



HAL
open science

Modélisation du profil émotionnel de l'utilisateur dans les interactions parlées Humain-Machine

Agnès Delaborde

► **To cite this version:**

Agnès Delaborde. Modélisation du profil émotionnel de l'utilisateur dans les interactions parlées Humain-Machine. Autre [cs.OH]. Université Paris Sud - Paris XI, 2013. Français. NNT : 2013PA112225 . tel-00958508

HAL Id: tel-00958508

<https://theses.hal.science/tel-00958508v1>

Submitted on 12 Mar 2014

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.



THÈSE

Discipline : INFORMATIQUE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE D'INFORMATIQUE

Présentée par

Agnes Delaborde

Modélisation du profil émotionnel de l'utilisateur dans les interactions parlées Humain-Machine

Soutenue le 19 décembre 2013

Sylvie Pesty	Professeur, Université de Grenoble, LIG-Magma	Rapporteur
Daniel Luzzati	Professeur, Université du Maine, LIUM-LST	Rapporteur
Bernadette Dorizzi	Professeur, Telecom Sud Paris, Département EPH	Examineur
Anne Vilnat	Professeur, Université Paris-Sud, LIMSI-CNRS	Examineur
Axel Buendía	Docteur, Responsable développement SpirOps AI	Examineur
Laurence Devillers	Professeur, Université Paris Sorbonne IV, LIMSI-CNRS	Directeur de thèse

Résumé

Les travaux de recherche de la thèse portent sur l'étude et la formalisation des interactions émotionnelles Humain-Machine. Au delà d'une détection d'informations paralinguistiques (émotions, disfluences,...) ponctuelles, il s'agit de fournir au système un profil interactionnel et émotionnel de l'utilisateur dynamique, enrichi pendant l'interaction. Ce profil permet d'adapter les stratégies de réponses de la machine au locuteur, et il peut également servir pour mieux gérer des relations à long terme. Le profil est fondé sur une représentation multi-niveau du traitement des indices émotionnels et interactionnels extraits à partir de l'audio via les outils de détection des émotions du LIMSI. Ainsi, des indices bas niveau (variations de la F0, d'énergie, etc.), fournissent des informations sur le type d'émotion exprimée, la force de l'émotion, le degré de loquacité, etc. Ces éléments à moyen niveau sont exploités dans le système afin de déterminer, au fil des interactions, le profil émotionnel et interactionnel de l'utilisateur. Ce profil est composé de six dimensions : optimisme, extraversion, stabilité émotionnelle, confiance en soi, affinité et domination (basé sur le modèle de personnalité OCEAN et les théories de l'*interpersonal circumplex*). Le comportement social du système est adapté en fonction de ce profil, de l'état de la tâche en cours, et du comportement courant du robot. Les règles de création et de mise à jour du profil émotionnel et interactionnel, ainsi que de sélection automatique du comportement du robot, ont été implémentées en logique floue à l'aide du moteur de décision développé par un partenaire du projet ROMEO. L'implémentation du système a été réalisée sur le robot NAO. Afin d'étudier les différents éléments de la boucle d'interaction émotionnelle entre l'utilisateur et le système, nous avons participé à la conception de plusieurs systèmes : système en Magicien d'Oz pré-scripté, système semi-automatisé, et système d'interaction émotionnelle autonome. Ces systèmes ont permis de recueillir des données en contrôlant plusieurs paramètres d'élicitation des émotions au sein d'une interaction ; nous présentons les résultats de ces expérimentations, et des protocoles d'évaluation de l'Interaction Humain-Robot via l'utilisation de systèmes à différents degrés d'autonomie.

Abstract

Analysing and formalising the emotional aspect of the Human-Machine Interaction is the key to a successful relation. Beyond and isolated paralinguistic detection (emotion, disfluences...), our aim consists in providing the system with a dynamic emotional and interactional profile of the user, which can evolve throughout the interaction. This profile allows for an adaptation of the machine's response strategy, and can deal with long term relationships. A multi-level processing of the emotional and interactional cues extracted from speech (LIMSI emotion detection tools) leads to the constitution of the profile. Low level cues (F0, energy, etc.), are then interpreted in terms of expressed emotion, strength, or talkativeness of the speaker. These mid-level cues are processed in the system so as to determine, over the interaction sessions, the emotional and interactional profile of the user. The profile is made up of six dimensions: optimism, extroversion, emotional stability, self-confidence, affinity and dominance (based on the OCEAN personality model and the interpersonal circumplex theories). The information derived from this profile could allow for a measurement of the engagement of the speaker. The social behaviour of the system is adapted according to the profile, and the current task state and robot behaviour. Fuzzy logic rules drive the constitution of the profile and the automatic selection of the robotic behaviour. These determinist rules are implemented on a decision engine designed by a partner in the project ROMEO. We implemented the system on the humanoid robot NAO. The overriding issue dealt with in this thesis is the viable interpretation of the paralinguistic cues extracted from speech into a relevant emotional representation of the user. We deem it noteworthy to point out that multimodal cues could reinforce the profile's robustness. So as to analyse the different parts of the emotional interaction loop between the user and the system, we collaborated in the design of several systems with different autonomy degrees: a pre-scripted Wizard-of-Oz system, a semi-automated system, and a fully autonomous system. Using these systems allowed us to collect emotional data in robotic interaction contexts, by controlling several emotion elicitation parameters. This thesis presents the results of these data collections, and offers an evaluation protocol for Human-Robot Interaction through systems with various degrees of autonomy.

Remerciements

La réalisation d'une thèse est une expérience unique : si elle représente la finalité d'un cycle d'étude, elle constitue également un apprentissage du travail de chercheur dans un environnement concret. Effectuer des expérimentations auprès d'utilisateurs potentiels, éplucher des pages et des pages d'études à la recherche de l'information manquante, présenter ses travaux devant la communauté, sont autant d'activités que j'ai réalisées avec un immense plaisir.

Je remercie tout d'abord les membres du jury pour le temps et l'attention portés à mes travaux, et pour leurs questionnements pertinents. Mes remerciements vont également naturellement au professeur Laurence Devillers, qui a su réunir tous les éléments pour que ces quatre années se déroulent admirablement bien, tant d'un point de vue professionnel que personnel. J'ai été heureuse de travailler au sein du LIMSI, auprès de personnes dynamiques et agréables. Je remercie pour leur sympathie l'équipe du thème « Dimensions affectives et sociales dans les interactions parlées » : travailler avec vous m'a permis d'apprendre beaucoup. Je remercie également les équipes des partenaires du projet, SpirOps et Aldebaran Robotics, avec lesquels j'ai eu l'occasion de collaborer.

Le soutien de ma famille et de mes proches m'a permis de conserver intacte la motivation de réaliser ce travail de recherche. Merci à Ludovic pour son soutien quotidien, son aide précieuse les jours de rush, sa présence rassurante. Merci à Dominique, qui a su me redonner courage à chaque fois que j'en ai exprimé le besoin, et qui m'a donné l'envie d'aller jusqu'au bout. Merci enfin à Marie, pour ses encouragements... percutants... lors de ma rédaction de thèse !

Table des matières

Résumé	3
Abstract	4
Table des matières	6
Introduction générale.....	10
1.1. Motivations et contexte	10
1.2. Objectifs de recherche	11
1.3. Considérations éthiques	12
1.4. Challenges et méthodologies.....	13
1.5. Apports de la thèse	14
1.6. Organisation du manuscrit	15
2. État de l'art	17
2.1. Systèmes d'Interaction Humain-Machine.....	17
2.1.1. Perception et interprétation de la communication non-verbale dans les Interaction Humain-Machine.....	18
2.1.2. Mémorisation et gestion de l'interaction à long-terme.....	21
2.1.3. Comportements du système.....	22
2.2. Théories sur les émotions	26
2.2.1. La paralinguistique	26
2.2.2. Les émotions : définitions	26
2.2.3. Les marqueurs affectifs.....	30
2.2.4. Les émotions dans l'interaction	30
2.3. Évaluation des systèmes d'Interaction Humain-Machine.....	31
2.3.1. Évaluation des performances du système.....	32
2.3.2. Évaluation psychologique et émotionnelle de l'utilisateur.....	33
2.3.3. Interaction Humain-Robot émotionnelle : que faut-il évaluer, et comment ?	35
2.4. Corpus d'interaction existants	35
2.4.1. Interaction Humain-Humain	36
2.4.2. Interaction Humain-Robot	36
2.5. Challenges actuels et pistes de recherche proposées	41
2.5.1. Challenges actuels de recherche	41

2.5.2.	Nos pistes de recherche	41
3.	Profil émotionnel et interactionnel de l'utilisateur	43
3.1.	Introduction	43
3.2.	Profil utilisateur et paralinguistique dans la littérature.....	44
3.2.1.	Profil utilisateur	44
3.2.2.	Profil de personnalité et paralinguistique.....	46
3.2.3.	Profil social et paralinguistique.....	50
3.3.	Nos travaux : Représentation émotionnelle et interactionnelle de l'utilisateur.....	54
3.3.1.	Le profil émotionnel et interactionnel : dimensions.....	54
3.3.1.1.	Dimensions émotionnelles	54
3.3.1.2.	Dimensions interactionnelles	56
3.3.2.	Le traitement multi-niveau des indices paralinguistiques.....	57
3.3.2.1.	Interprétation paralinguistique du signal audio.....	58
3.3.2.2.	Interprétation émotionnelle et interactionnelle des informations paralinguistiques	60
3.4.	Conclusion.....	63
4.	Comportements sociaux du système	65
4.1.	Introduction	65
4.2.	Stratégies de comportement du système dans la littérature.....	65
4.2.1.	Détermination des comportements sociaux du système.....	65
4.2.2.	Relation entre comportement social du système et profil utilisateur.....	67
4.3.	Nos travaux : Codification des comportements sociaux du système	68
4.3.1.	Comportements sociaux en fonction du contexte	68
4.3.2.	Comportements sociaux en fonction du profil utilisateur.....	70
4.3.3.	Expression du comportement social	72
4.3.3.1.	En situation d'assistance à personnes en perte d'autonomie	73
4.3.3.2.	En situation de jeu (sujets autonomes)	76
4.4.	Conclusion.....	78
5.	Expérimentations et corpus	79
5.1.	Introduction	79
5.2.	Étude des réactions émotionnelles utilisateur.....	81
5.2.1.	Interaction Adulte-Système – Corpus IDV-HS	81
5.2.1.1.	Protocole d'acquisition.....	81

