humain comme une communication réussie, alors que, de fait, des malentendus existent et n'ont simplement pas été explicités au cours de ce mode.

Sinon le robot est paramétré sur un effort conséquent afin d'agir proactivement. Il décrit sa perception en plus d'en requérir une confirmation, avec par exemple des phrases du type "Très bien, la bouteille s'appelle Eric". L'humain est ici plus à même de déterminer l'existence ou non d'erreur d'évaluation du monde par la machine. La compréhension mutuelle en est fortement améliorée.

Le résultat étonnant de cette étude vient du fait que les expérimentateurs humains semblaient plus souvent croire partager un common ground lors d'un faible effort collaboratif de la part du système que lors d'une collaboration étroite. Manquant d'informations claires sur ce que le robot a compris ou non, ils tendent par défaut à croire qu'un common ground a été établi. Néanmoins, il est évidemment plus efficace pour l'élaboration d'un dialogue constructif de procéder à une collaboration proactive de part et d'autre.

Le problème du grounding se pose d'autant plus que les humains et les robots ne sont pas dotés de systèmes de représentation symboliques similaires. Comme vu par exemple en première section, NAO utilise la mise en correspondance de graphes de vision et de dialogue, dont les nœuds, objets perçus, sont décrits à l'aide d'entités linguistiques et d'attributs. Le but in fine est que la machine soit capable de full grounding ou

connexion bidirectionnelle (entrante et sortante) entre le référent physique, son signe conventionnel ou label verbal, et son symbole dans le système cognitif.

B. Perspective taking

Cette partie peut être illustrée par les travaux de Feirreira et al [18]. Prenant place dans un environnement certes partagé mais dont l'ensemble des informations n'est pas accessible aux deux interlocuteurs, cette configuration rappelle celle du dialogue grounding. Cependant la démarche est proprement surpassée. Mettons-nous dans le cas de figure typique où le robot sait qu'un objet de la scène a été déplacé, alors que l'humain n'en a pas conscience. Dès lors, l'humain se retrouve possesseur d'une information erronée, avec laquelle le robot doit composer au mieux si une commande concernant cet objet est lancée.

Pour gérer ces différences de croyance, il faut le doter d'une capacité à prendre en compte que l'humain a un point de vue divergent du sien. Le processus de décision markovien partiellement observable offre un cadre de travail adapté à ces problématiques de gestion d'incertitude inhérente à l'information. Ce programme bayésien permet à l'agent de maintenir au cours du dialogue une certaine distribution des états de dialogue possibles, appelée état de croyance, alors qu'il interagit avec l'environnement via des algorithmes d'apprentissage par renforcement (RL). Le perspective taking consiste dès lors à modéliser chaque état de croyance de l'humain en estimant ce qui se passe dans son champ visuel et préhensible et, en même temps, à gérer un autre modèle de l'état du monde suivant les perceptions et raisonnements propres au robot. Par analogie avec les éléments déjà fournis, il s'agirait alors en quelque sorte de deux graphes de vision à faire coïncider.

Les auteurs ont donc choisi d'intégrer la notion de divergence dans le suivi des états de croyance, et y ont rajouté une commande qui informe l'interlocuteur, si besoin est, de son erreur, pour qu'il puisse poursuivre son objectif avec des données mises à jour. Ceci est testé en environnement virtuel sur des tâches de manipulation d'objets par contrôle immersif d'un avatar humain (déplacements, regard, interaction avec l'environnement). Grâce au module de raisonnement spatial SPARK, le robot peut à la fois alimenter dynamiquement sa base de connaissance à propos de l'environnement et détecter les gestes de l'utilisateur.

Quantitativement, les données sont encourageantes mais pas concluantes. En effet, les critères d'évaluation des diverses conditions posées, en termes de précision (taux de succès) et d'efficacité temporelle (peu de changements de cap), ne sont pas vraiment significatifs du fait d'intervalles d'incertitudes assez importants. Mais comme en témoigne l'extrait de conversation suivant figure 8, des différences comportementales sont remarquables. L'humain y est informé que des livres rouge et marron sont respectivement posés sur la table de chevet et la table de la cuisine. Ensuite, sans que l'humain le sache, les livres sont intervertis. Seul le robot est au courant de cette modification : il informe donc l'humain de la nouvelle configuration lorsque ce dernier souhaite récupérer le livre.

<p>R_1: Can I help you? U_1: Bring me the book on my bedside table</p> <p>R_2: The brown book is not on the bedside table, it has been moved to the kitchen.</p> <p>U_2: Ok, bring it to me</p> <p>R_3: I will give you the brown book that is on the kitchen table</p>	<p>R_2: Do you want the red one?</p> <p>U_2: No, the brown book.</p> <p>R_3: There is no brown book in your bedroom but there is one in the kitchen</p> <p>U_3: Are you sure? Well, bring me that one.</p> <p>R_4: I will give you the brown book that is on the kitchen table</p>
(a)	(b)

Fig. 8. Exemples de perspective taking

Les systèmes opérant par RL exécutent préférentiellement la commande plutôt que d'essayer de systématiquement vérifier sa validité par usage des actes « confirmer » et « informer » de la croyance divergente. Donc même si le problème de divergence des croyances n'est pas explicitement adressé (cas b), le module décisionnel contraint de traiter un niveau additionnel de malentendu (voir (b) de R_2 à U_3) est susceptible de faire aboutir la tâche (R_3 (a) = R_3 (b)).

Le suivi temps-réel des croyances d'un utilisateur par gestion du dialogue dans un cadre POMDP multimodal aide donc à planifier des tâches impliquant de fausses croyances sans nuire au déroulement des tâches ordinaires, et ce pour une expérience utilisateur plus naturelle et efficace. Le système reste à intégrer dans des robots pour confirmer ces résultats prometteurs au sein de configurations plus réalistes.

C. Turn Taking

Les humains apprennent, collaborent et délèguent à travers l'interaction sociale. Cette dynamique est gouvernée par le processus de turn-taking, déterminant les tours d'action "entreprenables" par les participants. Maîtrisé dès la petite enfance, cette pratique est symptomatique d'une qualité d'adaptabilité au cours de l'échange. Un robot qui empiète sur les tours de parole, manque de réactivité ou laisse à l'humain toute la charge de se s'adapter à la dynamique de la machine n'apportera pas un confort de dialogue exceptionnel. Laisser cette propriété émerger de la combinaison d'autres processus, par exemple avec la régulation des émotions et le comportement réactif chez le robot Kismet [19], revient à passer outre la somme des facteurs comportementaux et cognitifs impliqués dans l'acte en temps réel.

Il est plutôt intéressant de développer explicitement des contrôleurs individuels du phénomène et de les intégrer [20]. Cela permet de rendre pleinement compte des aléas conversationnels à rencontrer. Pourvu dans cette optique, le robot Simon [20], muni de l'architecture CADENCE (pour Control Architecture for the Dynamics of Embodied Natural Coordination and Engagement), génère un comportement multimodal hautement coopératif pour l'accomplissement d'une tâche, faisant preuve d'initiative, réciprocité, d'autonomie, de contrôle, de fluidité. Par réciprocité, on entend

que le robot est apte à modéliser donc prédire les besoins de son partenaire et à prendre des initiatives en conséquence. Cela est notamment permis par la détection de contingence, qui s'applique aux changements perceptifs produits en réponse aux propres actions de l'agent. Implémentée via des méthodes basées sur la vision ou l'audition, ce processus aboutit sur l'apprentissage de comportements sociaux et confère son rythme cyclique à l'interaction.

IV. DISCUSSION

La vision traditionnelle du robot conversationnel est celle du maître humain et de la machine esclave. Ainsi avons-nous longtemps restreint le robot à une obéissance mécanique suite à un ordre simple et direct, et ne l'avons autorisé à parler uniquement que pour indiquer son incompréhension et requérir des précisions le cas échéant. Parmi les premiers robots réellement capables de converser en langage naturel, nous pouvons citer le cas de MAIA [21] ou de RHINO [22], apparus dans les années 90. Le premier était un robot capable de porter des objets et d'obéir à des requêtes simples, tandis que le second servait de guide dans les musées et répondait aux questions d'utilisateurs. Fleurons technologiques de l'époque, leurs performances de dialogue laissaient cependant à désirer tant leurs limitations étaient importantes.

Tout d'abord, la taille de leurs jeux de commandes et de réponses était fixée et fortement limitée, faute d'une base de données suffisante. En outre, le seul acte de dialogue accepté était la requête, de même que la machine ne prenait pas d'initiative dans la conversation. De plus, ces programmes se trouvaient face à une incapacité de gérer le langage situé, le langage affectif et les fonctions coverbales. Enfin, il leur était également impossible d'apprendre de la discussion passée comme de planifier la prise de parole afin de l'utiliser pour un but précis.

Nous présentons ici une liste concise et non exhaustive des principaux défis auxquels se confronte aujourd'hui la recherche en robotique pour améliorer la performance conversationnelle.

A. Accepter plusieurs actes de dialogue

Les actes de dialogues sont définis par la fonction et

l'objectif d'un énoncé dans un discours. Par exemple, "Donne-moi le sel" est une requête. Searle a défini une classification d'actes de langage qui se séparent en cinq grandes familles : les assertifs, les directifs (dont fait partie la requête), les promissifs, les expressifs et les déclaratifs. La plupart des robots se contentent d'obéir à des actes de dialogue directifs mais prendre en compte d'autres actes de dialogue serait un avantage non négligeable dans la construction d'un dialogue amélioré.

B. Mixer l'initiative de dialogue

Un autre point d'évolution est celui de l'initiative de dialogue. Développée en section II.C, il s'agit ici de rendre la machine capable d'anticiper la conversation et de prendre de façon naturelle ou même de provoquer son tour de parole. Cela permet de rendre la conversation plus fluide et rend l'interlocuteur humain moins méfiant.

C. Situer le langage et répondre au symbole grounding

Traitée au cours des sections précédentes, la problématique posée est celle de la perception du monde qui nous entoure. En effet, ayant des perceptions et des systèmes de représentation différents, il est capital de pouvoir partager une vision commune avec des machines pour entretenir des dialogues productifs.

D. Gérer les émotions

La gestion des émotions est également un sujet phare de la robotique. Une bonne interprétation d'un discours passe en partie par la bonne appréhension de l'état émotionnel du locuteur.

E. Gérer la communication coverbale

De nombreuses recherches montrent que le coverbal est une des composantes majeures du dialogue. Traitée en section I.A, on a pu observer les effets bénéfiques de l'ajout de perception et de production de fonctions coverbales.

F. Gérer l'apprentissage

Parmi les nombreux points clés de l'interaction homme-machine, le fait de pouvoir apprendre du monde et des conversations fait partie des voies de progrès considérables qui permettront peut-être à la science-fiction d'aujourd'hui de devenir la réalité de demain.

G. Capacité à gérer le groupe

Enfin, présentée en dernier mais pas des moindres, la gestion de la multimodalité des groupes et des dialogues apparaît être l'un des principaux défis de ce domaine. En effet, cette branche prend en compte toutes les questions relatives à la séparation de sources, la gestion concurrente de plusieurs signaux, ou encore la génération de dialogues écrits, comme l'exige nos voies de communication privilégiée plus en avant.

Nous l'aurons compris, la robotique en matière d'IHR n'a qu'entamé ses balbutiements mais les promesses qu'elle formule sont merveilleuses. Cette étude peut également porter

à réflexion d'un point de vue éthique, en ce qui concerne cette quête d'une interaction homme-robot "plus vraie que nature", en ce sens que d'aucuns auront du mal à déterminer s'ils s'adressent à un humain ou un robot. Derrière un écran, la prouesse est déjà maîtrisée via des agents conversationnels assez convaincants, et des projets en cours à base de clones robotiques incarnés, dont celui intitulé Replika, tendent dans le sens de cette confusion. Les applications en sont multiples, de laisser une trace de l'utilisateur après son décès à gérer certaines tâches fastidieuses à notre place tel un secrétaire personnel [23]. L'enjeu de taille consiste en tout cas à fusionner intelligence conversationnelle et robot anthropomorphique, afin de créer des humanoïdes ultra-réalistes. L'humain reste toutefois un organisme social, émotionnel, créateur, relativement autonome... ce qui permet donc de conserver un délai d'avance non négligeable avant que les robots en soient de fidèles imitations, militant pour la reconnaissance de leurs droits ! D'autant plus que, bien que prompt à attribuer des traits spécifiques à son espèce à ce qui en est dépourvu et à s'y attacher, l'humain se sent mal à l'aise devant l'excès de réalisme d'une créature pourtant artificielle.

V. CONCLUSION

Au cours de ce papier, nous avons pu passer en revue le dialogue homme - machine et ses différentes composantes. Ce dernier requiert une coordination entre modules spécialisés de traitement des entrées acoustiques, processeurs sémantiques, générateurs d'énoncés, etc., et incluent la mise en œuvre de processus essentiels de résolution de référence, dialogue grounding, perspective taking, et autres turn taking, visant à

REMERCIEMENTS

Nous tenons à remercier Gérard Bailly et Sylvie Pesty pour leur aide dans les recherches fournies et l'écriture de cet article.

REFERENCES

- [1] F. Landragin et R. Lehoucq, "Je parle, donc je suis : le dialogue humain-machine est-il possible ?" Bifrost, la revue des mondes imaginaires, Ed. du Bérial, 2014, pp.180-186.
- [2] A. M. Turing, "Computing machinery and intelligence," vol. 59, No. 236, 1950, pp. 433-460.
- [3] W.J.M. Levelt, "Trends in cognitive sciences," 1999, Elsevier.
- [4] R. Lopez, "Introduction to dialogue systems," in *Spoken, multilingual and Multimodal Dialogue Systems: Development and Assessment*, West Sussex, England, 2005, pp. 21-24.
- [5] N. Mavridis, "A Review of Verbal and Non-Verbal Human-Robot Interactive Communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22-35, 2015.
- [6] R. Fang et al., "Embodied collaborative referring expression generation in situated human-robot interaction," *Proc. of the 10th Annu. ACM/IEEE Int. Conference on Human-Robot Interaction*, pp. 271-278, March, 2015.
- [7] J.Y. Chai, "Collaborative effort towards common ground in situated human-robot dialogue," *Proc. ACM/IEEE Int. Conference on Human-robot Interaction*, pp. 33-40, 2014.
- [8] G. Skantze et A. Hjalmarsson, "Towards incremental speech generation in dialogue systems," *Proc. of the 11th Annu. Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 1-8, 2010 Association for Computational Linguistics.

- [9] T. Baumann and Schlangen, D. (2012, June). The INPROTK 2012 release. In NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data (pp. 29-32). Association for Computational Linguistics.
- [10] M. L. Mauldin, "Chatterbots, tinymuds, and the turing test: Entering the Loebner prize competition," *AAAI*, vol. 94, pp. 16-21, Aug, 1994.
- [11] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36-45, 1966.
- [12] E. Guizzo, "IBM's Watson Jeopardy computer shuts down humans in final game," *IEEE Spectr.*, vol. 17, 2011.
- [13] R. Wallace, *The elements of AIML style*, Alice AI Found., 2003.
- [14] S. Wrede, "A user study on kinesthetic teaching of redundant robots in task and configuration space," *J. of Human-Robot Interaction*, vol. 2, no. 1, pp. 56–81, 2013.
- [15] B. D. Argall, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [16] C. L. Nehaniv et K. Dautenhahn, *Imitation and social learning in robots, humans and animals: behavioural, social and communicative dimensions*. Cambridge University Press, 2007.
- [17] N. Mavridis, "Grounded situation models for situated conversational assistants," Ph.D. dissertation, Massachusetts Institute of Technology, 2007.
- [18] E. Ferreira, «Users' Belief Awareness in Reinforcement Learning-Based Situated Human–Robot Dialogue Management," *Natural Language Dialog Systems and Intelligent Assistants*, Springer International Publishing, 2015, pp. 73-86.
- [19] C. Breazeal, "Emotion and sociable humanoid robots," *Int. J. of Human-Computer Studies*, vol. 59, no. 1–2, pp. 119–155, 2003.
- [20] C. Chao, "Timing multimodal turn-taking in human-robot cooperative activity," Ph.D. dissertation, Georgia Institute of Technology, 2015.
- [21] G. Antoniol et al., "Robust speech understanding for robot telecontrol," *Proc. of the 6th Int. Conference on Advanced Robotics*, pp. 205–209, 1993.
- [22] W. Burgard, "The Interactive Museum Tour-Guide Robot," *Proc. of the Fifteenth Nat. Conference on Artificial Intelligence (AAAI-98)*, 1998.
- G. Renouard. (2016, Dec 14). *De "Blade Runner" à "Westworld" : les robots doivent-ils nous ressembler?* [Online]. Available : <http://motherboard.vice.com/fr/read/de-blade-runner--westworld--les-robots-doivent-ils-nous-ressembler>.

La cobotique

Julien Mas
Univ. Grenoble-Alpes
Grenoble, France
julien.mas81@gmail.com

Romane Cecchi
Univ. Grenoble-Alpes
Grenoble, France
romane.cecchi@gmail.com

Abstract— La cobotique est un champ de la robotique en pleine expansion, car prometteur dans la pluridisciplinarité des problèmes qu'il est à même de traiter. Dans ce document nous présenterons la cobotique qui est une collaboration entre l'humain et le robot, en nous intéressant plus particulièrement aux cas où l'espace de travail et les tâches sont totalement partagés. Nous tenterons de présenter les différentes approches et composantes de cette interaction et quelques solutions techniques actuellement utilisées.

Keywords— *cobotique ; collaboration humain-robot ; coordination humain-robot ; intentionnalité ; communication*

I. INTRODUCTION

Alors que jusqu'à présent les robots travaillaient de manière autonome et isolée, le besoin de réaliser des tâches variées et changeantes a fait émerger une nouvelle discipline dans le champ des sciences robotiques du nom de robotique collaborative (cobotique).

Cette discipline se caractérise par un travail collaboratif et coordonné entre les robots et les humains, au sein d'un même espace de travail. Le principe est d'associer en temps réel les capacités d'un robot (force, précision, répétition...) et les compétences d'un être humain (savoir-faire, analyse, décision...). Son intérêt est d'accompagner ou d'aider le collaborateur dans ses fonctions ou dans des situations particulières. La cobotique possède des applications dans de nombreux domaines, comme celui de la santé, avec notamment des robots accompagnants pour personnes aveugles et des robots de soins aux personnes âgées, ou celui de la manutention, avec le besoin de déplacer des objets lourds et encombrants. La cobotique permet de soulager l'humain de tâches difficiles et également de pouvoir prendre plus de risques, avec par l'exemple des robots de secours dans des milieux dangereux. Mais la cobotique a aussi pour vocation d'améliorer la productivité de l'homme, en lui permettant de se consacrer aux tâches à forte valeur ajoutée. Ce domaine est complexe par le nombre de problèmes qu'il soulève tels que l'éthique, la sécurité ou la reproductibilité qui sont au cœur de ce champ de la robotique.

La collaboration humain-robot peut prendre un éventail de forme, passant d'une collaboration très simple et limitée à un espace et des tâches entièrement partagées. Nous nous focaliserons plus particulièrement dans cette revue sur le type de collaboration le plus complexe impliquant un partage total de l'espace de travail et des tâches à effectuer, ainsi qu'un couplage partiel ou indirect entre les agents (Fig. 1).

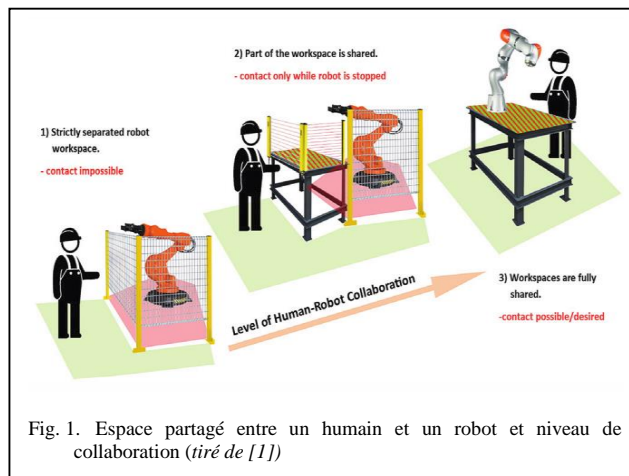


Fig. 1. Espace partagé entre un humain et un robot et niveau de collaboration (tiré de [1])

Nous tenterons dans la suite de ce document de présenter l'interaction homme-robot de façon qualitative et méthodique.

II. LA COLLABORATION

Les êtres humains et les robots qui collaborent sur une tâche commune forment une équipe. Cette équipe peut se définir comme un ensemble de partenaires possédant des compétences complémentaires qui s'engagent dans une action partagée avec un but commun. Même si la collaboration entre les humains s'établit plus ou moins naturellement, elle n'en nécessite pas moins de nombreuses compétences implicites. Les études basées sur la collaboration humain-humain permettent de définir les problématiques principales lors de la mise en place d'une collaboration humain-robot. Tout d'abord, il est essentiel de définir qui de l'humain ou du robot doit effectuer une tâche particulière afin qu'ils ne rentrent pas en conflit. Ensuite, il faut que l'humain et le robot puissent se coordonner dans leurs actions. Cela nécessite notamment que le robot connaisse les intentions de l'humain afin de planifier ses propres actions en conséquence. Enfin, il faut que le robot réussisse à se faire comprendre pour que la collaboration soit réellement efficace. Nous aborderons dans la suite de cet article l'ensemble de ces problématiques ainsi que les solutions mises en œuvre pour y répondre.

A. L'attribution des rôles

Lors d'une collaboration, chaque partenaire possède son propre rôle. Dans certains cas, le rôle est établi a priori par la nature de la tâche elle-même. Dans d'autres cas, une même

tâche permet différentes attributions. Deux sujets portant une table lourde par exemple peuvent collaborer en tant que partenaires égaux ou bien avoir une relation asymétrique dans laquelle l'un des deux sujets prend les décisions tandis que l'autre les exécute.

Lors d'un travail collaboratif humain-robot, l'idée qui a été développée par les chercheurs pour assigner les tâches est que le robot évalue lui-même l'action à exécuter. S'il est capable d'effectuer l'action, il suggérera de la faire, sinon il demandera de l'aide à l'humain. Si l'humain est capable d'effectuer l'action seul, la tâche lui sera assignée. Dans ce cas, c'est la nature même de la tâche qui détermine qui de l'humain ou du robot effectuera l'action [2].

Cependant, certaines tâches exigent la collaboration physique des deux partenaires (humain et robots). Dans ce cas, plusieurs méthodes d'attributions ont été conçues/pensées. L'une d'entre elles consiste en ce que l'agent humain génère des commandes que le robot exécute. Cette méthode basée sur une attribution asymétrique des rôles est celle qui a été le plus développée par les chercheurs jusqu'à présent. Néanmoins, d'autres approches basées sur une répartition plus égalitaire des rôles se sont développées depuis quelques années. Nous détaillons dans les sections suivantes ces deux méthodes d'attribution des tâches.

1) Une attribution asymétrique

Lorsque la tâche à effectuer nécessite la collaboration physique de l'humain et du robot, l'humain peut donner des ordres que le robot exécute. Une autre possibilité est que le robot détecte lui-même les actions de l'humain et s'y conforme. Dans les deux cas, la distribution des rôles n'est pas questionnée. Un nombre important d'études basées sur cette méthode asymétrique d'attribution des tâches ont été réalisées, en particulier pour des tâches nécessitant de soulever et transporter des objets lourds ou volumineux. Diverses plateformes à bras robotisé ont ainsi été développées [3], équipées de contrôleur pour détecter les intentions de l'opérateur humain [4], [5]. Les exosquelettes, qui doivent adapter leurs mouvements pour se synchroniser avec ceux de l'humain, utilisent également cette méthode d'attribution des tâches asymétrique. Plusieurs exosquelettes multiplicateurs de force ont été développés ces dernières années, en particulier pour des applications militaires [6]. Enfin, cette méthode englobe également les robots correcteurs de trajectoire ou de force [7] ainsi que les outillages d'assistance intelligents (*intelligent assist devices* ; IAD) [8], [9]. En effet, malgré leur nom, les IAD qui guident le mouvement d'un utilisateur en l'empêchant d'aller dans des directions ou régions indésirables de l'espace de travail peuvent être considérés comme robots « exécuteurs » car ils fournissent un soutien lors de l'action mais ne peuvent pas accomplir la tâche seuls.

2) Une attribution égalitaire

Lorsque deux humains effectuent une activité physique en collaboration, chacun peut alternativement s'emparer du contrôle de la tâche afin de prendre des décisions appropriées lors de la présence d'un obstacle par exemple. Cet échange de rôle se fait naturellement car l'humain sait que son partenaire peut connaître ou percevoir des choses qu'il ne sait pas ou ne voit pas. Afin d'améliorer la collaboration physique humain-

robot, une attribution égalitaire des rôles, où l'humain n'est pas seul à mener la tâche, semble donc nécessaire. Dans cette optique, P. Evrard and A. Kheddar [10], [11] ont implémenté une répartition flexible des rôles permettant à chaque partenaire d'osciller entre les comportements de meneur et de suiveur. Avec cette approche, chaque partenaire peut réclamer ou abandonner le contrôle de manière harmonieuse. Cette méthode a été utilisée lors d'une tâche de soulèvement de table entre un humain et un robot humanoïde. Le robot a pu adapter son comportement aux sujets humains qui ont changé leur rôle au cours de la tâche, mais les résultats obtenus ne sont pas encore très robustes. Dans une approche similaire, des chercheurs ont proposé un système où chacun des partenaires a la possibilité de basculer entre un comportement de meneur ou de suiveur lors d'une tâche de transport complexe [12]. Le robot est capable de prendre chacun des deux rôles, mais le passage d'un mode à l'autre n'est pas automatisé.

B. La coordination humain-robot

De nombreuses tâches impliquant une action conjointe, telle que déplacer une table, exigent la coordination des actions exécutées, et ce quel que soit le rôle de chacun des partenaires. Lors d'une action conjointe, humain et robot ont donc besoin de coordonner leurs actions dans l'espace et dans le temps pour aboutir à un but commun [13]. Pour cela, les chercheurs se basent essentiellement sur l'étude de la coordination des mouvements chez les humains [14]. Celle-ci soulève deux problématiques principales : (1) pour qu'il y ait coordination, les agents doivent comprendre les actions exécutées par l'autre agent et les buts associés. Ils peuvent ainsi sélectionner des objectifs et actions complémentaires (ou au moins qui ne sont pas en conflits) parmi plusieurs actions possibles. Cette problématique fait référence à la détection de l'intention qui sera traitée dans la section suivante (section C) ; (2) Les agents doivent coordonner leurs mouvements en temps réel, ce qui requière une estimation précise de la chronologie et de la trajectoire de ces mouvements.

La coordination en temps-réel peut impliquer une simple division de tâche (où chaque sous-tâche est attribuée à un seul agent) ou bien une coordination plus complexe (comme le transport d'une table par exemple). Dans les deux cas, les agents doivent accorder une attention particulière au timing des opérations. En effet, ce timing est important lors de la coordination. L. A. Suchman [15] a montré que si la réponse d'un système est retardée, les utilisateurs (pour lesquels le comportement du système n'est pas transparent) considèrent le manque de réponse en temps voulu comme une défaillance. Ainsi, les utilisateurs attendent une réponse rapide à leurs actions comme condition préalable à une action conjointe. Dans ce sens, J. Shah, J. Wiken, B. Williams, et C. Breazeal, [16] ont développé un système d'exécution permettant d'adapter les actions du robot au partenaire humain. Les auteurs montrent que ce système réduit le temps d'inactivité des humains, améliorant ainsi les performances de l'équipe.

C. La compréhension des intentions

Comme nous l'avons mentionné précédemment, un travail collaboratif efficace nécessite que chacun des membres de l'équipe soient informés de ce que les autres membres veulent

faire. En effet, les robots connaissant les objectifs et intentions de leur partenaire sont plus susceptibles de collaborer de façon productive pour mener à bien la mission commune. Il semble donc indispensable qu'un robot puisse prédire les intentions de son partenaire lors de sa collaboration avec un humain. En utilisant ces informations, le robot pourra ainsi être en mesure de prendre de meilleures décisions parmi toutes les actions possibles.

Prévoir les intentions du mouvement humain est un domaine émergent dans la recherche en robotique. Néanmoins, avec l'apparition des robots collaboratifs, de nombreuses recherches se sont focalisées sur la planification des actions du robot afin de réaliser une tâche avec l'humain. Prévoir l'intention du mouvement humain est un problème difficile combinant l'expérience dans le domaine de la robotique et de l'intelligence artificielle. Cela implique l'utilisation de capacités cognitives, comme la perception, le raisonnement, la prédiction, l'apprentissage ou encore la planification. De nombreuses approches ont été utilisées dans la littérature pour la reconnaissance de l'intention. Ces approches sont basées sur différentes méthodes probabilistes telles que des prédicteurs probabilistes [17], des chaînes de Markov cachées [18], des réseaux de neurones [19], ou encore des automates probabilistes [20]. Des travaux récents ont également utilisé le modèle probabiliste de la machine d'état pour prédire les intentions de l'humain en temps réel [21].

Une personne peut communiquer son intention soit délibérément par communication explicite, soit implicitement par des actions. Nous détaillons dans la suite les moyens utilisés par un être humain pour communiquer son intention à autrui, et les procédés mis en œuvre pour que le robot reconnaisse cette intention (Fig. 2).

1) La communication explicite

En utilisant la communication explicite, un individu s'assure que son partenaire ait reçu certaines informations au sujet de son intention. La communication explicite peut se produire par divers canaux, tels que la parole, le geste ou les signaux haptiques.

Pour les humains, la parole est un moyen très naturel de communication permettant de transmettre une énorme quantité d'informations. Cependant, extraire cette information est très complexe. Un signal acoustique capturé par un microphone doit être converti en un ensemble de mots par reconnaissance vocale. Pour la reconnaissance vocale, les données acoustiques capturées sont mises en correspondance avec le domaine fréquentiel via une transformée de Fourier, puis la séquence la

plus probable, suivie des mots les plus probables sont retrouvés avec des systèmes utilisant par exemple des modèles de Markov cachés. Si des informations plus complexes sont nécessaires, les séquences reconnues doivent être analysées pour comprendre la parole produite. Les approches les plus importantes concernant la compréhension de la parole sont les réseaux sémantiques [22] et la dépendance conceptuelle [23].

Une autre manière très fondamentale de communiquer l'intention est de produire des signaux visuels tels que des gestes et des expressions faciales. Les robots peuvent deviner les intentions à partir de la reconnaissance des gestes. Les gestes peuvent être classés en gestes manipulateurs et communicatifs (*ex.* gestes de pointage, signes primitifs et langage gestuel). Les gestes manipulateurs sont souvent faits inconsciemment, ils font donc partis de la communication implicite. Les gestes communicatifs quant à eux expriment une intention explicitement. Les modalités d'entrée pour les données gestuelles sont majoritairement les caméras. Plusieurs méthodes d'interprétation visuelle des gestes ont été développés dans la littérature [24], [25].

Des informations importantes sur l'intention de l'humain peuvent également dériver des gestes de la tête telles que la direction du regard ou des expressions faciales. En interprétant des expressions, les robots peuvent obtenir des informations sur l'approbation ou l'humeur. La direction du regard permet de prédire ce sur quoi l'humain se concentre [26].

Enfin, la communication tactile ou haptique est une autre façon de transmettre son intention. Par exemple, deux partenaires portant une table peuvent connaître la direction de déplacement du partenaire par les forces appliquées ou les couples [27]. Les angles et l'orientation peuvent également être utilisés comme information dans la communication haptique. Les humains et les robots peuvent non seulement collaborer haptiquement en manipulant des objets, mais aussi interagir directement pour se serrer la main par exemple.

2) La communication implicite

La communication implicite exige une interprétation importante. Un individu peut par exemple inférer les intentions de son partenaire en regardant ses actions (*ex.* gestes de manipulation) et son comportement. L'interprétation qui doit être faite est complexe et nécessite que le robot possède un modèle interne du partenaire humain. De nombreuses recherches se basent sur la reconnaissance des actions pour deviner les intentions [28]. Les signaux physiologiques telles que la fréquence cardiaque peuvent également être mesurés pour déduire le niveau d'approbation ou de stress [29], [30]. Ils

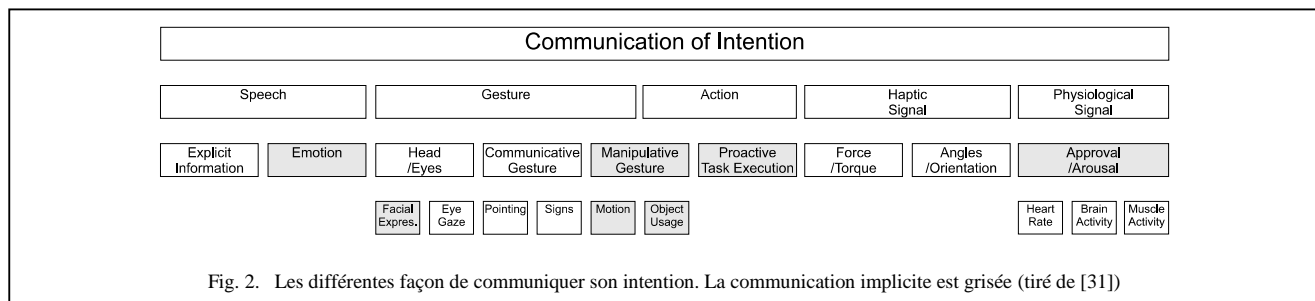


Fig. 2. Les différentes façon de communiquer son intention. La communication implicite est grisée (tiré de [31])

peuvent également renseigner sur la motivation ou l'implication de l'agent humain dans la tâche. Néanmoins, les mesures de signaux physiologiques nécessitent des dispositifs attachés à l'humain qui peuvent entraver le mouvement naturel.

Lorsque l'intention est estimée correctement et que tous les partenaires sont d'accord sur cette intention, une intention conjointe émerge. Sur la base de cette intention commune, des actions conjointes peuvent être coordonnées pour atteindre un objectif commun.

D. La communication des intentions

Toutes les études que nous avons vu jusqu'à présent s'intéressent à la compréhension de l'intention humaine par le robot. Néanmoins, la théorie de l'intention commune affirme qu'une collaboration robuste dans des environnements dynamiques exige un canal de communication ouvert pour coordonner le travail d'équipe et éviter les croyances divergentes [32]. Si cette théorie est applicable pour la collaboration humain-humain, elle devrait également l'être pour une collaboration humain-robot. Il semble donc nécessaire d'intégrer des moyens de communication au robot pour que l'humain puisse le comprendre et agir en conséquence.

La plupart des recherches dans ce domaine concernent le rôle du comportement verbal dans la coordination de l'activité conjointe. Néanmoins, le comportement non-verbal du robot semble avoir tout autant d'importance. En effet, l'étude menée en [33] montre l'intérêt de l'usage d'une communication non-verbale du robot, en mettant en place des gestes communicatifs tels que des hochements de tête ou des gestes de confusion pour permettre la communication de l'état mental du robot sans avoir besoin de le vocaliser. Alors qu'il envoie des informations en continu, l'agent humain est en mesure de détecter d'un simple regard à quel moment l'information transmise n'est pas bien interprétée. Un gain de temps réel a été mesuré lors de cette expérience autour du banc de test (Fig. 3).