5.2.1.2.	Annotation et scores.....	84
5.2.1.3.	Résultats.....	88
5.2.2.	Interaction Enfant-Robot – Corpus NAO-HR1	92
5.2.2.1.	Protocole d’acquisition.....	92
5.2.2.2.	Annotation et scores.....	95
5.2.2.3.	Résultats.....	98
5.2.3.	Conclusion	99
5.3.	Expressions émotionnelles en réaction aux comportements robotiques.....	100
5.3.1.	Interaction Enfant-Robot – Corpus NAO-HR1	100
5.3.1.1.	Résultats.....	100
5.3.2.	Interaction Adulte-Robot – Corpus IDV-HR.....	105
5.3.2.1.	Protocole d’acquisition.....	106
5.3.2.2.	Annotation et scores.....	110
5.3.2.3.	Résultats.....	113
5.3.3.	Conclusion	117
5.4.	Relation entre indices paralinguistiques et profil émotionnel et interactionnel	118
5.4.1.	Interaction Adulte-Robot et Adulte-Système	119
5.4.1.1.	Annotation perceptive du profil émotionnel et interactionnel	119
5.4.1.2.	Scores d’annotation	120
5.4.1.3.	Protocole pour une analyse statistique	122
5.4.2.	Conclusion	124
5.5.	Conclusion.....	124
6.	Implémentation et évaluation	126
6.1.	Introduction.....	126
6.2.	Implémentation	127
6.2.1.	Architecture du système d’interaction.....	127
6.2.1.	Représentation formelle des règles de logique floue	129
6.2.2.	Expression des comportements sociaux robot.....	132
6.3.	Mise en place d’un système semi-automatique avec sélection de comportement et mise à jour profil : expérimentation NAO-HR3	134
6.3.1.	Environnement de test.....	135
6.3.2.	Analyse des données obtenues	137

6.3.3.	Conclusion	140
6.4.	Évaluation d'un système d'Interaction Humain-Machine via un Magicien d'Oz à différents degrés d'autonomie	140
6.4.1.	Définition et intérêt du Magicien d'Oz	142
6.4.2.	Magicien d'Oz avec contrôle complet de l'expérimentateur	143
6.4.3.	Magicien d'Oz semi-automatique.....	145
6.4.4.	Système entièrement autonome.....	146
6.5.	Conclusion.....	148
Conclusion et perspectives.....		148
7.1.	Résumé.....	148
7.2.	Contributions de la thèse	149
7.2.1.	Le profil émotionnel et interactionnel de l'utilisateur.....	150
7.2.2.	Sélection du comportement social robotique	151
7.2.3.	Formalisation et implémentation de la gestion du profil et de la sélection du comportement.....	151
7.2.4.	Le Magicien d'Oz à différents degrés d'autonomie pour l'évaluation	152
7.3.	Les perspectives de travail et ouvertures	153
7.3.1.	L'annotation perceptive du profil émotionnel et interactionnel.....	153
7.3.2.	L'évaluation.....	153
7.3.3.	Vers une approche multimodale	154
7.3.4.	Vers une plus grande autonomie du système.....	154
Publications et présentations associées à la thèse		156
Bibliographie.....		157
Liste des Tableaux		171
Table des Figures		173
Annexes.....		177
Annexe I.....		177
Annexe II		188
Annexe III.....		196
Annexe IV.....		198
Annexe V		202

Introduction générale

1.1. Motivations et contexte

Dialoguer avec une machine peut se présenter sous différents aspects, des plus rudimentaires au plus élaborés. Faire comprendre ses intentions, ses désirs ou ses préférences au système peut se matérialiser par exemple sous la forme d'un clic indiquant si oui ou non l'utilisateur souhaite réaliser une action. Ce contexte est basique et pourtant, déjà, les développeurs de l'application peuvent être certains d'assurer une partie du confort de l'utilisateur en se penchant sur certains paramètres d'ergonomie tels que « le choix se présente-t-il à un moment favorable pour l'utilisateur ? », « le panel de réponses que nous lui proposons est-il en adéquation avec le contexte ? ». Ils peuvent également s'appuyer sur les habitudes de l'utilisateur, telles que « l'utilisateur souhaite-t-il que je lui pose cette question, ou préfère-t-il que le système prenne la main ? ». Dans un contexte aussi « simple », nous voyons déjà l'impact qu'a la prise en compte des caractéristiques de l'environnement utilisateur pour une utilisation plaisante, et donc durable, du système. Le cas de l'interaction avec un système doté d'intelligence artificielle élaboré (un avatar, ou un robot) fait surgir également d'autres préoccupations pour les designers. Plus le nombre de modalités d'interaction augmente, plus la modélisation de l'interaction se complexifie. Si le système sait reconnaître les gestes, il faut pouvoir les décrypter. Le système intègre la reconnaissance des émotions faciales ? Il faut savoir les reconnaître et correctement les interpréter. L'interaction par la voix ? Le comportement exprimé par la machine devra alors être cohérent avec l'intention sémantique de l'énoncé.

Parallèlement, plus le système semblera comprendre le raisonnement humain, plus l'utilisateur sera tenté de lui attribuer des caractéristiques humaines. Plus la machine mimera l'intelligence, plus l'utilisateur sera pointilleux et exigeant. Et plus la machine tentera de s'apparenter à un humain... Nous en venons bien sûr à la théorie de l'*Uncanny Valley* (Mori, 1970). Les réactions et l'apparence du système doivent être scrupuleusement étudiées : si le système tente d'être trop humain, la moindre erreur de conception le fera passer pour un « monstre ». Le défi de l'interaction Humain-Machine est alors immensément complexe. Le système doit tout d'abord *comprendre* l'utilisateur, puis ensuite *réagir* de façon compréhensible, mais en prenant garde toutefois de ne pas « en faire trop », au risque de « le faire mal », et de s'aliéner l'utilisateur. Il est donc important dans la réalisation d'un système Humain-Machine de rester humble dans notre compréhension de l'interaction humaine, sous peine de devoir faire face à des écueils insurmontables.

Ce travail de thèse s'est déroulé dans le cadre du projet robotique français Romeo¹, en application sur le robot Nao. Ce projet est labellisé par le pôle de compétitivité Cap Digital et financé par la Région Ile-de-France, la Direction Générale de la Compétitivité, de l'Industrie et des Services (DGCIS) et la Ville de Paris. Nao est un robot humanoïde bipède de cinquante-huit centimètres, capable d'interagir avec son environnement (micros, haut-parleurs, caméras, capteurs de pression et infrarouges, centrale inertielle). Même si la tête du robot n'est pas douée d'expressivité faciale, son corps pourvu de quatorze à vingt-cinq

¹ <http://projetromeo.com>

degrés de liberté lui permettent de représenter toute une gamme d'expressions émotionnelles. Ce robot est utilisé comme support d'étude pour le développement du futur robot humanoïde Romeo (un mètre quarante), dont l'application end-user principale sera l'assistance à domicile aux personnes en perte d'autonomie, mais également le divertissement (notamment avec des enfants dans le cadre du projet).

Outre les challenges de mobilité du robot, d'efficacité des capteurs et les contraintes liées à la puissance du système embarqué, le projet Romeo tente de répondre au défi de l'interaction intelligente Humain-Robot. Le robot doit être capable de dialoguer par la parole avec l'utilisateur sur du *small talk*, mais aussi de comprendre les émotions exprimées dans sa voix, et d'y réagir de façon idoine. Le robot Romeo évoluera en contexte réel : utilisé au domicile de l'utilisateur, en totale autonomie, il fera face et tentera de comprendre un contexte multimodal riche (visuel, tactile et audio). Afin de fluidifier le dialogue entre les différents modules, le système est doté d'une mémoire centrale (nommée ALMemory) et d'un système de prise de décision automatique développé par l'entreprise SirOps Artificial Intelligence². Ce système permet de représenter les comportements d'intelligence artificielle sous la forme d'une large quantité de *drives*³ prenant pour entrée les capteurs du système (par exemple l'état de l'environnement à l'instant t , l'état du système, etc.) et, à l'aide de règles de logique floue, de déterminer le meilleur comportement à adopter pour le système. Il est ainsi possible de faire évoluer conjointement une grande quantité de règles intervenant dans la prise de décision (SpirOps, 2005). L'implémentation des travaux réalisés au cours de cette thèse s'appuie sur ce système de prise de décision.

1.2. Objectifs de recherche

Lors de la conception d'un système d'interaction Humain-Machine, il est primordial de prendre en considération les composantes émotionnelles de la communication humaine. En effet, les affects sont étroitement liés au processus cognitif et à l'interaction sociale, comme le souligne notamment (Cosnier, 1987). La peur ou un haut niveau de stress, par exemple, pourront troubler, voire même empêcher, le raisonnement. La colère, même si elle n'est pas causée par, ni dirigée sur, l'interlocuteur, pourra modifier l'interaction. Ainsi, que le système soit implémenté dans un contexte de divertissement (robot joueur, avatar dans un jeu, etc.), d'aide à la personne (notamment *serious game*, robot assistant) ou d'enseignement (application d'e-learning, coaching), la qualité de l'interaction entre l'entité virtuelle et l'utilisateur et la bonne conduite de la tâche à réaliser – qui peut consister simplement en une interaction sociale plaisante – seront nettement facilités si le système est capable de prendre en compte, d'interpréter, et de répondre convenablement aux émotions exprimées par l'utilisateur.

Nous nous intéressons tout particulièrement à l'interprétation des informations paralinguistiques reconnues automatiquement par les outils de détection audio des émotions développés au laboratoire LIMSI-CNRS d'Orsay⁴. Ces outils permettent d'extraire du signal de parole des informations bas niveau telles que,

² SpirOps : <http://www.spirops.com>

³ Originellement des signaux du système nerveux central nécessaires à la survie. Ici, on peut les comprendre comme des unités de raisonnement.

⁴ LIMSI-CNRS : <http://www.limsi.fr>

notamment, les variations de la F0, d'énergie, de durée du temps de parole (Devillers *et al.*, 2005a). L'interprétation statistique de ces informations fournit des informations sur le type d'émotion exprimée (Joie, Colère, etc.), la valence (Positif, Négatif ou Neutre), la force de l'émotion, ou encore le degré de loquacité du locuteur (le ratio entre silence et temps de parole).

Un système de détection des émotions a été spécialement développé par le LIMSI-CNRS dans le cadre du projet ROMEO en utilisant les bibliothèques d'extraction d'indices bas niveau de Voxler⁵. Le robot Romeo aura ainsi potentiellement la capacité de détecter les émotions dans la voix de l'utilisateur. La question est de savoir comment il pourra exploiter au mieux ces informations. Brutes, elles signifient qu'à l'instant t , l'utilisateur était dans tel état émotionnel. Comment estimer l'impact de cet état émotionnel sur le comportement du robot ? En quoi cela peut-il – doit-il – modifier son comportement ?

Au delà d'une détection d'expressions paralinguistiques (émotions, disfluences,...) ponctuelles, il convient donc d'interpréter à plus haut niveau ces indices, et de fournir au système un modèle dynamique, et sur le long-terme, du profil de l'utilisateur. Le travail de cette thèse porte donc tout d'abord sur l'interprétation des émotions exprimées : que peuvent-elles révéler de l'individu ? Que signifient-elles dans l'interaction de cet individu avec son interlocuteur ? Quelles informations sont pertinentes et exploitables dans l'interaction avec le robot ? Toutes ces informations pourront alors permettre de déterminer, au fil des interactions, un profil émotionnel et interactionnel de l'utilisateur : ses tendances émotionnelles, et son attitude sociale par rapport à son interlocuteur robotique. Ce profil fournit ainsi un vecteur constitué de différentes variables propres à illustrer le niveau d'engagement de l'utilisateur dans son interaction avec le robot. Le robot aura donc une base sur laquelle appuyer sa sélection de comportement social, afin de satisfaire au mieux l'individu. Nous proposons dans ces travaux une approche sur l'interprétation des émotions exprimées oralement par l'utilisateur dans un système d'intelligence artificielle. Un renforcement par des indices issus d'une perception langagière et/ou gestuelle basé sur le même modèle améliorerait encore la pertinence du profil (cet aspect multimodal sera par ailleurs pris en compte dans la suite du projet ROMEO2).

1.3. Considérations éthiques

En 2010, le CNRS et l'INRIA rendent publique un rapport (Mariani *et al.*, 2009) visant à promouvoir le respect de l'éthique dans les Sciences et Technologies de l'Information et de la Communication (les STIC). Ce rapport propose des recommandations dans les différents aspects des STIC, et notamment dans la robotique.

Il est par exemple recommandé d'éviter les « effets d'annonce » visant à promouvoir des produits dont le niveau de développement n'est pas en mesure de fournir au public le résultat escompté. Cette question a été abordée notamment lors de notre première collecte de données avec des participants en perte d'autonomie, auxquels nous devons présenter une simulation de système de détection des émotions, mais également lors d'une expérimentation avec un même public, en interaction avec le robot NAO dirigé par un expérimentateur. Non seulement ces systèmes étaient introduits comme étant en test, et, bien que manipulés

⁵ Voxler : <http://www.voxler.eu>

par des expérimentateurs humains, nous nous sommes appliqués à ne pas présenter des systèmes au fonctionnement et à l'intelligence « parfaits ».

La conception de nos interfaces d'Interaction Humain-Machine s'appuie également sur un respect de la satisfaction de l'utilisateur, et non pas seulement sur des considérations d'ordre technique. Comme nous le verrons dans la description de nos travaux, nous nous sommes attachés à étudier le comportement du robot afin qu'il soit socialement désirable et animé d'intentions bénéfiques pour l'homme. Ces considérations, si elles peuvent paraître naïves de premier abord, sont en fait essentielles pour la détermination de l'espace de comportement des systèmes sociaux.

L'importance de l'éthique, notamment dans la robotique, est un sujet d'actualité auquel aucun chercheur ne peut se soustraire. Le projet ROMEO, par exemple, est soumis à un conseil constitué d'experts pluridisciplinaires (philosophes, spécialistes en neurosciences, etc.). Le domaine de l'assistance par plateforme robotique soulève également de nombreuses réflexions éthiques dans la communauté.

La réflexion éthique ne concerne cependant pas que l'action du robot sur l'utilisateur. Elle fait partie de chaque activité du chercheur, et conditionne la conception des protocoles d'enregistrement, d'acquisition, d'annotation, et d'utilisation des données. Il est important notamment de respecter l'anonymat des personnes ayant consenti à se faire enregistrer et à fournir des données personnelles, ainsi que de décorrélérer les interprétations réalisées sur ces données et l'identité de l'individu.

Les travaux présentés au cours de cette thèse ont été réalisés dans cet esprit.

1.4. Challenges et méthodologies

Déterminer le profil de l'utilisateur et les comportements désirables chez le robot, et les implémenter dans un système robotique, nécessitent de faire appel à différents champs disciplinaires. Ainsi, les travaux réalisés au cours de cette thèse sont ancrés sur les champs disciplinaires de l'intelligence artificielle, de la robotique sociale, la psychologie, la psychologie sociale, et la représentation des connaissances. Nous nous sommes donc posé différentes questions :

Quelles sont les réactions émotionnelles exprimées en interaction homme-robot, dans nos situations d'assistance et de jeu ? Afin d'étudier les différents éléments de la boucle d'interaction émotionnelle entre l'utilisateur et le robot, nous avons mis au point plusieurs systèmes présentant différents degrés d'autonomie : du système en magicien d'Oz pré scripté (comportements du robot définis à l'avance, insensibilité aux émotions du locuteur) au système d'interaction émotionnelle entièrement automatisé (perception automatique des émotions, mise à jour du profil et sélection des comportements automatiques), en passant par un système magicien d'Oz partiel (perception des émotions par l'expérimentateur et sélection du comportement et du profil automatiques).

Quelle relation entre émotions à court terme et interprétation sur le long terme ? Les travaux de la communauté en psychologie et en psycho-sociologie déterminent des relations entre la personnalité de l'individu, et les émotions qu'il peut potentiellement exprimer. Comme nous pouvons le lire dans les travaux de (Revelle et Scherer, 2009), « *[P]ersonality is to emotion as climate is to weather* » : l'émotion, étudiée sur le long terme, serait un indice de la personnalité de l'individu. En termes de relations sociales, les inflections de la voix, la durée de l'interaction, la prise du temps de parole, ont été diagnostiqués comme révélateurs de la

façon dont l'individu se positionne par rapport à son interlocuteur. (Spencer-Oatey, 1996) insiste notamment sur l'importance de la durée de la relation dans la détermination du lien social entre deux individus. Cette étude préliminaire nous oriente donc sur deux pistes : a) les informations paralinguistiques sont révélatrices de la personnalité et du positionnement social de l'individu, b) sans gestion du long terme, elles ne signifient rien sur le profil de l'individu.

Quelles sont les limites ? Nous nous retrouvons donc face à certaines limitations. Tout d'abord, les représentations et interprétations proposées dans les travaux de la communauté en interaction humain-humain font référence à des contextes riches et nuancés. Dans un système de détection automatique de l'émotion appliqué à la robotique, comment savoir ce qui a déclenché la colère de l'utilisateur ? S'agit-il de ce qu'a fait le robot, ou d'un événement extérieur non perçu ? Il convient alors de répertorier ce que nous pouvons mesurer et contrôler dans l'interaction, et de prendre position sur la façon de considérer les événements extérieurs à notre contrôle. Il convient alors de réduire notre interprétation de la personnalité et du lien social de l'utilisateur, à un profil des tendances émotionnelles et interactionnelles. Nous nous sommes également penchés sur la dimension éthique lors de la constitution du profil utilisateur – ce qui a notamment contribué à la réduction de l'étendue du profil : un robot à capacité de perception limitée peut-il décréter que l'utilisateur présente un haut degré de neuroticisme ? Et en ce qui concerne la sélection des comportements robotiques : le robot peut-il avoir un comportement négatif ?

Quelle implémentation ? Comme nous l'avons décrit précédemment, le système de prise de décisions du robot est régi par des règles de logique floue, implémentées dans le système SpirOps AI, participant à la sélection de comportement du robot Nao (et du futur Romeo). Il s'agit alors de déterminer les règles « expert » susceptibles de représenter l'utilisateur d'une façon compréhensible par le robot. Cette détermination inclut de prendre en compte toutes les entrées récupérables par le robot : les différentes informations reflétant l'état émotionnel de l'individu, l'historique de ces manifestations, le comportement du robot précédemment présenté à l'utilisateur, et l'état de la tâche en cours.

1.5. Apports de la thèse

L'enjeu principal de cette étude est donc tout d'abord l'interprétation viable des informations paralinguistiques extraites du signal audio, afin d'obtenir une représentation émotionnelle de l'utilisateur pertinente. L'interprétation des informations issues du signal audio, à court terme, permet d'obtenir des informations sur l'état affectif du locuteur. Une interprétation à moyen terme de ce signal peut renseigner sur l'interaction sociale, et l'interprétation à plus long terme fournirait le profil émotionnel et interactionnel du locuteur. Notre travail de thèse consiste donc à réaliser l'interprétation haut niveau des informations paralinguistiques extraites du signal audio, et de proposer un profil émotionnel et interactionnel susceptible de doter le système d'une représentation dynamique de l'utilisateur.

Dans l'étude de la boucle d'interaction entre l'utilisateur et le système, nous nous intéressons à la relation existant entre le profil utilisateur et les comportements exprimés par le système. Nous étudions les comportements sociaux possibles pour le robot, et délimitons un espace des comportements sociaux désirables selon le contexte d'interaction. Nous proposons également des règles pour la détermination du meilleur comportement possible en fonction du profil de l'utilisateur.