Certains robots sont également capables de transmettre leur intentionnalité par l'expression du visage [34] ou bien générer des gestes pour communiquer avec leurs partenaires [35]. De plus amples recherches ont été menées sur la façon d'améliorer la lisibilité d'un robot en utilisant des principes d'animation (comme la prévision et la réaction) et l'anticipation des actions,



Fig. 3. Léonardo et son espace de travail. Léonardo est un robot humanoïde expressif possédant 65 degrés de liberté, conçu pour l'interaction sociale et le travail d'équipe (tiré de [33]).

rendant le robot plus expressif et l'interaction plus fluide [36]. L'utilisation de la communication non verbale a également été étudiée par A. G. Brooks et R. C. Arkin [37]. Ils ont alors observé que la présence d'un canal de communication non-verbal affecte la façon dont les individus interagissent avec le robot. Enfin, M. Faria, A. Costigliola, P. Alves-Oliveira, and A. Paiva [38] ont montré que l'utilisation de la communication implicite basée sur des comportements et des actions associées à d'autres êtres (comme les animaux) permet aux individus de comprendre ce que le robot veut communiquer, même lorsque celui-ci manque de moyens de communication habituels. Ce résultat permet d'envisager la possibilité d'utiliser des robots non humanoïdes pour atteindre une communication similaire.

III. CONCLUSION ET PERSPECTIVES

Finalelement, les différents axes que nous avons abordés dans ce document nous permettent de caractériser l'interaction entre l'homme et la machine de façon à pouvoir aisément identifier la cause des problèmes rencontrés et également permettre une optimisation de l'interaction. Cette interaction est complexe et devient de plus en plus complète avec les innovations dans le domaine de la reconnaissance faciale et celui de la reconnaissance des gestes.

C'est par la compréhension de l'interaction entre les opérateurs que la cobotique peut se développer. Et c'est grâce à l'introduction de modèles d'interactions de plus en plus complet que l'on sera à même d'exécuter des tâches nécessitant un plus grand nombre d'opérateurs. La communication est au cœur de cette perspective d'un plus grand nombre d'agent car les moyens de communiquer entre robots et humains ne sont pas encore exactement les mêmes. C'est par l'harmonisation de cette communication que la cobotique pourra atteindre de nouveaux horizons.

REFERENCES

- [1] A. Vysocky and P. Novak, "Human robot collaboration in industry," *MM Science Journal*, pp. 903-906, 2016.
- [2] G. Hoffman and C. Breazeal, "Collaboration in human-robot teams," in *Proc. of the AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA, 2004*.
- [3] K. Kosuge and Y. Hirata, "Human-robot interaction," in *Robotics and Biomimetics, 2004. ROBIO 2004. IEEE International Conference on*, 2004, pp. 8-11.
- [4] T. Wojtara, M. Uchihara, H. Murayama, S. Shimoda, S. Sakai, H. Fujimoto, *et al.*, "Human-robot collaboration in precise positioning of a three-dimensional object," *Automatica*, vol. 45, pp. 333-342, 2009.
- [5] J. Stückler and S. Behnke, "Following human guidance to cooperatively carry a large object," in *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, 2011, pp. 218-223.
- [6] A. M. Dollar and H. Herr, "Lower extremity exoskeletons and active orthoses: challenges and state-of-the-art," *IEEE Transactions on robotics*, vol. 24, pp. 144-158, 2008.
- [7] O. Khatib, "Mobile manipulation: The robotic assistant," *Robotics and Autonomous Systems*, vol. 26, pp. 175-183, 1999.
- [8] M. A. Peshkin, J. E. Colgate, W. Wannasupphorasit, C. A. Moore, R. B. Gillespie, and P. Akella, "Cobot architecture," *IEEE Transactions on Robotics and Automation*, vol. 17, pp. 377-390, 2001.
- [9] Q. Zeng, E. Burdet, and C. L. Teo, "Evaluation of a collaborative wheelchair system in cerebral palsy and traumatic brain injury users," *Neurorehabilitation and Neural Repair*, 2008.

- [10] P. Evrard and A. Kheddar, "Homotopy switching model for dyad haptic interaction in physical collaborative tasks," in *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, 2009, pp. 45-50.
- [11] P. Evrard and A. Kheddar, "Homotopy-based controller for physical human-robot interaction," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, pp. 1-6.
- [12] A. Bussy, P. Gergondet, A. Kheddar, F. Keith, and A. Crosnier, "Proactive behavior of a humanoid robot in a haptic transportation task with a human partner," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 962-967.
- [13] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends in cognitive sciences*, vol. 10, pp. 70-76, 2006.
- [14] G. Pezzulo and H. Dindo, "What should I do next? Using shared representations to solve interaction problems," *Experimental Brain Research*, vol. 211, pp. 613-630, 2011.
- [15] L. A. Suchman, *Plans and situated actions: The problem of human-machine communication*: Cambridge university press, 1987.
- [16] J. Shah, J. Wiken, B. Williams, and C. Breazeal, "Improved human-robot team performance using chaski, a human-inspired plan execution system," in *Proceedings of the 6th international conference on Human-robot interaction*, 2011, pp. 29-36.
- [17] F. Stulp, J. Grizou, B. Busch, and M. Lopes, "Facilitating intention prediction for humans by optimizing robot motions," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, 2015, pp. 1249-1255.
- [18] T. Bandyopadhyay, K. S. Won, E. Frazzoli, D. Hsu, W. S. Lee, and D. Rus, "Intention-aware motion planning," in *Algorithmic Foundations of Robotics X*, ed: Springer, 2013, pp. 475-491.
- [19] H. chaandar Ravichandar, A. Kumar, and A. Dani, "Bayesian human intention inference through multiple model filtering with gaze-based priors," in *Information Fusion (FUSION), 2016 19th International Conference on*, 2016, pp. 2296-2302.
- [20] Y. Jiang, H. Koppula, and A. Saxena, "Modeling 3D Environments through Hidden Human Context," 2015.
- [21] V. Dutta and T. Zielinska, "Predicting the Intention of Human Activities for Real-Time Human-Robot Interaction (HRI)," in *International Conference on Social Robotics*, 2016, pp. 723-734.
- [22] W. A. Woods, "Understanding subsumption and taxonomy: A framework for progress," *Principles of Semantic Networks: Explorations in the representation of knowledge*, pp. 45-94, 1991.
- [23] R. C. Schank, "Conceptual dependency: A theory of natural language understanding," *Cognitive psychology*, vol. 3, pp. 552-631, 1972.
- [24] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, pp. 677-695, 1997.
- [25] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1991, pp. 237-242.
- [26] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, and K. Ikeuchi, "Flexible cooperation between human and robot by interpreting human intention from gaze information," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, 2004, pp. 846-851.
- [27] K. Kosuge, M. Sato, and N. Kazamura, "Mobile robot helper," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, 2000, pp. 583-588.
- [28] A. F. Bobick, "Movement, activity and action: the role of knowledge in the perception of motion," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 352, pp. 1257-1265, 1997.
- [29] D. Croft, "Estimating intent for human-robot interaction," in *IEEE International Conference on Advanced Robotics*, 2003, pp. 810-815.
- [30] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby, "Anxiety detecting robotic system-towards implicit human-robot collaboration," *Robotica*, vol. 22, pp. 85-95, 2004.
- [31] A. Bauer, D. Wollherr, and M. Buss, "Human-robot collaboration: a survey," *International Journal of Humanoid Robotics*, vol. 5, pp. 47-66, 2008.
- [32] P. R. Cohen and H. J. Levesque, "Teamwork," *Nous*, vol. 25, pp. 487-512, 1991.
- [33] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 708-713.
- [34] S. Sosnowski, A. Bittermann, K. Kuhlentz, and M. Buss, "Design and evaluation of emotion-display EDDIE," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 3113-3118.
- [35] Y. Hattori, H. Kozima, K. Komatani, T. Ogata, and H. G. Okuno, "Robot gesture generation from environmental sounds using inter-modality mapping," 2005.
- [36] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: improving robot readability with animation principles," in *Proceedings of the 6th international conference on Human-robot interaction*, 2011, pp. 69-76.
- [37] A. G. Brooks and R. C. Arkin, "Behavioral overlays for non-verbal communication expression on a humanoid robot," *Autonomous Robots*, vol. 22, pp. 55-74, 2007.
- [38] M. Faria, A. Costigliola, P. Alves-Oliveira, and A. Paiva, "Follow me: Communicating intentions with a spherical robot," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, 2016, pp. 664-669.

La robotique à l'épreuve des émotions

Du robot jouet au clone robotisé

Jane Bourgninaud
Master CNA
Grenoble INP Phelma
Grenoble, France

Julien Miranda
Master CNA
Grenoble INP Ense3
Grenoble, France

Résumé— Abstract : Dans cet article la question de l'utilité des émotions pour les robots compagnons sera posée, afin d'étudier quelques robots choisis qui représentent des partis-pris dans l'expression des émotions : de Nao et Pepper au Geminoid DK en passant par KOBIAN-R, de nombreux modèles très différents poursuivent un but commun : proposer à l'homme un compagnon d'un nouveau genre, qui saura exprimer des émotions, afin de donner une nouvelle dimension aux interactions homme-machine.

Keywords—Robots sociaux, IA, Intelligence, Compagnon, Emotion.

I. INTRODUCTION

L'expression des émotions par des robots semblables à celles des êtres humains représente un large champ exploratoire dans le domaine de la robotique actuelle. Les enjeux sont à la fois de pouvoir reconnaître et catégoriser les émotions humaines d'une part, ainsi que de pouvoir créer des robots dont la face est dotée d'assez de degrés de liberté et d'un contrôle suffisamment précis pour pouvoir générer des émotions crédibles et facilement reconnaissables.

Cet article a pour objectif de réaliser une vue d'ensemble de quelques travaux qui existent sur le sujet de la transmission des émotions par des robots. Il abordera dans un premier l'intérêt de l'expression des émotions en robotique, ainsi que la classification des émotions couramment utilisée, celle de Paul Ekman. Les robots compagnons, porteurs d'une fibre affectueuse, seront à l'honneur de la seconde partie. Ensuite nous étudierons le robot KOBIAN-R, et surtout l'algorithme d'implémentation utilisé dans (8) pour performer leur étude sur la reconnaissance interculturelle des émotions générées par un robot. Enfin, notre étude se portera sur les robots anthropomorphiques en tentant de répondre à la question posée dès les années 1970 par l' "Uncanny Valley" : Jusqu'à quel point un robot peut-il être ressemblant à l'homme sans que cela ne produise une gêne ?

II. EXPRESSION DES EMOTIONS CHEZ LES ROBOTS SOCIAUX : UTILITE ET METHODE

1. Utilité des émotions pour des robots sociaux

Les robots sociaux ont commencé à apparaître autour des années 1930. Humanoïdes, zoomorphiques ou anthropomorphiques, quelle que soit leur morphologie, il s'agit de robots capable d'analyser et de reproduire le comportement humain.

L'intégration de l'expression des émotions dans ce type de structure est essentielle notamment parce que les hommes ont tendance à communiquer avec elle de la même façon qu'ils le feraient avec une autre personne (2). Il est donc primordial que le robot puisse s'exprimer de façon non verbale à l'image de l'être humain.

En outre, les émotions produites par un robot peuvent avoir trois intérêts pour le constructeur (2). Elles facilitent la croyance en l'interaction humain-robot, indispensable pour des robots sociaux. Elles permettent de produire à l'utilisateur un retour (feedback) sur l'état actuel de robot, ses objectifs ainsi que ses intentions. Enfin, elles permettent un contrôle des mécanismes de la machine.

En effet, concernant ce dernier aspect, les émotions peuvent permettre de rendre compte du mode de comportement dans lequel se situe le robot mais aussi de déclencher l'apprentissage ou l'adaptation. Pour les robots à morphologie zoomorphique par exemple, une des approches est d'utiliser les mimiques animales connues comme la demande de nourriture, la reconnaissance du danger (peur), comme indice de l'état du robot régulant ainsi ses demandes et son comportement.

2. Utilité du langage corporel, des mouvements et des gestes

Le langage corporel humain est utilisé de façon récurrente, consciente ou non, pour acquérir des informations supplémentaires concernant l'expression des émotions des êtres humains. Il se constitue d'un certain nombre de gestes et de mouvement qui s'additionnent pour former le langage non verbal.

Brezel et Fitzpartick (6), ont montré que les hommes percevant l'ensemble des gestes comme un apport sémantiquement riche qu'on leur demande d'y prêter attention ou non. Le regard ainsi que la direction du corps sont notamment deux indices qui renseignent sur le lieu d'attention de la personne.

Le langage non verbal est également primordial dans le cas de la réalisation de robots zoomorphiques. L'intérêt de ces robots est de parvenir à imiter de façon réaliste les animaux domestiques que peuvent être le chien ou le chat par exemple de façon à établir une relation de l'homme à l'animal. Ce type de robot est souvent appelé "robot compagnon". L'interaction de l'homme avec ce genre de robot est plus facilement modélisable car beaucoup plus simple que les interactions existantes entre deux êtres humains.

Pour les robots humanoïdes, une utilisation réaliste des expressions faciales constitue la première étape pour parvenir à transmettre une émotion de façon non verbale.

3. Catégorisation des expressions faciales

Les émotions sont des concepts humains dont chacun peut faire l'expérience au cours de sa vie. Elles sont exprimées de différentes manières par les acteurs du corps, porteurs du langage non verbal. Parmi eux, les expressions faciales restent des indices majeurs de la détection des émotions.

En 1972, Paul Eckman propose une classification des émotions primaires (1) basée sur une étude réalisée sur des colonies en Papouasie Nouvelle Guinée. Selon lui, six émotions : la peur, le dégoût, la tristesse, la joie, la surprise et la colère sont des émotions universelles reconnues par tous, quel que soit les cultures, à travers des indices faciaux.

En 1978, aidé par Wallace Friesen, il développe le FACS, "Facial Action Coding System", qui permet l'encodage des émotions en fonction des différents degrés de liberté présents dans le système facial humain. Il repose sur la répertoriabilité de 46 unités d'action, qui correspondent chacune à la détente ou la contraction d'un ou plusieurs muscles menant à une action faciale spécifique. Ils sont numérotés de 1 à 46. Chaque émotion est exprimée systématiquement à l'aide d'un ensemble d'unités d'action spécifiques. Par exemple, la joie est exprimée à l'aide d'une remontée des joues (UA 6) et d'un étirement du coin des lèvres (UA12). Dans le cas de la description des émotions, ces actions peuvent être pondérées par une lettre entre A et E exprimant l'intensité de l'action, A étant l'intensité minimale. D'autres unités d'actions sont également prises en compte, treize d'entre elles concernent les mouvements de la tête, et sept d'entre elles concernent le mouvement des yeux.

Plusieurs autres approches ont pu être proposées (2) pour permettre la catégorisation des émotions notamment une vision qui implique une échelle continue basée sur l'excitation et la valence (3). D'autres propositions tendent à combiner la vision segmentée de Paul Eckman avec une approche continue.

L'ensemble des concepts proposés par Paul Eckman montre qu'il est possible de catégoriser les différentes émotions, notamment les six émotions primaires, en utilisant les différents degrés de liberté du visage. En robotique, cela suggère la réalisation de robots humanoïdes capables d'arranger la disposition de leurs traits, de façon à exprimer diverses émotions.

4. Evolution des expressions faciales pour des robots humanoïdes

Les expressions faciales réalisées par les robots humanoïdes d'aujourd'hui manquent souvent de réalisme. En effet, elles sont limitées par le design mécanique des systèmes ainsi que leur contrôle. Les transitions entre deux expressions distinctes sont notamment souvent abruptes ce qui contraste avec les expressions effectivement réalisées par les êtres humains.

Les entités primaires utilisées pour exprimer ces émotions sont la bouche et les lèvres, les joues, les yeux, les sourcils et le front. Le réalisme des émotions chez des robots humanoïdes dépend souvent du nombre de ces entités ainsi que du nombre de degrés de liberté que le constructeur leur octroie.

Le robot "Sparky" par exemple (4), possède trois entités : les lèvres, les yeux et les sourcils, ce qui lui permet d'exprimer un faible nombre d'expressions relativement basiques. Tandis que le robot Felix (5), qui est un robot dont la structure a été réalisée à base de LEGO, possède uniquement deux lèvres et deux sourcils, mais il est capable d'exprimer les six émotions universelles ainsi qu'un certain nombre de combinaisons. En comparaison, le robot Kimset (2), possède quinze activateurs, et la plupart d'entre eux sont capables de travailler conjointement pour afficher une émotion particulière.

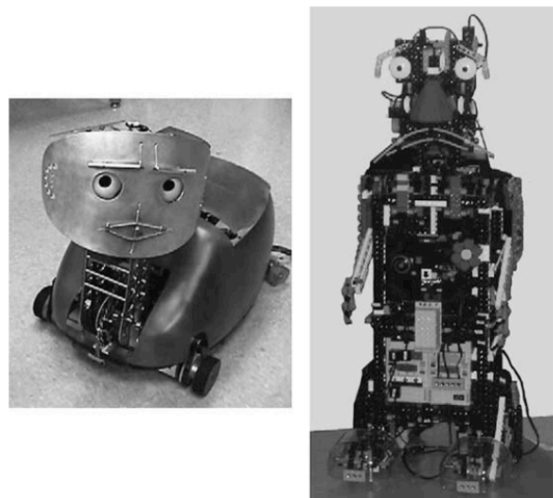


Figure 1 : Les robots Sparky (à gauche) et Felix (à droite)

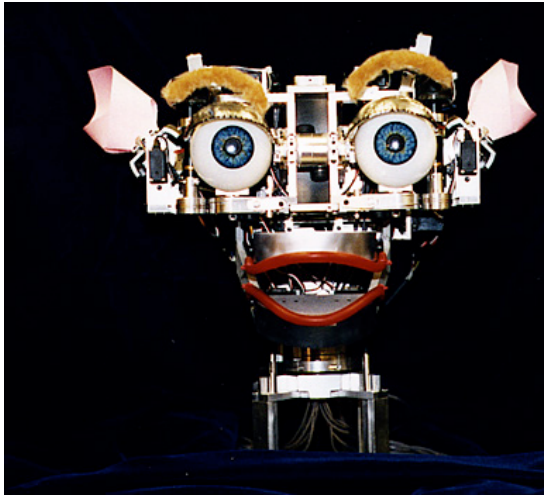


Figure 2 : Le robot Kimset

D'autres robots encore, possèdent des artifices relatifs à l'être humain, comme des dents, de la peau ou des cheveux qui les rendent particulièrement réalistes. Néanmoins, le problème de la variation abrupte des activateurs entre deux émotions demeure sur ce genre de produit, rendant leur réalisme tout à fait relatif lorsque le visage est en mouvement.

III. LES ROBOTS COMPAGNONS : DE L'UTILE VERS L'AFFECTIF

Il existe d'ores et déjà des robots compagnons (commercialisés ou en voie de l'être). Ces robots ne disposent pas nécessairement de moyens d'expression des émotions techniquement complexes, via par exemple de nombreux degrés de liberté sur le visage, mais sont conçus pour dégager une impression globale de bienveillance vis-à-vis de leurs interlocuteurs humains. Ainsi, de tels robots ont vocation à évoluer dans un environnement composé d'un ou plusieurs humains, avec lesquels il doivent alors échanger. La communication non verbale étant primordiale dans tous types d'échange, ces robots se doivent d'adopter des attitudes cohérentes afin d'éviter d'être rejetés dans un premier temps, puis de susciter une certaine confiance. Dans notre cadre nous étudierons en particulier deux robots capable d'exprimer, dans une certaine mesure et sans disposer de nombreux degrés de libertés au niveau du visage, des émotions interprétables par un humain.

A travers ces robots se construit un défi de taille, marqué par un parti pris fort : il s'agit d'exprimer des émotions, de communiquer au-delà du verbal, sans pour autant copier l'apparence humaine. Ainsi, les robots Nao et Pepper conçus par l'entreprise Française Aldebaran Robotics (rachetée depuis

et désormais renommée SoftBank robotiks), et leur concurrent Japonais Asimo, sont les symboles de robots compagnons capables de traduire un certain degré d'émotion d'une façon novatrice : ni animaux de compagnie, ni interlocuteurs humains, ils s'érigent comme une alternative intéressante à de nombreux égards, ouvrant la voie à de nouvelles perspectives pour le conseil, l'accueil ou l'accompagnement des personnes. En effet, ces robots présentent l'avantage vis-à-vis d'un animal de compagnie de partager de nombreuses choses avec l'être humain, inaccessible à ce dernier : ils sont conçus pour pouvoir effectuer des tâches d'ordinaire réservées aux hommes, et sont dotés de paroles, ce qui les rend plus proches sur certains aspects cognitifs. S'ils n'égalent pas l'être humain dans ses capacités, ils en évitent également certains écueils : en particulier, ils peuvent se montrer d'excellents confidents car ils sont dépourvus de la fâcheuse tendance de l'homme à juger son prochain.

1. L'apparence

Ainsi cette catégorie de robot ne cherche pas à faire illusion, et personne ne pourrait s'y tromper : il ne s'agit pas d'une copie robotisée d'un être humain, malgré le caractère humanoïde qu'ils partagent (deux bras, deux jambes, et une tête focalisant l'attention). En effet, leur apparence se rapproche plus de personnages de dessins animés pour enfant : les lignes courbes arrondies, au niveau du corps et de la tête, les yeux disproportionnés, les couleurs douces (dominance du blanc) très éloignées de la peau humaine, leur petite taille (58 cm pour Nao, environ 120 cm pour Pepper et Asimo), sont autant d'éléments pensés pour prodiguer à ces robots un caractère amical et bienveillant antérieur même à toute interaction à proprement parler, en particulier pour les enfants et les personnes âgées, qui représentent un segment de marché très prometteur, car c'est au travers de ces populations que l'arrivée, quasi-inéluctable des robots dans notre vie quotidienne devrait s'amorcer.

2. Caractère affectif

Néanmoins, Nao se différencie de son concurrent Asimo par son orientation résolument portée sur l'affect et l'émotion. En effet, alors qu'Asimo est doté de nombreuses fonctionnalités pratiques, Nao et Pepper sont totalement centrés sur le partage d'émotion et la création d'une relation affective avec leur interlocuteur. Ainsi les différentes descriptions de ces deux robots compagnons véhiculées par leurs concepteurs sont construites autour du champ lexical de la bienveillance : il ne s'agit plus d'un objet technologique dont on pourrait froidement détailler les caractéristiques techniques, mais d'un compagnon adapté et adaptatif, capable d'œuvrer pour le confort et le bien-être de son interlocuteur.



Figure 3 : Robot Nao avec une personne âgée

Cette nouvelle voie tracée par les deux compagnons s'inscrit en totale opposition (peut-être de manière paradoxale) avec la méfiance suscitée par les nouvelles technologies numériques en général et l'intelligence artificielle en particulier. En effet, la raison d'être de ce nouveau type de robots, s'éloignant de la plupart des images véhiculées par la science-fiction et fantasmées par l'opinion, n'est pas la performance calculatoire, ni le dépassement des capacités de l'être humain, mais bien la cohabitation, la complémentarité, et l'accompagnement de celui-ci en toute quiétude. C'est ainsi que peut s'instaurer ce qu'on pourrait appeler une relation de confiance, basée sur la capacité de ces robots à exprimer ce que l'humain perçoit comme des émotions auxquelles il associe une dimension affective. De cette confiance naît alors une certaine complicité entre l'humain et le robot, qui tient moins aux tâches effectuées par la machine qu'à sa manière de les réaliser.



Figure 5 : Robot Pepper dans un environnement d'interaction

3. Mouvement, gestuelle : quelles capacités pour un robot compagnon ?

Au-delà l'aspect purement esthétique, les capacités de mouvements de ces robots se doivent de lui permettre d'agir dans l'environnement au moins aussi bien qu'un animal de compagnie (déplacement, interaction avec l'environnement), tout en étant suffisamment fluides et « naturels » pour ne pas briser l'impression de bienveillance chaleureuse voulue. C'est dans le cadre de cette double exigence que se place dont Nao.

On remarque que, au-delà de toute la connectivité qui en fait un compagnon pratique au demeurant, de nombreux degrés de libertés sont dédiés au mouvement des jambes et des bras. En robotique la problématique d'un robot bipède est complexe comparée aux robots roulants, c'est néanmoins le choix qui a été fait pour Nao, contrairement à Pepper, qui peut donner l'illusion d'être dotée de jambes mais ne se déplace pas en marchant.

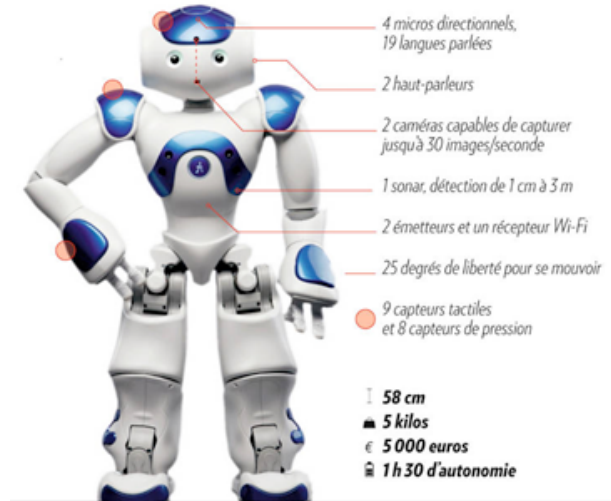


Figure 4 : Détails du robot Nao

IV. LE ROBOT KOBIAN-R

Le robot KOBIAN-R (7) est une adaptation du robot KOBIAN qui est un robot humanoïde bipède capable de produire des expressions faciales. La tête du robot contient 24 degrés de liberté visible sur la figure 7. La taille de cette tête est similaire à celle d'une femme adulte et son poids est de 1,7 kg. Le front est également doté d'une technologie basée sur l'utilisation de diode électro lumineuse permettant de changer sa couleur. Cependant, il est d'une épaisseur assez fine pour permettre le mouvement des sourcils à l'aide d'une technologie magnétique.

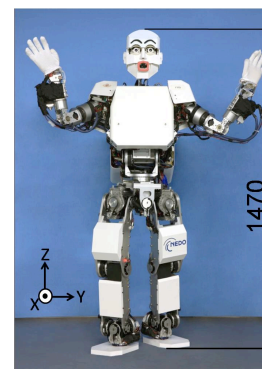


Figure 6 : Le robot KONIAN-R

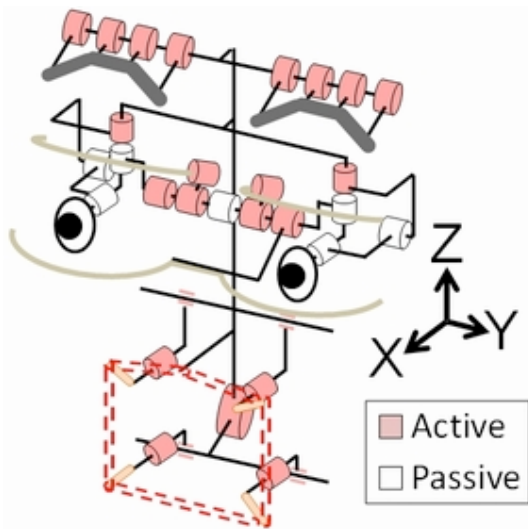


Figure 7 : Kobian - R, Représentation des 24 degrés de liberté de la face du robot

1- Générer des expressions faciales

Une façon d'implémenter ce genre de robot est décrite dans l'étude (8). Comme tout robot, pour fonctionner il nécessite une entrée, une sortie et une équation permettant de passer de l'une à l'autre. Dans la configuration proposée par l'auteur, l'entrée est un vecteur comportant les informations liées à l'émotion que l'on souhaite réaliser et la sortie est un ensemble de variables correspondant aux angles à adopter par les différents degrés de liberté du robot.

L'état émotionnel que l'on veut reproduire constitue la donnée d'entrée. Ils sont donc classifiés à l'aide d'une version améliorée de la roue des émotions de Plutchik présentée sur la



Figure 8 : Roue des émotions

figure 8 . Cette roue décrit aussi bien les émotions que les actes de communication.

La sortie du système constitue l'ensemble des angles correspondant aux degrés de liberté du robot équivalent à une réplique émotionnelle. Comme KOBIAN-R est capable de produire un très grand nombre de combinaison, il a fallu déterminer un ensemble de configurations pertinentes. Pour cela, les auteurs se sont appuyés sur les données présentées dans le FACS, tout en maximisant le nombre de configuration et en minimisant le pour chacune, le nombre de variable impliquées.

Pour établir la relation entre l'entrée et la sortie du système, il convient de déterminer les corrélations qui existent entre ces deux ensembles de données. L'état émotionnel est donc converti en un vecteur à 6 composantes, une pour chaque partie du visage et une pour le nez. Un polynôme de degrés trois est utilisé pour catégoriser l'état émotionnel. Les données sont ensuite classifiées à l'aide d'un linéaire de Fisher. La sortie de ce classifieur constitue la réplique émotionnelle souhaitée. A l'aide de la méthode du plus proche voisin, cette réplique est rapprochée de la réplique robotique la plus semblable contenant les angles réels réalisables par le robot.

Enfin, l'asymétrie par rapport à la ligne médiane du visage est un facteur déterminant dans l'expression des émotions prise en compte dans ce modèle. Elle est introduite moyennant deux valeurs de seuil, Theta U (Upper) et Thêta L (lower). Si la valeur absolue d'une des composantes du vecteur d'entrée est comprise entre ces deux valeurs de seuils, une notification est activée. Pour chaque partie du visage, une notification activée entraîne le calcul de deux vecteurs de sortie (un pour chaque hémisphère).

2- Expressions faciales et facteur culturel

Les auteurs n'ont pas cherché simplement à implémenter le système du robot KOBIAN-R pour vérifier son fonctionnement, le but de leurs investigations était de générer les expressions faciales dans des versions différentes selon les différentes cultures.

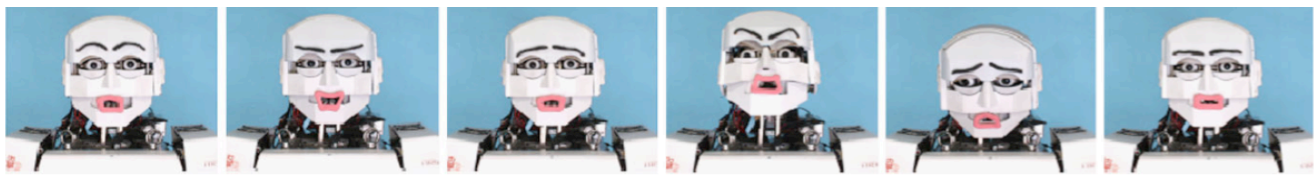


Figure 9 : Quelques exemples de simulation. De gauche à droite : admiration, malice, appréhension, reproche, remord, ennui.

Cet aspect de la recherche entre en considération dans l'implémentation au niveau de la conversion des états émotionnels en vecteurs. En effet, le vecteur représente la position des 6 caractéristiques de la face du robot, il faut donc préalablement déterminer comment traduire l'état émotionnel à travers les positions des attributs de l'homme que sont le nez, la bouche, les yeux et les sourcils. Ce qui équivaut à faire la relation entre une émotion exprimée sur un visage et son interprétation. Pour ce faire les auteurs ont utilisé deux tables de classification : les séries d'entraînement ouest et japonaise (pour l'Europe et l'Asie).

Ces deux séries existent sous la forme de tableaux qui associent à chaque état émotionnel, l'expression correspondante. Pour la culture européenne, (la série ouest), les corrélations ont été réalisées en s'appuyant sur diverses études. Cependant, pour établir les corrélations relatives à la culture asiatique, les auteurs ont demandé à des illustrateurs asiatiques de réaliser des dessins exprimant les états émotionnels en question. Ils ont extrait les données de ces dessins.

Pour mener leurs expériences à bien, les auteurs ont proposé à des participants de toutes nationalités confondus de reconnaître les expressions faciales du robot via un site internet.

Durant la phase préliminaire, les auteurs ont voulu vérifier la reconnaissance d'expressions complexes provenant de la série de test relative à la culture européenne. Ils ont proposé aux internautes de d'assigner à des photos du robot une émotion parmi les six proposées. Une photo d'une expression neutre du robot était également montrée. Les expressions étaient complexes. Aucune des six émotions de base que sont la joie, la colère, la surprise, le dégoût, la tristesse et la peur n'ont été utilisée car elles ont servies lors de l'apprentissage du robot. Les résultats de reconnaissance pour les égyptiens sont regroupés dans le tableau 1.

Les résultats de cette première approche montrent que le taux de reconnaissance moyen des émotions et de 68,8%. Il est plus faible lorsqu'il s'agit de reconnaître des actes de communication. On pouvait en effet s'attendre à ce que plus les émotions à transmettre soient complexes, plus leur reconnaissance est difficile et culturellement dépendante. La figure 10 montre que la reconnaissance de certaines émotions diffère selon les cultures. Les égyptiens peinent à reconnaître l'appréhension, alors que les asiatiques sont en défaut sur la reconnaissance de la malice. La conclusion des auteurs quand à cette étude est que l'interprétation d'une expression complexe peut être variées et culturellement dépendante. Ce phénomène est d'autant plus présent que l'expression est complexe. D'où la nécessité de développer un modèle prenant

en compte les différences culturelles dans la production d'expressions faciale.

Expression	Taux de reconnaissance moyen	Expression	Taux de reconnaissance moyen
Appréhension	7,7%	Soulagement	30,8%
Ennuie	65,4%	Malice	38,5%
Amour	50,0%	Incrédulité	42,3%
Admiration	46,2 %	Gratitude	7,7%
Remord	73,1%	Pitié	32,3%
Espoir	38,5%	Reproche	84,6%

Tableau 1 : Reconnaissance des expressions par les sujets égyptiens

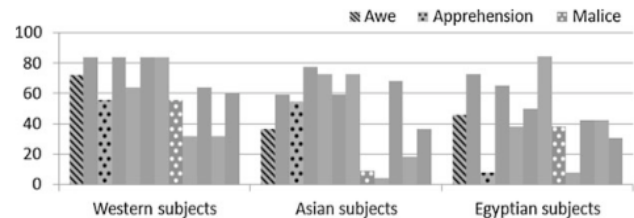


Figure 10 : Pattern de la reconnaissance des expressions en fonction des groupes culturels. Mise en évidence des trois expressions dont le taux de reconnaissance est bas dans un groupe et haut dans les deux autres

Dans une deuxième étude, les auteurs ont cherché à étudier l'effet de l'asymétrie sur la reconnaissance des émotions. Ils ont donc proposé aux internautes l'exercice suivant : Dans la première partie, les participants devaient choisir la photo la plus appropriée parmi les deux proposée pour décrire l'émotion proposée parmi l'ennui, le dégoût, l'incompréhension et le désaccord. Une de ces photos présentait des asymétries, l'autre non. Dans une deuxième partie, les auteurs cherchaient à étudier l'effet des asymétries sur la reconnaissance de la joie. Ils ont donc demandé de choisir entre deux images (l'une asymétrique, l'autre non) celle correspondant le mieux à cette émotion. Cinq asymétries étaient proposées, une pour chacune des caractéristiques du visage (sourcils, paupière, bouche).

Les résultats de cet exercice tendent à montrer que l'asymétrie est utilisée pour caractériser des expressions de valence négatives et cet aspect est observable sans différence culturelle notable.

3- Importance des degrés de liberté dans la réalisation des expressions faciales.

Dans cette étude, les auteurs ont cherché à évaluer les différences culturelles qui peuvent survenir quant à la reconnaissance d'expressions faciales complexe. Pour ce faire, ils ont utilisé le robot KOBIAN-R qu'ils ont implémenté avec un algorithme complexe. Cette étude nous montre que la réalisation d'expressions complexes nécessite une précision dans le positionnement des attributs de la face du robot. Les 24 degrés de liberté du robot KOBIAN-R sont donc un atout majeur pour atteindre ce but. Néanmoins, ils peuvent également permettre de réaliser des expressions irréalistes, qui n'auraient pas leur pendant humain. C'est pourquoi le contrôle de ces degrés de libertés par un algorithme précis et performant est également un point essentiel pour performer cette étude.

V. REPRODUIRE L'APPARENCE HUMAINE : UN PAS VERS L'EMOTION ?

Les différents robots présentés précédemment réussissent à susciter l'émotion grâce à un aspect visuel, à des gestes, à une attitude. Une question se pose alors : si un robot avait une l'apparence d'un humain, pourrait-il alors exprimer des émotions de manière plus pertinente encore, ou s'agit-il seulement de la quête du mythe de Pygmalion rendue atteignable par les avancées technologiques ?

De nombreux scientifiques s'intéressent à la création de robots Androïdes ressemblant à l'homme à s'y méprendre. Visuellement impressionnants, ces robots qui présentent pour caractéristiques principales une peau extrêmement réaliste et des cheveux et yeux aussi vrais que nature. L'un des derniers modèles de ce type provient du Danemark : dans [11], les chercheurs s'intéressent à l'influence de la peau du robot, ultra réaliste. Ils démontrent que la sensation haptique familière (absente sur les robots jusqu'ici présentés) joue un rôle dans l'établissement d'une relation de confiance (primordiale pour les robots compagnons). La sensation du toucher est liée à la fois à des processus conscients et inconscients, dans l'interaction de l'homme avec le robot, et peut jouer un rôle émotionnel fort. Ainsi ces nouveaux robots, à l'apparence trompeuse se rapprochent de plus en plus de l'homme dont ils deviennent plus difficilement différenciables.

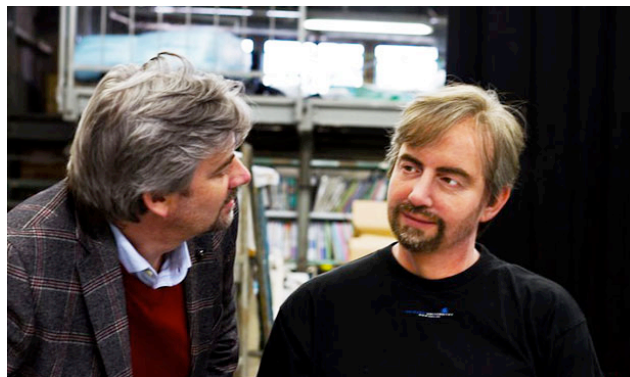


Figure 11 : Geminoid DK : un robot plus vrai que nature

Cependant c'est précisément cette ressemblance accrue qui donne plus de poids aux différences qui demeurent : cet effet est appelé « Uncanny Valley » (en français « la vallée de l'étrange » depuis l'apparition de cette expression en 1970 sous la plume de Masahiro Mori [12]). Plus les robots ressemblent à l'homme, plus ils en deviennent effrayant. Cet écueil, évité par les robots introduits plus haut (qui se distinguent tous clairement d'un être humain réel), est potentiellement un frein à l'établissement d'une relation de confiance entre l'homme et un robot compagnon qui lui ressemble et c'est ainsi que de telles considérations sont sujettes à un vif regain d'intérêt au vu des prouesses techniques réalisées dans ce domaine.

La description de ce phénomène [13] repose sur le fait que contrairement à beaucoup de processus de la vie courante, l'affinité ressentie par une personne n'est pas modélisable par une fonction monotone. En effet, si un humain sera plus à l'aise avec un robot bipède possédant deux bras qu'avec un robot très éloigné de l'apparence humaine, pousser la ressemblance beaucoup plus loin conduit non pas à une confiance accrue, proche de celle suscitée par un être humain réel, mais à une méfiance prononcée. Le terme de « vallée » correspond au creux dans la représentation de l'affinité fonction de la ressemblance, schématisée ci-dessous :

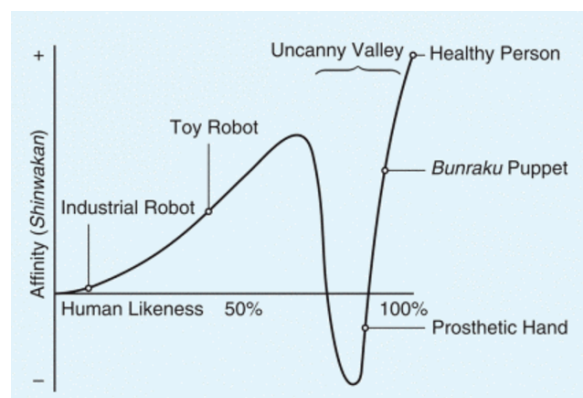


Figure 12 : Schématisation de la relation entre ressemblance et affinité provoquée

Au-delà de la simple esthétique, la notion de mouvement est également d'une grande importance dans l'optique de l'expression des émotions, comme cela a été évoqué au travers d'autres exemples, que ce soit dans l'instauration d'un climat propice à la réception des émotions ou dans l'expression des émotions en elles-mêmes. Ainsi la vallée de l'étrange se dédouble et le creux s'accroît : c'est toute la problématique des microexpressions faciales par exemple qui se retrouve dans la méfiance du trop réaliste :

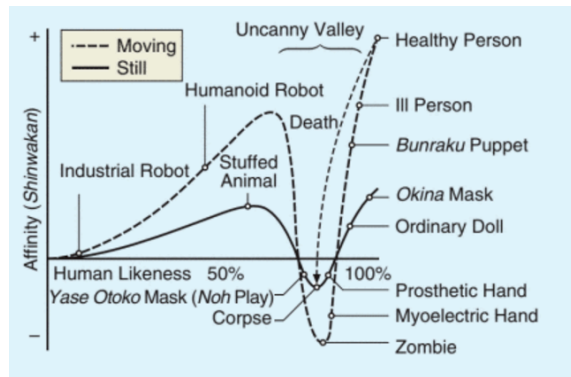


Figure 13 : Les mouvements accentuent la « Uncanny Valley »

Dans la quête des émotions, la ressemblance extrême n'est donc pas salutaire tant qu'elle n'évite pas le piège du rejet suscité par les entités si proches de l'homme qu'elles en deviennent terrifiantes.

VI. CONCLUSION

A la lumière des considérations précédentes, il apparaît clairement que l'expression des émotions par les robots est un domaine de recherche très actif tant il ouvre des perspectives nouvelles et répond à un réel besoin industriel. Ainsi, si les robots sont capables d'exprimer de plus en plus d'émotions au niveau du visage, il est à noter que, à l'image des dessins animés qui les traduisent sans volonté de réalisme, ces émotions peuvent être véhiculées par d'autres biais. En particulier, les robots compagnons sont en passe de réussir le pari d'instaurer avec un humain un climat de confiance basé sur une forte dimension affective, et ce sans avoir recours à de complexes mécanismes de reproduction d'expressions faciales nécessitant de nombreux degrés de liberté. L'évolution vers une prépondérance du caractère affectif et émotionnel des

robots, marquée par ces deux pistes importantes, et exacerbée par la quête quasi-mystique du robot indissociable de l'homme, (alors que les dernières avancées tendent à prouver que la clef des émotions ne réside pas dans le mimétisme) marque l'avènement de nouveaux robots s'écartant de leur racine étymologique (robotika signifie travail, labeur en tchèque). Ces problématiques sont également particulièrement intéressantes dans le cadre de travaux tels que ceux de Damasio [11] : l'esprit, la conscience et l'intelligence ont besoin des émotions pour fonctionner, et l'expression des émotions par les robots sera ainsi une étape cruciale dans la quête d'une intelligence artificielle 'forte', sans quoi l'homme sera condamné à ne créer que des Phineas Gage automatisés.

REFERENCES

- [1] Ekman, P. (1999). Basic Emotions In T. Dalgleish and T. Power (Eds.) *The Handbook of Cognition and Emotion* Pp. 45-60.
- [2] Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3), 143-166.
- [3] Thorpe, W. H. (1956). Learning and instinct in animals.
- [4] Scheeff, M. (2000). Experiences with Sparky, a Social Robot, in proceedings of the Workshop on Interactive Robotics and Entertainment (WIRE-2000). K. Elissa, "Title of paper if known," unpublished.
- [5] Cañamero, L., & Fredslund, J. (2001). I show you how I like you-can you read it in my face?[robotics]. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, 31(5), 454-459.
- [6] Breazeal, C., & Fitzpatrick, P. (2000, November). That certain look: Social amplification of animate vision. In *Proceedings of the AAAI fall symposium on society of intelligence agents—the human in the loop*.
- [7] Kishi, T., Otani, T., Endo, N., Kryczka, P., Hashimoto, K., Nakata, K., & Takanishi, A. (2012, October). Development of expressive robotic head for bipedal humanoid robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4584-4589). IEEE.
- [8] Trovato, G., Kishi, T., Endo, N., Zecca, M., Hashimoto, K., & Takanishi, A. (2013). Cross-cultural perspectives on emotion expressive humanoid robotic head: recognition of facial expressions and symbols. *International Journal of Social Robotics*, 5(4), 515-527
- [9] Cassou-Noguès Pierre, « Le robot qui vous veut du bien », *Multitudes*, 1/2015 (n° 58), p. 210-215
- [10] A. Damasio. *L'erreur de Descartes* (1995)
- [11] Elizabeth G. Dougherty,, Henrik Scharfe, Initial Formation of Trust: Designing an Interaction with Geminoid-DK to Promote a Positive Attitude for Cooperation, Volume 7072 of the series Lecture Notes in Computer Science pp 95-103
- [12] M. Mori, "The uncanny valley", *Energy*, vol. 7, no. 4, pp. 33-35, 1970.
- [13] M. Mori, K. F. MacDorman and N. Kageki, "The Uncanny Valley [From the Field]," in *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98-100, June 2012.

Les Comportements non-verbaux dans les interactions Homme-Machine

Roman Piekut

Master 2 Recherche en Sciences Cognitives

INP Grenoble

FRANCE

Roman.Piekut@grenoble-inp.org

Bertrand Metral

Master 2 Recherche en Sciences Cognitives

INP Grenoble

FRANCE

Bertrand.Metral@grenoble-inp.org

Abstract — Les comportements non-verbaux contribuent fortement aux interactions humaines. Si l'on souhaite entreprendre un développement de la robotique au point où des interactions homme-machine seraient courantes, il semble essentiel de mesurer si oui et comment, des comportements non-verbaux peuvent contribuer à cette interaction. La complexité des indices et signes impliqués dans la communication humaine, affectant la difficulté de leur programmation robotique, relève du fait que l'interaction se situe simultanément à différents niveaux. En outre, la problématique des comportements non-verbaux ne relève pas uniquement de leur production à l'attention de l'interlocuteur, mais aussi de ceux émis par l'interlocuteur et de leur interprétation. Nous verrons ainsi comment ces comportements peuvent affecter la perception des suggestions du robot, mais aussi leur influence sur l'interaction. Enfin, nous aborderons la dimension contextuelle, et comment celle-ci peut conditionner la prise en compte des informations par l'agent robotique.

Keywords — *comportements non-verbaux ; interaction ; persuasion ; structure sociale ; robotique*

I. INTRODUCTION

Les signes non-verbaux forment avec la parole des combinaisons non-congruentes, c'est-à-dire que les gestes sont globaux et synthétiques, tandis que la parole est un système sémiotique organisé. Les gestes accompagnent souvent la parole, qu'ils précisent partiellement ou complètent (Ekman et Friesen [1]). Ils sont impliqués dans la cohésion du discours voire parfois aident à l'organiser (Goldin-Meadow [2]). Le corps peut faire passer un message de nature différente que les mots que l'on prononce. Alors que le sens des mots est souvent relativement explicite, celui des gestes peut-être implicite et témoigner d'une très grande variabilité culturelle (Hall [3]). Les individus ne réalisent pas obligatoirement les nombreuses informations véhiculées et leur traitement échappe, pour une grande part, à la conscience. La communication non verbale ajoute une dimension supplémentaire au message, parfois en contradiction avec le langage parlé. Ainsi, au regard des nombreux signes intentionnels et indices contingents qui constituent le comportement non-verbal, il convient de s'interroger sur les conséquences que leur prise en compte peut avoir sur l'interaction entre un individu et un agent robotique.

II. SIGNES NON-VERBAUX ET CONTEXTE COMMUNICATIF

Dans son article, Mavridis [4] a décrit les différents indices verbaux, mais surtout non verbaux dont la maîtrise contribuerait à la tâche à accomplir par l'agent robotique. On retrouve notamment la problématique du regard dans l'interaction entre un individu et l'agent robotique. Cet indice non-verbal semble être particulièrement important lorsque l'agent robotique doit participer activement à une interaction, dans le but d'accomplir un objectif avec un individu. La gestion du regard relève aussi d'une certaine importance selon l'auteur, lorsqu'il s'agit de gérer une situation à plus de deux interlocuteurs : le regard semble alors primordial dans la gestion des tours de parole. Dans une problématique similaire, Mavridis souligne l'importance des indices dits « backchannel ». Ces indices contribuent à confirmer au locuteur (i. e., un individu) l'intérêt que l'auditeur (i. e., l'agent robotique) porte à sa communication. Enfin, un autre type d'indices non-verbaux, les gestes déictiques, est présenté par l'auteur. Ces gestes constituent un pointage vers l'interlocuteur, vers un objet présent dans la situation, ou encore vers soi-même. Ils indiquent la direction vers laquelle se trouve le référent. Ils se différencient ainsi d'autres types de gestes co-verbaux, comme les gestes représentationnels, performatifs, discursifs, interactifs, énonciatifs ainsi que les gestes dit de cadrage. Ces actions ont généralement un but illustratif, permettant de faciliter la communication en ajoutant de manière plus ou moins volontaire des informations au discours. Il apparaît essentiel d'implémenter une bonne maîtrise de l'utilisation de gestes déictiques pour accroître la persuasion de l'agent robotique vis-à-vis de l'individu. Ceci est particulièrement important lorsque l'agent robotique a pour fonction d'indiquer une direction (e.g. dans un service d'accueil).

Que ce soit pour un service d'ordre privé ou public, une des fonctions assignées à l'agent robotique est d'assister le sujet humain. L'article de Chidambaram, Chiang et Mutlu [5] étudie l'influence que les signes non verbaux peuvent avoir dans la persuasion d'un agent robotique lors de son interaction avec un sujet. La tâche durant laquelle avait lieu l'interaction était une résolution de problème : « the Desert Survival Problem ». Dans cette tâche, le sujet devait organiser de manière

hiérarchique, les matériels dont il pensait avoir le plus besoin s'il se trouvait dans un désert. L'interaction de l'agent robotique avec le sujet était constituée de suggestions prodiguées par le robot, sur le classement des objets effectué par le sujet. Ces suggestions étaient formulées *a posteriori* et allaient à l'encontre de la décision du sujet. Selon les conditions expérimentales, le robot produisait ou non des indices non verbaux. Ceux-ci pouvaient être vocaux (i. e., changements de ton dans l'expression verbale de l'agent) ou bien corporels (i. e., proximité, regard et gestuelle). Dans cette étude, les auteurs ont testé 3 hypothèses. La première prédisait que les participants seraient plus réceptifs aux suggestions de l'agent robotique si celui-ci produisait des indices non-verbaux (vocaux ou corporels) que lorsqu'il n'en produisait pas. La seconde prédisait que les participants seraient plus réceptifs si l'agent produisait des indices non-verbaux corporels par rapport à des indices non-verbaux vocaux. La troisième prédisait qu'un effet de genre devrait se manifester : les femmes seraient plus sensibles aux indices non-verbaux corporels que les hommes. Les résultats de cette étude semblent confirmer les deux premières hypothèses. Les participants étaient effectivement plus sensibles aux suggestions de l'agent robotique, lorsque des indices non-verbaux étaient utilisés, et cet effet était plus important pour les indices corporels plutôt que vocaux. Cependant, la manipulation expérimentale met en place des situations fortement contrastées. En effet, l'utilisation ou l'absence d'indices non-verbaux, mais aussi l'opposition entre indices vocaux ou corporels, ne permettent pas de cibler précisément la contribution de ces différents indices. Plus particulièrement, les indices corporels regroupent 3 types d'indices (i. e., proximité, regard et gestuelle) sans pour autant préciser dans quelle proportion ces indices participent à la qualité persuasive des suggestions de l'agent robotique.

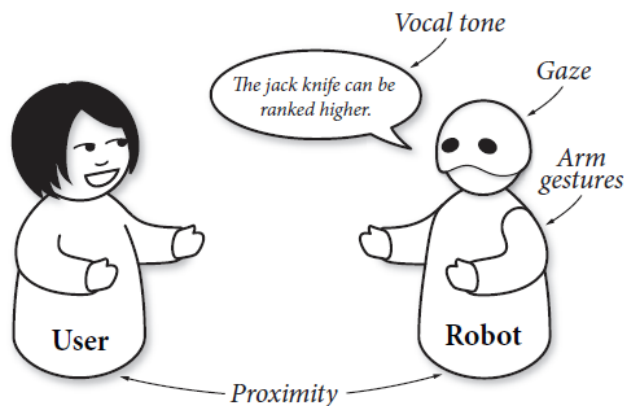


Fig. 1. Interaction humain robot telle que présentée dans [5], présentant les différents comportements non-verbaux d'une certaine importance.

Cet article suggère donc qu'en interagissant efficacement avec un sujet, l'agent robotique parviendrait avec crédibilité à accomplir son rôle, en réduisant le risque que le sujet en

interaction ignore les compétences programmées de l'agent robotique.

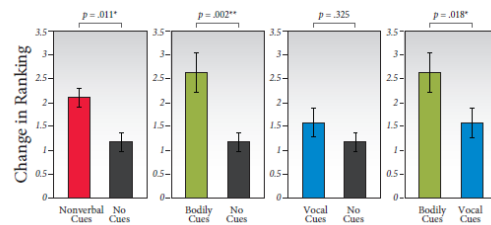


Fig. 2. Résultats présentés dans [5], relatifs au changement du classement des objets dans le sens des suggestions du robot, en fonction des conditions (plus la valeur est élevée, plus le sujet se conforme aux suggestions).

Il est donc pertinent de déterminer, parmi ces indices non-verbaux corporels, lesquels seraient les plus à même de participer à la capacité du robot à interagir efficacement avec son interlocuteur.

Ces articles [4] et [5] soulignent l'importance que peut revêtir l'utilisation appropriée des signes non verbaux dans un contexte communicatif à des fins de persuasion. Un agent tel que le robot doit pouvoir, malgré la considération portée par l'individu sur cet agent, être à même de persuader son interlocuteur. Les indices non-verbaux considérés dans [4] font principalement référence l'attention portée à la communication et à l'interaction. Si leur implémentation doit permettre à l'agent robotique de gérer son interaction avec un individu, il est aussi d'autres paramètres à prendre en compte. En effet, dans une interaction, l'individu et l'agent robotique ne sont pas les agents passifs de la situation, ils en sont les acteurs et ils façonnent cette situation d'interaction sociale. Lors d'interactions, les interlocuteurs communiquent aussi sur leur statut. Dans leur ouvrage, Ellyson et Dovidio [6] ont ainsi recensé les nombreuses études qui traitent des problématiques de statut, de hiérarchie et de dominance sociale. Il est donc légitime de s'intéresser aux indices non-verbaux selon leur influence sur cet aspect de l'interaction sociale.

III. COMPORTEMENTS NON-VERBAUX ET CONSTRUCTION SOCIALE DE L'INTERACTION

Dans l'article de Gurtman [7], l'auteur propose un circumplex de la relation interpersonnelle, propre à décrire les rapports qu'entretiendraient différents individus en fonction de leur personnalité. Ce circumplex oppose diamétralement l'hostilité à l'affabilité, et la dominance à la soumission. Etant donné que la dimension relevant de l'hostilité ne semble pas cohérente avec le développement d'un robot interactif, nous nous focaliserons sur la dimension opposant la dominance à la soumission. Dans l'étude de Tiedens et Fragale [8], les auteurs se sont intéressés à l'adaptation d'un individu selon les signes non-verbaux que son interlocuteur exprime. Plus précisément, les auteurs ont observé si face à un comportement de domination ou de soumission, les interlocuteurs avaient plutôt tendance à imiter ce comportement, ou bien à adopter un comportement complémentaire (i. e., comportement de domination en réponse à une soumission, ou *vice-versa*). Le

sens véhiculé par ce comportement, est transmis par un ensemble de signes non-verbaux. Ces signes peuvent contribuer à donner une impression d'expansion de l'individu (e. g., occupation de davantage d'espace), auquel cas ces signes sont perçus comme un comportement de domination. Ils peuvent aussi contribuer à donner une impression de constriction (e. g., bras croisés, dos voûté) auquel cas ils sont perçus comme un comportement de soumission. La tâche que les sujets ont réalisée consistait en une interaction dont le but était de décrire à un partenaire, trois photos qui se trouvaient derrière lui sans que celui-ci ne puisse se retourner pour les voir. Le partenaire était un compère dont les signes non-verbaux traduisaient un certain comportement. Le sens véhiculé par ce comportement pouvait être celui d'une expansion, d'une constriction ou être neutre. Les auteurs ont relevé, en référence à cette typologie, le comportement adopté par le sujet. Ils ont également mesuré le degré de confort du sujet par rapport à la tâche et par rapport à la situation. Les résultats semblent indiquer que les sujets ont davantage tendance à adopter un comportement complémentaire à celui exprimé par le compère, mais aussi que lorsque le comportement complémentaire est adopté, le sujet rapporte un degré de confort plus élevé par rapport à la situation. Il semble ainsi y avoir un réel intérêt à prendre en compte les types de comportements (e. g., dominance ou soumission, ou encore introverti et extraverti) lorsqu'une interaction humain-robot est envisagée. Par exemple, dans le cas d'un robot de service ou encore d'un robot « social », il serait bénéfique pour la réalisation de sa tâche que ce robot, adapte ses comportements non-verbaux de manière à signifier un caractère complémentaire à celui identifié chez l'individu. Ce comportement adapté induirait un sentiment de confort pour l'individu vis-à-vis de cette interaction. D'autres chercheurs se sont intéressés à comment, au-delà du confort, la complémentarité du comportement peut jouer sur la manière dont l'individu perçoit l'agent robotique.

Dans l'étude de Lee, Peng, Jin et Yan [9], les auteurs se sont intéressés à la personnalité perçue du robot, lorsque celui-ci exprime différents comportements. Dans cette étude, les auteurs ont manipulé comportements verbaux et non-verbaux exprimés par AIBO™, un robot « social » développé par SONY. La concaténation de ces comportements vise à induire chez le robot, un caractère extraverti ou introverti. Les auteurs ont ainsi formulé quatre hypothèses. La première est que lorsque le robot manifestera des comportements induisant un caractère particulier, les sujets seront capables de l'identifier comme une personnalité (H1). La seconde est que lorsqu'une personnalité est perçue, les sujets appliqueront dans l'interaction des règles sociales adaptées à cette personnalité (H2). La troisième hypothèse est que lorsque le sujet identifie une certaine présence sociale du robot, cela modulera les réponses sociales en direction du robot (H3). Enfin, la quatrième hypothèse est que la tendance qu'auront les sujets à établir une forme de relation sociale avec le robot sera liée à la présence sociale perçue du robot ainsi qu'aux réponses d'ordre



Fig. 3. AIBO™, le robot social développé par SONY, utilisé dans [9].

social envoyées par le sujet à l'attention du robot (H4). Les sujets qui participent à l'expérience ont été identifiés en tant que personnalité extravertie ou introvertie. La tâche à laquelle ils ont participé, consistait en une interaction de 25 minutes avec le robot. Durant cette interaction, les participants pouvaient produire des commandes verbales parmi les 17 proposées par les examinateurs. Chacune de ces commandes amenait une réponse du robot, qui était différente selon le caractère attribué à ce robot. A l'issue de l'interaction, les auteurs faisaient compléter aux sujets un questionnaire en ligne, où ils renseignaient l'attractivité, la personnalité et l'intelligence perçue du robot, ainsi que la présence sociale qu'ils avaient ressentie.

Les résultats confortent les hypothèses des auteurs. Les sujets étaient bien capables d'identifier une personnalité chez le robot (H1), ce qui les amenait à appliquer des règles sociales à leur interaction (H2). Concernant H2, les résultats confortent ceux de Tiedens et Fragale présentés dans [8], concernant un effet de complémentarité : les sujets introvertis indiquaient trouver le robot extraverti plus intelligent, et *vice-versa*. Concernant H3, la présence sociale perçue du robot s'avère être un bon prédicteur de l'appréciation de l'interaction avec le robot, de l'attraction sociale envers le robot, et de l'intelligence perçue du robot. Concernant H4, il y a une covariance significative de la tendance à former une relation sociale avec le robot lorsqu'une présence sociale est perçue.

Malgré le fait que le robot utilisé, AIBO™, ressemble à un chien, on peut suggérer que les gestes non-verbaux contribuent à la dimension sociale de l'interaction. Ceux-ci semblent étayer la considération que l'individu peut porter au robot, en lui attribuant des caractéristiques plus « humanisantes », telle que l'intelligence, la présence, ou un trait de personnalité comme l'introversion/extraversion. L'importance de cette considération sociale, bien que temporaire dans certains cas, facilite l'échange entre différents protagonistes durant l'interaction. On peut suggérer que l'utilisation par l'agent robotique de comportements non-verbaux adaptés, au regard du comportement de l'individu et fondés sur les signes qui lui seraient complémentaires, faciliterait à la fois l'interaction humain-robot, mais aussi l'acceptation du robot. Cette acceptation laisse à penser qu'en attribuant des caractéristiques sociales au robot, l'individu s'ouvre à la possibilité d'interagir moins formellement avec l'agent robotique voire de façon plus personnelle. Pécune, Ochs et Pelachaud [10] ont ainsi proposé un modèle conceptuel de

l'attitude sociale dans une relation entre humain et machine. Ce modèle social met en avant le souci de la relation désirée et de la relation ressentie par l'individu et l'agent robotique. Les auteurs soulignent ainsi que si une intercompréhension peut être élaborée entre les interlocuteurs, elle n'implique pas forcément une bonne entente entre les protagonistes. Lorsque ces relations sont basées sur le principe de la complémentarité, il s'avère que la relation sociale établie est plus durable. Cependant, si l'agent robotique doit pouvoir jouer un rôle permettant à l'interaction d'évoluer d'une forme publique à une forme privée, on peut se demander si sa capacité à s'adapter ne devrait pas être aussi fonction de cette distinction entre contexte public et privé.

IV. SPHERE PUBLIQUE, SPHERE PRIVEE ET TYPE DE TACHE

L'intérêt de la prise en compte de la relation sociale devrait aussi amener à moduler la considération de l'état émotionnel de l'individu. Dans l'étude de Jakob, Manstead et Fischer [11], les auteurs se sont intéressés à la manière dont l'environnement de l'individu (i. e., les autres individus qui l'entourent) peut moduler sa réaction émotionnelle face à une situation. La tâche à laquelle les sujets ont participé, consistait en la lecture de scénarios, dont le but était d'induire certaines émotions (i. e., colère, joie, tristesse, dégoût, peur, et un changement abrupte d'émotion). Les auteurs ont ensuite demandé aux sujets, pour chaque situation, d'indiquer dans quelle mesure ils trouvaient plaisant d'être plus ou moins entouré (i. e., seul, avec un ami, avec un étranger). Les auteurs ont aussi demandé aux sujets d'indiquer l'intensité de l'émotion induite, s'ils se trouvaient dans la situation décrite par le scénario. Les premiers résultats indiquent que lorsque les individus auraient ressenti de la colère, ils auraient trouvé plus plaisant d'être seul plutôt qu'entouré par des amis ou des étrangers. En revanche, lorsqu'on regarde le niveau de l'émotion ressentie en fonction des situations, la joie et la colère auraient été plus intenses en situation publique, tandis que la tristesse, le dégoût, la peur et un changement d'émotion auraient été plus intenses en situation privée. Ces résultats suggèrent donc que le seuil de prise en compte de l'émotion manifesté par un individu, doit être adapté selon que la situation relève de la sphère privée ou publique. Les individus auraient tendance à manifester de manière plus intense certaines émotions en situation publique, auquel cas le robot devrait être programmé pour en réduire le seuil pour ne pas être trop facilement saturé. Cette observation est aussi valable pour les émotions plus volontiers exprimées dans la sphère privée.

Si l'agent robotique doit prendre en compte l'émotion de l'individu en fonction du contexte, la manifestation ou non de cette prise en compte, par des signes non-verbaux, peut, elle aussi, être contextuelle. Par exemple, il est peut-être des situations où l'assistance du robot devrait jouer un rôle minimal, et d'autres situations où cette assistance peut s'avérer essentielle. Dans l'étude de Balzarotti, Piccini, Andreoni, et Ciceri [12], les auteurs ont formulé deux hypothèses. La

première est que si l'ordinateur mettait en évidence sa prise en compte de l'état émotionnel du sujet, alors le sujet produirait

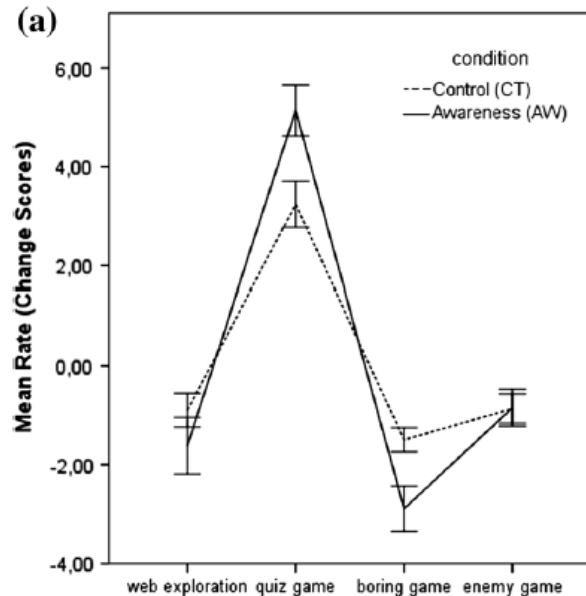


Fig. 4. Résultats présentés dans [12], du changement moyen de la quantité de signes non-verbaux produit par le sujet selon la tâche, en fonction de leur prise en compte indiquée ou non indiquée par l'ordinateur.

davantage de signes non-verbaux en accord avec cette prise en compte. La seconde hypothèse est que le type de tâche modulerait cette signalisation à l'attention de l'ordinateur. La tâche à laquelle les sujets participaient, consistait en une série de quatre types de jeux : une exploration de site web, un jeu banal (i. e., « Boring Game », un avatar à déplacer pour collecter des prix), un jeu avec ennemi (i. e., « Enemy Game », un jeu similaire au précédent, mais en fuyant un ennemi), et un quizz. Les deux premiers jeux sont donc conçus pour induire l'ennui, la frustration, tandis que les deux derniers sont conçus pour être émotionnellement exigeants. Les résultats ne valident que partiellement les hypothèses des auteurs. Ces résultats montrent la tendance qu'on les sujets à produire plus de signes non-verbaux durant la tâche de quizz, lorsque l'ordinateur a indiqué prendre en compte l'état émotionnel du sujet. Les résultats indiquent aussi moins de signes non-verbaux pour la tâche « Boring Game », lorsque l'ordinateur a indiqué prendre en compte l'état émotionnel du sujet, comparé à la condition où aucune indication en ce sens n'était fournie.

L'absence d'interaction significative serait due, selon les auteurs, au postulat concernant les deux autres tâches, que les auteurs ont envisagées comme émotionnellement engageante pour « Enemy Game » et émotionnellement neutre pour la navigation sur une page web.

Quand bien même l'agent robotique utilisé dans [12] est simplement un ordinateur, le résultat reste cohérent pour les

agents artificiels. Ce résultat semble indiquer que même en sachant que la prise en compte des comportements se fait par un agent artificiel, cela suscite chez l'individu une tendance à communiquer plus aisément dans les situations potentiellement exigeantes. De plus, concevoir l'adaptation de l'agent robotique au regard de la tâche que l'individu doit réaliser, peut permettre de réduire la prise en compte systématique des signes non-verbaux de l'individu, lorsque la situation ne l'exige pas, et donc d'alléger l'algorithme à l'origine du comportement du robot.

V. DISCUSSION

Le futur de la robotique semblent être orienté vers le concept du robot en tant qu'assistant voire compagnon. Pour ce faire, si la compréhension des messages verbaux est essentielle, il apparaît dans les études présentées que ce qui peut être véhiculé par les comportements non-verbaux l'est aussi. Toutefois, nous avons pu voir que ce traitement des comportements non-verbaux peut être appréhendé selon différents points de vue, selon la tâche, le contexte, et le type d'interaction désirée entre l'agent robotique et l'individu. Si les études [5], [8], [9] et [10] soulignent un certain bénéfice de l'utilisation adéquate de ces comportements non-verbaux dans l'interaction, les études [11] et [12] amènent à considérer ce bénéfice au regard des situations contextuelles. En effet, si l'on conçoit un agent robotique de service publique, dont le comportement le rend très familier, cet effet en théorie bénéfique, pourrait être modulé par le contexte et susciter un certain malaise chez l'individu. Il est en effet des situations où l'on s'attend à interagir de manière formelle avec son interlocuteur, et non à se sentir à l'aise comme dans une situation privée. L'ensemble des résultats de ces études est donc à considérer sous le prisme de l'expérimentation. Il est autrement fort probable que la tentation d'appliquer, sans souci des nuances contextuelles, des effets en apparence simples, placerait l'agent robotique au sein de la « vallée de l'étrange ». Ce concept souligne en effet l'étrangeté dramatique des robots conçus dans un souci de ressemblance à l'humain. Cette problématique s'incarne dans celle de la pluridisciplinarité, qui recouvre le monde de la robotique. Dans cet article, nous nous sommes efforcés de présenter la complexité que peut revêtir une interaction entre un humain et un robot. Cette complexité se traduit par une incertitude et une variabilité qui imposent un ajustement continu dans les relations interpersonnelles. C'est toute la difficulté de la robotique contemporaine, de simuler cette flexibilité, mais d'approcher cette difficulté pourrait permettre de mieux comprendre le fonctionnement humain en tant qu'être social.