Nos études et expérimentations, ainsi que notre collaboration avec les partenaires du projet ROMEO, nous permettent de formaliser et d'implémenter la constitution automatique du profil et la sélection de comportement robotique. Nous avons réalisé plusieurs collectes de données avec des utilisateurs cible, en contexte d'application proche du contexte final.

Les différents systèmes que nous avons mis au point pour ces expérimentations nous permettent de réaliser l'évaluation de certaines composantes de l'interaction émotionnelle entre l'utilisateur et le robot. Nous proposons différentes architectures modulaires permettant de réaliser une évaluation à l'aide de systèmes à différents degrés d'autonomie, incluant à la fois des modules automatisés et des modules gérés en Magicien d'Oz.

1.6. Organisation du manuscrit

Notre manuscrit présente cinq grandes parties.

Tout d'abord, nous présentons Chapitre 2 un état de l'art des systèmes d'Interaction Humain-Machine, des théories sur les émotions et sur l'évaluation des Interaction Humain-Machine, et un tour d'horizon des corpus de données interactionnelles existants. Dans ce chapitre, nous nous intéressons aux travaux de la communauté quant à la perception et à la modélisation des émotions dans les systèmes d'Interaction Humain-Machine : quels sont les éléments de perception utilisés dans la communauté ? Comment sont-ils modélisés ? Comment sont-ils évalués ? Nous résumons ensuite les challenges actuels de recherche et proposons nos pistes de recherche situées dans ces axes de recherche.

Nous présentons ensuite (Chapitre 3) une représentation émotionnelle et interactionnel de l'utilisateur. Nous organisons la présentation comme suit : tout d'abord, un état de l'art des travaux de la communauté orienté spécifiquement sur le profil utilisateur, et l'utilisation de la paralinguistique dans la représentation de la personnalité et des caractéristiques d'interaction sociale de l'utilisateur. Nous exploitons ensuite cette base d'étude afin de présenter nos propres travaux sur la constitution automatique du profil émotionnel et interactionnel de l'utilisateur à partir d'indices paralinguistiques extraits du signal audio.

Le chapitre 4 suivant détaille nos analyses sur les comportements sociaux du système. Dans un premier temps, nous présentons les études de la communauté sur lesquelles nous nous fondons, au sujet de la détermination des comportements sociaux du système et leur relation avec le profil utilisateur. Nous présentons ensuite nos travaux pour déterminer les comportements du robot en fonction de l'utilisateur et du contexte, et sur l'expression du comportement en fonction du contexte applicatif et du public.

Au cours de ces travaux de thèse, nous avons mené différentes expérimentations visant à la collecte de données en conditions d'interaction réelle. Nous présentons ces différentes études, ainsi qu'une analyse des données permettant d'étudier différentes composantes de la boucle d'interaction émotionnelle entre l'utilisateur et le robot, à la lumière de nos travaux présentés dans les chapitres 3 et 4.

Le dernier chapitre est consacré à l'implémentation et à l'évaluation de notre système d'Interaction émotionnelle Humain-Robot. Nous présentons l'architecture du système, ainsi que la représentation formelle des règles de logique floue présidant à la constitution du profil émotionnel et interaction de l'utilisateur ainsi que de la sélection automatique du comportement robot. Nous montrons également la réalisation des comportements sociaux du robot sur le robot NAO. Nous présentons ensuite une expérimentation mettant

en œuvre notre système en semi-automatique (entrée des émotions exprimées à la main, et sélection profil et comportements automatiques), dans le cadre d'un scénario original d'interaction ludique entre le robot et des enfants. Dans la dernière partie du chapitre, nous proposons un protocole d'analyse d'un système d'Interaction Humain-Robot par l'utilisation de systèmes modulaires à différents degrés d'autonomie. Cette technique repose sur la mise en œuvre de modules autonomes ou dirigés en Magicien d'Oz, et permet de réaliser une évaluation technique, mais également d'évaluer la cohésion sociale des comportements et la satisfaction de l'utilisateur.

Chapitre 2

2. État de l'art

Tout d'abord, nous présentons l'état de l'art de la perception de la communication non-verbale dans les systèmes d'Interaction Humain-Machine, et l'interprétation réalisée par ces systèmes. Ces applications peuvent traiter de l'interaction Humain-ACA ou Humain-Robot. Nous nous intéressons également aux modalités de perception des émotions utilisées dans ces applications : la gestuelle, la voix, le lexique, ou encore l'état de la tâche en cours. L'exploitation de ces informations permet d'obtenir des mesures d'attention portée mutuellement par les interactants, d'engagement dans l'interaction, mais également une représentation de la personnalité. Nous organisons cette étude selon une typologie présentant un degré plus ou moins haut « d'intelligence » du système, via la quantité d'informations exploitées par le système et le niveau d'interprétation réalisé.

Notre étude se spécialisant sur l'expression des émotions, nous nous concentrons ensuite sur la définition de la paralinguistique en général, et plus particulièrement des émotions et leurs manifestations à plus ou moins long terme, ainsi qu'aux marqueurs affectifs (également nommés « *affect bursts* » en anglais). Nous présentons également les différents modèles émotionnels, et différentes analyses portant sur les émotions dans l'interaction.

La troisième partie de l'État de l'art est consacrée à l'évaluation des systèmes d'Interaction Humain-Machine. L'évaluation est une étape primordiale pour valider le bon fonctionnement du système, et son adéquation avec la tâche finale visée. Nous répertorions les différentes métriques et méthodes proposées par la communauté, tant du point de l'évaluation du système que de l'évaluation du profil psychologique et émotionnel de l'utilisateur.

Nous nous intéressons ensuite aux corpus de données d'interaction existants dans la communauté, car l'étude des interactions (pour la conception de systèmes de règles expert ou de modèles statistiques) doit s'appuyer sur un ensemble de données réalistes. Nous déterminons alors les corpus proches de nos contextes applicatifs finaux et de notre public cible sur lesquels nous pourrions potentiellement appuyer nos études.

Pour conclure, nous résumons les challenges actuels dans l'Interaction Humain-Machine, et proposons nos pistes de recherche.

2.1. Systèmes d'Interaction Humain-Machine

A moins de viser l'anthropomorphisme, le but en Interaction Humain-Machine n'est pas de réaliser un système qui soit copie conforme de l'être humain. Tout d'abord, ce but n'est pas atteignable pour des raisons techniques (ressources technologiques – notamment en embarqué sur un robot –, connaissance de l'être humain imparfaite, représentation du monde fastidieuse, coût), mais également pour des raisons utilitaires et selon nous éthiques : l'être virtuel ne peut remplacer l'être humain. La chaleur de l'interaction, partager des expériences et un passé communs, tout cela n'est pas dans les attributions d'un agent virtuel. Celui-ci doit aider et accompagner l'utilisateur humain dans des situations de la vie quotidienne, et cela d'une façon aussi

agréable, intelligente, intuitive et ergonomique que possible. Ainsi, s'il n'est pas attendu de la machine qu'elle ressemble en tous points à l'humain, il est cependant impératif qu'elle partage certains codes de la communication interpersonnelle humaine.

Un système traitant de l'interaction émotionnelle entre un humain et un robot peut se décomposer classiquement comme suit :

Un **module de perception** chargé de collecter les informations provenant de l'utilisateur. Il peut s'agir des signaux relevés par les capteurs de pression, de son, de vidéo, voire même des entrées clavier. Ces informations peuvent être interprétées en termes de contenu langagier (un ordre donné au robot), d'émotion (l'utilisateur se sent mal), d'intensité (grande colère), d'intention (si par exemple toucher la tête du robot signifie « éteins-toi »), ou d'identité du locuteur (par la voix, le visage, par code, etc.),

Un **système émotionnel** en charge de traiter et coordonner ces différents éléments d'entrée, et de leur permettre d'impacter le système interne du robot. Par exemple, l'interprétation des différentes entrées, traitées sur le long terme grâce à une gestion d'un historique, peuvent permettre de mettre à jour l'état émotionnel courant du robot,

Un **module comportemental** contribue alors à la sélection du comportement à adopter. Cette sélection peut se fonder sur une myriade de facteurs différents, coordonnés par un système de prise de décision automatique. Par exemple, le système émotionnel peut transmettre des informations poussant le module comportemental à choisir de lever le bras pour saluer un utilisateur connu et amical. Mais ce module peut également recevoir des ordres inhibiteurs provenant d'autres sources, tels que la gestion de l'énergie de la machine : dans un tel cas, ce module aurait à traiter conjointement l'information « économiser l'énergie ». Dans un autre cas de figure, où le robot serait doté d'un état émotionnel, il pourrait être maussade et refuser de saluer l'interlocuteur humain,

Des **effecteurs** en charge de répercuter cette décision interne. Les actions à effectuer sont bien évidemment dépendantes des capacités motrices du robot, et codés selon le message interactionnel à transmettre. Dans notre exemple de la salutation à l'utilisateur, le robot pourrait alors hocher la tête, saluer par la voix ou émettre un son, faire un geste, ou encore juste éclairer ses yeux par un code couleur.

La gestion des perceptions réalisées par le système mène généralement à une interprétation en termes de profil utilisateur, et peut avoir un impact sur le système émotionnel et comportemental du système. Nous étudions dans cette section les différentes approches menées par la communauté en Interaction Humain-Machine, et la façon dont le système peut mémoriser les informations perçues et exprimer des comportements sociaux. Nous nous intéressons également aux différents modes d'expression du comportement du système.

2.1.1. Perception et interprétation de la communication non-verbale dans les Interaction Humain-Machine

De très nombreuses équipes de chercheurs se penchent sur la façon la plus appropriée de traiter les perceptions extérieures du système en Interaction Humain-Machine. L'étude de l'interaction du système avec son environnement peut se décliner en un grand nombre de sous-domaines : l'étude des moyens de percevoir le monde (les technologies liées aux capteurs, micros, caméras, etc.), des conditions de perception (environnement bruité, quantité d'informations à traiter), de la détection automatique de certains

phénomènes plus ou moins haut niveau (présence, source, parole, émotion...), l'interprétation de ces phénomènes à différents niveaux (contenu lexical, profil émotionnel, personnalité, intentions, identité, etc.), et l'évaluation de l'impact du système sur son environnement.

Dans notre étude, nous nous intéressons tout particulièrement aux informations non-verbales détectées par le système, et la façon dont celles-ci sont interprétées par le système, à plus ou moins haut niveau, et vont impacter éventuellement le comportement en sortie. Nous présentons donc un tour d'horizon des systèmes d'interaction, en opérant une progression selon la quantité d'informations extérieures traitées, ainsi que sur le niveau d'interprétation réalisé à partir de ces entrées.

Les équipes de l'université de Stanford et du MIT (Jung *et al.*, 2013) présentent une démarche exploratoire de la coopération entre un humain et un robot via l'expression de « *backchannel* »⁶ de la part du système. Ils proposent un protocole d'expérimentation comparative afin de déterminer l'impact du backchannel sur le travail d'équipe Humain-Robot, et évaluer la performance du robot. Un système de gestion de tâches, basé sur un automate à états finis, établit des priorités en fonction des ordres de l'utilisateur et de sa connaissance du contexte de la tâche. L'expression du *backchannel* est réalisée dès que le locuteur a parlé. La perception de l'environnement s'appuie sur une reconnaissance de la parole et une localisation de l'humain, afin de fournir un *backchannel* oral et gestuel aux moments appropriés, et en se tournant vers l'utilisateur. Les entrées récupérées par le robot dans ce contexte et leur interprétation ne sont pas la priorité. Nous voyons que cette étude d'Interaction Humain-Robot s'oriente principalement sur l'importance du comportement social robotique pour l'engagement de l'utilisateur et l'estimation de la performance du système.

(Stiefelhagen *et al.*, 2004) présentent une étude menée avec le robot ARMAR (Institut de Technologie de Karlsruhe), fondée sur deux perceptions : l'audio et la vidéo. Leurs études visent à doter le robot de systèmes de reconnaissance du dialogue spontané, de traitement du dialogue multimodal et de perception visuelle de l'utilisateur. Cette étude s'appuie donc sur l'application de modèles statistiques pour une interprétation de la parole naturelle de l'utilisateur et une perception visuelle des gestes de pointage de la part de l'utilisateur, et de l'orientation de la tête. La fusion des indices audio et vidéo permet une analyse multimodale du discours. Ces informations sont présentées comme pouvant potentiellement mener à une interprétation plus haut niveau, telle que l'analyse de l'intention de l'utilisateur dans une situation de dialogue avec le robot. La richesse de l'étude présentée consiste donc en la quantité d'informations perçues dans le monde extérieur, et d'une analyse à moyen niveau de ces informations.

Les études menées par (Castellano *et al.*, 2010b) présentent une interaction entre le robot iCat et des enfants dans le cadre d'interactions ludiques et éducatives. Un enfant joue aux échecs avec le robot. La perception réalisée automatiquement par le robot s'appuie sur une étude de l'état du jeu, et vient impacter le système interne du robot (composé d'une humeur et réactions affectives) afin de lui faire générer des

⁶ Mécanisme de communication utilisé par un locuteur en interaction humain-humain constitué de postures et de signaux non verbaux, destinés à faire comprendre à l'interlocuteur son propre degré d'attention et d'audition active.

comportements affectifs visibles sur son visage. Le système proposé ne se focalise donc pas sur une grande quantité d'informations extérieures perçues, mais sur une analyse extrêmement poussée de l'état de la tâche en cours. En effet, la réaction émotionnelle du robot s'appuie sur une attente vis-à-vis de l'action de l'enfant en jeu, et d'une réaction émotionnelle concordante afin d'aider l'enfant à mieux comprendre le jeu et de maximiser son engagement. Il s'agit donc d'une interprétation haut-niveau d'indices contextuels minimes.

(Ochs *et al.*, 2012) proposent un modèle d'agent de dialogue rationnel capable d'exprimer de l'empathie. Le système s'appuie sur une analyse du discours de l'utilisateur exprimé dans un contexte d'interaction dialogique avec l'agent. Les actes de communication et les émotions annotés manuellement sur ce discours permettent d'identifier les circonstances dans lesquelles une émotion peut se produire dans une situation d'Interaction Humain-Machine, à la lumière des théories sur les processus d'évaluation cognitive. Le système réalise des inférences sur les buts et les intentions de l'utilisateur via les actes de communication formulés, et décide de son type de réaction empathique en fonction d'une estimation de la réussite ou non de ces buts. Cette étude s'appuie sur l'interprétation approfondie des mécanismes de réflexion propres à l'énonciation des actes de communication. Le raisonnement du système ne s'appuie donc pas sur une grande quantité d'informations multimodales, mais sur une interprétation à haut niveau des composantes du dialogue transcrit afin de réaliser une prédiction de l'état émotionnel de l'utilisateur.

L'architecture EGO (Fujita *et al.*, 2001) présentée en 2001 par une équipe japonaise et exploitée dans de nombreuses études et applications ultérieures, propose un système capable de réaliser une association cognitive entre des symboles (par exemple, le nom d'un objet) et une expression émotionnelle. L'architecture s'appuie sur une perception visuelle et audio des événements extérieurs, et, grâce à un principe de mémoire associative, permet de faire fluctuer l'état émotionnel du robot, avant de déclencher des mécanismes de régulation homéostatique. Ce système très riche présente une interprétation des stimuli extérieurs (couleur, parole, présence) et son impact sur un système émotionnel du robot développé, via des mécanismes de compréhension haut niveau. Cette architecture a permis notamment la constitution d'un modèle mental pour le robot constitué d'un espace mental, d'une humeur, d'une équation sur les émotions, et d'une personnalité pour le robot (Itoh *et al.*, 2005).

Les travaux de (Lapeyre *et al.*, 2013) de l'équipe Flowers de Bordeaux présentent une nouvelle plateforme robotique bio-inspirée, implémentée sur le robot Poppy, permettant de gérer l'interaction entre un robot mobile bipède et un être humain. Des algorithmes permettront au robot d'adapter son déplacement en fonction du comportement et des préférences de l'utilisateur. Les travaux réalisés sur cette plateforme permettraient donc une interprétation haut niveau de données d'entrées potentiellement visuelles, auditives, et tactiles afin de représenter l'utilisateur.

Dans le cadre du projet EmotiRob visant à mettre au point un robot compagnon pour des enfants hospitalisés, (Duhaut, 2010) propose un modèle computationnel bio-inspiré qui traite un événement d'entrée (analysé depuis la parole, la prosodie ou la vision), et déclenche une modification dans l'état interne du robot, tant par une analyse directe de l'évènement (réaction physiologique), que par une analyse plus poussée influant sur le choix de comportement général du robot (interprétation cognitive). L'auteur propose également des pistes pour une compréhension à différents niveaux de l'énoncé exprimé par l'utilisateur, fournissant des informations sur la teinture émotionnelle du discours et sur les possibles références et

implications sous-entendues. Toutes ces informations visent à la constitution d'un système robotique doué de personnalité et réalisant une importante adaptation à l'utilisateur.

(Buendia et Devillers, 2012) proposent un schéma bio-inspiré pouvant interpréter différents signaux extérieurs (telles que des informations sur le visage, l'identité, sur la parole, les émotions) pour gérer une interaction sociale à long terme entre un utilisateur et un robot. Ce système s'appuie sur la préservation de la cohérence et de la continuité du dialogue suivant des théories en sociologie (par exemple la théorie des faces de Goffman), sur des processus d'évaluation cognitive des événements, et une représentation inspirée des neurosciences présentant un treillis d'hypothèses d'interprétation possibles. Les auteurs proposent ainsi une architecture permettant de concevoir un robot capable de comprendre le concept de mensonge, et sachant montrer de la compassion. Cette étude s'appuie donc sur une interprétation à très haut niveau des signaux sociaux d'entrée.

En conclusion, nous remarquons deux principales tendances de recherche dans la communauté Interaction Humain-Machine : les équipes se focalisant essentiellement sur la détection mono- et multimodale des stimuli extérieurs et leur interprétation en éléments informatifs à moyen niveau (acte de communication, mouvement de pointage, émotion, etc.), ou sur l'interprétation de ces informations à moyen niveau en une représentation haut-niveau (profil, intentions...). Notre étude se concentrera principalement sur l'interprétation haut niveau des signaux, ainsi que sur l'impact des comportements du système dans la boucle d'Interaction Humain-Robot.

2.1.2. Mémorisation et gestion de l'interaction à long-terme

Le concept de mémorisation en Interaction Humain-Machine est fondamental pour déterminer la façon dont l'interaction peut être gérée à long terme. Il s'agit en d'autres termes de spécifier les indices pertinents à stocker et la façon dont ils doivent être stockés par le système, mais également d'estimer leur utilité pour des réutilisations futures.

Le système de mémoire associative, tout d'abord, est un mécanisme humain visant à établir des liens, ou des connections, entre plusieurs éléments. Les études de (Itoh *et al.*, 2005) s'appuient sur l'hypothèse que si un humain mémorise une information en étant dans un certain état d'humeur, alors il retrouvera plus facilement cette information en étant à nouveau dans cette même humeur. L'association consiste donc ici à appairer une information et un état émotionnel. Les auteurs proposent alors un modèle de mémoire co-associative fondé sur l'utilisation de réseaux de neurones, et une implémentation appariant des concepts et des valences (tel que « Pomme » associé comme un souvenir plaisant pour la catégorie « Rouge »). Dans une étude menée avec le robot WE-4RII, une balle rouge déclenchait alors une expression émotionnelle de Joie chez le robot, car la couleur rouge était associée à une émotion plaisante.

Ce principe a également été appliqué dans le cadre de l'étude de (Sawada *et al.*, 2004), sur le robot QRIO. Une mémoire à court terme stocke les résultats obtenus par les modules de perception, et une mémoire à long terme gère l'association entre les informations perçues et l'état interne du robot. Par exemple, la gestion à long terme peut appairer un nom avec un objet identifié ou une voix, et modifier l'état interne associé à l'objet cible. Une composante du modèle permet également la gestion de l'homéostasie du système, afin que le robot revienne toujours à un état interne spécifique. Les auteurs présentent des

expériences au cours desquelles le robot, auquel on applique par exemple une forte pression sur le bras, associe le visage de l'utilisateur à la sensation négative de douleur.