RÉFÉRENCES

- [1] P. Ekman, et W. V. Friesen, "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding," *Semiotica*, 1969, vol. 1(1), pp. 49-98.
- [2] S. Goldin-Meadow, "Using our hands to change our minds," *WIREs Cogn Sci*, 2017, vol. 8.
- [3] E.T. Hall, *La dimension cachée*, Paris: Seuil, 1971
- [4] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Robotics and Autonomous Systems*, v.63 n.P1, pp. 22-35, January 2015.
- [5] V. Chidambaram, Y.-H. Chiang, et B. Mutlu, "Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues." *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI'12, ACM, NY, USA, 2012, pp.293-300.
- [6] S. L. Ellyson, et J. F. Dovidio, "Power, dominance, and nonverbal behavior," New York: Springer-Verlag, 1985.
- [7] M. B. Gurtman, "Exploring Personality with the Interpersonal Circumplex," *Social and Personality Psychology Compass*, 2009, vol 3, pp. 601-619.
- [8] L. Z. Tiedens, et A. R. Fragale, "Power moves: Complementarity in dominant and submissive nonverbal behavior," *Journal of Personality and Social Psychology*, 2003, 84, pp. 558-568.
- [9] K. M. Lee, W. Peng, , S.-A. Jin, et C. Yan, "Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction," *Journal of Communication*, 2006, vol. 56, pp. 754-772.
- [10] F. Pecune, M. Ochs, et C. Pelachaud, "De l'intercompréhension à la complémentarité des relations sociales humain-machine," in *Vers une communication Homme-Animal-Machine ? Contribution interdisciplinaire*, M. Grandgeorge, B. Le Pévédic, F. Pugniere-Saavedra, C. Jost, France. EME, 2015.
- [11] E. Jakobs, A.S.R. Manstead, et A.H. Fischer, "Social context and the experience of emotion," *Journal of Nonverbal Behavior*, 1996, vol.20, pp. 123-142.
- [12] S. Balzarotti, L. Piccini, G. Andreoni, et R. Ciceri, "'I Know That You Know How I Feel': Behavioral and Physiological Signals Demonstrate Emotional Attunement While Interacting with a Computer Simulating Emotional Intelligence." *Journal of Nonverbal Behavior*, 2014, vol .38, pp. 283-299.

Cadres d'apprentissage pour les interactions homme-robot

Nicolas ROUDIER

Master 2 Recherche en Sciences cognitives
Grenoble INP, France
nicolas.roudier@phelma.grenoble-inp.fr

Ali SAGHIRAN

Master 2 Recherche en Sciences Cognitives
Grenoble INP, France
ali.saghiran@phelma.grenoble-inp.fr

Résumé—Le développement du domaine de la robotique et particulièrement des interactions hommes-machines ont contraint les fabricants à abandonner les méthodes de programmations traditionnelles « en amont » pour des méthodes d'apprentissages rendant plus comptes des conditions d'utilisation des robots. Ces méthodes se basent sur la démonstration des tâches à effectuer par des humains, qui seront ensuite reproduites par des robots. Cependant, ces méthodes reposent sur des organisations temporelles strictes, dans lesquelles le robot et l'humain vont effectuer des actions puis recevoir des retours. Une alternative serait de s'inspirer de la relation mère-enfant, dans laquelle ces cadres temporelles s'organisent d'eux même au fil des interactions entre les participants. Cet article décrit les dernières avancées dans ce domaine.

Mots clés : *Interactions humains-robots, Machine Learning, Programming, Apprentissage robot, Pragmatic Frames*

INTRODUCTION

Avec l'avancée rapide de l'automatisation et le développement de robots polyvalents dans certains secteurs d'activité, il est devenu apparent que l'on ne puisse plus se contenter d'effectuer la programmation des robots en amont, dans les usines de productions. En effet, les robots modernes ne sont plus destinés à être isolés dans des chaînes de montages mais à interagir avec des êtres humains dont le comportement est parfois difficile à prédire. La simple écriture d'un programme convenant à tous les cas de figure possible de l'utilisation d'un robot est donc extrêmement difficile.

Il est donc devenu primordial de trouver des méthodes de programmations simples et rapides pour gagner en productivité. Une méthode pour résoudre ce problème est de doter le robot de la capacité d'adapter ses actions aux tâches qui lui sont confiées, par le biais d'un instructeur humain montrant l'enchaînement correct d'action. Il existe plusieurs manières de réaliser cet enseignement.

La manière la plus simple d'appliquer cette méthode d'enseignement est d'essayer d'imiter les gestes de l'instructeur humain : c'est l'apprentissage par imitation. L'inconvénient de ce type d'apprentissage est qu'il ne permet pas des comportements très complexes, et n'est principalement

utile que dans des chaînes de montages, dans lesquelles les gestes répétitifs sont effectués par des robots

Un autre type d'apprentissage est l'apprentissage par renforcement inverse. Celui-ci s'inspire de la méthode d'apprentissage par renforcement, qui consiste à établir une « fonction de récompense » qui correspond au degré de complétion de la tâche en cours, et que le robot devra essayer de maximiser par ses actions. Cet apprentissage « hédoniste » présente cependant l'inconvénient de devoir créer cette fonction de récompense, ce qui n'est pas toujours facilement effectuable. Pour pallier aux inconvénients de l'apprentissage par renforcement, la technique de l'apprentissage par renforcement inverse a été développée par Pieter Abbeel en 2004 [1]. Cette technique reprend les principes du renforcement, mais en permettant au robot d'extraire la fonction de récompense par lui-même en observant un humain effectuant la tâche. Le robot ne va ainsi pas simplement imiter les gestes de l'humain, mais va extraire le but de la tâche, et essayer de l'atteindre par un enchaînement de gestes différents si nécessaire.

Cependant, le point commun entre ces types d'apprentissage est que les sessions sont organisées de façon fixe temporellement : un laps de temps est dédié à l'observation de l'humain par le robot, puis un autre laps de temps est dédié à l'observation du robot par l'humain, qui va alors corriger le robot si nécessaire dans un autre laps de temps. Or, les sessions d'apprentissage chez le nourrisson ne sont pas organisées de façon strictement régulière, mais de façon adaptative choisie par interaction entre le nourrisson et le tuteur.

Dans cet article, nous aborderons donc le thème de l'organisation des sessions d'apprentissage d'abord chez le nourrisson, puis les applications possibles d'une organisation non-régulière temporellement des sessions d'apprentissage chez les robots.

I. APPRENTISSAGE CHEZ L'HUMAIN

La relation mère-enfant est un exemple naturel de partage de connaissance. Il pourrait donc être utilisé pour créer des protocoles d'enseignements utilisables par des robots.

Cette acquisition de connaissance chez le nourrisson commence à s'effectuer avant la naissance et évolue au fur

et à mesure du développement de l'enfant. Cependant, les interactions dont il va être question dans cet article sont les interactions conscientes, dans lesquelles le nourrisson ainsi que la personne responsable s'organisent pour échanger de l'information.

Ces interactions s'effectuent de façon multimodale, en associant les informations venant de plusieurs sens pour créer une représentation de l'environnement. L'enseignant peut à la fois énoncer des mots mais aussi effectuer des gestes pour décrire ses intentions.

Jérôme Bruner propose dans ses travaux [2] l'idée de l'existence de formats temporels réguliers utilisés par l'enseignant lors de l'apprentissage, appelés Cadres Pragmatique (*Pragmatic frames* en anglais). Il cite l'exemple d'une mère lisant un livre d'image à son enfant. Celle-ci va d'abord pointer une image du doigt en lui demandant ce qu'il voit (« What's that ? »). L'enfant a alors l'opportunité de répondre, puis la mère va commenter de façon positive ou négative tout en répétant la réponse attendue (« That's right/wrong, it's a *pineapple* ! »). Ce type d'organisation temporel est courant et n'est pas utilisé que pour l'apprentissage, mais aussi pour les jeux, impliquant ainsi des fonctions cognitives différentes.

Tuteur	Enfant
Question + Pointage	
	Réponse
Evaluation + Annonce de la bonne réponse	
Temps	

Figure 1 - Exemple de cadre pragmatique

La « syntaxe » de ces cadres pragmatique n'est pas standardisée et varie beaucoup selon les cultures, voir même selon les individus. Les enfants dépendent cependant d'une syntaxe stable, notamment concernant la position de l'information à apprendre.

L'avantage de ces cadres pragmatiques est qu'ils sont auto-générateurs : ils s'organisent de façon autonome au fur et à mesure des interactions entre le tuteur et l'enfant. Cela fait d'eux un choix privilégié pour une utilisation dans le cadre de l'éducation de robots, retirant ainsi le besoin de programmer de façon stricte l'organisation des sessions d'éducation.

II. CADRES PRAGMATIQUES & APPRENTISSAGE CHEZ LES ROBOTS

Cette partie se base principalement sur les travaux de Vollmer A.L. et al. en lien avec l'apprentissage et

l'interaction homme-robot dans les cadres pragmatiques [3], et forme donc une synthèse du papier.

A. Caractéristiques d'un cadre pragmatique de l'apprentissage

Un cadre pragmatique est un cadre pour l'apprentissage verbal ou non-verbal, mais requiert une forme d'interaction entre le maître (humain) et l'apprenti (robot) comme a été décrit précédemment. Ce mode fusionne donc durant l'apprentissage perception et action chez l'agent apprenti.

Afin de formaliser ce concept de cadre pragmatique (*pragmatic frame*), Vollmer définit 4 principales caractéristiques :

(1) La syntaxe du cadre d'apprentissage comme l'ensemble des actions verbal et non-verbal (comportements, prosodie) qui le décrit. Cela permet à l'apprenti d'identifier les étapes, ainsi que le début et la fin de l'apprentissage.

(2) La signification du cadre d'apprentissage (*meaning*) qui rassemble l'ensemble des opérations cognitives que le cadre impose. Par exemple, l'enseignement de la lecture de la mère décrit précédemment, requiert une segmentation d'images, une classification puis une association de labels. La signification fait référence donc aux techniques d'apprentissage automatique impliquées.

(3) Le contenu d'apprentissage est simplement l'information qui est transmise lors de l'apprentissage.

(4) La fente dans le cadre pragmatique est l'emplacement temporel de la transmission du contenu à apprendre. Cet emplacement est fixé donc par la syntaxe du cadre pragmatique suivi.

L'intérêt d'étudier les cadres pragmatiques d'apprentissage utilisés jusqu'à maintenant est de faire le point sur les approches utilisées en termes de *machine learning*, ainsi que ce que l'interaction implique ou requiert sur l'apprentissage général de l'agent (en termes de compétences sensori-moteurs ou linguistiques).

B. Exemples d'implémentation

L'objectif de cette partie est d'illustrer certaines applications et approches d'apprentissage qui incluent des interactions homme-robot et les cadres pragmatiques d'apprentissage employés.

Pour prendre en compte la perspective d'interaction homme-robot, nous montrerons un exemple de chaque catégorie d'approches en suivant la catégorisation proposée par Cuayáhuil [4] :

Les approches d'apprentissages forment globalement 3 catégories :

(1) L'apprentissage passif (mode le plus basique) dans lequel l'agent n'agit pas durant la session d'apprentissage et suffit d'observer le maître. Cette classe englobe donc les apprentissages supervisés et non-supervisés :

Calinon et al. [5] : les auteurs de cet article décrivent l'expérience de faire apprendre à un robot à nourrir une poupée (*Robota Doll*) avec une cuillère (cf. Figure 2). L'objet de l'apprentissage se résume en des mouvements appris par des démonstrations kinesthésiques par l'intermédiaire de représentations probabilistes (Modèle de Markov cachés pour encoder les mouvements et Régression en mélange de gaussienne pour l'action). L'apprentissage commence par initialisation (via un bouton), l'expérimentateur fournit au bras robot des démonstrations kinesthésiques liant le point de départ (assiette) et arrivée (bouche de la poupée). La fin de l'apprentissage se fait via appui sur un bouton. Le robot ne connaît initialement que les points de départ et d'arrivée et ne fait qu'intégrer les données de la démonstration à ses modèles de représentations.



Figure 2. Apprentissage par démonstration kinesthésique [5].

(2) L'apprentissage par exploration est l'approche où l'information n'est pas explicitement fournie au robot mais ce dernier y accède par sa propre exploration comme dans les approches d'apprentissage par renforcement. Le comportement est donc guidé par des fonctions de récompense que l'agent cherche à maximiser. Cette classe elle-aussi contient deux types de méthodes, un type dans lequel la fonction de récompense interne au système et un type dans lequel elle est externe et liée au *feedback* du maître après chaque action :

Steel et Kaplan [6] : Dans cette étude, l'objectif est de faire apprendre l'association image-label à un robot AIBO à travers une interaction sociale avec un maître humain. L'apprentissage se formule selon la séquence suivante :

(H : Homme, R : Robot)

H : « *Ballon* », « *Qu'est-ce que c'est ?* » ou « *Est-ce un ballon ?* »
R : (*Pointe l'objet, puis répond*) « *Ballon ?* »
H : (*feedback*) « *Bien !* », « *Oui* » ou « *Non, Ballon* »

Cette approche montre bien une interaction entre l'homme et le robot et demande trois tâches au robot en parallèle (acquisition de la connaissance, remémoration et comparaison de l'entrée avec les connaissances antérieures). On considère donc qu'elle contient trois cadres pragmatiques.

Grizou et al. [7] : Dans cet article, on a appris à un bras robot d'empiler des blocs sous formes de tours dans plusieurs positions et avec différents scénarii. L'utilisateur fournit un *feedback* durant chaque scénario et des instructions pour chaque mouvement. Le robot emploie le sens de ce retour dans sa fonction de décision (fonction de maximum de vraisemblance dans des modèles de décision de Markov). La méthode est très singulière parce qu'elle permet au système d'apprendre une partie du cadre pragmatique, suivant la forme des instructions qui est choisie par l'utilisateur et qui sont inconnues initialement par le robot.

(3) L'apprentissage actif dans lequel le robot mène l'interaction avec l'utilisateur (maître) en lui demandant l'information :

Cakmak et Thomaz [8] : Dans cette étude, la machine demande à l'utilisateur, après qu'il effectue deux démonstrations d'un mouvement particulier (mettre céréales dans un bol, ajout de sel à une salade...), des questions prédéfinies selon la tâche pour proposer une réalisation proche et vérifier la validité auprès de l'utilisateur. Ce dernier fournit un retour à la machine. Sauf que dans ce travail, aucun algorithme d'apprentissage n'a été implémenté car l'objectif n'était que d'évaluer les participants et l'interaction qu'ils ont eue avec la machine.

III. DISCUSSION

De manière générale, les travaux ayant suivi un cadre pragmatique durant l'apprentissage montrent une rigidité imposée et un manque de liberté en termes de structure. L'utilisation demande donc une préparation et une connaissance préalable de l'utilisateur et de la structure exacte suivie par le robot. Dans la plupart des études d'après Vollmer [3], les programmeurs préparent et enregistrent les entrées à « transmettre » au robot.

Par ailleurs, l'inférence statistique nécessaire par l'algorithme d'apprentissage ralentit l'exécution, mais il est aussi possible de l'accélérer en incorporant de l'information à travers les interactions sociales. Si l'on considère l'exemple de Calinon et al., sans informations à propos de l'intention du maître le robot rencontrera des difficultés et sera plus lent pour la déterminer (il doit trouver une information dans un ensemble très grand). L'utilisation donc de différents cadres pragmatiques d'apprentissage associés à chaque type d'information à transmettre permet un apprentissage plus rapide.

En comparant les cadres utilisés dans les études mentionnées avec l'apprentissage naturel, nous observons que l'interaction est très limitée vu que le langage et l'interaction naturels sont largement plus flexibles et dynamiques. De plus,

l'apprentissage de plusieurs tâches est requis dans un même cadre lors d'une interaction naturelle. Bruner [2] montre aussi que le maître ajuste la variabilité et la difficulté des tâches à apprendre par rapport aux performances de l'apprenant (à travers l'exemple de la mère et l'enfant).

Parmi les défis à relever, On peut citer les idées principales proposées par Vollmer et al. [3] :

(1) La gestion de plusieurs cadres pragmatiques prédéfinis ce qui peut être programmé de manière simpliste pour permettre à l'utilisateur de basculer d'un cadre à un autre, ou même au robot de proposer certains cadres particuliers d'apprentissage. Il restera un défi plus fondamental pour résoudre complètement ce point en définissant les modèles de représentations nécessaires au robot pour comprendre la syntaxe et la signification du cadre pragmatique.

(2) Il est aussi intéressant de permettre à l'utilisateur comme au robot la possibilité de choisir le cadre pragmatique le mieux adapté à la situation et le plus efficace.

(3) Les robots doivent être aussi capables d'apprendre de nouveaux cadres pragmatiques inexistant dans leur répertoire. Il sera nécessaire de décider quand un nouveau cadre est proposé puis pouvoir cerner les objectifs, la syntaxe et la signification du nouveau cadre pragmatique et de proposer des méthodes d'apprentissage au moment de « l'injection » de l'information. L'exemple développé par Grizou et al. [7] montre en effet une éventuelle piste en employant sur des méthodes probabilistes (Vraisemblance, Expectation-Maximisation) pour déduire le sens des consignes de l'utilisateur, et apprendre simultanément.

CONCLUSION

Dans cet article, Nous avons essayé de faire le point sur un ensemble non exhaustif d'approches d'apprentissage chez les robots qui gardent l'aspect interactif et miment l'apprentissage naturel. Ces méthodes montrent des structures particulières et des invariants qui peuvent être généralisées dans plusieurs cas (apprentissage sensori-moteur, langage naturel)

L'implémentation de ces cadres pragmatiques simule principalement les méthodes utilisées dans le développement et l'apprentissage de l'Homme. Cela peut poser certains défis techniques liés au moyen d'interaction (langage usuellement, développement parallèle de la communication et des facultés

cognitifs) mais peut en revanche aider à architecturer l'apprentissage et faciliter l'acquisition de l'information principale [9].

RÉFÉRENCES

- [1] Pieter Abbeel, Andrew Ng, (2004) "Apprenticeship learning via inverse reinforcement learning." In 21st International Conference on Machine Learning (ICML).
- [2] Bruner J.S., Comment les enfants apprennent à parler, Col. Actualités pédagogiques, Retz, Paris, 1987.
- [3] Vollmer, A.-L. Wrede, B. Rohlffing, K. and Oudeyer, P.-Y, (2016). "Pragmatic frames for teaching and learning in human-robot interaction: review and challenges", *Frontiers in Neurobotics*.
- [4] Cuayáhuitl H. (2015) "Robot learning from verbal interaction: a brief survey," in *Proceedings of the AISB Workshop on New Frontiers in Human Robot Interaction* (Canterbury, UK).
- [5] Calinon S., D'halluin F., Sauser E. L., Caldwell D. G., Billard A. G. (2010). Learning and reproduction of gestures by imitation. *IEEE Robot. Autom. Mag.* 17,
- [6] Steels L., Kaplan F. (2002). Aibos first words: the social learning of language and meaning. *Evol. Commun.* 4, 3–32.10.1075/eoc.4.1.03ste
- [7] Grizou, J., Lopes, M., and Oudeyer, P.-Y. (2013). "Robot learning simultaneously a task and how to interpret human instructions," in *2013 IEEE Third Joint - International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (Osaka: IEEE), 1–8
- [8] Cakmak M., Thomaz A. L. (2012) "Designing robot learners that ask good questions," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (Boston, MA: ACM), 17–24.
- [9] Rohlffing, K. J., Wrede, B., Vollmer, A.-L., and Oudeyer, P.-Y. (2016), An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames, *Frontiers in Psychology*, 7, 470

Architectures cognitives pour l'interaction homme-robot

Gabriela Vives

Master 2 Cognition Naturelle et Artificielle
Grenoble INP - Phelma

`gabriela.vives@phelma.grenoble-inp.fr`

Nolwenn Martin

Master 2 Cognition Naturelle et Artificielle
Grenoble INP - Phelma

`nolwenn.martin@phelma.grenoble-inp.fr`

Abstract—Cet article s'intéresse aux architectures cognitives susceptibles de modéliser une interaction sociale Homme-Robot aujourd'hui. D'abord une présentation de l'architecture cognitive est nécessaire pour comprendre en quoi elle pourrait répondre aux problématiques de l'interaction Homme-Robot. Ensuite des architectures cognitives récentes, appliquées à ce type d'interaction, sont présentées, un exemple visant à modéliser l'apprentissage par observation et interaction étant plus amplement détaillé. Néanmoins ces exemples montrent la limitation des architectures cognitives connues à ce jour pour l'interaction Homme-Robot : elles ne prennent pas en compte toutes les modalités nécessaires à une interaction complète avec l'Homme et sont encore aujourd'hui difficilement implémentables. C'est pourquoi l'architecture cognitive ne doit pas être considérée comme la solution unique à l'interaction sociale Homme-Robot.

Index Terms—Architecture cognitive, HRI sociale.

I. INTRODUCTION

La robotique est présente dans de nombreux domaines de notre société. Aujourd'hui, un des problèmes majeurs est la mise en place de robots intelligents, et plus particulièrement socialement intelligents. C'est à quoi s'attaque l'interaction homme-robot (HRI). Il existe actuellement plusieurs approches dans le but de créer des robots sociaux. Une d'entre elles se base sur les architectures cognitives, un domaine de la cognition ayant de multiples applications. Nous distinguons deux types d'architectures cognitives : celles qui traitent de l'intelligence artificielle telles que SOAR, et celles plutôt inspirées des neurosciences qui cherchent à modéliser les processus de la cognition comme ACT-R. SOAR (State, Operator And Result), est une architecture cognitive symbolique, créée par John Laird, Allen Newell et Paul Rosenblum à partir de 1983. Son objectif est de créer des représentations de la connaissance et de les utiliser de manière appropriée pour l'apprentissage et la résolution de problèmes [1]. ACTR (Adaptive Control of Thought - Rational), développée par John Robert Anderson est une architecture bio-inspirée qui se base sur une modélisation de la représentation en deux temps : déclarative et procédurale. La connaissance procédurale est représentée sous forme de productions, alors que la connaissance déclarative est représentée sous forme de modules inter-connectés [2]. Ces architectures en ont inspirées de nombreuses autres,

qui s'emploient aujourd'hui à modéliser de façon la plus complète possible les interactions sociales humaines et à les implémenter en robotique. Cet article présente un aperçu des dernières architectures récemment mises en place pour traduire une interaction sociale humain-robot et en propose une revue.

II. L'ARCHITECTURE COGNITIVE, QU'EST CE QUE C'EST ?

A. Définition générale

Une architecture cognitive est un processus calculatoire artificiel qui tente de simuler le comportement d'un système cognitif (généralement celui d'un humain), ou qui agit intelligemment sous respect d'une certaine définition. Le terme architecture implique une approche qui tente de modéliser les propriétés internes du système cognitif représenté et non seulement le comportement extérieur. En d'autres termes, l'architecture cognitive est une représentation simplifiée visant à modéliser des processus psychologiques ou intellectuels. De manière générale, elle décrit la structure et le fonctionnement d'un modèle cognitif computationnel. Cela permet d'effectuer une analyse de la cognition et du comportement à des niveaux multiples, les résultats pouvant alors être généralisés à d'autres domaines [3]. La figure 1 décrit les principaux composants de toute architecture cognitive (modules mémoire, décision, perception, apprentissage) répartis sur trois niveaux : traitement, interprétation et inférence .

B. La définition décortiquée.

Une architecture cognitive s'emploie à décrire le fonctionnement interne d'un agent et la façon dont des comportements peuvent être générés, d'où les termes cognition et comportement. Ceci s'effectue à des niveaux multiples : les aspects exogènes (interaction avec l'environnement) et endogènes (fonctionnement du cerveau) de la cognition étant modélisées et des dimensions temporelles variables sont considérées. Finalement, une architecture cognitive s'inscrit dans un domaine général, s'attachant à modéliser tantôt une tâche précise, mais également des modes opérationnels plus larges qui doivent pouvoir s'appliquer à diverses situations [3].

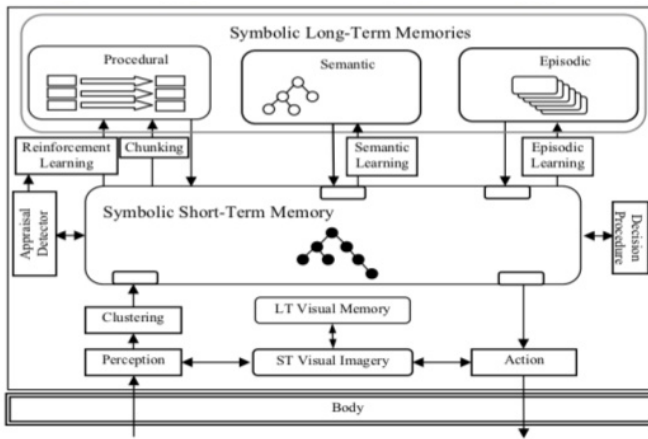


Fig. 1. Schéma descriptif de la structure d'une architecture cognitive. [4].

C. Les aspects clés de l'architecture cognitive.

Il existe trois aspects de l'implémentation d'une architecture cognitive qui représentent les grandes lignes respectées par tous les acteurs du domaine. D'abord une telle architecture doit permettre de représenter la connaissance, les structures de la mémoire, le rapport entre perception sensorielle et motricité, les symboles et leurs structures (associatives, relationnelles, etc.). Ensuite il est nécessaire que cette architecture indique le moyen de contrôle de ces structures. Par exemple, comment sont sélectionnées les actions ? Quels processus sont arbitraires ? Comment est géré le traitement des données ? Finalement une architecture cognitive doit aussi expliquer les interfaces sensorielles et motrices utilisées, ainsi que leur organisation et le processus de pré-traitement des données extérieures. Au sein de ces grands principes, chacun choisit une implémentation spécifique qui lui permettra de répondre à certains critères. La question que nous nous posons alors est de savoir si les architectures cognitives conviennent à la HRI sociale, et si certaines ont fait leur preuves dans ce domaine aujourd'hui [3].

III. LES ARCHITECTURES COGNITIVES POUR LA HRI SOCIALE DE NOS JOURS.

La HRI sociale pose actuellement plusieurs problèmes, notamment en terme de modélisation de l'esprit humain. Il est en effet difficile de produire un robot autonome, socialement intelligent, qui puisse être accepté "comme un homme" par l'être humain. L'interaction homme-robot demande au robot de pouvoir percevoir les états de l'homme (attention, émotions, inférence de modèles mentaux), d'avoir des connaissances sociales (règles, contexte, sens commun), de pouvoir communiquer (verbalement et non verbalement - avec le corps), de procéder à des interactions sur le long

terme (personnalisation, apprentissage à vie, gestion de faits inconnus) et d'agir avec les humains (action, reconnaissance de comportements, tâches collectives, *intention reading* - capacité à orienter, diriger son intention et suivre celle des autres). Plusieurs approches computationnelles tentent alors de répondre à ces attentes [5]. Il existe aujourd'hui principalement trois approches théoriques à la base de la conception de modèles pour la HRI sociale qui s'inspirent principalement de l'homme : la cognition humaine par modélisation de certains phénomènes, des modèles complètement fondés sur la cognition humaine tels que ceux inspirés des processus neuronaux, ou une approche fondée à la fois sur la cognition et ses limites, permettant l'implémentation de compétences sous contraintes au robot [5]. Le choix de l'homme comme point de référence implique trois conséquences importantes : il existe un recoupement entre les compétences perceptives et motrices entre l'homme et le robot, l'homme peut prévoir le comportement du robot et la dimension temporelle d'action du robot doit être la même que celle de l'homme [5].

A. Les livrables nécessaires à la HRI sociale.

Afin d'agir de façon similaire à l'humain, notamment lors des interactions d'ordre sociales, le robot doit savoir quoi faire et quand le faire. En effet, l'interaction pose des problèmes d'ordre temporel, le robot se devant de répondre et/ou d'agir rapidement pour que le lien homme-robot ne se rompe pas. Il est également important pour le robot de parler et/ou d'agir de la bonne façon pour que l'échange reste cohérent. Ainsi l'interaction reste naturelle. Toute la difficulté de la HRI sociale est d'ailleurs là : comment implémenter un système qui puisse déterminer le moment approprié pour suggérer ou effectuer une action représentative de ce qu'il sait de la personne avec laquelle il est en train d'interagir ? [3]

D'autre part, il est nécessaire de fournir au robot un ensemble de capteurs et de fonctions motrices afin qu'il puisse récupérer les informations de son environnement et interagir avec lui. En quelque sorte, il développe ainsi son système perceptif. Il faut alors pouvoir implémenter un système multi sensori-moteur capable de gérer les informations qu'il reçoit et les fonctions motrices qu'il possède, dans leur ensemble, tout en faisant le lien entre eux. [3]

Ces deux livrables constituent les fondements de la HRI sociale actuelle, mais posent encore beaucoup de problèmes pour l'implémentation de robots socialement intelligents.

B. Architecture cognitive ou Intégration fonctionnelle.

Deux approches existent pour l'implémentation de robots socialement intelligents : l'architecture cognitive et l'intégration fonctionnelle. L'intégration fonctionnelle permet de construire des systèmes généraux à partir de caractéristiques

particulières. Elle considère les fonctions - cognitives et computationnelles - nécessaires à l'implémentation de systèmes artificielles une par une. L'ensemble du système est alors construit par intégration des différentes fonctions voulues. La figure 2 résume le processus d'intégration fonctionnelle.

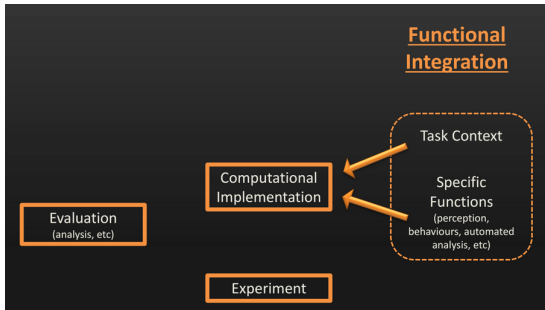


Fig. 2. Schéma illustratif de l'intégration fonctionnelle [3].

Au contraire, l'architecture cognitive est fondée sur des processus généraux issus de l'ensemble des domaines des sciences cognitives, telles que la psychologie et les neurosciences, pour implémenter des fonctionnalités spécifiques aux systèmes. La figure 3 illustre l'architecture cognitive.

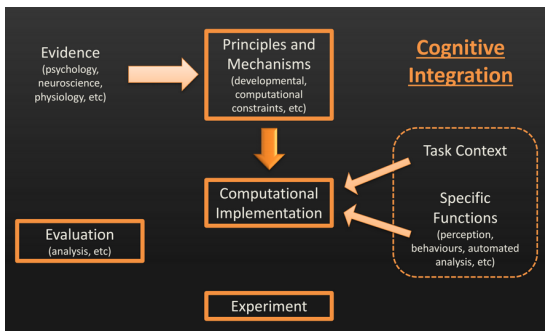


Fig. 3. Schéma illustratif de l'architecture cognitive [3].

Malgré leurs approches opposées, les deux processus ont les mêmes objectifs et fournissent des résultats semblables. Après définition computationnelle du système, celui-ci peut être expérimenté. Les résultats fournissent un retour sur le processus d'implémentation et permettent ainsi l'amélioration du système, telle qu'en témoigne la figure 4.

L'intégration fonctionnelle a fournie des démonstrations et des moyens d'investigation intéressants, répondant à des questions scientifiques précises. Cependant, elle ne décrit pas formellement les processus implémentés, ayant alors des conséquences fondamentales sur la manière dont le système fonctionne. [3] Pour sa part, l'architecture cognitive permet d'intégrer et de valider, des données théoriques et quantitatives issues des sciences humaines empiriques telles que la psychologie et les neurosciences et représentatives de

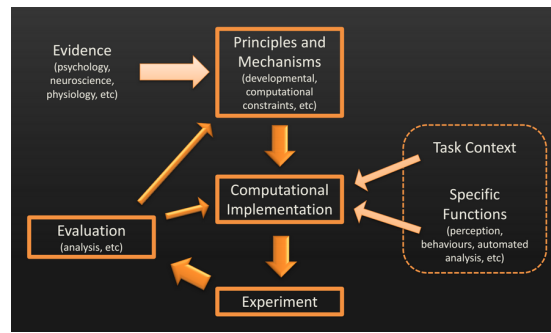


Fig. 4. Schéma illustratif de la mise en place d'un système computationnel répondant aux attentes de la HRI sociale [3].

principes cognitifs. En formalisant ces principes, elle facilite la comparaison de différents systèmes, une tâche plutôt complexe. D'autre part, les résultats expérimentaux peuvent mettre en avant la nécessité d'intégration de données supplémentaires issues des autres domaines des sciences cognitives et les corrections nécessaires à apporter. Ainsi, cette approche permet l'amélioration des modèles d'architectures cognitives et donc des systèmes robots résultants [3]. C'est pourquoi, l'architecture cognitive semble être la meilleure approche computationnelle existant aujourd'hui pour l'application à la HRI sociale.

Cependant, l'architecture cognitive est plutôt introspective, se focalisant principalement sur des compétences individuelles. Or le domaine de la HRI demande au contraire une interaction constante avec d'autres systèmes et ne réside pas dans l'individu. Il est donc nécessaire d'arborer une approche holistique. La solution réside potentiellement dans l'itération des comportements individuels. En effet ceux-ci sont fortement impliqués dans les processus d'interaction qui agissent également sur l'individu, telle qu'en témoigne la figure 5. La section suivante présente un ensemble de travaux en architecture cognitive évoluant dans cette optique.

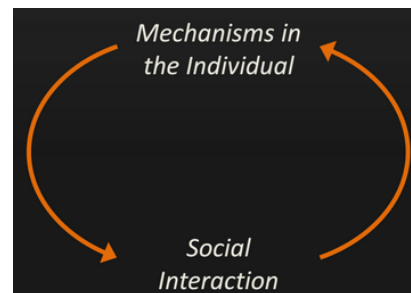


Fig. 5. Implication mutuelle de l'individu et de son interaction [3].

C. Des architectures cognitives récentes pour la HRI sociale.

De nombreuses architectures cognitives sont aujourd'hui développées pour des applications à la HRI sociale. Cette section présente un aperçu de trois des derniers travaux de l'année 2016, présentées lors du Workshop de Christchurch (Nouvelle Zélande) de cette même année[5]. Ils s'intéressent à différentes fonctions cognitives intervenants dans le cadre d'une interaction sociale : les tâches collaboratives dans l'apprentissage et dans de simples actions conjointes, ainsi que le phénomène d'adaptation de comportement selon contexte (*Alignment*). Ces exemples donnent un aperçu des avancées actuelles en architecture cognitive pour la HRI sociale.