(Hoffman et Breazeal, 2006) s'appuient également sur la mémoire associative pour présenter un modèle gérant une relation entre des actions et des concepts, en fonction de poids. Par exemple, le souvenir perceptif d'un chat en train de miauler peut être fortement corrélé au souvenir de l'action de tirer la queue du chat, et connecté de façon plus faible à d'autres souvenirs plus généraux (une odeur, un lieu).

Le principe de mémoire affective, proposé par (Thomaz *et al.*, 2005), est un modèle dynamique inscrit dans le cadre des études sur la référence sociale. La référence sociale consiste à se faire une idée de l'émotion à exprimer par rapport à un nouvel événement, en fonction des émotions que d'autres auront exprimées dans une telle situation. La mémoire affective s'attache donc à observer les réactions émotionnelles d'un référent, de les associer à un concept, et d'exprimer cette même émotion en retour lorsque le concept surgit. Ce système est ainsi présenté comme un mécanisme d'apprentissage de l'environnement réalisé par le robot, et de constitution automatique d'ensembles de souvenirs émotionnels pour le robot.

Une composante d'un système de mémorisation est également la capacité à oublier certaines informations, afin notamment de ne pas surcharger la mémoire et d'optimiser le processus cognitif du système. Le cadre du projet DEEP (Bossier *et al.*, 2007; Sterlin, 2007) vise à doter des Personnages Non Joueurs de capacité dialogique prenant en compte l'expérience, les émotions et la personnalité. Le processus de mémorisation développé évalue la pertinence d'un « ticket » (la représentation d'un stimulus) en fonction de différents niveaux, dont nous pouvons citer par exemple : si l'intensité de l'état émotionnel au moment du stockage du ticket est forte, si le thème exprimé dans le ticket est proche d'un thème pour lequel l'agent porte un intérêt, si le ticket est récent, si les informations portées par le ticket sont crédibles pour l'agent. Toutes ces informations permettent d'évaluer dynamiquement la pertinence de conserver en mémoire cette information. En fonction de la pertinence calculée, le processus d'oubli peut s'engager de deux façons : soit le souvenir est oublié de façon intégrale (suppression du ticket), soit il est opéré une dénaturation progressive de l'information portée par le ticket (via une fusion du souvenir, par exemple).

2.1.3. Comportements du système

Dans l'Interaction Humain-Robot, les comportements exprimés par l'interface (qu'il s'agisse d'un robot, ou des « *knobots* » tels que des agents conversationnels animés, etc.) sont étroitement liés à des caractéristiques de l'utilisateur. Ces caractéristiques peuvent être récupérées grâce à des capteurs (par exemple visuels, tactiles, audio...), puis interprétés (reconnaissance de la parole, de l'émotion) et intégrés à une représentation dynamique de l'utilisateur. L'étude des types de comportements potentiellement exprimés par le système est fondamentale dans l'élaboration d'un système d'Interaction Humain-Machine.

Dans la communication entre un humain et un robot, au-delà des attentes d'efficacité du système, – c'est-à-dire par exemple que le robot mène à bien la tâche qui lui a été confiée, et que son utilisation est intuitive et ergonomique –, un robot social se doit de partager les codes de communication interpersonnelle de l'être humain (Duhaut et Pesty, 2011; Feil-Seifer *et al.*, 2007), de manière à répondre de façon efficace aux messages de l'utilisateur. Ainsi, le robot sera plus à-même d'établir et de maintenir une relation naturelle et socialement acceptable. Exprimer de l'empathie représente par exemple un de ces moyens de communication naturels.

Un exemple de comportement : l'empathie

Nous trouvons différentes définitions de l'empathie dans la littérature, parmi lesquelles (Davidson *et al.*, 2003; Davis, 1983; Tapus et Matarić, 2007). La littérature présente également l'empathie sous la dénomination générale d'alignement émotionnel. Toutes ces théories s'accordent cependant généralement sur le fait que l'empathie peut être considérée comme la compréhension de l'état émotionnel de l'autre, et que l'expression de la réponse affective est similaire à l'état reconnu. En conséquence, implémenter l'empathie dans un système signifie que l'expression émotionnelle de l'utilisateur doit être détectée, que le contexte doit être compris, et que le système doit être capable d'exprimer une réponse émotionnelle compréhensible par l'utilisateur.

L'empathie est un mécanisme très utilisé dans la communauté (Damm *et al.*, 2011; De Vignemont et Singer, 2006; Thomaz *et al.*, 2005), car il maximise l'engagement de l'utilisateur humain, et aide à l'accomplissement de la tâche. Par exemple, un robot joueur apparaîtra plus amical aux yeux de l'utilisateur s'il présente un comportement empathique (Leite *et al.*, 2013) ; un robot destiné à accompagner des personnes âgées pour des exercices physiques sera considéré plus agréable et proche de l'utilisateur (Fasola et Matarić, 2012). L'empathie peut également être utilisée dans un contexte thérapeutique (Tapus et Matarić, 2008) afin de s'harmoniser avec les attentes sociales de l'utilisateur. Comme autre application de l'empathie, (Ochs *et al.*, 2012) proposent un agent rationnel de dialogue empathique, qui déduit le type et l'intensité d'une émotion potentiellement ressentie par l'utilisateur, d'après une compréhension des actes de communication exprimés. Cette interprétation s'appuie sur le modèle émotionnel OCC (Ortony *et al.*, 1988) et sur les théories de l'évaluation cognitive, et affecte un agent géré par un système de type BDI (« Belief », « Desire » et « Intentions » décrit en 2.2.2) qui va sélectionner l'expression empathique la plus appropriée. Le mécanisme social qu'est l'empathie est donc considéré comme un moyen puissant pour renforcer la relation entre le robot et l'utilisateur humain.

Cependant, comme (Tapus et Matarić, 2008) l'indiquent, exprimer de l'empathie n'est pas toujours le comportement le plus pertinent de la part du robot. En effet, le profil de l'utilisateur, sa personnalité et ses préférences sociales ont un impact sur la façon dont il appréhende l'empathie : cela peut lui sembler soit rassurant (dans un contexte thérapeutique par exemple), soit agaçant si l'utilisateur attend plutôt des comportements proactifs de la part du robot (dans un contexte d'urgence, par exemple, où le robot devrait prendre une décision).

Les attentes de l'utilisateur en matière de comportement robotique

Les attentes de l'utilisateur en ce qui concerne le comportement du robot varient donc en fonction de différents paramètres. Dans la littérature, nous repérons cinq grands axes permettant la définition d'un profil utilisateur conditionnant potentiellement les attentes en matière de comportement du robot : la tranche d'âge et le sexe, l'habitude des nouvelles technologies, la personnalité de l'utilisateur, le rôle du robot, et sa physiologie.

- L'âge et le sexe

Des études menées sur la propension à l'animisme chez l'enfant, et sur les différences garçon-fille indique que le système peut être perçu soit comme un simple appareil technique (notamment chez l'adolescent masculin), ou au contraire générer des attitudes maternelles (chez la petite fille). La perception

évoluerait avec l'âge, notamment sur la représentation des besoins et capacités présumés du robot. Nous pouvons citer les travaux menés par (Kanda *et al.*, 2004; Mahdhaoui, 2010; Okita et Schwartz, 2006).

- L'habitude des nouvelles technologies

Les habitudes de l'utilisateur en matière de nouvelles technologies peuvent conditionner la façon dont l'utilisateur interagit avec le système. En effet, en fonction du degré de nouveauté que ce système représente pour l'utilisateur, les réactions et les attentes peuvent différer (Granata *et al.*, 2010; Ray *et al.*, 2008).

- La personnalité

La personnalité de l'utilisateur joue un rôle essentiel dans l'estimation des attentes de l'utilisateur en termes de comportements robotiques. (Aly et Tapus, 2013), par exemple, s'appuie sur l'hypothèse qu'un utilisateur préférera un robot qui exprime des comportements propres à un type de personnalité proche de la leur.

- Le rôle du robot

Le rôle attendu de la part du robot conditionne fortement les attentes de l'utilisateur. L'étude et la conception de robots assistants notamment, les « *Social Assistive Robot* » (SAR), est un champ disciplinaire particulièrement actif. La robotique d'assistance vise un public en perte d'autonomie ou hospitalisé, mais également des personnes atteintes d'ASD (« *Autism Spectrum Disorders* »). Nous citons notamment les travaux de (Dautenhahn et Werry, 2002; Feil-Seifer et Mataric, 2008; Feil-Seifer *et al.*, 2007; Kidd *et al.*, 2006; Stiehl *et al.*, 2009; Tapus et Mataric, 2006; Wada *et al.*, 2005; Wainer *et al.*, 2006).

Le robot compagnon, ou robot joueur, est également un domaine d'étude de la communauté. Nous pouvons citer notamment les travaux sur l'iCat (Castellano *et al.*, 2010a) et les travaux sur l'adaptation des enfants en interaction avec un robot humanoïde (Nalin *et al.*, 2012).

- La physionomie

En dernier point, l'aspect visuel du robot conditionne également la représentation que l'utilisateur se fera du robot. L'Uncanny Valley est une théorie initiée par (Mori, 1970), expliquant que la ressemblance d'un système avec un humain n'évolue pas conjointement avec la familiarité que l'on éprouve pour ce système. Au contraire, la courbe mettant en rapport l'évaluation de l'humanité et la familiarité éprouvée présente une altération (une vallée de l'étrange) dans laquelle nous retrouvons tous ces systèmes qui nous mettent mal à l'aise, qui nous semblent étranges et dérangeants. Il s'agit d'une interprétation personnelle, une évaluation des systèmes qui tentent de mimer l'être humain mais ressemblent seulement à des « monstres ». La physionomie, – et donc le degré d'anthropomorphisme –, du robot implique également des attentes en termes d'interaction sociale et d'intelligence (Duffy, 2003; Goetz *et al.*, 2003; Pesty et Duhaut, 2012; Von der Pütten et Krämer, 2012).