Dans le cadre de tâches collaboratives, l'homme et le robot doivent pouvoir partager un savoir commun. Chacun doit adapter son comportement par rapport à l'autre afin d'être compréhensible. Dans un cadre éducatif, celui qui apprend doit pouvoir partager son savoir et l'enseignant doit pouvoir s'adapter pour s'assurer de l'acquisition correcte des compétences. Ceci requière donc une adaptation mutuelle. En particulier, le robot doit pouvoir modéliser ses propres états mentaux perçus par les autres. Il doit par exemple pouvoir se demander : "L'humain a-t-il bien compris que j'(robot) avais besoin de cette objet pour réaliser la tâche, ou dois-je lui expliquer à nouveau?". L'architecture cognitive développée par Alexis Jacq et al.[6], *Mutual Modelling* (MM), effectue une généralisation de la théorie de l'esprit (ToM) qui décrit la capacité à attribuer des états mentaux et des connaissances aux autres (la représentation que l'on se fait de l'autre en fonction des émotions, gestes, paroles que l'on perçoit). En effet, lors d'interactions, l'homme est constamment en train de collecter et d'analyser des informations pour rester attentif et conscient du comportement (émotions, buts) de ses collaborateurs. MM, modèle bayésien, caractérise alors l'effort d'un agent à modéliser les états mentaux d'un autre, dans un contexte éducatif. La figure 6 illustre cette architecture.

Dans le même cadre, Sandra Devin et al.[7] expose les compétences spécifiques nécessaires à un robot autonome pour une interaction conjointe avec l'homme (*Joint Action*), ainsi que les possibilités d'intégration de telles compétences dans une architecture cognitive. Selon la psychologie, les agents doivent **partager un but et un plan d'action (intentions) commun** pour y parvenir. Pour cela, ils doivent pouvoir **négoier** les actions à mettre en place pour le plan. Il faut également qu'ils puissent **se représenter** et **se percevoir** l'un l'autre lors de l'exécution du plan (attention commune, observation des actions, représentation des compétences de l'autre (ToM)). Enfin, il doivent être **coordonnés** involontairement (coordination involontaire telle que la synchronisation naturelle du mouvement de rocking-chairs côte à côte) ou volontairement (changement d'attitude volontaire vis à vis de l'autre). Ces compétences doivent

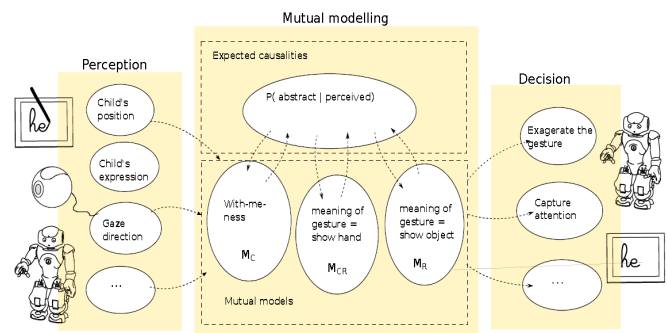


Fig. 6. Schéma illustratif de l'architecture cognitive MM [6]. Les cadres jaunes représentent les parties principales, les ellipses blanches des modules. Le schéma illustre une situation d'incompréhension des gestes du robot par un enfant en train d'apprendre. Le robot perçoit les expressions et la position de l'enfant. Ces données sont traitées par MM qui modélise ensuite la situation d'incompréhension. A partir de ce modèle, le robot doit prendre une décision. Afin de résoudre ce problème, une des choix possibles du robot est d'exagérer les mouvements.

alors être appliquées à un robot. Ceci demande une certaine adaptation des compétences, le robot n'ayant pas les mêmes capacités qu'un homme et l'homme ne se comportant pas de la même façon avec un robot qu'avec un autre être humain. Par exemple, les capteurs fournissent des informations sur la géométrie et la sémantique de l'environnement et des algorithmes d'identification (observation et raisonnement) du comportement humain (plan recognition) permettent d'inférer les buts du partenaire. Ces compétences doivent alors être intégrées dans une architecture cognitive pour aspirer à une application concrète dans un système humanoïde. Pour cela, l'architecture doit contenir les composantes **Situation Assessment** pour analyser la situation (reconnaissance du contexte), **Mental State Management** pour gérer les états mentaux de l'homme (construction et maintien de leur cohérence), **Communication for Joint Actions** pour permettre la communication verbale et non verbale, **Intention Prediction** pour déterminer les buts du robot, **Shared Plan Elaboration** pour l'élaboration du plan commun (acquiescer) et **Shared Plan Achievement** pour son exécution. Cette dernière sollicite **Action Achievement** qui modélise les actions du robot et rassemble les informations de **Human Actions Monitoring** qui interprète les mouvements et reconnaît les actions de l'homme. La figure 7 illustre un modèle d'architecture potentielle, incluant les compétences essentielles pour effectuer une tâche conjointe.

Dans une même optique, Paul Baxter et al. [8], propose une architecture s'inspirant de la neuropsychologie qui plaide de plus en plus en faveur d'un rôle distributif et associatif de la mémoire comme base de la cognition (mémoire associative), *The Distributed Associative Interactive Memory* (DAIM). Cette architecture s'applique au phénomène par lequel un

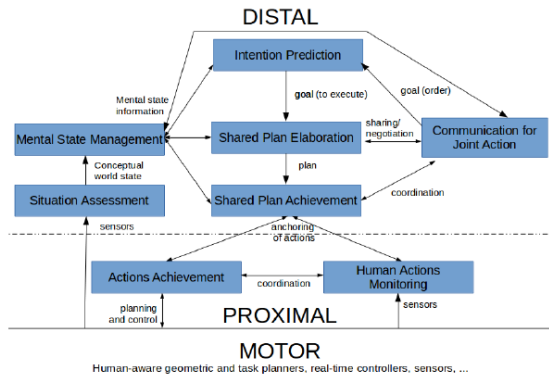


Fig. 7. Schéma illustratif d'une hypothétique architecture cognitive pour des actions conjointes [7]. Motor, Proximal et Distal représentent les 3 niveaux de l'architecture correspondant respectivement à l'intention conjointe, l'exécution du plan, la coordination cohérente des gestes pour l'exécution, dans l'espace-temps. Les carrés bleus représentent les différentes composantes essentielles à l'architecture, réparties dans leurs niveaux respectifs.

individu modifie son comportement pour "l'aligner" à celui des autres selon la situation (Alignment), un des aspects fondamentaux de l'interaction sociale. Fondée sur un ensemble de principes de l'architecture cognitive ayant déjà fait leurs preuves dans d'autres domaines, DAIM met en évidence comment un tel phénomène peut être implémenté chez un robot dans le cadre de la HRI sociale. DAIM caractérise le comportement de l'agent humain en temps réel à partir des informations sensorielles de l'interaction qui lui sont fournies. DAIM étant intégrée dans le robot, le comportement de ce dernier est alors contrôlé en temps réel de façon autonome. Le robot peut alors s'adapter aux données de son environnement. Il peut donc interagir avec l'homme tout en enregistrant des données, lui permettant de conserver une interaction fluide et cohérente. La figure 8 illustre la boucle de contrôle du robot. DAIM permet donc une caractérisation des styles d'interaction des individus. D'autre part, suite à son intégration dans un robot Nao, les résultats expérimentaux témoignent en faveur d'une application de ce type d'architecture à la cognition en général.

La connaissance des mécanismes cognitifs sociaux est essentielle à la mise en place d'architectures cognitives intégrées pour la HRI (raisonnement probabiliste, raisonnement spatial, lien perception-action, signalement de l'action). Ces exemples montrent à quel point les architectures cognitives jouent un rôle crucial dans la HRI sociale actuelle, et la qualité de représentation des fonctions cognitives qu'elles modélisent. Cependant, il est aujourd'hui encore compliqué de les implanter directement dans des robots pour une mise en application concrète. La partie qui suit, présente de façon

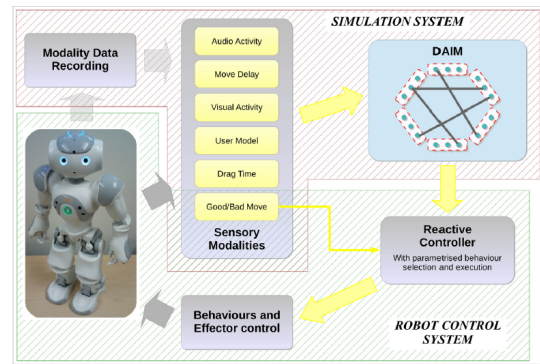


Fig. 8. L'influence de DAIM dans l'architecture contrôle du comportement du robot [8]. La partie supérieure rouge représente le système computationnel, la partie inférieure verte, le système de contrôle du robot. C'est 2 parties sont intégrées distinctement dans le robot. DAIM s'adapte en fonction des données sensorielles en entrée, et utilise les informations apprises pour agir en temps réel sur le comportement contrôlé du robot par l'intermédiaire du *Reactive controller*. Ce composant est utilisé de façon à part entière pour contrôler le comportement du robot pendant l'interaction où a lieu la collecte de données. Une analyse de la cohérence du mouvement du robot (*Good/Bad move* dans les modalités sensorielles permet un retour direct sur ce composant et distinct de DAIM - les 2 parties étant séparées. Ceci permet d'ajuster plus rapidement le comportement du robot en cas d'erreur, que par l'intermédiaire de DAIM.

détaillée, une architecture cognitive pour la perception et l'apprentissage du robot, développée sur la base de la HRI, et qui a été réellement expérimentée sur un système robotisé.

IV. UNE ARCHITECTURE COGNITIVE POUR LA PERCEPTION ET L'APPRENTISSAGE DES ROBOTS, FONDÉE SUR LA HRI.

A. Présentation et objectifs

Dans cette partie, nous proposons de développer un exemple d'architecture cognitive permettant un apprentissage par l'observation et l'interaction du robot avec des humains. Cette architecture a été développée par Uriel Martinez-Hernandez, Luke W. Boorman, Andreas Damianou et Tony J. Prescott afin d'élaborer un système intelligent capable d'apprendre par observation et par interaction avec l'Homme [9]. Ils sont partis du constat que les tentatives de concevoir un robot qui apprend, n'étaient basées que sur l'observation et donc rendaient l'apprentissage passif. Or l'interaction est une capacité cruciale pour améliorer la perception et les processus d'apprentissage dans un environnement sans cesse changeant. C'est ainsi que cette architecture cognitive a vu le jour. Elle se compose de plusieurs modules, notamment un module dédié à la mémoire bio-inspirée. Cette mémoire artificielle reçoit des informations multimodales via les entrées sensorielles du robot, puis décide du comportement qu'exécutera la robot en réponse aux stimuli.

B. Le modèle SAM

Cette architecture cognitive est basée sur un modèle déjà existant : SAM (Synthetic Autobiographical Memory). Ce modèle se compose de différents modules connectés entre eux visibles sur la figure 9. Le modèle SAM reçoit en entrée une information perceptive multimodale (visuelle, sonore, toucher) dans un format approprié. Une fois l'information traitée (ce ne sera pas l'objet de notre exposé), la sortie la plus probable est choisie en fonction de la situation. Le robot pourra prendre en compte la reconnaissance des gestes et visages humains ainsi que leur prédiction afin de prendre cette décision.

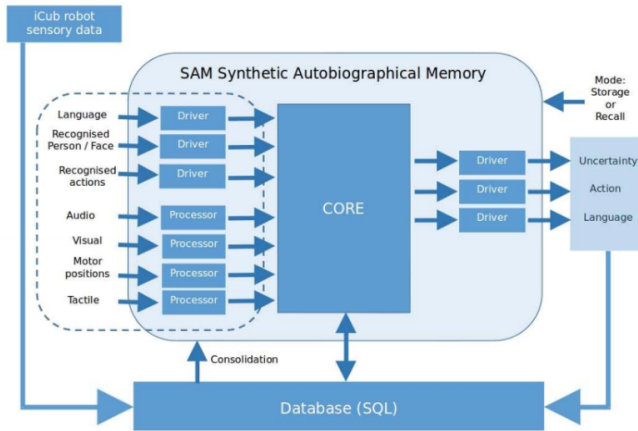


Fig. 9. Architecture cognitive pour l'apprentissage par interaction humain-robot [9].

La première étape dans l'interaction humain-robot est la détection du visage humain par le robot. Ainsi un module réalisant cette étape est implémentée selon ce que prévoit le modèle SAM. Pour apprendre via la reconnaissance de visages, un modèle de ce qui est perçu par le robot est stocké. Lorsque un visage est reconnu, l'information est communiqué au module qui gère la parole dans le but de commencer une interaction avec l'humain. Cette architecture cognitive permet également un apprentissage l'observation des gestes de l'homme. Une fois celui-ci repéré face au robot, ses bras sont détectés et leurs gestes sont suivis. Les actions que le robot peut reconnaître sont basés sur le mouvement des mains : la manipulation d'objets comme attraper, poser, tirer vers soi, pousser ainsi que les gestes de bienvenue. Les observations faites par le robot sont apprises selon le modèle SAM puis stockées pendant le processus d'apprentissage. Par la suite le robot sera capable de reconnaître ces actions apprises et de les associer à un comportement précis. Des exemples de ce que renvoie le module de reconnaissance d'actions intégrées dans l'architecture cognitive peut être vu figure 10.

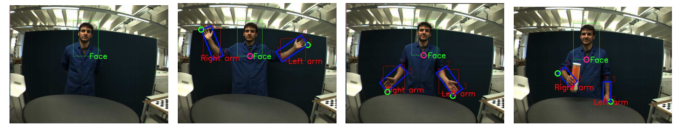


Fig. 10. Quelques exemples de ce que renvoie le module de reconnaissance d'actions intégrées dans l'architecture cognitive [9].

Outre la reconnaissance des visages humains et de leurs actions, le robot est capable de communiquer verbalement pour une interaction homme-robot plus naturelle. Le robot est programmé pour commencer une interaction verbale avec l'humain lorsqu'il reconnaît certaines phrases, préalablement implémentées, comme "bonjour" ou "qui es-tu?". Après observation et apprentissage lors de l'interaction, il utilise les modules de rappel de son architecture cognitive pour répondre à des énoncés spécifiques.

C. Implémentation

Cette architecture est intéressante car elle a été concrètement implémentée sur une plate-forme robotique, l'humanoïde iCub. Celui-ci est une plate-forme robotique complexe et open-source ayant des modalités perceptives multiples et conçu expressément pour la recherche. Les entrées visuelles permettent au robot de détecter, reconnaître et suivre le visage et les bras de l'humain, les entrées auditives permettent l'interaction vocale. iCub est aussi doté de systèmes de contrôle de la fixation du regard et du mouvement du cou. Cela permet de modéliser une interaction sociale qui soit plus naturelle que pour un robot immobile. La figure 11 montre le robot iCub apprenant les actions et le visage de l'humain qui se place face à lui. Cette implémentation sur le robot humanoïde iCub s'est avérée concluante sur la reconnaissance de visages humains, ainsi que sur l'observation et l'apprentissage d'actions effectuées par l'homme. Néanmoins nous observons que les architecture cognitives actuelles sont souvent centrées sur un objectif, comme ici l'apprentissage par observation et interaction. Il est impossible que l'interaction homme-robot soit modélisée dans sa totalité et c'est en cela que les architectures cognitives trouvent leurs limites aujourd'hui. Dans la prochaine section présentent les limitations et perspectives des architectures cognitives connues à ce jour.

V. PERSPECTIVES

A. Les avantages que peut apporter cette structure théorique à la HRI.

L'architecture cognitive possède de nombreux avantages pour la HRI sociale. Elle permet de réunir tous les aspects d'un agent autonome dans une base théorique très cadrée et de comparer des systèmes computationnels de contrôle complexes. Étant donné l'approche très cadrée autour de

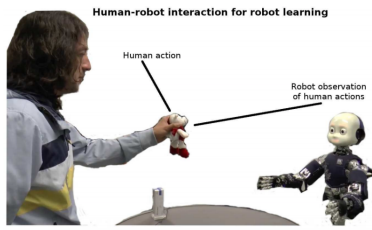


Fig. 11. Le robot iCub apprend les visages et actions de l'homme via son architecture cognitive [9].

principes bien définis de l'architecture cognitive, elle peut également apporter des résultats probants permettant une généralisation des données de la HRI à d'autres domaines des sciences cognitives. D'autre part, d'un point de vue pratique, la HRI sociale permet de produire des systèmes plus fiables en rendant compte des limites du comportement humain.

B. L'architecture cognitive, oui mais...

Cependant, même si l'architecture cognitive semble être une solution à la HRI sociale, elle n'est peut-être pas la seule. L'intégration fonctionnelle en témoigne d'ailleurs, comme complément pour la recherche dans l'ensemble des domaines des sciences cognitives, nécessaires à la HRI sociale. D'autre part, beaucoup d'aspects fondamentaux et pratiques de la HRI sociale sont encore méconnus et nécessitent d'être compris tels que la réaction de l'homme face au robot qui varie selon le contexte et la variété des domaines d'application. Ces connaissances sont nécessaires à la formation d'architectures cognitives efficaces. Par conséquent, il n'existe pas encore de modèles véritablement fonctionnels, permettant de se questionner sur leur véracité et leur falsification potentielle. Aujourd'hui, la plus part des architectures cognitives appliquées à la HRI sociale s'inspirent d'architectures pré-existantes, ayant fait leur preuve, pour tenter une modélisation de la cognition, sans chercher à les réfuter. Il serait donc intéressant de s'y confronter dans un second temps. D'autres systèmes font également actuellement leur apparition, comme la modélisation cognitive computationnelle de l'équipe de neuro-cybernétique de Philippe Gaussier de l'ETIS basée l'organisation corticale [10], les hypothèses étant vérifiées en partenariat avec des psychologues et neurologues.

VI. CONCLUSIONS

L'architecture cognitive est un domaine général qui modélise la cognition humaine de façon computationnelle. Elle est utilisée dans de nombreux domaines des sciences cognitives. Notamment, ces dernières années, un certain nombre d'applications à l'interaction homme-robot ont été développés dans le but de résoudre des problèmes clés de la HRI, tels que la mise en place d'un robot socialement intelligent. En effet,

malgré son caractère introspectif, l'architecture cognitive possède de véritables avantages pour la HRI sociale. Elle produit des systèmes formalisés fondés sur des principes cognitifs généraux et permet leur conception concrète. D'autre part, en prenant compte les résultats de la cognition naturelle elle permet de modéliser et d'expliquer de nombreux comportements. C'est pourquoi beaucoup de recherches sont actuellement en cours et ne cessent d'apporter des résultats probants en faveur de l'architecture cognitive pour la HRI sociale tels qu'en témoigne les exemples présentés dans cet article.

Comme l'illustre l'exemple développé "Une architecture cognitive pour la perception et l'apprentissage des robots, fondée sur la HRI", l'architecture cognitive offre donc de vastes perspectives à la HRI sociale, mais de nombreuses questions restent encore à résoudre. D'un point de vue théorique, on peut se demander jusqu'où faut-il décrire la cognition pour pouvoir produire des comportements de robots naturels. Ou encore, quels sont précisément les principes nécessaires à l'implémentation d'une architecture cognitive permettant la modélisation d'un comportement social ? Cette dernière trouve d'ailleurs en partie réponse dans l'étude récente de Devin et al. [7], présentée en amont, qui résume les compétences nécessaires pour effectuer des tâches collectives, un des aspects fondamentaux de l'interaction sociale. Ces compétences sont d'ailleurs généralisables à l'interaction sociale dans son ensemble. D'un point de vue pratique, il serait intéressant de déterminer les meilleurs mécanismes computationnels pour implémenter ces principes, ou encore de se demander jusqu'où faut-il pousser l'implémentation pour pouvoir expérimenter l'architecture sur un système réel. En effet, il reste encore beaucoup à faire ! En dépit de ses nombreux avantages et des améliorations, il reste difficile de tester les architectures cognitives sur des robots. Les résultats manquent d'applications concrètes. C'est pourquoi, malgré l'apport considérable de l'architecture cognitive, peut-être n'est-ce pas l'ultime solution à la HRI sociale !

REFERENCES

- [1] J. E. Laird, *The Soar Cognitive Architecture*. MIT Press, Apr. 2012.
- [2] G. Trafton, L. Hiatt, A. Harrison, F. Tamborello, S. Khemlani, and A. Schultz, "ACT-R/E : An Embodied Cognitive Architecture for Human-Robot Interaction," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 30–55, Mar. 2013.
- [3] P. E. Baxter, "Cognitive architecture for social hri, social hri summer-school plymouth university - uk," 2013.
- [4] B. K. Maruthi, "Cognitive architecture," 2013.
- [5] G. T. Séverin Lemaignan, Paul Baxter, "Workshop on cognitive architectures for social hri 2016, hri conference, christchurch - new zeland," 2016.
- [6] A. Jacq, W. Johal, P. Dillenbourg, and A. Paiva, "Cognitive Architecture for Mutual Modelling," Feb. 2016.
- [7] S. Devin, G. Milliez, M. Fiore, A. Clodic, and R. Alami, "Some essential skills and their combination in an architecture for a cognitive and interactive robot," 2016.

- [8] P. E. Baxter, J. de Greeff, and T. Belpaeme, "Cognitive architecture for human–robot interaction : Towards behavioural alignment," *Biologically Inspired Cognitive Architectures*, vol. 6, pp. 30–39, Oct. 2013.
- [9] Martinez-Hernandez, Uriel, et al., "Cognitive architecture for robot perception and learning based on human-robot interaction." *Citeseer*.
- [10] P. Gaussier, "Etis ([http ://www.etis.ensea.fr/neurocyber/web/en/](http://www.etis.ensea.fr/neurocyber/web/en/))."

Socially assistive robotics and elderly people

Edwige Laflaquière, Marie Palma, Annalisa Paroni

Cognitive Sciences, Grenoble Institute of Technologies

[e.laflaquiere,mariepalmam,lisa.paroni}@gmail.com](mailto:{e.laflaquiere,mariepalmam,lisa.paroni}@gmail.com)

Abstract— Socially assistive robotics is a new domain that explores the interaction between humans and robots. Little knowledge is available, but preliminary studies seem to show that we can take advantage from this domain to provide better healthcare conditions to the elderly through designing and implementing robots that serve as companion to these people. As shown by quite a few studies, companion robots have beneficial effects on elderly people on the psychological, physiological and social level. However many aspects still need clarification, such as how to integrate this kind of robots in healthcare situations and the role this robot are supposed to take in the interaction with humans.

Keywords—robotic; social robots; assistance; elderly people

I. INTRODUCTION

Since the 16th century, the inventors are interested in the articulated robots, in particular with the appearance of the intelligent machines. These machines offer a service increasingly more precise, and their use tends to be democratized: new technologies, formerly booked with industries, extend at the present time to the field from health. These technologies, functioning in autonomy like certain robots or not, have always the role to improve quality of life of the users, to decrease the constraints and to optimize behaviors by them. For example, of the robot-surgeons surgeons of the University of Hopkins assist, or can also replace some their gestures, like the robot STAR, Smart Tissue Autonomous Robot. In parallel, the populations are done increasingly growing older, because in particular of progress of medicine and science. The Public health must then face new challenges: to find average equipment and financial to accompany a proportion always growing dependent elderly people. A new way of looking after and of accompanying is then being written. The researchers in robotics are interested in these problems and work in collaboration with the health professionals, in order to propose to the user technologies of assistance, coming to support the human assistances which often have suddenly missed in structures like health care facilities.

This work makes a statement on the various situations in which these technological assistants are used, their effectiveness from a qualitative and quantitative point of view. Then, we will see in what these tools can be useful to find a better place in society.

A. ROBOTICS HISTORY

Robotics is closely related to progress of mechanics. Humans has built objects articulated for approximately 1600 years before JC It is at that time in Egypt that the first clepsydras are carried out, or clock with water. Then, the clock industry gave rise to the first automats. In 1738, Grenoble-native engineer Jacques de Vaucanson creates an animal automat of form. This articulated duck has very advanced functions for the time: de Vaucanson tried to reproduce the digestive system of the animal, it is then able to swallow its food, then a mechanism of convolutions of pipes digests food. The interest here is to reproduce a physiological behavior, this automat does not have yet a role of interaction with humans.

The automats gradually give way to the robots, i.e. machines which have sensors enabling them to collect information on the environment, in order to give an answer thanks to their driving bodies. It is the case of the electric dog of Hammond and Miessner (1915), which is attracted by the light, thanks to the technique of animal phototropism. In 1940, Grey Walter builds cybernetic tortoises able to locate sources of light and to circumvent them. These animal robots are not yet robots of service, i.e. their purpose is not to help humans. Little by little, the techniques improve and the robots take a human appearance: they are given eyes and mouth, and emotions thanks to mimicry which they can produce. So they are integrated gradually in an interaction with humans. It is the case of robots such as Nao, Pepper (Bank Software, 2006) and iCub (universities European), which express their emotions via mimicry of their faces. Then, the scientists re-use the robot in order to interact with the elderly person. Thus the Paro robot, playing a social role, is used in workshops of animation and individual relational therapy for people in health care centers.

B. SOCIALLY ASSISTIVE ROBOTS

This kind of robots, i.e. robots that provide assistance to human users through social interaction, are classified as Socially Assistive Robots (Feil-Seifer & Mataric, 2011 cited in [1]) and the field that studies and develops this kind of

social robots is Socially Assistive Robotics (SAR). Not surprisingly, this is an interdisciplinary field that combines multiple disciplines such as robotics of course, but also engineering, medicine, psychology and communication. More specifically, SAR is an area of robotics at the intersection of assistive robotics (i.e. aiding human users through interaction with robots) and socially interactive or intelligent robotics (i.e. socially engaging human users through interactions with robots) [1]. Socially assistive robots must have specific characteristics in order to be classified as such. First and foremost, the social component of the interaction between a robot and a human user serves the purpose of helping the latter. In order for this to be possible, the robot must be capable of perceiving its environment, interact with human users, display appropriate social cues, and effectively communicate with human users (Okamura, Mataric & Christensen, 2010 cited in [1]).

As an interdisciplinary field, SAR provides a wide range of real and potential applications. One of those is healthcare, and more specifically mental healthcare. Rabbitt and colleagues [1] point out three main roles socially assistive robots have served in clinical research and work up to date: companion, therapeutic play partner and instructor.

Playing the role of therapeutic partners, the socially assistive robots mostly aid children in practicing or building clinically relevant skills, or can be integrated to cognitive training therapy for the elderly. As instructors, they can provide direct instruction and supervision to patients or clients engaged in relevant treatment activities (basically motor activities). However, socially assistive robots are more commonly used as companions for elderly people. In this instance, the robot takes the place of a real animal mimicking its functioning. The resulting therapy is one that is analogous to pet therapy.

C. INTERACTION

Depending on the role they take, the interaction of the robot with the human user changes. Sibata & Wada [2] identify four categories of human-interactive robots that are used for psychological enrichment in terms of their relationship with the human users:

1. Performance robots: they execute movements that express meanings to humans (mostly for fun). The movements they perform are pre-programmed and mostly repetitive. As a result, this kind of robot is not very interactive with the human user;
2. tele-operated performance robots: they are controlled remotely by a hidden operator;
3. operation, building, programming and control robots: mostly used for education (“edutainment”);
4. Interactive autonomous: they connect with the human user in the physical world and use verbal and nonverbal communication depending on the functions of the robot. The interactions between the robot and the human user are mostly personal.

A. ROBOTIC ASPECT

1) OPERATION

Social interaction is governed by what we feel or perceive. In order to create an interactive robot, we need to assign it facial expressions, mimics, etc that can be interpreted by its interlocutor. Psychologists [3] highlight six major facial emotions in humans, called the “big 6”. They correspond to universal emotions. There are shades in these emotions, forming complex emotions in between. We can distinguish: happiness, sadness, disgust, fear, anger and surprise. Even if these emotions are held by the face, they can also be seen through gesture, prosodic characteristics or attitude.

It’s easy to imagine a robot with facial expressions and body gesture, as for example EmotiRob.

However, prosodic feature seems to be more complex to pursue. This feature has been experimented with Emox robot and dependent elderly people that gradually tend to lose the ability to communicate with their nurses or even their families. Using everyday robots at home improve their communicative skills, and help them through something called “socio-affective glue” to outdo seclusion [4].



Figure 1 – Emox at LIG lab [4]

In order to interact with humans, Breazeal and Scassellati [5] identified four important characteristics a social robot shall have:

1. It must be credible as a life-like system
2. It must act on a consistent way
3. It must be able to express internal states such as BDI
4. It must be able to quickly react to unexpected events

Robotic research has studied the fact that the physical shape of the robot plays an important role in social interaction [6]. It’s complicated to imagine having a conversation with a robot if we are not able to localize its head, or its eyes for example.

Anthropomorphism is one’s tendency to rationalize the social interaction, by assigning human characteristics or BDI to a non-human agent. “[...] the humanoid form best helps us interact with machines because of the fact that we anthropomorphize. But this does not necessarily mean that one

should build seamlessly perfect-looking human-machines a s anthropomorphism allows roboticists be heuristic in robot design” [6]. We will make a review of robots used in geriatrics and highlight the fact that actually, non-humanoid robots are the most used and experimented in the field. Maybe the animal form is easier to implement in terms of costs and operation, because there are less degrees of freedom to implement. As a matter of fact, animal robots tend to be more attractive for now.

2) MEDIATOR ROBOTS IN GERIATRICS

Elderly people may suffer from multiple mental diseases. In particular, depressive disorders are common among elderly people in nursing homes [2]. Pet therapy is often exploited to alleviate depressive symptoms among these people. However, it is not always feasible or ideal to have a real animal interacting with them [7]. Mediator robots offer a new way of support for dependent people, especially in hospitals or nursing homes. There is a lot of literature on mediator robots, since this kind of robots is often used to encourage interaction or to act as animals.

In Japan since 2001, the Paro robot is used to decrease anxiety and alleviate depression [2,9] in nursing homes. Owing great acceptance and results, Paro was quickly imported to Europe and is now used in many structures, not only by elderly. Such robots are particularly adapted to hospital mediation because researches shown that the benefit of animal mediation and robot mediation were comparable. It’s known that only a few hospitals or nursing homes allow residents to have animals, yet their presence is beneficial to elderly [8]. Another study [9] found no significant differences between the robotic dog AIBO and a real dog concerning overall attachment.

Benefits provided by animals with elderly have been shown in [8]: the study shows that animal-assisted therapy do have a positive impact on the elderly, in assisted living facilities. Indeed, this kind of therapy with an animal had the advantage to motivate elderly people, their mood was better. Moreover, they anticipated the animal’s visits: before the animal came, they were full of joy.

As Paro, NeCoRo (Neco means “robot” in Japanese) behaves as a pet. He can’t move, but is able to feel emotions of his user, to which he can respond with eye, ear and brow movements. Beside the fact that NeCoRo can answer to the user with mimics, he can also start communication thanks to his caterwauling. An analyzer of vocal spectra also allows him to detect the humor of the owner according to the frequencies of its voice. Then, this robot reproduces the animal behavior, while avoiding the risks bound to the presence of a real animal at home: risks of unpredictable behavior, allergies etc.

Robots have now a role of aid to individuals. It is in this purpose that Robear was built, the robot humanoid with the head of a bear. His big and friendly head reassures the elderly. He has a direct role of assistance to the elderly because it is capable of lifting and moving disabled persons up to 140 kg without difficulties. His numerous sensory sensors allow him to adapt themselves to the position, and to the weights of the patient to have the safest and more precise gestures. Robear is particularly useful for patients living in

rural areas, or for an increasingly aging population and to multiply the offer of care.

Moreover, Philipps Research developed iCat, a prototype of an emotionally intelligent user-interface robot. Its aim is to develop an intuitive interaction between him and the users. He is a cat-like robot character. He can move his head and some parts of the face as eyebrows, eyes, eyelids, lips. Despite having such a few degrees of freedom, iCat can express its emotions, reflecting its internal states according to its interlocutor. We can see in the figure below that not all the big 6 are represented with iCat, because fear and disgust are not shown. These two emotions are very pejorative, and maybe they are not adapted to social interactions.

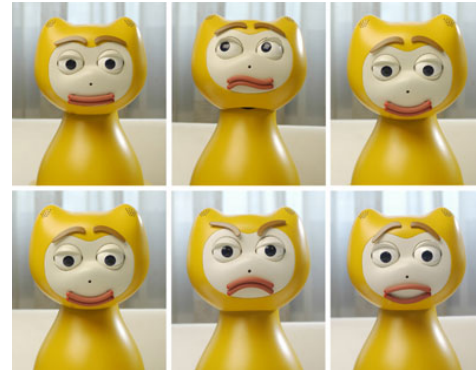


Figure 2: iCat facial expressions [10]

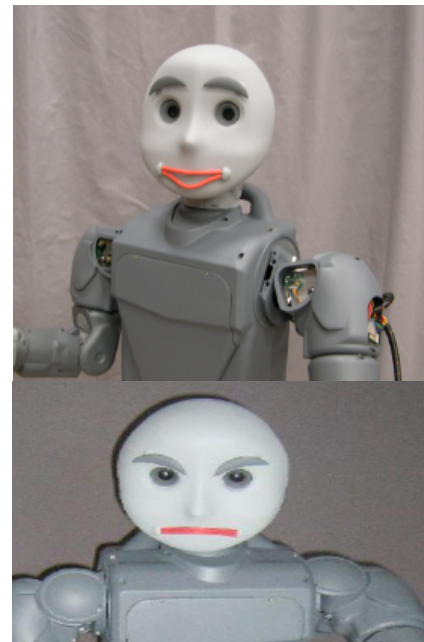


Figure 3 : Bandit II facial expression @Interaction lab

Scientists have here crossed psychology and technology. iCat has been used as a game buddy for children, or as a TV assistant. Actually, children preferred to play with iCat rather

than to play alone, with a computer. Here is highlighted the need for people to have a companion animal. They finally concluded that iCat and his emotions have a significant effect on how users interact with robots. They prefer to interact with an intelligent robot, because they prefer to communicate with someone else. That's also why SAR are important for elderly. With their facials emotions, people feel close to them, and this kind of robots try unconsciously to establish a relationship.

This kind of relationship can be distant when users don't know what the robot is capable of, but after can be turned into a trust relationship. Indeed, this SAR can react to users' emotions, voice, mood, but will remain as a pet. They would never have the capacity to judge the elderly, that's why they can be very helpful for people who have difficulties to elaborate words and who are afraid of what other people think.



© 2014 Aldebaran

Humanoid robots are actually not assistive enough to be useful in psychological training like animal robots. Their application tend to be more related to the nursing homes staff, helping with the grooming tasks or reminding pill-taking. Actually, these assistive robots act more like computers or digital schedules that socially assistive robots. Kompai for example is an assistive robot developed in 2015 which can alert dedicated services if the senior person falls. These robots are indeed extremely limited, and humanoid robotic faces some issues directly linked with the lack that the field is still very young and unused.