Ces cinq paramètres nous permettent de constater qu'il existe une corrélation entre les caractéristiques de l'utilisateur et les comportements attendus de la part du robot.

L'expression d'émotions et d'intentions

Instaurer une interaction entre l'humain et le robot implique également que ce dernier dévoile, par certains moyens, son état interne courant. Par exemple, cet état peut être représenté par le degré de

satisfaction du robot, sur un modèle mental de type (Miwa *et al.*, 2003). Un robot doté d'une personnalité (tel que décrit dans l'étude de (Goetz et Kiesler, 2002)), ou de tout élément laissant supposer des mécanismes cognitifs ou émotionnels, affecte la représentation mentale de l'utilisateur vis-à-vis du robot, ainsi que son implication dans l'interaction. Même si le robot n'est pas doté de personnalité à proprement parler, la façon d'exprimer ses désirs et ses intentions, par exemple via l'expression d'émotions (Breazeal, 1999), peut amener l'utilisateur à répondre aux besoins du robot (tels qu'apporter de la distraction ou du confort) et ainsi augmenter l'efficacité de l'interaction. Un comportement conciliant ou désobéissant aux injonctions de l'utilisateur aura également un impact sur les expressions émotionnelles de ce dernier (Batliner *et al.*, 2004). Les mécanismes de référence sociale (Thomaz *et al.*, 2005) ou de synchronie interactionnelle (Chetouani, 2011) visent également à la production de comportements émotionnels susceptibles de stimuler l'engagement de l'utilisateur.

Les expressions d'un comportement robotique social ont été explorées également dans les travaux de (Buendia et Devillers, 2012), au travers de l'utilisation de techniques de dialogues avancées, dans le but de concevoir un robot comprenant le concept du mensonge et présentant de la compassion. Ces travaux s'appuient sur l'étude de la continuité dans le dialogue par la préservation de la « face » de l'interactant (Goffman, 1974), ainsi que sur les processus d'évaluation des événements émotionnels proposées par (Scherer *et al.*, 2001) et une représentation des événements sous forme de treillis d'hypothèses fondé sur un modèle bio-inspiré.

Il existe donc, dans les choix réalisés par la communauté, différentes façons de réaliser l'expression du comportement dans l'interaction sociale entre un utilisateur et le système.

Les différentes modalités d'expression

Au cours de leur étude sur les attentes des utilisateurs en termes de robotique, (Ray *et al.*, 2008) indique notamment l'importance de la sélection de modalité selon le public cible. En effet, selon certaines caractéristiques propres à l'utilisateur (parmi celles mentionnées plus haut), la compréhension du système pourra paraître plus ou moins intuitive. Par exemple, une personne n'ayant jamais utilisé d'ordinateur pourra avoir des difficultés à interagir avec une interface nécessitant l'utilisation d'une souris ou d'un clavier. Également, une personne déficiente visuelle sera plus touchée par les expressions lexicales et émotionnelles dans la voix du robot, que par une expression physique ; un enfant pourra être plus sensible à des postures.

Différentes modalités peuvent être ainsi envisageables pour l'expression de comportement d'un système, et nécessitent d'être évaluées selon le contexte applicatif et le public ciblé. Tout d'abord, l'expression d'un comportement s'appuyant sur le lexique dans le but de rendre le système plus naturel. C'est notamment le cas avec l'utilisation des *backchannels*, et des *conversational fillers*, qui permettent que le système ne reste pas figé lors de l'écoute de l'utilisateur (Bickmore et Cassell, 2005; Kanda *et al.*, 2007; Poppe *et al.*, 2010; Shiwa *et al.*, 2008). D'autre part, des postures peuvent induire une intention, ou l'expression d'une émotion (Beck, 2011; Beck *et al.*, 2010; Breazeal, 2002; Magnenat-Thalmann *et al.*, 2008). L'expression des comportements sociaux du système peut également être exprimée via la sélection de la distance à l'utilisateur (Pedicca et Vilhjalmsson, 2009), ou l'expression d'états mentaux par l'utilisation d'actes de conversation multimodaux (Rivière *et al.*, 2012). (Vinciarelli *et al.*, 2012) proposent une vue d'ensemble très intéressante de la synthèse de comportements sociaux dans l'Interaction Humain-Machine.

2.2. Théories sur les émotions

Comme nous l'avons mentionné précédemment, les informations paralinguistiques jouent un rôle prédominant dans l'interaction sociale. Elles peuvent venir appuyer le contenu linguistique dans un but emphatique, mais également le contredire, dans une tournure humoristique ou ironique. Instinctivement comprises par l'être humain (hormis cas pathologiques), elles peuvent présenter des manifestations complexes, telles que les émotions mélangées, sur lesquelles l'accord interpersonnel est délicat (Devillers *et al.*, 2005a). Nous présentons dans cette section les définitions du paralinguistique, et des émotions telles qu'utilisées couramment par la communauté. Nous décrivons également les modélisations de l'interaction émotionnelle Humain-Humain et Humain-Machine.

2.2.1. La paralinguistique

Le paralangage, terme introduit par (Trager, 1958), regroupe tous les phénomènes de communication employés seuls ou pour compléter un discours. Par exemple, les gestes, les mimiques, les expressions non verbales, les fluctuations vocales, sont tous autant de moyens pour transmettre une intention communicative, un désir, ou l'expression d'un état interne (état de santé, humeur, émotion).

Les expressions paralinguistiques sont exprimées soit par la voix, ou par la gestuelle, la posture, l'expression faciale. (Schuller *et al.*, 2013) se penchent principalement sur le paralangage dans la voix, et notent la distinction entre la paralinguistique incrustée dans le message linguistique (une toux, un éclat de rire), et les manifestations superposées au message, telles qu'un pitch élevé pour dénoter l'anxiété.

L'étude du paralangage est un champ disciplinaire particulièrement actif, qu'il s'agisse d'études purement linguistiques ou d'études sur la reconnaissance automatique de ces manifestations en traitement des langues. Nous citerons par exemple les travaux de (Campbell, 2007) sur les rires, (Devillers *et al.*, 2005a) sur les émotions dans la prosodie, (Pal *et al.*, 2006) pour l'analyse des pleurs, ou encore (Aarsand et Aronsson, 2009) dans leur étude sur les exclamations involontaires (les *response cries*).

Notre étude se concentre sur les états émotionnels.

2.2.2. Les émotions : définitions

La difficulté de la description formelle d'une émotion est bien connue de la communauté. Comme (Scherer, 2005b) le mentionne, il convient de déterminer tout d'abord avec précision ce que l'on entend par « émotion ». S'agit-il du processus affectif menant à l'expression d'une émotion ? D'un état à un moment donné ? De traits impliquant plutôt l'expression d'émotions sur la durée ? Différentes approches permettent de catégoriser les états émotionnels. Nous présentons une distinction en termes d'expressions émotionnelles ponctuelles, et les états émotionnels à plus long terme. Nous développons également sur les principaux modèles émotionnels.

L'émotion

Deux approches principales permettent de décrire l'émotion en tant que phénomène ponctuel : l'approche catégorielle, et l'approche dimensionnelle.

L'approche catégorielle consiste à définir les émotions comme un ensemble d'étiquettes représentant l'émotion à un instant donné. Selon les auteurs, plusieurs catégorisations sont possibles pour les émotions, citons notamment :

- Les émotions primaires

La catégorisation proposée par Ekman (Ekman, 1992a, 1992b) s'appuie sur les émotions Colère, Joie, Tristesse, Peur, Dégoût et Surprise, considérées comme étant primaires et universelles.

- Le circomplexe des émotions

La catégorisation par similarité de (Plutchik, 1980) représentée en un circomplexe (Figure 1) définissant les émotions en thèmes (Colère, Joie, Dégoût, Tristesse, Surprise, Peur, Anticipation, Confiance).

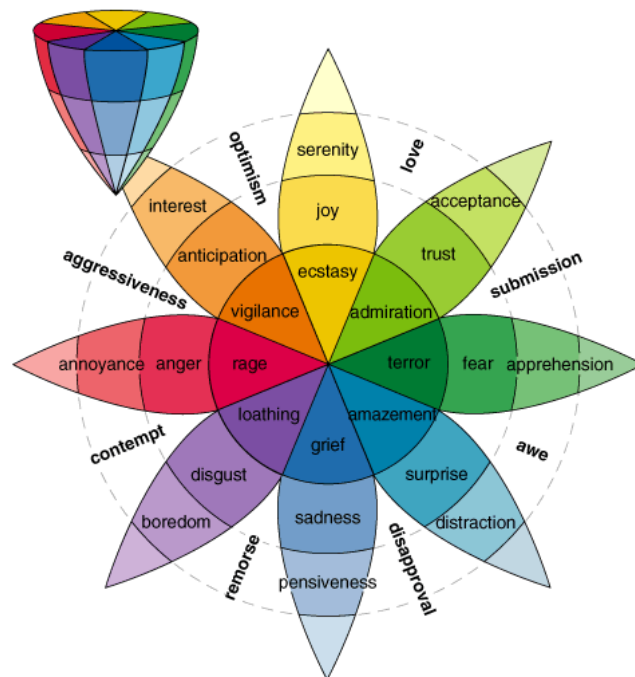


Figure 1. Roue des émotions de (Plutchik, 1980), composée d'émotions prototypiques desquelles sont dérivés les états mixtes.

- Les émotions morales

Les émotions morales décrivent les affects considérés comme reflétant une prise de position à caractère social. Citons par exemple le Dégoût, l'Embarras, le Mépris, ou la Fierté (Davidson *et al.*, 2003) (voir Figure 2), ou considérés du point de vue de leur caractère réactif à un événement (Charaudeau, 2000).

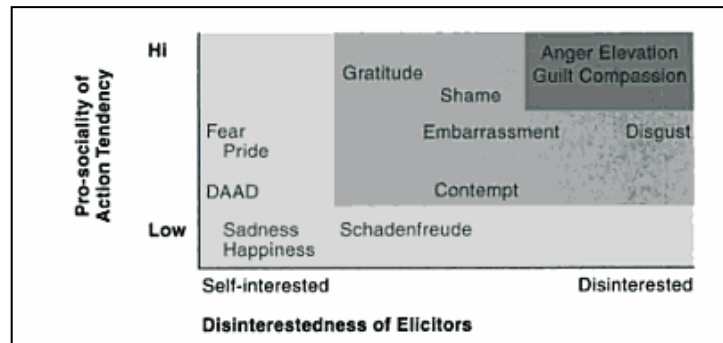


Figure 2. Les émotions morales selon (Davidson *et al.*, 2003), réparties sur deux critères : l'intérêt de l'individu, et le caractère prosocial.