3) PERFORMANCE EVALUATION

In order to evaluate the performance of socially assistive robots in geriatrics or with a larger public, many studies were completed. The performance is investigated via quantitative and/or qualitative measures. In general, quantitative measure are less common, as they are more difficult to perform and less accessible. The studies that carried out quantitative objective measures exploited mostly MRI (surface-based morphometry (SBM)) [2], EEG and salivary and urinary hormone levels (see [1]). On the contrary, qualitative and

more subjective measures are easier to obtain, and questionnaires are widely used in several studies, to enquire the effects of socially assistive robots both on the elderly people and the care assistants (see [1]).

We decided to focus on questionnaires. Quite recently, a new evaluative toolkit has been developed to measure several aspects of a socially assistive robot [10]. This toolkit is based on the technology acceptance model (TAM) by Davis in 1989. TAM provides several scales measuring utility and ease of use, it has been extended to social versions as UTAUT. Researchers created a questionnaire to evaluate several constructs as anxiety, facilitating conditions, trust...



Figure 4 – Cap Digital FUI ROMEO project (2009-2012)

This table presents constructs and their definition. Some constructs are strongly correlated as for example: Intention to use and actual usage.

TABLE 1 – MODEL OVERVIEW

Code	Construct	Definition
ANX	Anxiety	Evoking anxious or emotional reactions when using the system.
ATT	Attitude	Positive or negative feelings about the appliance of the technology.
FC	Facilitating conditions	Objective factors in the environment that facilitate using the system.
ITU	Intention to use	The outspoken intention to use the system over a longer period in time.
PAD	Perceived adaptability	The perceived ability of the system to be adaptive to the changing needs of the user.
PENJ	Perceived enjoyment	Feelings of joy or pleasure associated by the user with the use of the system.
PEOU	Perceived ease of use	The degree to which the user believes that using the system would be free of effort
PS	Perceived sociability	The perceived ability of the system to perform sociale behavior.
PU	Perceived usefulness	The degree to which a person believes that using the system would enhance his or her daily activities
SI	Social influence	The user's perception of how people who are important to him think about him using the system
SP	Social presence	The experience of sensing a social entity when interacting with the system.
Trust	Trust	The belief that the system performs with personal integrity and reliability.
Use	Use/Usage	The actual use of the system over a longer period in time

Figure 5 – Extract of the UTAUT implementation with elderly people [10]

Following questionnaires form a base to evaluate the robot integration in a social context.

The senior answers in a grid where the possible answers are:

- 0 : I disagree
- 10 : I totally agree

TABLE 2 – QUESTIONNAIRE

ANX	1. If I should use the robot, I would be afraid to make mistakes with it
	2. If I should use the robot, I would be afraid to break something
	3. I find the robot scary
	4. I find the robot intimidating
ATT	5. I think it's a good idea to use the robot
	6. the robot would make my life more interesting
	7. It's good to make use of the robot
FC	8. I have everything I need to make good use of the robot.
	9. I know enough of the robot to make good use of it.
ITU	10. I think I'll use the robot during the next few days
	11. I am certain to use the robot during the next few days
	12. I'm planning to use the robot during the next few days
PAD	13. I think the robot can be adaptive to what I need
	14. I think the robot will only do what I need at that particular moment
	15. I think the robot will help me when I consider it to be necessary
PENJ	16. I enjoy the robot talking to me
	17. I enjoy doing things with the robot
	18. I find the robot enjoyable
	19. I find the robot fascinating
	20. I find the robot boring
PEOU	21. I think I will know quickly how to use the robot
	22. I find the robot easy to use
	23. I think I can use the robot without any help
	24. I think I can use the robot when there is someone around to help me
	25. I think I can use the robot when I have a good manual.
PS	26. I consider the robot a pleasant conversational partner
	27. I find the robot pleasant to interact with
	28. I feel the robot understands me.
	29. I think the robot is nice
PU	30. I think the robot is useful to me
	31. It would be convenient for me to have the robot
	32. I think the robot can help me with many things
SI	33. I think the staff would like me using the robot
	34. I think it would give a good impression if I should use the robot.
SP	35. When interacting with the robot I felt like I'm talking to a real person
	36. It sometimes felt as if the robot was really looking at me
	37. I can imagine the robot to be a living creature
	38. I often think the robot is not a real person.
	39. Sometimes the robot seems to have real feelings
Trust	40. I would trust the robot if it gave me advice.
	41. I would follow the advice the robot gives me.

Figure 6 – Extract of the UTAUT toolkit questionnaires [10]

The results of the questionnaire are scores for each constructs. This performance evaluation was tested on elderly from 65 to 94 with the iCat socially assistive robot, and revealed which construct is significant to build a socially assistive robot.

Even if evaluation through questionnaires seems to lack scientific grounding, this technic is the most widespread and easy to use, especially when the targeted public is the elderly. We decided to make a quick review of the UTAUT model and its applications because it is complete and adapted to socially assistive robot.

Another study uses a clinical protocol to evaluate changes in senior people with dementia after interaction with Paro in nursing homes [11].



Figure 7 – Paro robot [11]

This study also analyses quantitative data, evaluating the mean time of interaction between Paro and the elderly.

III. DISCUSSION

A. BENEFICIAL EFFECTS OF SOCIALLY ASSISTIVE ROBOTS

From a general point of view, the studies conducted so far have identified improvements on multiple domains as a result of using socially assistive robots, and more specifically companion robots with the elderly. Rabbitt and colleagues [1] give an overview of quite a few studies that investigated the effects of integrating socially assistive robots in the everyday life activities of elderly people. The cited studies found that, as a consequence of the interaction with such robots, beneficial effects were reported from a psychological, a physiological and a social point of view.

From a psychological perspective, a study reported that both the elderly participants and the nursing staff benefited some improvements: on the one hand, the elderly participants showed improved feelings and improved ability to overcome stress; on the other hand, the nursing staff experienced lower stress levels since less supervision was required by the elderly people while interacting with the robot. Other studies found reduced depression, better feelings, improved general atmosphere in the health service facility due to increased social interaction among the elderly people and with the caregivers.

In turn, from a physiological point of view, this psychological better state improved the reactions of the subjects' vital organs to stress.

In particular for dementia, Rabbitt and colleagues [1] cite some dedicated studies that investigated the effect of robot therapy and robot assisted activity on elderly patients with dementia. As a result of interacting with the robots, patients showed behavioral improvements. In particular, a study using Paro reported better cortical neurons activity, and the effect

was found to be even more prominent for those patients who particularly liked Paro.

Another study [11] that investigated the effect of Paro on a comparable dementia population also found an improved neuronal activity, reduction of anxiety levels measured by skin conductivity and cardiac rhythm.

As regarding therapeutic robots, Kim and colleagues [2] conducted a controlled trial to investigate the effect of cognitive training on cortical thickness. They found that cognitive training provided attenuation of age related cortical thinning in the fronto-temporal association cortices and the group that integrated a robot in the entrainment showed less thinning than classical cognitive training, thanks to the fact that robots provide multi-domain training as opposed to the one specific domain that classical training usually provide.

B. LIMITATIONS OF THE STUDIES

As mentioned above, more objective quantitative measures are to be gathered. Not only are they paramount for research purposes, but also they play a fundamental role together with qualitative measures to justify the price socially assistive robots have. In a world where progressively decreasing budgets are available in healthcare, managing the financial budgets is crucial and socially assistive robots are very expensive, hence the need of solid evidence, both qualitative and quantitative, of their beneficial effects.

Also, the high manufacturing costs of companion robots prevent general accessibility to this technology which in turn prevents their widespread use. As a consequence, it is important that through further research the manufacturing costs are lowered, making it possible for more healthcare facilities to afford buying and integrating companion robots in their services.

Further research is also needed to investigate the best and most appropriate ways of introducing a socially assistive robot, and specifically a companion robot, depending on the culture of the user. Furthermore, it is still to be clarified how a companion robot is to be used to produce the best possible effects and consequently to fully exploit its beneficial potential for the care of elderly people.

The role of a companion robot is not completely defined yet: it is important to correctly place it in the relationship with an elderly person and with the healthcare staff. A socially assistive robot is not meant to substitute any human caregiver, nor any real animal per se, instead it is meant to integrate the human action, to help providing a better therapy, to help overcoming the problems the interaction with a real animal presents. Hence, it is important to better understand the social mechanisms underlying the interaction between the human user and the socially assistive robot to provide all the users the knowledge needed to correctly integrate this kind of robot in their lives.

Lastly, the use of robots that integrate components as e.g. cameras or microphones and that may be connected to the internet might be dangerous, since such devices may be exposed to hacking. It is then important to make sure that all the data are rendered safe from those possible attacks.

C. CONCLUSIONS

Although SAR is a very recent domain of research, socially assistive robots and in particular companion robots are showing promising positive psychological, physiological and social effects on elderly people.

Evidence is mostly based on qualitative measures. More quantitative evidence is needed, although this kind of measures is more difficult to collect.

Also, the role of socially assistive robots remains unclear and uncertain. Most people are unfamiliar with these new non-human beings and lack of knowledge often generates fear. What is certain and important to communicate is that these robots are not meant to replace any human in any step of the healthcare process or in everyday life. They are meant to help. What is not perfectly defined yet is the kind of relationship the human user should have with them.

In conclusion, socially assistive robots are promising, but have yet to find their place at the side of the human, just as pets had to find theirs when they first were domesticated.

REFERENCES

- [1] Rabbitt SM, Kazdin AE, Scassellati B: Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology review*, 2015;35:35:48.
- [2] Shibata T, Wada K: Robot Therapy: A New Approach for Mental Healthcare of the Elderly - A mini-review. *Gerontology*, 2011;57:378-386.
- [3] Kim GH, Jeon S, Im K, Kwon H, Lee BH, Kim GY, Jeong H, Hua NE, Seo SW, Cho H, Noh Y, Park AE, Kim H, Hwang JW, Yoon CW, Kim HJ, Ye BS, Chin JH, Kim JH, Suh MK, Lee JM, Kim ST, Choi MT, Kim MS, Heilman KM, Jeong JH, Na DL: Structural Brain Changes after Traditional and Robot-Assisted Multi-Domain Cognitive Training in Community-Dwelling Healthy Elderly. *PLoS ONE* 2015;10(4).
- [4] Sasa Y, Aubergé V: Caractéristiques prosodiques de la "glu socio-affective" de l'interaction face à face : un robot compagnon médiateur d'un habitat intelligent pour personnes âgées. 3rd SWIP-Swiss Workshop on Prosody, 2014.
- [5] Breazeal C, Scassellati B : How to build robots that make friends and influence people. *Intelligent Robots and Systems*, 1999. IROS '99. Proceedings.
- [6] Duffy BR: Anthropomorphism and Robotics.
- [7] Kidd CD, Taggart W, Turkle S: A Sociable Robot to Encourage Social Interaction among the Elderly. *Proceedings 2006 IEEE International Conference on Robotics and Automation*.
- [8] Foster EB: Effect of the Human-Animal Bond on Nursing Home and Assisted Living Facility Residents, 2005. University of Tennessee Honors Thesis Projects. http://trace.tennessee.edu/utk_chanhonoproj/852
- [9] Banks MR, Willoughby LM, Banks WA : Animal-assisted therapy and loneliness in nursing homes : use of robotic versus living dogs. *J Am Med Dir Assoc*, 2008 Mar;9(3):173-7.
- [10] Heerink M, Kröse B, Evers V, Wielinga B: Measuring acceptance of an assistive social robot: a suggested toolkit. *The 18th IEEE International Symposium on Robots and Human Interactive Communication*, 2009.
- [11] Donose GP, De Poix A, Floriot A, More C, Seknazi A: Evaluation en EHPAD de la technologie et l'intelligence artificielle du robot PARO chez les résidents âgés avec démences modérés à sévères et troubles de comportement

Integration of Episodic Memory in Companion-Robots to Enhance Social Interactions

Fanny BROQUA

Engineering Student in Cognitive Sciences
Grenoble Institute of Technology - Phelma
Grenoble, France
fanny.broqua@gmail.com

Argitxu CALDICHOURY

Engineering Student in Cognitive Sciences
Grenoble Institute of Technology - Phelma
Grenoble, France
argi.caldi@gmail.com

Abstract— In this paper, we address the question of implementing an episodic memory into companion-robots to improve the quality of Human-Robot interaction. First we describe the different types of human memory to conclude that episodic memory is the most important one to develop, mainly because it concerns early stages of memory loss. Then, we illustrate the benefits of integrating this memory system for social interactions and healthcare. Finally, we discuss the psychological effects and ethical issues.

Keywords— *Human-Robot Interaction, Human Memory Sytems, Episodic Memory, Social Interactions, Healthcare*

I. INTRODUCTION

Human-Robot Interaction (HRI) is a young discipline that has been developing over the past few years with the appearance of complex robots and people's exposure to these robots in their daily lives [1].

Interactions with social robots increasingly become a part of our daily lives, whether they are short or long-term ones. With the first one, robots only have three minutes to adjust to the user to avoid rejection (ex. Spontaneous interactions in public spaces with limited time), while with the second one, there is an evolution of the human-robot relationship.

For long-term interaction, it is therefore essential that robot companions can interact with us multiple times and over a long period of time. Right now, some robots are only performing predefined rigid sets of repetitive behaviors, which leads to a decrease of interest among users because of the lack of interaction [2].

To match our social and cognitive needs and increase the quality of the interaction, the challenges of long-term robot companions are therefore to be more familiar and natural. They should also have the ability to create and develop affection with their users to maintain engaging interpersonal relationships.

In psychology, the memory is the faculty of the mind to encode, store and retrieve past experiences. Without it, the learning and development of language, relationships and personal identity would be impossible. Therefore, memory is

vital to human being. The same principle applies to long-term robot companions [2], which is why memory is thought to be the solution to enhance the HRI.

II. STATE OF ART

A. *Creating a human-inspired memory*

a) *Human memory systems*

Memory is the ability of our brain to retain information in order to use it later and the human memory is divided into three systems according to a major criterion that is the duration of retention of the information, themselves divided into sub-systems (Figure 1):

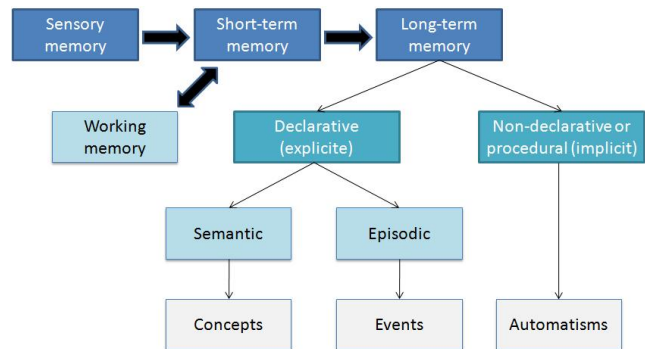


Figure 1 : Memory systems

• A sensory memory:

It corresponds to the ability to remember, for a very short period of time (from a few milliseconds to one or two seconds) the sensory information that reach us most of the time unconsciously using our 5 senses (hearing, smell, touch, sight and taste). It allows remembering images or noises without realizing it and so makes us possible for example to remember a specific picture during our childhood or the taste of the cake our grandmother made for us. Sensory memory is

essential for the storage of information in the next memory system addressed.

• A short-term memory:

The short-term memory (or working memory) is the memory of the present. It can hold information for a few seconds or even a few tens of seconds and we solicit it all the time, for example to remember a phone number just the time to write or compose it or when a face encountered on the street remains in memory for a short time. The storage capacity of short-term memory is limited to about seven data, which is the "magic number" that corresponds to the number of elements that can be memorized simultaneously in the short term [3].

In most cases, the neurobiological mechanisms associated with working memory do not allow the long-term storage of this type of information: their memory is quickly forgotten. Nevertheless, there are interactions between the working memory system and those of the long-term memory. They allow the memorization of certain events and, thus, to remember old memories when in front of present situations, in order to adapt better.

• A long-term memory:

With long-term memory, the retention, assimilation and transcription of information are unlimited and can last a lifetime [3]. All our memories, our learning and all the information that we have made conscious effort to remember are recorded and stored in long-term memory. However, unlike the hard disk of a computer for example, long-term memory does not store a copy of the perceived and/or manipulated information but instead it interprets, recomposes and reorganizes them.

Long-term memory itself can be divided into three memory systems: declarative (or explicit) and non-declarative (or implicit).

The implicit one, which corresponds to the procedural memory, is unconscious and is the memory of automatisms [3]. It is a very reliable type of memory that allows people to remember a sequence of specific gestures for years – even if they are not used for a long time – without having to relearn each time. For example when cycling, swimming, driving, tying shoelaces, etc., the movements are made without conscious control and the neural circuits are automated.

On the contrary, the declarative one is accessible to consciousness and expressed through language and in the form of images [3]. We find in this memory the general (semantic type) and specific (episodic type) knowledge.

- Semantic memory allows the acquisition of general knowledge about ourselves (our history, our personality) and the world around us (geography, politics, nature, social relations, etc.) [4]. It is the

memory of “definitive knowledge” and concerns personal data accessible to our conscience that we can express. This memory is necessary for learning and using a language.

- Episodic memory is a system that receives and stores information about episodes or temporally dated events (autobiographical events) [4] as well as the spatio-temporal relations that unite them. We access it when we remember the past events experienced with their context in a precise way (date, place, emotional state, etc.). Episodic memory concerns for example the memory of events that took place several days back (Where was I last Sunday?) or objects seen a few minutes ago (Where did I put my keys?).

The age-related decline in memory has been increasing over the past decades but not all memory systems previously discussed are affected in the same way [5]. Indeed, procedural and semantic memory systems are partially or totally spared by the effects of age while the ability to manipulate information in working memory and episodic memory (Where, When and How and event happened) is more vulnerable to the effects of aging. In particular, the decline of episodic memory seems to be one of the fundamental characteristics of cognitive aging [5].

The difficulties of episodic memory are those that are most emphasized by the elderly themselves. Overall, mainly two types of processes involved in the functioning of episodic memory are affected: associative and strategic processes. The latter correspond to the cognitive operations implemented during the encoding and the retrieval steps of memorization – discussed further in II.A.b) – while the associative processes correspond to the mechanisms allowing to associate different characteristics of an information within the same memory representation [5]. Those can intervene both during the encoding and the retrieval steps.

Episodic memory being thus the most vulnerable memory system, we think that to enhance the quality of the long-term HRI, it becomes necessary to implement a human-like episodic memory into robot-companions. This would allow creating a solid bond between the robot and its users as the robot would be able to understand the needs of the user and also help him in his daily life. This would apply to elderly users with a decline in memory but also other cases discussed further in II.B.

b) Human-like memory characteristics

When constructing an episodic memory, design decisions such as functional and architectural requirements, must be taken into account. While functional requirements are derived from the definition of episodic memory itself, the architectural requirements result from the decision to include this episodic memory system into architecture [6].

The episodic memory system's process can be divided into three steps that will form the architecture of the implemented memory into robot-companions: the encoding, the storage and the retrieval (Figure 2). Based on the work of Nuxoll (2007)[6], taken over in 2010 by Kasap and Magnenat-Thalmann [7], these three steps can be defined as following.

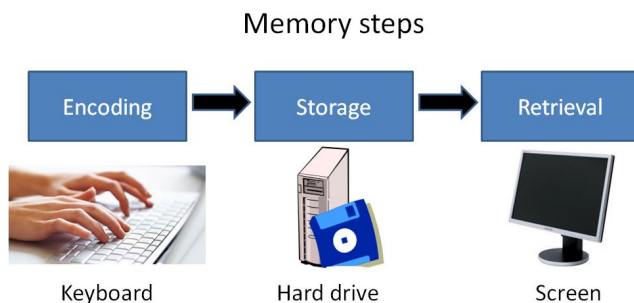


Figure 2 : Memory steps illustration with a computer

- **Encoding:** it consists in processing information to make it a true souvenir. With it, it is possible to give meaning to the information. It answers the question: *how a memory is captured and stored?*[6]

- **Storage:** it corresponds, as its name indicates, to the storage of information in the appropriate area of the brain. Some factors may modify this retention of information, in particular, the greater the emotional burden the longer the storage will be. The storage time remains unlimited unless you do not reuse this information regularly. To talk about storage in memory, we often talk about consolidation which refers to the period during which the brain automatically repeats information, without us noticing, until it is sufficiently anchored in our memory to be retained for a long time. In this step, it is sought to determine: *how a memory is maintained?*[6]

- **Retrieval:** this step is about retrieving information already stored whenever it is needed. At the time the event is remembered, the different constitutive elements of the souvenir that is concerned are "reassembled". It deals with: *how a memory is retrieved?*[6]

The encoding and storage phases are really important because if one of them is badly performed, this means that there will be a problem in the retrieval phase leading to what is called a blackout.

These three high level steps could be even more detailed and this way we would get a larger set of steps [6]. Indeed, Nuxoll has come up with this idea where the larger set of steps includes a framework of all the points where a major design decision has to be made in the implementation of an episodic memory. Along with that, Nuxoll defined functional requirements (Architectural, Automatic, Auto-noetic and Temporally Indexed) and architectural requirements (Task Independent, Low Resource Demand, Non-Invasive

Integration) [6] to be taken into account in the design and implementation of the artificial human-like episodic memory into robots. He also affirmed that cognitive capabilities such as Sensing, Learning and Reasoning have to be implemented in order to get a better HRI.

Kasap and Magnenat-Thalmann (2010) [7] came up with a similar idea, that is to create an episodic memory-based long-term affective interaction between a robot-companion and its user. Their model is based on face and speech recognition and its architecture is given as an example on Figure 3.

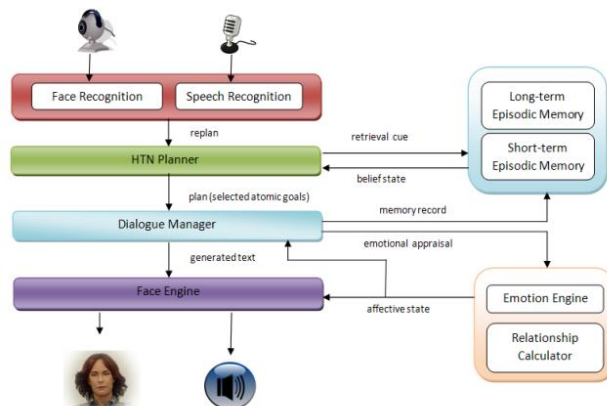


Figure 3 : Architecture of the interaction [7]

Several models [2], [8], [9], [10] of architecture are currently being developed and within the next few years, we'll probably be able to use them on a daily basis for social interactions as developed in the next part of the article.

B. Episodic memory for the improvement of social interactions

a) *Memory prosthesis for elderly users*

Memory loss is common among elderly people; it causes independency loss, leading to health and safety risks. The EU project ACCOMPANY is currently working on a robot as a memory prosthesis, which is capable of episodic memory visualization, that could be used to help the elderly remembering past events of their human-robot interactions.

The EU population is estimated to be 50% over 65 years old around 2060 [9]; this research is therefore significant for the years to come. Its goal is to allow older people to live at home for as long as possible and to decrease the rapid development of diseases such as Alzheimer and Dementia.

Home care robots should be able to help elderly people in their daily lives to improve their health conditions and strengthen their independency in the foreseeable future. Indeed, with the increase of the average age of a population, the number of people with dementia will rise from 66 million in 2030 to 115 million in 2050 [9]. The main issue of dementia

is a long term memory loss at early stage or in other words an episodic memory loss: the robot's memory visualization is there to recover the loss by interacting with the human concerned.

b) Companion dialogue

Social robots can also be used in aspects of everyday life and thus, communication between human and robot is very important. Gregor Sieber and Brigitte Krenn, two researchers from the Australian Research institute for Artificial Intelligence, are currently working on enhancing companion dialogue by improving the episodic memory of the robot [11]. Their study consists in developing a communication model which connects episodic memory, semantic memory and dialogue: robotic mental and memory models must therefore be linked to their expressive behaviour. To create an episodic memory element, they use RDF-based data stores to implement representations of the cognitive model knowledge and the robot's task knowledge. The Resource Description Framework that is used is a standard model for data exchange on the Web. It owns features that facilitate data merging even though the underlying schemas differ and it supports schema evolution over time without requiring all the data consumers to be modified.

They have achieved to build a companion capable to analyse a present situation with one stored in their memory. The robot's « brain » is a recollection of information and preferences of the user. The dialogue between the agent and the user is based on past interactions they have had which enables the robot to constantly select the next dialogue depending on what is said. Plus, the RDF knowledge representation graphs allow the agent to have a grounded knowledge of its experiences.

c) Emotional memories for humanoid robots

M.S Ryoo, Yong-hoSeo, Hye-Won Jung and H.S. Yang from the Artificial Intelligence and Media Laboratory of Korea [12] are developing an affective dialogue communication system which consists in modelling an emotional memory for Humanoid Robots. Indeed, sharing emotional experience with a human strengthens the relationship. They created AMI, a human-like system which is able to start a dialogue with humans based on emotional memories. The framework was based on the human brain structure: the model consists in 5 sub-systems (Figure 4) which allow AMI to preserve explicit memory. A memory system is integrated as it improves AMI's social skills as well as communication capacities: the robot is able to represent itself and its communication partner in a

current situation. This system also allows AMI to have an emotional memory of topics and users.

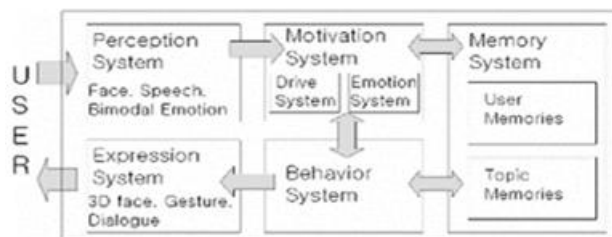


Figure 4 : Framework of the affective system that enables dialogue interaction [12]

III. DISCUSSION

Social robots have already been developed, such as the Paro robot (Figure 5) created by Wada and Shibada [13], [14], [15] which is a seal-shaped robot capable of responding to certain stimuli (for example touch, sound and light) and act on its own due to its internal specifications. This therapeutic robot, which is used in care houses, has a somewhat memory, as it can recognise frequently used words (for example its name) and thus adapt its behaviour according to it. Results of using Paro showed that it could lead to a reduction of stress among elderly users.



Figure 5 : Paro the seal-shaped robot [16]

Conversational robots that are capable of communicating with humans also exist. Robovie [17] (Figure 6), a humanoid social robot, has been created to engage conversation with its user, by asking simple child-like questions and socially greeting.



Figure 6 : Robovie, the humanoid social robot [16]

These robots are capable of enhancing interactions with a human being, and even more so for the long-term relationship. Nonetheless, an important factor is missing in both of those examples: memory implantation.

As said in part II., memory is the key to a stronger human-robot bond for it will allow the robot to have a behavioural coherence and credibility, thus enhance the perception of intelligence and quality of the interaction with the robot [16]. Furthermore, adding a memory to a robot will mean that the robot will be able to adapt and shape itself according to a user, so that each user can have an individualized companion. This personal bond will increase the human's attachment to its agent and thus reinforce its social purpose.

The study of the cultural, behavioural, and sociological aspects of spatial distances between individuals allowed Koay et al. [18] to discover that a trust bond which makes the users more at ease with the interaction relies simply on proxemics and directed responses of the robot according to the user's expectations. Indeed, a complex robot with many perceptual capabilities isn't necessary for the user to feel in control of its robot and thus, in a trusting environment.

IV. OTHER ISSUES

A. *Psychological effects*

The human-robot interaction has a positive effect on human's psychosocial capacities. Indeed, a study from H. Robinson et al. [19] shows that a 12-week intervention of Paro in an elderly care-home, with a twice a week interaction, enabled to decrease the participants overall loneliness compared to a control group (which had not participated in the interactions). Not only did Paro have a positive effect on each resident's mood, but it also helped to achieve a better atmosphere in the entire test group (Figure 7). As loneliness fades away, other aspects of their lives improve as social interaction is necessary to maintain a healthy mind.

The companion robot is comparable to a live animal which has the same effects on people, though a robot is more suited for care homes. Indeed, Paro is capable of triggering conversations between elderly people as it is a shared interest. However, this positive effect on the residents could be strengthened by a tighter bond between the robot and the human: what Paro is missing is an episodic memory. In fact, having a memory and communication skills would enable the robot to join a conversation between two residents and lead the discussion according to what the users prefer or have in common. Nonetheless this would mean that the robot should be able to adapt to each resident which is a complex aspect of robots' capacities. There is still a long way to go.



Figure 7 : The care home residents interacting with Paro [19]

B. *Ethical issues*

Every scientific study needs to raise the question about ethical issues: what are the ethical consequences of having a daily interaction with a companion robot? First, it is the strength of the human-robot bond that is a concern as previous technological devices have been subject to such an attachment, as in the case of the Tamagotchi in the 90's [20]. One might ask if an over-attachment is healthy, and if it will after a while decrease social interactions like mobile phones or other personal objects. Indeed, a home-care robot is present in the user's everyday life and will therefore create an intimate and strong bond, even more so if it is capable of personalized interaction.

Ideally, these new memory implanted robots should be able to live with their owners without any supervision. Thus, the issue of security comes up and it concerns not only the handle of the device by adults as well as children, but also the protection of one's data collection in the robot's memory: where is it stored? How private is it? Ethical memory models have been set up to organize and control digital artefacts that autonomously record and store the user's data which is then used for future human-robot communication. For further information, consult [21].

V. CONCLUSION

In this paper we have presented the different types of human memory that one can describe (semantic, short-term and long-term memory) and thus concluded that an episodic memory implementation could be the next step to develop companion home care robots. We have also seen that memory can be used to increase health conditions, and the human-robot dialogue.

Home care robots have already helped residents to improve their quality of life as they help to soothe loneliness and depression, and to better social interactions. A few models have been developed to direct the dialogue between the robot and the user so that the robot structures its responses according to the data it has stored about its owner: the interaction quality is improved by personalizing the interaction. Finally, integrating emotions in the robot's memory records also enables to strengthen the bond with its user. Nonetheless, there are some ethical issues concerning the addiction to the device and security that need to be taken in consideration.

Developing an episodic memory for companion robots sound promising and seems to have a lot to offer, but the road is still long to achieve a perfect memory implemented robot, so that many studies are determined to enhance current work for a better future.

REFERENCES

- [1] S. Daselaar and R. Cabeza, "Age-Related Decline in Working Memory and Episodic Memory," 2013.
- [2] W. C. Ho, K. Dautenhahn, M. Y. Lim, P. A. Vargas, R. Aylett, and S. Enz, "An initial memory model for virtual and robot companions supporting migration and long-term interaction," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, pp. 277–284.
- [3] V. La Corte, "Systemes de mémoire et distorsions mnésiques: approches neuropsychologique et neurophysiologique," Université Pierre et Marie Curie-Paris VI, 2012.
- [4] E. Tulving, "Episodic and semantic memory 1," *Organ. Mem. Lond. Acad.*, vol. 381, no. 4, pp. 382–404, 1972.
- [5] D. TOURS, "Déclin de la mémoire épisodique au cours du vieillissement normal: étude des médiateurs du déficit des processus stratégiques," Université de Clermont-Ferrand, 2010.
- [6] A. M. Nuxoll, "Enhancing intelligent agents with episodic memory," Citeseer, 2007.
- [7] Z. Kasap and N. Magnenat-Thalmann, "Towards episodic memory-based long-term affective interaction with a human-like robot," in *19th International Symposium in Robot and Human Interactive Communication*, 2010, pp. 452–457.
- [8] M.-L. Sánchez, M. Correa, L. Martínez, and J. Ruiz-del-Solar, "An Episodic Long-Term Memory for Robots: The Bender Case," in *Robot Soccer World Cup*, 2015, pp. 264–275.
- [9] W. C. Ho, K. Dautenhahn, N. Burke, J. Saunders, and J. Saez-Pons, "Episodic memory visualization in robot companions providing a memory prosthesis for elderly users," 2013.
- [10] W. C. Ho, K. Dautenhahn, M. Y. Lim, and K. Du Casse, "Modelling Human Memory in Robotic Companions for Personalisation and Long-term Adaptation in HRI," in *BICA*, 2010, pp. 64–71.
- [11] G. Sieber and B. Krenn, "Towards an episodic memory for companion dialogue," in *International Conference on Intelligent Virtual Agents*, 2010, pp. 322–328.
- [12] M. S. Ryoo, Y. Seo, H.-W. Jung, and H. S. Yang, "Affective dialogue communication system with emotional memories for humanoid robots," in *International Conference on Affective Computing and Intelligent Interaction*, 2005, pp. 819–827.
- [13] K. Wada and T. Shibata, "Robot therapy in a care house—its sociopsychological and physiological effects on the residents," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 2006, pp. 3966–3971.
- [14] K. Wada and T. Shibata, "Robot therapy in a care house—results of case studies," in *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp. 581–586.
- [15] K. Wada and T. Shibata, "Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 972–980, 2007.
- [16] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 291–308, 2013.
- [17] A. M. Sabelli, T. Kanda, and N. Hagita, "A conversational robot in an elderly care center: an ethnographic study," in *Proceedings of the 6th international conference on Human-robot interaction*, 2011, pp. 37–44.
- [18] K. L. Koay, D. S. Syrdal, M. L. Walters, and K. Dautenhahn, "Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study," in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 2007, pp. 564–569.
- [19] H. Robinson, B. MacDonald, N. Kerse, and E. Broadbent, "The psychosocial effects of a companion robot: A randomized controlled trial," *J. Am. Med. Dir. Assoc.*, vol. 14, no. 9, pp. 661–667, 2013.
- [20] L.-R. Bloch and D. Lemish, "Disposable Love The Rise and fall of a virtual pet," *New Media Soc.*, vol. 1, no. 3, pp. 283–303, 1999.
- [21] P. A. Vargas *et al.*, "Advocating an ethical memory model for artificial companions from a human-centred perspective," *AI Soc.*, vol. 26, no. 4, pp. 329–337, 2011.

Learning by Teaching: a Paradigm for the Use of Robots in Education

Erika Godde and Martina Jurcova,
M2 Sciences Cognitives, Phelma Grenoble INP, Grenoble, France
erika.rassat@free.fr
mmjurcova@gmail.com

Abstract—In the present paper, we provide a state-of-the-art review of the use of Nao robot within *Learning by teaching* paradigm for young children education. After a general presentation of the *Learning by teaching* paradigm, we briefly present the two most advanced studies in this framework, using Nao robot for English vocabulary learning by Tanaka et al. at University of Tsukuba in Japan and using Nao for handwriting learning by Dillenbourg's group at Ecole Polytechnique Fédérale de Lausanne in Switzerland. Afterwards, we develop in more details the technical aspects necessary for robot's design, control and interaction with children. Finally, we present the social dimensions to be implemented and remaining challenges in order to enable efficient long-term child-robot interactions in educational context.

Index Terms—robots, education, *Learning by teaching*, Nao, child-robot interaction, engagement, long-term interaction

I. INTRODUCTION

The use of robots in the education field is not recent. They have been used for a few decades, mostly to teach programming skills to university students. However the recent development of attractive, easy-to-use, and above all, very affordable humanoid robots, allowed the domain a huge leap forward during the last decade. Robots are now used from kindergarten to master's degree in a wide range of areas.

Mubin et al. published in 2013 a review of the applicability of robots in education [1]. They identified the main areas using robots as a teaching help. The most displayed areas are still science, technology, engineering and mathematics (STEM). But robots are as well more and more used in language learning, and also in some more confidential fields as presented below. In scientific fields, the mechanical and electronical kits are the most represented robot types, especially in higher grades and at university. Humanoid robots recently emerged for programming learning with older students, but also with younger children. The use of humanoid robots helped extending the application field of robots in education, using them not only as a "kit" but also as a social asset. Mubin et al. present in their review most of the paradigms used with robot in a classroom. In fact the paradigm depends on the assigned role for the social robot. It could be:

- a teacher: in a "one-to-many" interaction, the robot may help the students to remember their vocabulary.
- a tool: in a "one-to-many" interaction, students program the robot or play games with it.
- a peer: often in a "one-to-one" interaction, a student and a robot solve a problem together, or the robot says "well-done" when the student gives a good answer.
- a tutoree: a protégé, in a "one-to-one" interaction, the student teaches the robot to do something or to answer correctly to some questions.

That last paradigm, named *Learning by teaching*, particularly drew our attention. We have therefore chosen to focus on *Learning by teaching* a robot, and to review the recent works conducted on it.

In the first section, we will define the *Learning by teaching* paradigm, explaining its origins and benefits for the students. We will also review robots assets in this paradigm. We will next present the actual works of two teams : one of them engaged in a language learning project and the other one in improving the handwriting of children. We will then develop more in details some of the technical aspects that need to be taken into account when designing a "learning by teaching a robot" situation. Finally we will discuss the issues raised by the perspective of that kind of long-term relationship between a child and a robot at school.

II. Learning by teaching

A. Presentation of the paradigm

The *Learning by teaching* (LbT) is a well known paradigm in education sciences since the 18th century and the beginning of the "Education Nouvelle" movement in France. It's still used today in many classrooms all over the world with, for example, Montessori and Freinet pedagogies. Those pedagogies are based on active methods, like *Learning by doing*, and cooperation, like *Learning by teaching*. These methods are enabled by mixing children of various ages and skills. [2]

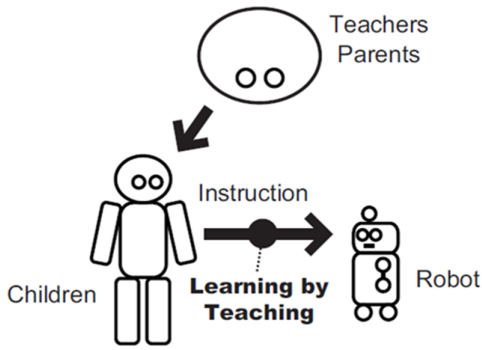


Fig. 1. The schema of LbT using a robot [5].

In the *Learning by teaching* paradigm, a less-skilled child, the tutoree, asks an advanced child, the tutor, to teach him a concept he doesn't master yet. The tutor teaches the tutoree the new concept, answers his questions, corrects his achievements. If the benefits for the tutoree are obvious, they are in fact even greater for the tutor. *Learning by teaching* proved indeed many benefits for the tutor. In fact, teaching a tutoree implies a better grasp of material through rephrasing, explaining connections, answering questions, searching for new ways of teaching, adapting to the tutoree's demands and difficulties.

Not only teaching implies a deeper understanding and a more effective learning of the taught subject, it also benefits to learning in general. The teaching activities make the tutor think of how to teach and therefore how to learn. This metacognition activity improves the tutor's learning skills. [2]

Finally this paradigm increases as well the self-esteem of the tutor and his motivation. A study of Chase et al. on the use of the *Learning by teaching* with a teachable agent shows that the tutor makes more efforts to teach his tutoree than to learn for himself, this is called the "protégé effect" [3]. Both examples developed in the second section rely on this effect. [4]

B. Specific benefits of Learning by teaching a robot

The LbT can also be adapted to be used with a robot as a peer. The general principle can thus be schematized as in Figure 1.

Chase et al. conducted a study in 2009 on learning by teaching a teachable agent (TA) [3]. They developed a program named Betty's Brain. 8th-grade students used Betty's Brain to teach a character to answer questions about fever mechanisms. This study reveals that students tend to make more effort and to spend more time working when teaching the agent than learning for themselves. They had more motivation to learn for their agent

than for themselves. It seems that the "protégé effect", increasing the benefits of peer teaching, works just as well with a teachable agent.

Chase et al. went further and showed more specific benefits of the use of a teachable agent (TA):

- Teaching a TA, instead of a peer student, has the benefit of making the learning process explicit. The student has the opportunity to see what the TA "thinks". This possibility enhances the metacognition process mentioned above.
- The TA acts as an egoprotective buffer in case of failure. When the TA fails to answer a question, students believe that their teaching is only partly responsible. In fact, they consider the TA as sentient and partly responsible too. This responsibility sharing has a protective effect on students feelings. So instead of dropping their efforts, they increase them to provide the necessary adjustments. The introduction of a TA in the LbT paradigm allows to modify completely students behaviours in case of failure, increasing action and willingness.
- Making the learning process explicit and increasing reaction in case of failure seems to help developing the incremental theory of teaching : believing work improves intelligence.
- Finally learning by teaching a TA benefits more to lower achieving students. In Chase et al.'s study, the lower achieving students working with a TA had better performances than higher achieving students who learn for themselves. Indeed, in a traditional classroom, the lower achieving children are most of the time the tutorees in the LbT paradigm. It's hard for some children to find lower achiever than them to tutor. That can have a great impact on their self-esteem. Using a TA solves this problem by always having a lower achiever, with a skill level adapted to the teaching student. [4]

All the benefits of teaching a TA could also be found in teaching a robot. Moreover robots have even new assets compared to TA. Tanaka and al. began to study the benefits of teaching a robot with 2-year-old children [6]. In fact, learning by teaching a robot is particularly adapted for young children. The robot presents the same benefits as a TA, but provides in addition, embodiment, possibility of using haptic sense and active methods, like learning by doing. All these benefits make the relationship easier between young children and robot, enabling the use of learning by teaching with an agent accessible to young children. Moreover, introducing a robot in the LbT paradigm allows the use of non verbal communication, proven important for establishing a "protégé" relationship. Using a robot increases the motivation and the attention of lower achieving children. [4]

III. EXAMPLES OF LEARNING BY TEACHING A ROBOT IN CLASSROOM

Two most advanced studies focused on robots in education of young children based on LbT paradigm and using Aldebaran's

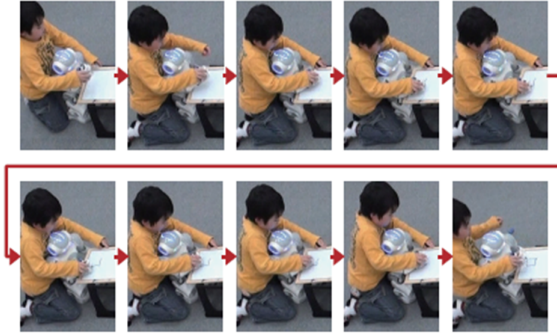


Fig. 2. Direct teaching : child helps Nao step-by-step by guiding his hand, Tanaka's team [5].

Nao robot are presented in the following section. These studies were conducted by Tanaka et al. from University of Tsukuba in Japan and by Dillenbourg's group at Ecole Polytechnique Fédérale de Lausanne (EPFL) in Switzerland, respectively on English vocabulary learning [6] [7] [5] and on handwriting [4] [8] [9].

A. A care-receiving Nao robot helps learn English vocabulary, University of Tsukuba, Japan

1) *Background of the experiment* : The first study was part of a RUBI project, investigating the robotics in early childhood education [6]. LbT paradigm was used to improve learning of English words through a drawing game. Direct teaching approach was used : children taught robot by directly moving its hand to produce the targeted movement for a given word, as depicted in Figure 2. The main experiment was introduced in a class of 3 - 6 year old children.

2) *Design of the experiment*: In [5], the experiment consisted of 2 parts : (1) feasibility of spontaneous caretaking behaviour among children older than 3 years, already shown for younger children. (2) test of the contribution to child's learning reinforcement. Familiar and natural conditions for children were assured for the experiment. All the activities were recorded with two camcoders. The first part of the experiment was conducted during 6 days with 19 participants. The Aldebaran Robotics' Nao was introduced into common classroom games. Various learning behaviours were tested and "caretaking behaviour" was analysed from camera recordings. A successful spontaneous caretaking towards "weaker agent" was reported. Furthermore, the most adopted behaviour was touching, which motivated the choice of direct teaching in the follow up of the study.

In the next step, a more challenging learning task was used to study the efficiency of the reinforcement and teaching methods used by children, consisting of three parts:

- pre-test: the child was associating a graphic card to a word card shown and pronounced by the teacher until four unknown verbs were identified.
- test: four rounds of verb-learning game were conducted, two randomly chosen rounds involved also the robot. Main steps consisted in : (1) a card containing one of the four unknown words was picked up by the teacher, (2) if the child did not know the correct answer, the teacher said it by repeating the English word, (3) if the card was marked with CRR, the experimenter proceeded with the robot in the same way as with the child right before : the robot would show an incorrect movement, ask for correction and perform the correct answer, in case the robot was not involved, the teacher would proceed to next retrieval with the child, (4) next time the same word was picked up, the correction would be performed by the child instead of the experimenter, (5) a free interaction of 10 minutes between the child and the teleoperated robot took place.
- the post-test showed that children learnt more efficiently during rounds with the robot. Moreover, haptic behaviour was assessed as a good predictor of interaction quality.

Indeed, frequently used teaching forms freely adopted by the child were direct teaching, gesturing and vocal teaching. In their follow-up study, [10] English words describing various shapes (triangle, circle, square...) were learnt. More importantly, the learning factor was incorporated to avoid frustration or irritation due to the fact that Nao would not improve even after child's instructions. In conclusion, the possibility to teach the robot and to observe the correct answers given by the robot seemed helpful for children. Also, the learning ability of the robot seemed to enhance child's engagement. More task-relevant interactions with the robot were observed when the robot had either learning ability or was giving correct answers (as opposed to the third condition without learning ability).

3) *Identified problems and challenges*: However, an important variability in results was observed, which is due to the free interaction period when the number of opportunities for learning could not be controlled. Also, older participants seemed to learn more from observing correct answers from the robot than younger children, however higher statistics would be needed to conclude on this aspect.

The use of robots as teachable agent is still quite a novel approach and reveals yet unanswered practical concerns, mainly linked to the safety. However, children showed good interaction with robots and the latter is considered as perspective tool for educational purposes.

B. Improving handwriting with Nao Robot, EPFL, Switzerland

1) *Introduction:* In CHILI Lab at Ecole Polytechnique Fédérale de Lausanne in Switzerland, Nao robot has been adapted to educational purposes of handwriting learning. The general setup used in the study is shown in Figure 3.



Fig. 3. Nao robot in interaction with one of the children during handwriting tasks [9].

The first studies were conducted in cooperation with primary schools and occupational therapy clinic in Switzerland, all in all with about 80 children aged 5 - 8 [4], exploring one-to-many interaction. The aim was the validation of robot's use in real educational environment through criteria such as technical validation, acceptance, sustained engagement and adopting of the role of a teacher by the children in short-term interaction [4]. Two individual case studies were performed, too. For the latter, Nao robot was programmed for needs of particular students in difficulties (self-confidence and handwriting), using one-to-one interaction over a more extended period of time (one session per week over about 5 weeks). The principle consists of using Nao robot as the "bad writer" which can be taught by the child. The targeted benefits for the child are multiple:

- practise handwriting, focusing particularly on letters that are critical for the particular child and individual approach thanks to the possibility to adapt the pace to the individual performance.
- improve child's motivation thanks to the "protégé effect" as explained previously.
- improve child's self-esteem.

The main motivational challenges that had been identified were:

- keeping engagement and motivation of the student over extended span of time (pace of improvement individually adapted, rewarding feeling, self-esteem, robot's reactions).
- credible teaching situation.

And the main technical challenges for this project included:

- believable handwriting
- easiness of parametrisation of the robot
- evolution of learning adapted to the child
- acquisition, recognition and learning from hardly legible examples given by children with handwriting impairments
- full autonomy of the robot which should be used by non-experts in robotics or programming (teachers, therapists)

2) *Design of the experiment:* In this section, one-to-one case studies design are in focus, as these were conducted over longer period of time compared to group studies. Face-to-face arrangement of child and Nao robot was adopted first [4]. Two tablets, a camera and an experimenter are also involved. One tablet is used for the selection of the letter or the word to be learnt. The camera is used to monitor the session and assess real-time engagement of the child. The handwriting is based on a probabilistic algorithm where human motion trajectories are seen as observations issued from a probabilistic distribution of possible motion trajectories [11], indeed, trajectories with low cost are more likely to be seen [12]. Learning a word consists in turn-taking where robot produces a word and the child corrects it. The words to practice may be suggested by the teacher, knowing the specific difficulties of the child.

Main outcomes of the study are:

- the well maintained engagement of children during the handwriting activity with Nao robot.
The used metrics were the "with-me-ness" measure and the number of provided demonstrations as well as time spent to perform each demonstration. The "with-me-ness" measures the proportion of time when the child focuses on the task. [9]. Briefly, it was assessed from the attentional focus derived from the field-of-view using an open-source real-time vision-based head pose estimator and computation of the with-me-ness. The visual field of attention was approximated to 40° cone spanning from head sellion. However, a potential intersection of objects (implying that one would be obscured by another) was not taken into account, thus all objects would be considered in focus. Figure 4 shows the principle of child's sellion tracking used for real-time estimation of with-me-ness.
- natural management of the role of teacher. With a minimal explanations of the task and context, children took the role of the teacher, managed the turn taking and progression of the activity. Furthermore, a free feedback provided on robot's performance was positively correlated with quality of robot's learning and no child adapted only playful behaviour with the robot. Children understood that consistency was needed to facilitate robot's learning (convergence to targeted performance) and developed protégé relation.

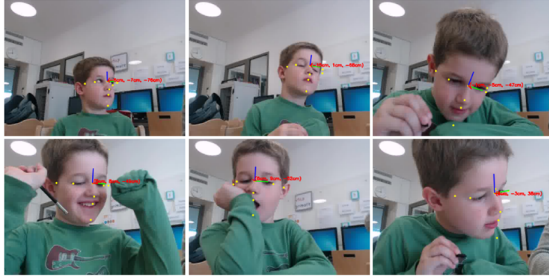


Fig. 4. Head pose results on images captured during a field experiment. Detection of face features (and therefore, estimation of the pose) is robust to significant occlusions and face rotations [9].

3) *Outcomes and challenges:* Approval for the use of the robotic technology at school thanks to the confidence about its educative and social value was given. Nevertheless, further studies on end-purpose are needed.

The main remaining difficulty is the evaluation of the quality of the interaction reflecting the psychological impact of child-robot interaction. A direct use of a questionnaire might reflect child's willingness to please experimenter rather than his/her actual behaviour and perception of the interaction. Three main types of engagement were distinguished in previous studies : social, task-specific and social-task [13] and Jacq et al. focused on the persistence of "protégé effect" as they considered it to be critical for the engagement in the LbT framework [14].

IV. TECHNICAL ASPECTS OF TEACHABLE ROBOTS FOR EDUCATIONAL PURPOSES

In this section, we develop some technical aspects, such as the robot's design, its autonomy, the learning implementation and some ways to enhance the engagement of the children. All these aspects appeared to be central for using LbT paradigm with robots in the education field in the 2 examples presented in the previous section.

A. Robot's design

The first important issue to take into consideration in order to be able to create a relationship between a child and a robot is the design of the robot.

In 2004, Woods et al. conducted a study about kid's view on robot design [15]. They found out that children have consistent opinions about the feelings or understandings of the robot, and that their physical aspect influences those opinions. 10-year-old children

were asked to attribute two emotions (sadness, happiness) and six behaviours (friendly, aggressive, shy, bossy, angry, frightened). They concluded that children felt discomfort in front of a human-like robot. They're more comfortable with half human, half machine robot. The typical friendliest robot as a cartoon-like face, exaggerated facial features, colored emotion (bright for happiness, dull for sadness), a gender (they prefer female) with a whole body participating in the communication (importance of legs and arms). Both teams working on this paradigm (CHILI Lab team and Tanaka's team) chose to work with the humanoid Nao robot [4] [5]. First of all, the design of the Nao robot (Aldebaran) matches the specificities cited above. It also has a safe structure (lightweight plastic) easy to use with children in close interaction. Moreover, it's an affordable robot for schools to acquire in years to come. Finally, both teams also underline its ease of deployment in the field. The Tanaka's team also developed an application for English learning on a Pepper robot (Aldebaran) in a one-to-many interaction with 6-year-old children [16]. Pepper was chosen for his humanoid design, mostly for his ability to express emotions, his capacity to communicate, a wide range of autonomous behaviours, such as speech, and an emotional recognition function. It also has a smooth way to move, well-adapted for young children.

Furthermore, humanoid robot offers several interface designs. CHILI Lab chose to use a tactile tablet in the CoWriter project [4]. This interface enables the child to correct himself before sending his instruction to the robot. Tanaka et al. chose to interact by touching and guiding robot's movement [5]. This method enables the child to use haptic sense, known to be very important in young children learning. Touching the robot appears as an important parameter in the construction of the child-robot relationship. Finally the Tanaka's team also used talking in its study with Pepper [16]. Even if the talking interface has not been actually completely mastered yet, it could be useful for many applications in the future.

B. Robot's control

The ways to control the robot vary within the studies, it depends on the purpose of the experiment, finality being to have a complete autonomous robot. To study children reactions to some of the robot's behaviour, both teams used the wizard of oz (WOz) control: an experimenter pilots the robot making him react to the child's behaviour. The children may know, or not, that the robot is guided by a human-being [5] [17] [10] [18]. It allows the experimenters to explore a variety of behaviours and to decide which ones are the most relevant for the paradigm and will be implemented to get an autonomous robot.

The Tanaka's team uses a comparison of 2 care-receiving robots (CRR) to implement Nao's learning abilities [18]. They used two robots in the classroom: one autonomous, implemented based on previous observations, one in WOz, adapting so that the children

are more interested in the WOz robot than in the autonomous. Then a new autonomous robot is implemented to behave as the previous WOz robot including the new relevant behaviour. Afterwards, this new autonomous robot is compared with a WOz robot, etc.

CHILI Lab plans to use a mixed initiative to implement some social behaviours such as guessing and reacting to child's emotions [19]. In that case, an autonomous robot proposes a behaviour and the experimenter agrees or disagrees with it. Thus the robot learns the "appropriate" behaviour. It is an intermediate step between the WOz robot and the fully autonomous one.

Finally, the *Learning by teaching* paradigm is aimed to be used with a fully autonomous robot, with an a priori definition of the robot's behaviour, like in the CoWriter project [4]. Building this kind of robot requires nevertheless enough previous studies to get the right design.

C. Learning abilities and adaptation to user's performances

Tanaka's team study on robots learning abilities in LbT showed that among 3 ways of learning (expert, medium and slow minded robot) the children showed more engagement and better performances with the medium one [10].

In fact the robot's behaviour must match the child's idea of a robot with difficulties so it can cause the "protégé effect" and engage the child in tutor/tutoree relationship [4]. The child needs indeed to think that the robot "needs" him to be invested, eliminating the expert robot, but he also needs to feel he's efficient in teaching, eliminating the slow-minded robot which doesn't progress.

To have an optimal effect on child's performances, the robot must learn neither too fast, nor too slowly. Indeed, a too fast learner may act as an expert robot: it won't be believable for the child that the robot has difficulties. In contrary, a robot which learns too slowly will be seen as a slow-minded robot and may discourage child's effort to teach. That's why the robot must use child's feedback to evolve. Figure 5 shows an example of the evolution of handwriting production and convergence of robot's production towards child's demonstrations.

Several studies on human teaching style has been done about this issue, for example Knox et al. in 2012 [20]. Most of them are conducted to improve machine learning, but it could also be of interest to improve the adaptive behaviour of the robot to the tutor. The robot could recognize a teaching style and adopt the corresponding behaviour to enhance the skills of the tutor.

Another way to adapt the robot to the teaching style of the student is by improving the feedback to the child teaching and the feedback on the performance. If a robot displays behaviours in order to show its understanding, the tutor can adapt his teaching to the state of understanding of the tutoree. It implies that the robot must be able to recognize the behaviour of the student, analyze it in the teaching context and adopt a corresponding

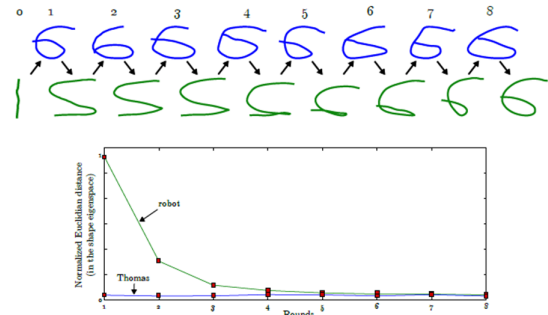


Fig. 5. Demonstrations provided by one of the participating children, Thomas, (in blue) and corresponding shapes generated by the robot (in green). After 8 demonstrations, Thomas decided that robots "6" was good enough, and went to another character [14].

behaviour showing he understands, or not, the child. So the cognitive architecture of the robot must include a second order mutual modelling [14].

D. Engagement in a one-to-one interaction

The engagement measures to what extent the human is with the robot during an interactive task. Engagement is essential for the success of all learning tasks. The aim here is to design an application that makes the child "want to use it".

Several paradigms are used to study the engagement, e.g. non verbal cues in video, gaze and head direction measures, trunk posture, but all these measures are processed a posteriori. The CHILI Lab has defined a new way of measuring the engagement in real-time called with-me-ness [9]. The with-me-ness is measured by "comparing the attentional focus of the human with the expected a priori targets of attention elicited by the task at end". This online indication of the child engagement would be very useful for the robot to adapt his behaviour to the child, enhancing the engagement when needed.

Several factors have been proved to influence child's engagement :

- turn-taking: like in a peer conversation, the robot should know when it's time to act (or speak) and time to listen to the tutor [9].
- an unexpected behaviour of the robot, like making mistakes or disobeying the child, learning fast or slow, etc. This change of behaviour helps maintain child's attention on him. Lemaignan et al. showed in their "You're doing it wrong..." study that this unexpected behaviours must not be too anthropomorphic to enhance the engagement [17].

- spatial arrangement: Johal et al. tried several spatial arrangements for the CoWriter project (side-by-side and face-to-face). It seems that the side-by-side arrangement gives a better with-me-ness measure, even if there is no real change in the writing performance [8].
- robot's reaction to the child's feedback on his performance. This reaction has not been implemented yet but the experimenters observed that the children seemed to wait for a reaction of the robot when they gave it a feedback, whether a positive or a negative one [19].

Those many works show the importance of social robotics to enhance the engagement, particularly in long-term relationship as discussed below.

V. DISCUSSION: SOCIAL CHALLENGES OF A LONG TERM CHILD-ROBOT INTERACTION IN EDUCATION

Long-term studies conducted in realistic environments [21] are still quite rare since: (1) they are much more laborious and time-consuming than studies focusing on a single human-robot interaction, (2) studies over extended periods requires high degree of autonomy of robots, reached only recently. Essential aspects influencing the success of a long-term interaction using LbT paradigm in education are the acceptance of robot by users (children, teachers, therapists, parents, etc), development of child's affective bonding with robot, definition of the role of the adult and the conservation of child's motivation.

A. Acceptance of the humanoid robot

From child's perspective, complex mechanisms enhancing social and emotional interaction between children and robots are essential for their acceptance [21], since these enable the creation of "protégé effect" - the basis for affective bonding and task-relative engagement of children and thus for LbT paradigm [4], as we mentioned in sections II and III. From adults' perspective, clear specifications of use of the robot are needed based on consensus between all parts (teachers, parents, therapists, etc.) in order to (1) avoid any negative effects on cognitive development of children and (2) avoid negative repercussions on adults position and relationship with children. The balance between benefits and risks, revealing mainly questions of safety and efficiency, is always dynamic and it changes over time and culture, therefore a continuous communication between developers and the end users is recommended [7]. Social and intellectual framework for the use of robotics technology taking into account not only the desired outcome (learning reinforcement), but also the cultural and philosophical particularities of given societies. Indeed, this "care-receiving" robot configuration seems to be more acceptable than "care-taking" robots and maintain the essential role of a

human "care-giver" who is available all along the child-robot interaction, who designs and controls the interaction of children with robots [7].

B. Role of the adults within the child-robot interaction

Adults played role of educational scenario designer as well as session facilitator - suggesting possible corrections, suggesting trying another task, prompting to evaluate robot's performance, facilitating possible shortcomings during the interaction. Further studies on child's and teachers' perception of teacher's role in this LbT configuration using teachable robots are required. For instance, the child rarely turned to the teacher (experimenter) during handwriting sessions, possibly implying that the task was easy to understand and meaningful so the child could interact with the robot alone [8]. However, the relationship and the attachment to the traditional human caregivers is important [22]. Care-receiving robots should support and assist human care-givers, stimulate their creativity and alleviate the physical burden [7]. Therefore, it is also important that the robotics technologies are easy to use by instructors, not only developmentally adapted for children [23].

C. Development of affective bonds between child and robot

As mentioned before, the development of affective bonding is critical in protégé effect creation and maintenance. However, there might be potential effects on affective development of more vulnerable children and other unknown and unanticipated emotional effects on children [22]. Development and sustainability of affective bonding between young children (less than 2 years) and robots was shown to be possible over 6 months period. Indeed, touch was found to be a good predictor of interaction quality and social connectedness in early social interactions. Therefore it was also thought of as a useful reinforcement signal for learning activities [6]. Most body parts presented an initial increase of touching (interest for a novelty) and then a decrease (habituation), except for hands that presented a steady increase. These developed haptic behaviours were similar to the ones towards their human peers. The complexity of social and emotional interactions seems to be essential for creation and long-term duration. Too simple mechanisms lead to the habituation effect and children's abandon of the interactions.

D. Conservation of motivation in long-term interactions

Credible social behaviour were found to be indispensable to raise children's attention and motivation for interaction with the robot, for at least two weeks since the main motivation for children to interact with the robot was "becoming friend" with it [24]. Playfulness and tangibility played a major role in maintaining child's engagement during educational activities, since hands-on way shows to be more engaging [23]. Learning

ability with adapted dynamics (not too slow, not too fast) of the robot using child's instructions showed positive effect on learning outcome. In general, addition of novel behaviours over time and integrating and behaving with regards to child's feedback seems to be essential to maintain a long-term relation, all the more that educational framework allows participants to quit the interaction if they wish [21].

E. Age effect on the perception of robot

Younger children (under 2 years) maintained their interest longer over the sessions than older ones. Progressively, their understanding evolved, passing from a "moving object" to a "social agent" perception and empathic interactions started to emerge [24]. Presumably, younger children have no expectations about robot's capacities [5] [7]. For older children, more complex, diverse and dynamic behaviour seems to be needed. Otherwise, a less human-like embodiment can be adopted for certain applications, since less "cognitive human-likeness" is expected from a more caricatured embodiment.

Finally, for long-term studies in general, more precise specification of "long-term" is needed with regards to the number of interaction sessions and the duration of each session. Furthermore, it was suggested that only the studies conducted after the familiarization period should be considered for long-term studies [21].

VI. CONCLUSIONS AND PERSPECTIVES

The present article focused on the integration of Nao robot as teachable agent for LbT paradigm into real class environment. First, the principles of *Learning by teaching* paradigm was presented in general terms and the advantages of using a humanoid robot as teachable agent in LbT paradigm were recalled. Then, a focus on the two most developed studies, conducted by Tanaka et al., University of Tsukuba and Dillenbourg et al. at EPFL, Switzerland, were presented. These have been exploring the application of Nao robot using LbT paradigm for primary school children, respectively for English vocabulary learning and handwriting improvement (linked to self-confidence strengthening). Next, the technical aspects concerning robot's design, control, interaction and engagement were developed. Finally, social aspects identified as crucial for a successful long-term interaction were presented, showing that the implementation of social dimensions (emotional, affective bonding, etc.) presents multiple challenges for further development, as mentioned in the following paragraphs.

The use of robots for long-term social interactions in naturalistic environment, for instance their use in education, is a recent area. Thus, the existing studies have still often exploratory character and were performed with a limited sample of participants [21]. Perspective issues for studies with Nao were identified by Tanaka's and Dillenbourg's teams as follows:

- development of reactions of robot according to the feedback from the child and also according to the fairness of the feedback (better for a good score than for a poor score) [4].
- development of the ability to empathise with users, since empathy was identified as essential for creation and development of human social interactions.
- definition of "conceptual with-me-ness" as proportion of the time that child spends on the attentional target expected by the robot for the task in question [4] that could be integrated as a feedback for Nao robot that could in turn adapt a strategy to regain child's attention [9].
- improvement of attention estimation by adding heuristics to detect and ignore the erroneous face detection and adding a pupil tracking on top of head pose [4].
- performing studies with larger sample size [23] [4] [7].

More generally, robots could potentially bring together the best aspects of both human tutorees and teachable agents by [2]:

- presenting the optimal helpful learning style.
- expliciting their learning process.
- adapting to individual needs of every child to allow the child to benefit even more from LbT, more generally, development of robot's memory would contribute to flexibility and personalisation of its behaviour depending on the user (child) they are interacting with at a given moment [21].
- recognising and adapting to tutor's style and thus provide an individualized approach that was identified as key for efficient learning.
- using the knowledge of teaching style of a child to understand child's understanding of the subject and his/her approach, difficulties to guide their efforts in a personalized non-threatening way, to question less covered aspects (probably not well understood by the child).
- increasing the involvement of children thanks to appealing physical embodiment and social interactions which were evaluated by the participants as more supportive, enjoyable to work with, providing deeper and more attractive social presence than virtual agents without physical embodiment.
- effectively using the non-verbal communication channels (pointing, gestures, posture, personal space, touch) which are essential in human communication [24].
- providing large data sets with natural teaching techniques adopted by children during LbT sessions and thus allow for characterization of the most natural teaching styles.
- providing adapted feedback to maintain rewarding and engaging feeling of the child.

Other perspective applications for younger children were suggested, such as:

- learning of reading, sports and crafts activities.

- multiplying the one-to-one training that was identified as more efficient in several cases and impossible with normal teacher - student ratios.
- and especially using it in special education.

Another critical aspect revealing several challenges is the hardware design which depends partially on cultural context of the application [2] [7]. An optimum between the rich physical capabilities (important number of liberty degrees, believability) and affordability seems to be essential to enable widespread use. A possibility of a single personality (software features) adapted for a specific student and various possible embodiments for various application activities may be a possible strategy. Or a combination of a physical part and a virtual part of robot's "body" may be a solution to optimize its cost. Furthermore, other recommendations, such as the limitation of screen time per day (less than 2 hours, for instance, for American Academy of Pediatrics, 2003, Barlow 2007) may impose additional limitations.

Finally, an important issue during studies with children concerns the collection and analysis of the data, as well as the use of robust metrics for the analysis, since the interactions need to take place in realistic environment and there is also a huge variability of interaction behaviours among children.

REFERENCES

- [1] O. Mubin, C. J. Stevens, S. Shahid, A. Al Mahmud, and J.-J. Dong, "A review of the applicability of robots in education," *Journal of Technology in Education and Learning*, vol. 1, pp. 209–0015, 2013.
- [2] J. Werfel, "Embodied teachable agents: Learning by teaching robots," in *Intelligent Autonomous Systems, The 13th International Conference on*, 2013.
- [3] C. C. Chase, D. B. Chin, M. A. Opezzo, and D. L. Schwartz, "Teachable agents and the protégé effect: Increasing the effort towards learning," *Journal of Science Education and Technology*, vol. 18, no. 4, pp. 334–352, 2009.
- [4] S. Lemaignan, A. D. Jacq, D. Hood, F. Garcia, A. Paiva, and P. Dillenbourg, "Learning by teaching a robot: The case of handwriting," Tech. Rep., 2016.
- [5] F. Tanaka and S. Matsuzoe, "Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning," *Journal of Human-Robot Interaction*, vol. 1, no. 1, 2012.
- [6] F. Tanaka, A. Cicourel, and J. R. Movellan, "Socialization between toddlers and robots at an early childhood education center," *Proceedings of the National Academy of Sciences*, vol. 104, no. 46, pp. 17 954–17 958, 2007.
- [7] F. Tanaka and T. Kimura, "The use of robots in early education: A scenario based on ethical consideration," in *RO-MAN*, 2009, pp. 558–560.
- [8] W. Johal, A. Jacq, A. Paiva, and P. Dillenbourg, "Child-robot spatial arrangement in a learning by teaching activity," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 533–538.
- [9] S. Lemaignan, F. Garcia, A. Jacq, and P. Dillenbourg, "From real-time attention assessment to with-me-ness in human-robot interaction," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 2016, pp. 157–164.
- [10] S. Matsuzoe and F. Tanaka, "How smartly should robots behave?: Comparative investigation on the learning ability of a care-receiving robot," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 339–344.
- [11] H. Yin, P. Alves-Oliveira, F. S. Melo, A. Billard, and A. Paiva, "Synthesizing robotic handwriting motion by learning from human demonstrations."
- [12] K. Dvijotham and E. Todorov, "Inverse optimal control with linearly-solvable mdps," in *Proceedings of the International Conference on Machine Learning (ICML)*. ICML, 2010, p. 335342.
- [13] H. O'Brien and E. Toms, "What is user engagement ? a conceptual framework for defining user engagement with technology," *Journal of the American Society for Information Science and Technology*, vol. 56, no. 6, pp. 938–955, 2008.
- [14] A. Jacq, F. Garcia, P. Dillenbourg, A. Paiva *et al.*, "Building successful long child-robot interactions in a learning context," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 239–246.
- [15] S. Woods, K. Dautenhahn, and J. Schulz, "The design space of robots: Investigating children's views," in *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*. IEEE, 2004, pp. 47–52.
- [16] F. Tanaka, K. Isshiki, F. Takahashi, M. Uekusa, R. Sei, and K. Hayashi, "Pepper learns together with children: Development of an educational application," in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*. IEEE, 2015, pp. 270–275.
- [17] S. Lemaignan, J. Fink, F. Mondada, and P. Dillenbourg, "You're doing it wrong! studying unexpected behaviors in child-robot interaction," in *International Conference on Social Robotics*. Springer, 2015, pp. 390–400.
- [18] S. Matsuzoe, H. Kuzuoka, and F. Tanaka, "Progressive development of an autonomous robot for children through parallel comparison of two robots," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 2014, pp. 244–245.
- [19] A. Jacq, W. Johal, P. Dillenbourg, and A. Paiva, "Cognitive architecture for mutual modelling," *arXiv preprint arXiv:1602.06703*, 2016.
- [20] W. B. Knox, B. D. Glass, B. C. Love, W. T. Maddox, and P. Stone, "How humans teach agents," *International Journal of Social Robotics*, vol. 4, no. 4, pp. 409–421, 2012.
- [21] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.
- [22] N. Sharkey, "The ethical frontiers of robotics," *Science*, vol. 322, pp. 1800–1801, 2008.
- [23] A. Sullivan and M. U. Bers, "Robotics in the early childhood classroom: learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade," *International Journal of Technology and Design Education*, vol. 26, no. 1, pp. 3–20, 2016.
- [24] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," *Human-computer interaction*, vol. 19, no. 1, pp. 61–84, 2004.

Influence of Human Personality Traits on Trust in Human-Robot Interactions

Pantecouteau, Hugo
Artificial and Natural Cognition Master
PHELMA
Grenoble, France

Passera, Brice
Artificial and Natural Cognition Master
PHELMA
Grenoble, France

Abstract—In the field of human-robot interaction (HRI), it is tempting to look for an optimal combination of parameters which fosters the efficiency and the comfort during the interaction. However, it is unlikely that all the users will perceive, trust and react the same way in an identical human-robot interaction situation. This is one of the reason why studies attempt to find tendencies in the users' behavior towards robots according to their personalities, for example, by looking for correlations between some personality traits of the users and their trust in the robot. Using questionnaires, trust games, measures of the distance to the robot or users' reactions to unusual requests, some studies correlated extroversion, emotional stability and agreeableness with trust or aspects connected to trust. With further studies in this field, personality traits could be integrated in the robot software in order to adapt the robot's behavior during the interaction.

Keywords—trust; robots; personality; proxemics; trust game; five factors model; Eysenck personality inventory;

I. INTRODUCTION

With the growing interest in robots and the will to expand their usage, some new questions arose in the field of human-robot interaction (HRI). Robots are being used in more and more situations involving a human interaction. From the robotic arm on car assembly lines to robots developed to be able to interact and assist humans in a variety of tasks. The medical field has seen the apparition of robots in surgery with human controlled robotic arms allowing increased accuracy in the procedure. The US army has been using mine-clearing robots with more and more autonomy. Lately, we have seen an upsurge of robot usage in social situations aiming at human assistance. One of the most advertised situations is the use by the French railway company (SNCF) using Pepper Robot as hostesses in train stations. Social robots are also used in relations with the elderly as a companion and even a healthcare assistant programmed to remind them to take their medication or exercise. In these social situations, the developers are faced with new challenges. While developing for assembly line robots with highly accurate set of movements the ability to control every aspect of the environment is crucial. However, when developing a robot aimed at human interaction the

environment and external factors cannot be as controlled, creating challenges in technical development but also in acceptance of the robot by the user. These factors influenced researchers to first work on technical reliability to improve human acceptance of the system. And more recently, to investigate the different factors suspected to influence the quality of the HRI. One of these factors was trying to establish a quality relation with the robot by evaluating the impact of trust in the interaction. A meta-analysis by Hancock *et al.* [1] first looked at the influence of trust on automation and reviewed the literature to provide a first model presenting the main factors influencing trust in automation (fig. 1). Building on this work, a new meta-analysis by Schaefer *et al.* in 2014 [2] explored the factors influencing the development of trust and their specificities regarding the domain of HRI. This latest meta-analysis also expanded on the influence of the human factor which was under represented in previous work mainly due to the lack of studies.

A. Definition of trust

Researchers have extensively investigated trust in human interactions. However, no universalized definition of the concept has emerged highlighting its complexity. The absence of a consensus can also be explaining by the wide variety of disciplines looking into it (e.g. psychology, sociology, economics). In the field of HRI, trust is still a new and underrepresented line of research. Therefore, definitions used in HRI are translated from the automation or human-computer interactions (HCI) fields [3]. However, extending these definitions to HRI may not be entirely relevant. Indeed, robots differ from automated systems and computer interactions by their ability to move, and physically interact with humans. Moreover, in HRI, humans can freely move around and touch the robot. These characteristics are not present in other automation interaction and could influence the implications of trust in a HRI context.

HRI is not different than any other field of research and exposes a multiplicity of definitions. Schaefer *et al.*'s meta-analysis listed 32 human-robot trust definitions from 282 trust definitions. The most referenced one is "trust in automation is the attitude that an agent will help to achieve an individual's

Human Operator	Automation	Environment
Human Traits +* Age +Gender +Ethnicity +Personality Human States Fatigue Stress Attentional control Cognitive Factors +* Understanding the automation +* Ability to use automation +* Expectancy of automation Emotive Factors +Attitudes toward automation Confidence in automation +Satisfaction with automation +* Comfort with automation	Features * Degrees of automation +* Appearance * Mode of communication Capability +* Errors in automation +* Automation behavior(s) Quality/Accuracy * Cueing/Feedback/Alarms	Task-Related +* Proximity +Risk

Note: * represents experimental and + represents correlation findings.

Fig. 1. Factors influencing trust in automation

goals in a situation characterized by uncertainty and vulnerability” [4]. From these definitions, they isolated the most repeated terms used to define human-robot trust: expectations (15), risk/vulnerability (14), confidence (12), and predictability (5). They also analyzed studies from the three fields of human-machine interaction. In these studies, the level of trust is revealed to influence the reliance on the system and to modify the quality of the interactions. Over- or under-reliance caused by trust can result in a dissatisfying human-machine interaction [1]. Schaefer *et al.* proposed a model based on the factors influencing trust in HRI. These factors of trust development are: automation based —its features and capabilities—, environment based (i.e. factors related to the task, such as the risk involved) and human based factors.

B. Human based factors influencing trust

These human factors are often overlooked and relegated to a secondary focus under the capacities and technical abilities of the automation system or robot. However, as we have discussed, the human operator is generally considered to be one of the most influencing factors in these interactions. When investigating the HRI the human factor is even more crucial. The embodied system interacts with humans with from a vast array of individual’s characteristics susceptible to influence trust development in the robot. The argument is also that the human operator is the one trusting the technology. Therefore, the factors influencing its ability to trust the robot must be considered as highly important. In their model, Scheafer *et al.* addressed the human factors as the human states, the cognitive and emotive factors as well as the human traits (fig. 1). The cognitive factors are, ability to use automation, understanding of automation and expectancy. For example, they found the capacity to learn is negatively correlated with trust [5] as well as a positive correlation of the perceived usefulness of the system with trust[6]. They divided the emotive factors in attitudes toward automation, confidence in automation,

satisfaction with automation and comfort with automation. For example, a study found as satisfaction improves so does trust in the system. Comfort in automation is explored as the familiarity and proximity level with the operator. The last component of human factors is regarding human traits such as gender, age, ethnicity. Last, we will develop throughout the present work the influence of personality traits as a component of the human factors.

C. Personality traits

We decided to focus this work on the influences of personality traits on trust in HRI. While investigating, we realized that personality trait as a factor of influence on trust is rarely explored and yet the studies done provide interesting results on the influences of personality archetypes on trust. Most studies presented in this work use the Five Factors Model (FFM), also named the Big Five or OCEAN model of personality traits. Built around a factor analysis, five dimensions have been identified to be five different factors describing personality. These five factors are [7]:

- Openness to experience: imaginativeness, broad-mindedness and artistic sensibility;
- Agreeableness: likeability and friendliness;
- Conscientiousness: dependability, being organized, and will-to-achieve;
- Neuroticism/Emotional stability: adjustment vs anxiety;
- Extraversion: activity and sociability.

The FFM is the most commonly used model in academics however some research presented in this work use an older model of personality created by Eysenck based around biological measures linked with traits characteristics [8].

Eysenck's model is based around two major traits corresponding to the extraversion and neuroticism of the Big Five. Some examples of biological measures used to assess the neuroticism trait are the heart rate or blood pressure that Eysenck links to the activation of the sympathetic nervous system characterizing. For him, people with low threshold of activation are more vulnerable to negative affect and get easily nervous or upset. Eysenck's model is organized around two axes corresponding to the positioning of the subject in both traits extraversion/introversion and emotional stability/instability. In his later work, Eysenck added the psychoticism/socialization trait which he describes as the susceptibility to aggression and the possibility of having a psychotic episode. Throughout this review of literature, we will explore the influence of these personality traits on trust regarding HRI.

II. HOW TO ASSESS TRUST AND PERSONALITY TRAITS?

A. The assessment of personality traits

The two main models of personality used in these studies are the Five Factors Model or 'Big Five' [9] or the Eysenck Personality Inventory [10]. The questionnaires present numerous claims about how we react in specific situations or how we perceive ourselves and the subjects have to score each of the sentence by a number representing how much they agree with it (see the sample in fig. 2). These questionnaires are derived in numerous adjusted or shortened versions [11], [12].

B. The assessment of trust

The most common method to measure trust is a subjective one: self-reported measures like questionnaires. For example, the Checklist for Trust between People and Automation [13] which measures trust between humans and machines, or more robot-specific questionnaires such as the Schaefer 40-item scale [14], or some items of the de Ruyter Social Behavior Questionnaire [15]. In the Schaefer scale, each item starts with 'What % of the time will this robot...' and the user can answer on a scale between 0% and 100%, with 10% intervals. Some of

these questionnaires can be found in [14]. Naturally, as the concept of trust can be divided into various components, some questionnaires assess only one aspect of trust, for example 'propensity to trust' for the Automation-Induced Complacency Potential Rating Scale [16] which is a stable trait, unique to the individual and independent of the experiences.

Economic trust games are used to assess the level of trust between human subjects, like the investment game [17]. In this game, the player 1 has an amount of money M . Then, the player 1 has to decide how much money he gives to the player 2 knowing that: a) the amount X will be tripled before the player 2 get it so the player 2 will get $3X$; b) X can be zero. Eventually, the player 2 has to give back an amount of money Y to the player 1, which can also be zero. Berg et al. observe that a large part of participants in player 1 position decide to give money even during the first trial ($X > 0$). This is considered as trust towards player 2, though it has been claimed that the non-zero transfers could be explained by altruistic behaviors instead of trust. To test this hypothesis, Brulhart *et al.* [18] conducted a study using the investment game but the participants in player 2 position were initially either rich or poor. Participants in player 1 position did not give a significantly different amount of money to the poor player 2 than the rich player 2. They concluded that this behavior can be considered as trust. The investment game is also used for human-robots interactions. However, as pointed out by Salem *et al.* [3], 'these methods only model very specific trust situations related to monetary gain or loss' so they are not easily generalizable. Moreover, the situations designed in these studies are far from a natural and realistic environment, regarding human-robot interaction situations, for example with a home companion robot. To address this issue, some studies take place in a domestic environment and try to use various types of domestic tasks.

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

Disagree strongly 1	Disagree a little 2	Neither agree nor disagree 3	Agree a little 4	Agree strongly 5
------------------------	------------------------	---------------------------------	---------------------	---------------------

I see Myself as Someone Who...

- | | |
|--|--|
| <input type="text"/> 1. Is talkative | <input type="text"/> 23. Tends to be lazy |
| <input type="text"/> 2. Tends to find fault with others | <input type="text"/> 24. Is emotionally stable, not easily upset |
| <input type="text"/> 3. Does a thorough job | <input type="text"/> 25. Is inventive |
| <input type="text"/> 4. Is depressed, blue | <input type="text"/> 26. Has an assertive personality |
| <input type="text"/> 5. Is original, comes up with new ideas | <input type="text"/> 27. Can be cold and aloof |

Fig. 2. Sample from the Big Five Inventory[33]

As in between-humans interactions, proxemics can be used to study the level of trust between agents, that is to say analyzing how agents use the space around them during their interaction. Regarding HRI situation, it can consist in measuring the physical distance between the agents during the interaction. Hall described the personal spatial zones, based on the correlations between the nature of the agent allowed in each zone [19]. For example, during a conversation between friends, the agents tend to be in a range of 0.45m to 1.2m from each other, a zone called 'Personal zone', whereas the range for a conversation to non-friends is 1.2m to 3.6m (the Social zone). It has been shown that these zones are influenced by cultural factors. Some studies focused on the minimum physical distance the subjects let a robot approach them, depending on robot features or human states or traits. They often measure the distance when the subjects approach the robot too. Such studies, even if not specifically designed to test an hypothesis about trust, can test if the users treat social robot as a normal social agent and they can indirectly give some material about various components of trust.

A relatively recent way to assess trust placed in a robot is an objective one: observing the behavioral responses of subjects when a robot addresses them unusual requests [20]. The unusual requests can consist in revocable/irrevocable material damage or breaches in privacy. The measure is the number of unusual requests the subjects followed. This can be compared across task types or across various features of the robots or human states or traits.

III. MAIN RESULTS

A. *Explicit links between personality traits and trust*

To the best of our knowledge, only three studies focused explicitly (but never exclusively) on the link between trust and personality traits. It is very interesting to note that this small panel of studies illustrates very well the diversity of approaches and investigated aspects.

Merrit and Ilgen [21] designed a task when participants have to inspect luggage for dangerous items and decide whether it is necessary to search further or not, with accuracy and speed constraints. The participants were told that they could use a 'Automated Weapons Detector' to help them, as an automated virtual assistant, whenever they want. The machine was described in terms of competence (error rate), dependability (tendency to break down), predictability (same answer for the same item or not) and level of information given to the user about the process ('acquiring', 'scanning', 'processing'). They were told that these characteristics varied whereas it was not the case. After the description and a 1-minute demonstration, questionnaires were administered to the subjects: 12 items from the Automation-Induced Complacency Potential Rating Scale [16] which is a 5-point Likert-type scale ranging from 'strongly disagree' to 'strongly agree' about propensity to trust machines in general. They also created a 6-item scale assessing the initial trust in the Automated Weapon Detector based on the Muir's theory [22]. The same

questionnaire was also administered after the task, to evaluate the experience-based trust, and a 10-item scale inspired of the Five Factors Model was used to assess extroversion score. First, they found that a high score of extroversion is correlated with a higher propensity to trust ($\beta=0.2$). They also found a better correlation between propensity to trust machines in general and initial trust in the Automated Weapon Detector than between propensity to trust machines and experience-based trust which is informative about the various components of trust.

Haring et al. [23] used the investment game preceded by a phase of interaction between the subjects and an android robot (a short conversation, a simple request 'Move the box' and an unusual request 'Touch my hand'). They used the Eysenck personality scale (personality traits), the Godspeed questionnaire (perception of the robot by the subject) and proxemics (the subjects had to pick up a chair and move it where they want to sit). The study shows a positive correlation between the extroversion score and the amount of money given to the robot ($r=0.44$). Given that the subjects were in player 1 position (they give money first) and the android robot was in player 2 position (it gives money after the first transfer), this result suggests that the more the subjects are extroverts, the more they trust the robot. Moreover, this amount is similar to the one observed in the initial studies using the investment game with humans [17] which suggests that the android robot may be considered as a normal social agent.

Salem et al. [3] conducted a study that involved a non-humanoid mobile robot in a realistic domestic environment (see the map in fig. 3), which welcomes the subject, makes him/her comfortable and addresses usual ('Would you be so kind as to help me setup the table?') and unusual requests ('Please pour the orange juice from the bottle into the plant on the windowsill'). They combined a subjective assessment, the Godspeed questionnaire for the perception of the robot and a scale based on Madsen and Gregor work for the trust (among others) with an objective assessment, the number of unusual requests followed by the subject. They also administered the Ten Item Personality Inventory, a personality scale based on the Five Factors Model [12]. They showed positive correlations between extroversion and various items about the perception of the robot: 'Anthropomorphism' ($r_s=0.39$), 'Human nature' ($r_s=0.47$), 'Uniquely Human' ($r_s=0.31$) and 'Psychological closeness' ($r_s=0.38$). These results mean that the subjects with a higher score of extroversion tend to anthropomorphize the robot more, to attribute human traits and uniquely human traits to the robot more and feel closer to the robot than those with a lower score. They also found positive correlations between emotional stability (lowest score of neuroticism) and the following items: 'Animacy' ($r_s=0.39$), 'Likability' ($r_s=0.43$), 'Anthropomorphism' ($r_s=0.38$) and 'Psychological closeness' ($r_s=0.42$). Contrary to the previous studies about extroversion and trust, they did not find any correlation between a personality trait and the rate of unusual request (objective measure) or the trust questionnaire.

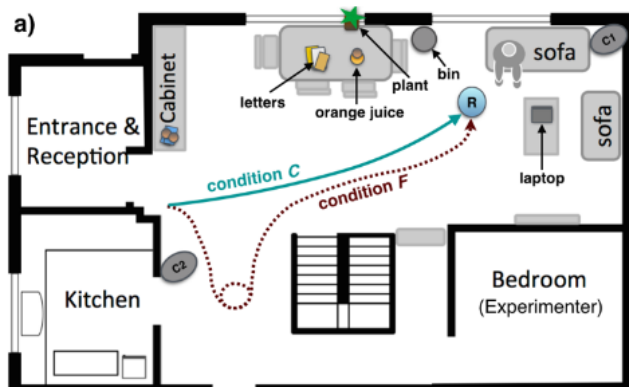


Fig. 3. Map of the house and path followed by the robot (from Salem *et al.*)

B. Interesting indirect clues to study the link between personality traits and trust

The other studies do not focus explicitly on the link between trust in robots and personality traits but offer some interesting results that can be useful to explore this link. The first series of experiments deals with the distance tolerated by the subjects. Walters *et al.* [24], [25] found that 40% of the participants approach the robot (mechanistic appearance) to a distance less than 0.45m, matching with the 'Close intimate' and 'Intimate' personal zones described by Hall. If a human stranger approaches to this distance, most people would feel uncomfortable or threatened. They then wondered whether these participants consider the robot as a normal social agent. They tried to explain the observed variance by analyzing 12 traits of Eysenck personality scale (they did not use the scale because their shortened version distorted the reliability of the scale). They found correlations between accepted distance and one group of factors. Given that they did not follow the Eysenck scale, the group of factors does not match any category suggested by Eysenck so they named it 'Proactiveness' according to its subfactors which were Creativity, Impulsiveness, Excitement-seeking and General Activity Level. Proactiveness was shown to be positively correlated with distance which means that the more a subject is 'proactive', the less he/she gets close to the robot. This result is surprising, as impulsive and 'excitement-seekers' could have been expected to be curious and excited about this new type of interaction. Takayamu and Pantofaru [26] used a personality questionnaire based on the Five Factors Model and a task where the subjects had to indicate when they started to feel uncomfortable while a mechanistic robot gets close to them. Another condition was similar to this one but the participants were told that the robot was teleoperated. There was a last condition where the subjects approach the robot and stopped where they feel they had to stop. They found a negative correlation between the personality trait of agreeableness and distance: the more agreeable the subjects were (according to the definition of the Five Factors Model), the closer they approach the robot. They also found a positive correlation

between neuroticism and distance in the teleoperated robot condition.

Another series of experiments deals with the link between the personality of the users and their preferences for virtual, mechanistic or humanoid appearance of the robot. Syrdal *et al.* [27] and Walters *et al.* [28] found that extroversion and emotional stability were negatively correlated with the preference for a mechanistic appearance. Moreover, Looije *et al.* [29] found that the more conscientious the subjects are, the more they prefer a text interface to a physical robot.

IV. DISCUSSION

A. Tendencies highlighted by the studies

Despite the small number of studies focusing on the link between personality and trust or aspects related to trust, some convergent tendencies seem to emerge from this work (see fig. 4). The results will be discussed here separately for each personality trait. The most frequently mentioned trait is extroversion. Indeed, extroversion score was shown to be positively correlated with the score of a "propensity to trust" questionnaire (not exclusively about social robots) and with the amount of money given to an android robot during the investment game. These correlations suggest that the more the user is extrovert, the more he tends to trust the robot. This is consistent with the findings about trust between humans: the more a person is extrovert, the more he/she tends to trust the other people [30], [31]. Moreover, the subjects tend to give the same amount of money than in the historical experiment with only human participants. They may consider it as a normal social agent. This is supported with the results showing a positive correlation between the extroversion score and the tendency to attribute human traits to the robot, to feel close to it and to prefer a humanoid appearance than a mechanistic one. However, the diversity of experimental approaches reveals some inconsistencies across the results. Indeed, Salem *et al.* did not find any correlation between any personality trait (using Five Factors Model) and the tendency to follow or not follow the unusual requests sent by the robot. Does it really conflict with the other results? Before considering this result as a conflict one, we can wonder if the number of unusual requests followed by the subjects is a good way to assess trust. We can also wonder if it assesses the same aspect of trust than the investment game of the various questionnaires. For example, in the procedure with unusual requests, the subjects may feel no responsibility for the actions they perform following an unusual request of the robot, not to mention that the consciousness of the experimental environment around may bias the reactions. Another inconsistency with the previous results is the positive correlation found between 'proactiveness' and the distance to the robot when the subjects approach the robot. This trait is not referenced in the Eysenck's model or in the Five Factors Model. Indeed, the researchers tried to shorten the Eysenck questionnaire but when they checked the validity, they realized that it created a distortion. They then decided to conduct a factor analysis on each item. This is how they were brought to put together the items 'Impulsiveness',

‘Creativeness’, ‘Excitement-seeking’ and ‘General level activity’ into the the ‘Proactiveness’ trait. They found that the more a subject is ‘proactive’, the less he/she tends to approach the robot at a close distance. To our opinion, it is a surprising result, as we could have been expected impulsive and creative people to be curious about this new type of interaction, not to mention the aspect of ‘excitement-seeking’. Even if it seem to be a surprising result and this trait may seem to be linked with extroversion, it does not perfectly fit with the extroversion as described by the Five Factors Model or the Eysenck’s model, so it is difficult to discuss it further.

Emotional stability (the contrary of ‘neuroticism’) was found to be negatively correlated with the distance to the robot allowed by the participants when they know that the robot is teleoperated. This means that the more the subjects were emotionally stable, the more they tend to tolerate the teleoperated robot close to them. This is not surprising, given that emotional stability is the contrary of neuroticism and we can expect people with high neuroticism score to be more anxious of this type of situations (a teleoperated robot coming to them).

Eventually, one study found a negative correlation between agreeableness and distance to the robot when the subjects approach it, which is consistent with what we could have been expected, given that agreeableness is linked with prosocial behaviors. Given that the studies mentioned in this review involve environment-related factors such as proximity and type of task, or robot-related factors such as appearance, the influences of personality traits on trust cannot be studied in an efficient way without considering them as strongly connected to the contextual variables.

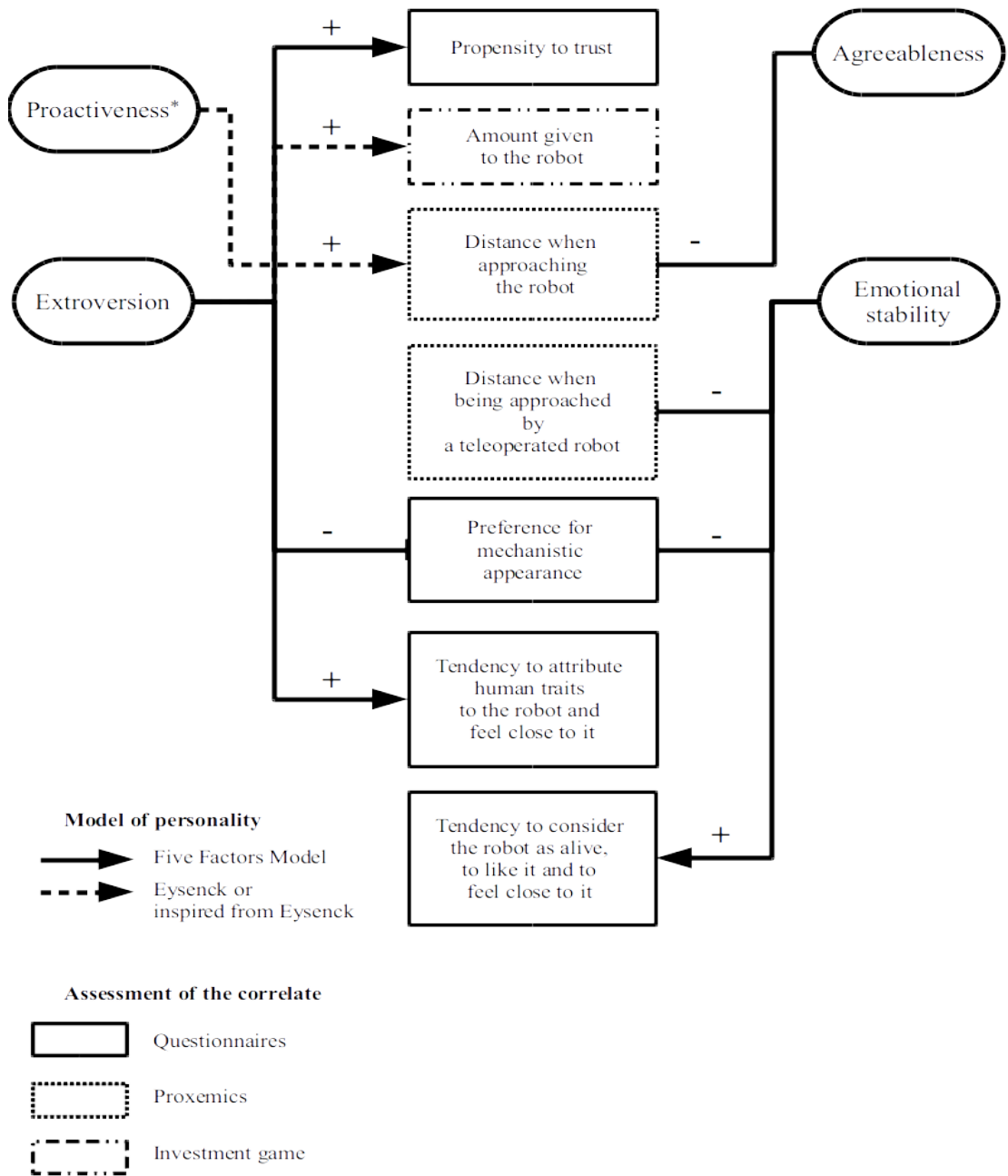
B. Perspectives

Following this work, we highlighted the importance of the operator’s personality. However, the small amount of data does not allow us to reach any definite conclusion. Moreover, a number of data presented in this paper only represent a secondary hypothesis of the experiments. A first step to validate these results would be replicating the experiments with personality traits as a dominant variable. Considering the Big Five model, some traits of personality have not been investigated regarding their effects on trust. Further studies need to be done to address this issue. Some results provided demonstrate similarities between the influence of personality traits with trust in HRI and trust in human interaction. Further investigation of these relations could be interesting for the evaluation of trust in HRI. Moreover, while investigating trust in HRI, we observed that some human factors described by Schaefer et al. are tightly linked or influenced by personality trait, for example attitude towards automation and comfort with automation are linked with the level of extroversion. These observations could lead to use personality traits as a predictor towards certain human-associated factor influencing trust in HRI. New studies could also focus either on the influence of personality traits in development of trust or on the influence on the ability to trust a pre-existing agent. The study of the

development of trust affects robot aimed at long term relationship with the user whereas the ability to trust and existing agent (i.e. in a “one-time-interaction”). For example, studies could be conducted to investigate the influence of personality traits on usage, propensity to use and confidence in the information delivered by Pepper in the hostess situation experimented by the French railway company.

C. Future implementations

This line of research could be used in a way to adapt the social robot to its user. Indeed, if the implementation of robots in nursing homes and their utilization with isolated elder persons becomes more frequent, using personality traits to configure the robot, could be a facilitator in the development of trust between the robot and its user. Personality traits are easily evaluated and it is possible to implement in the robot software a questionnaire assessing the main personality traits affecting the development of trust. From the answers provided, the software could adapt its behavior to its user’s personality. For example, after the questionnaire, if the user is qualified as introvert, the robot could keep a safe distance when interacting with the user, therefore lowering the level of discomfort created by its presence and allowing an optimal development of trust. In “one-time-only” situations, the robot could adapt itself to the cultural environment. A cross-cultural social study [32] performed personality test around 56 countries giving them a rough idea of the repartition of personality archetypes throughout geographical regions. Based on studies like this, the robot software could adapt itself to the average personality archetype therefore adopting behaviors in accordance with the cultural profile of the country. Such level of adaptation could greatly improve HRI trust development. As we stated before in this context, it is the human trusting the machine. The better the robot adapts its behavior and attitude towards the human, the easier trust will develop. Moreover, access to a potential user’s personality traits could be used as a predictor of the robot’s efficiency in its mission. For example, in the healthcare assistant situation, the personality archetype could predict the reaction towards the technology by the person and, in some cases, the robot could be counterproductive. In conclusion, we presented in this study the emerging field of the personality trait as a factor influencing trust. We presented interesting results paving the way for a greater implication of personality traits in trust development for HRI.



*Proactiveness is the name given by the researchers to a group of factors including 'excitement-seeking', 'general level activity', 'creativity' and 'impulsiveness'

Fig. 4. Summary of the correlations found across the studies

REFERENCES

- [1] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 53, no. 5, pp. 517–527, 2011.
- [2] K. E. Schaefer *et al.*, "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction," Jul. 2014.
- [3] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, 2015, pp. 141–148.
- [4] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004.
- [5] L. Wang, G. A. Jamieson, and J. G. Hollands, "The Effects of Design Features on Users' Trust in and Reliance on a Combat Identification System," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 55, no. 1, pp. 375–379, Sep. 2011.
- [6] H. Lee, J. Kim, and J. Kim, "Determinants of success for application service provider: An empirical test in small businesses," *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 9, pp. 796–815, Sep. 2007.
- [7] A. E. Poropat, "A meta-analysis of the five-factor model of personality and academic performance," *Psychol. Bull.*, vol. 135, no. 2, pp. 322–338, Mar. 2009.
- [8] H. J. Eysenck, "Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm," *Personal. Individ. Differ.*, vol. 12, no. 8, pp. 773–790, 1991.
- [9] L. R. Goldberg, "An alternative 'description of personality': the big-five factor structure," *J. Pers. Soc. Psychol.*, vol. 59, no. 6, p. 1216, 1990.
- [10] H. J. Eysenck and S. B. G. Eysenck, *Manual for the Eysenck Personality Questionnaire: (EPQ-R Adult)*. Educational Industrial Testing Service, 1994.
- [11] T. Sato, "The Eysenck Personality Questionnaire Brief Version: Factor Structure and Reliability," *J. Psychol.*, vol. 139, no. 6, pp. 545–552, Nov. 2005.
- [12] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr., "A very brief measure of the Big-Five personality domains," *J. Res. Personal.*, vol. 37, no. 6, pp. 504–528, Dec. 2003.
- [13] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Towards an Empirically Determined Scale of Trust in Computerized Systems: Distinguishing Concepts and Types of Trust," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 42, no. 5, pp. 501–505, Oct. 1998.
- [14] K. E. Schaefer, "The perception and measurement of human-robot trust," University of Central Florida Orlando, Florida, 2013.
- [15] B. de Ruyter, P. Saini, P. Markopoulos, and A. van Breemen, "Assessing the effects of building social intelligence in a robotic interface for the home," *Interact. Comput.*, vol. 17, no. 5, pp. 522–541, Sep. 2005.
- [16] I. L. Singh, R. Molloy, and R. Parasuraman, "Automation-induced 'complacency': Development of the complacency-potential rating scale," *Int. J. Aviat. Psychol.*, vol. 3, no. 2, pp. 111–122, 1993.
- [17] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity, and social history," *Games Econ. Behav.*, vol. 10, no. 1, pp. 122–142, 1995.
- [18] M. Brühlhart and J.-C. Usunier, "Does the trust game measure trust?," *Econ. Lett.*, vol. 115, no. 1, pp. 20–23, Apr. 2012.
- [19] H. Edward and others, "The hidden dimension," *Doubleday Gard. City*, vol. 14, pp. 103–124, 1966.
- [20] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, "The benefits of interactions with physically present robots over video-displayed agents," *Int. J. Soc. Robot.*, vol. 3, no. 1, pp. 41–52, 2011.
- [21] S. M. Merritt and D. R. Ilgen, "Not all trust is created equal: Dispositional and history-based trust in human-automation interactions," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 50, no. 2, pp. 194–210, 2008.
- [22] B. M. Muir, "Trust between humans and machines, and the design of decision aids," *Int. J. Man-Mach. Stud.*, vol. 27, no. 5–6, pp. 527–539, 1987.
- [23] K. S. Haring, Y. Matsumoto, and K. Watanabe, "How do people perceive and trust a lifelike robot," in *Proceedings of the World Congress on Engineering and Computer Science*, 2013, vol. 1.
- [24] M. L. Walters *et al.*, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, 2005, pp. 347–352.
- [25] M. L. Walters *et al.*, "The Influence of Subjects' Personality Traits on Predicting Comfortable Human-Robot Approach Distances," in *Proceedings of Cog Sci 2005 Workshop: Toward Social Mechanisms of Android Science*, 2005, pp. 29–37.
- [26] L. Takayama and C. Pantofaru, "Influences on proxemic behaviors in human-robot interaction," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 5495–5502.
- [27] D. S. Syrdal, K. Dautenhahn, S. N. Woods, M. L. Walters, and K. L. Koay, "Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance," in *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, 2007, pp. 86–92.
- [28] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst, and K. L. Koay, "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Auton. Robots*, vol. 24, no. 2, pp. 159–178, Feb. 2008.
- [29] R. Looije, M. A. Neerinx, and F. Cossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 6, pp. 386–397, 2010.
- [30] A. M. Evans and W. Revelle, "Survey and behavioral measurements of interpersonal trust," *J. Res. Personal.*, vol. 42, no. 6, pp. 1585–1593, 2008.
- [31] C. Shikishima, K. Hiraishi, and J. Ando, "Genetic and environmental influences on general trust: A test of a theory of trust with behavioral genetic and evolutionary psychological approaches," *Jpn. J. Soc. Psychol.*, vol. 22, no. 1, pp. 48–57, 2006.
- [32] D. P. Schmitt, J. Allik, R. R. McCrae, and V. Benet-Martinez, "The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations," *J. Cross-Cult. Psychol.*, vol. 38, no. 2, pp. 173–212, Mar. 2007.
- [33] O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," *Handb. Personal. Theory Res.*, vol. 2, no. 1999, pp. 102–138, 1999.

Index alphabétique

- Agent conversationnel, 1
- Apprentissage, 29
 - automatique, 29
 - développemental, 29
- Apprentissage actif, 53
- Architecture cognitive, 33

- Cadres pragmatiques, 29
- Cobotique, 11
- Collaboration homme-robot, 11
- Communication homme-robot, 11, 24
 - non-verbale, 11, 24
 - verbale, 11
- Comportement non-verbal, 24
- Confiance, 62

- Dialogue, 1

- Education, 53
- Emotions, 16

- Interaction homme-robot, 29, 33, 47, 53
 - enfant, 53
 - long-terme, 47, 53
- Interaction située, 1
- Interaction sociale, 47

- Mémoire épisodique, 47

- Personnalité, 62
- Personnes âgées, 41
- Persuasion, 24
- Proxémique, 62
- Psychothérapie, 41

- Robot social, 11, 16, 24, 29, 41, 47
 - assistant, 11, 41
 - compagnon, 16, 47
 - médiateur, 41

- Tour de parole, 1

- Vallée de l'étrange, 16, 53



Année universitaire 2016-2017
Grenoble, France