- Les émotions complexes

Les émotions complexes, ou émotions mélangées (Devillers, 2006; Douglas-Cowie *et al.*, 2005; Ekman, 1992a; Sander et Scherer, 2009) sont des émotions non-prototypiques que l'on observe fréquemment en conditions réelles. Il s'agit d'émotions primaires, mixtes ou morales se produisant soit simultanément, soit séquentiellement, afin de représenter une émotion combinée.

L'approche dimensionnelle décompose les états émotionnels en plusieurs facteurs interconnectés. Notamment, les travaux de (Russell et Mehrabian, 1977) proposent une description de l'émotion en termes d'Intensité, de Plaisir, et de Dominance (Pleasure, Arousal, Dominance). Faire varier ces trois valeurs permettrait de recouvrir tout l'ensemble des émotions verbales exprimées.

Les « émotions » sur le long-terme

L'implication des manifestations émotionnelles (qu'il s'agisse de l'étude de l'émotion exprimée ou de l'émotion ressentie) dans les phénomènes à plus long terme a été largement étudiée par la communauté. Nous nous penchons par exemple sur les travaux de (Scherer et Bänziger, 2004), qui s'appuient sur une catégorisation des manifestations émotionnelles (dans leur sens le plus vaste) selon différents critères : leur intensité, leur durée, leur synchronisation, si elles sont dépendantes d'un évènement, leur relation au mécanisme d'évaluation cognitive (Scherer et Peper, 2001), leur rapidité d'évolution, et leur impact sur le comportement. Nous retrouvons en Figure 3 une représentation de la durée des manifestations émotionnelles présentées dans les travaux de (Scherer et Bänziger, 2004).

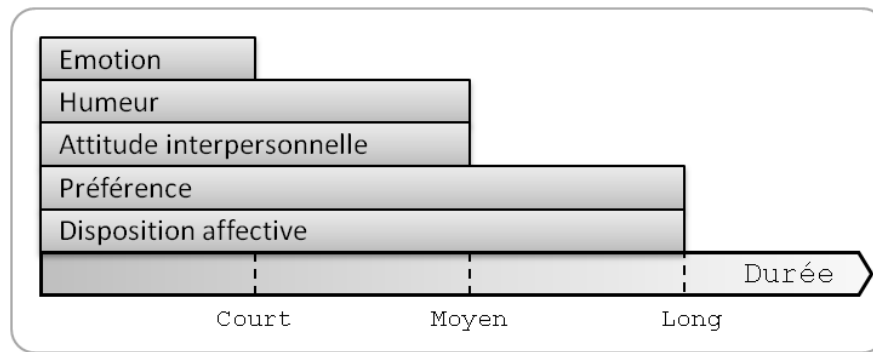


Figure 3. Différence de durée des différents types d'états affectifs. D'après (Scherer et Bänziger, 2004).

Les manifestations émotionnelles sur le plus long terme ont également été évoquées sous d'autres termes et classifications. Nous citerons notamment :

- Les traits émotionnels

Les traits émotionnels (Revelle et Scherer, 2009) sont caractérisés par la propension d'un individu à éprouver certaines émotions plus fréquemment que d'autres.

- Le tempérament

(Charaudeau, 2000) décrit le tempérament comme les tendances ou les inclinations d'un individu à présenter des comportements émotionnels de façon récurrente. Ces tempéraments sont représentés par des adjectifs, parmi lesquels : Colérique, Irascible, Amoureux, Peureux, Haineux.

- La personnalité

L'étude de la personnalité est un champ disciplinaire très actif. Citons principalement les travaux réalisés sur les cinq grands facteurs de la personnalité (le *Five-Factor Model*) exploité dans les modèles des Big Five et OCEAN (Hofstee *et al.*, 1992; McCrae *et al.*, 1986). La personnalité est définie comme un ensemble d'affects, de comportements, de représentations cognitives et de désirs susceptibles de définir un individu.

La modélisation émotionnelle

La connaissance du contexte d'émergence, enfin, permet d'appréhender encore mieux l'émotion exprimée par l'individu. En d'autres termes, il s'agit de déterminer quel type d'évènement a déclenché l'émotion. Comme représenté dans le modèle émotionnel OCC (Elliott, 1992; Ortony *et al.*, 1988), il est possible d'évaluer les conséquences d'un évènement sur les buts de l'agent, sur les actions qu'il réalise, ou sur les objets de l'environnement, et d'anticiper ainsi les expressions émotionnelles potentielles de l'agent. Cette connaissance du contexte d'émergence peut permettre de faciliter la compréhension du message émotionnel.

L'évaluation cognitive décrite par (Lazarus *et al.*, 1980), puis par (Scherer, 1989), met en relation notre perception et notre évaluation d'un évènement comme étant la cause d'une série de mécanismes cognitifs visant à réagir émotionnellement à cet évènement. (Scherer, 2005a) propose une évaluation séquentielle de la pertinence de l'évènement vis-à-vis des buts du locuteur, suivie par une analyse des implications de cet évènement, puis une vérification des capacités de faire face à cet évènement, en enfin une analyse de la compatibilité de l'évènement avec ses propres standards (moralité et buts). (Gratch et Marsella, 2004) ont développé un modèle informatique fondé sur les théories de l'évaluation cognitive.

En intelligence artificielle, le modèle cognitif dit « BDI » (pour « Beliefs », « Desires » et « Intentions », soit « croyances, désirs et intentions ») représente un cadre formel pour évaluer une réaction rationnelle due à une perception de l'environnement, en fonction des états mentaux du locuteur. La Théorie de l'Interaction Rationnelle proposée par (Sadek, 1991) présente trois attitudes mentales primitives proches du modèle BDI : la croyance, le choix et l'incertitude. Le modèle de (Ochs *et al.*, 2006) met en œuvre un agent rationnel qui s'appuie sur cette théorie.

2.2.3. Les marqueurs affectifs

Les marqueurs affectifs, nommés « affect bursts » en anglais, sont des phénomènes de discours venant interrompre la continuité de la parole, et se présentant sous la forme d'expressions émotionnelles courtes et non-verbales. Nous retrouvons par exemple les rires, les hésitations (« euh »), les toussotements (lorsqu'ils sont porteurs de sens), les pleurs. Ces manifestations peuvent donc se présenter sous la forme de sons ou d'interjections non verbales.

D'après (Scherer, 1994), et tel que reporté par (Schröder, 2003), les *affect bursts* peuvent être catégorisés en termes d'effet *push-pull*. Les manifestations conséquentes à un effet de *push* sont des réponses à des événements physiologiques. Elles portent donc un caractère plutôt universel, et présentent de fortes différences interpersonnelles. Dans le cas des réponses à un effet de *pull*, il s'agit d'expressions dont la symbolique est propre à une culture donnée, et pour lesquelles les différences interpersonnelles seront moindres dans les cultures données.

Suite à une étude sur la reconnaissance des *affect bursts* actés hors contexte en allemand, (Schröder, 2003) propose une classification en fonction du message émotionnel transmis. L'*affect burst* pourra donc relever d'une expression d'Admiration, de Menace, de Peur, d'Agacement, de Mépris, de Colère, de Dégout, d'Exaltation, d'Ennui ou de Soulagement selon l'étude de Schröder.

2.2.4. Les émotions dans l'interaction

Les états émotionnels ressentis et exprimés ont un impact dans l'interaction entre humains, et permettent à chaque interactant de se positionner par rapport à son interlocuteur. Les travaux de (Sacks, 1992) en Analyse Conversationnelle ont démontré que l'expression affective réalisée au cours d'une interaction peut être évaluée comme un niveau d'engagement, et que leur discours s'adapte de façon prospective et rétrospective afin de constituer un ensemble d'actions « faisant conversation ». Selon (Goffman, 1983) et (Brown et Levinson, 1987), les choix du locuteur sont contrôlés par des variables sociales et affectives. Chaque interactant décide alors soit de faire converger le système vers les dispositions de l'autre, soit de le faire diverger. Nous présentons dans cette section des travaux menés dans la communauté sur l'analyse des expressions émotionnelles au sein de la boucle d'interaction.

Les travaux de (Mohammad et Nishida, 2010) proposent deux mécanismes de réaction émotionnelle dans l'interaction. Le premier mécanisme est causé par les changements de l'environnement provoqués par un agent, qui vient modifier la façon dont l'autre agent perçoit la situation, et donc son état interne. Le second mécanisme met en jeu la prise en compte directe du comportement de l'agent pour la réévaluation de l'état de l'autre agent, indépendamment de son impact sur l'environnement.

L'implication affective (Vaudable *et al.*, 2010) est un processus visant à harmoniser l'interaction entre les interlocuteurs. Elle implique qu'un acte de communication génère des émotions chez chaque interlocuteur. L'implication affective se décompose en deux dimensions : l'Induction et la Rétroaction. L'Induction consiste à provoquer une réaction émotionnelle chez l'interlocuteur, tandis que la Rétroaction est la réaction à un événement produit dans les tours de parole précédents. L'implication affective mesure également le niveau d'intérêt exprimé par les locuteurs, allant de l'intérêt simple (machinal), en passant par la sympathie (incluant la politesse), jusqu'à la surenchère émotionnelle.

La synchronie interactionnelle (ou synchronie interpersonnelle) est un processus dynamique qui impacte chez les locuteurs leur gestuelle, leur façon de s'exprimer, leurs émotions exprimées, ou encore leur regard. L'analyse de la synchronie consiste à identifier l'influence que chaque interactant exerce sur l'autre, et la façon dont ceux-ci adaptent leurs comportements. Se synchroniser avec l'interlocuteur nécessite d'avoir recours à des mécanismes tels que l'attention conjointe, l'empathie, ou encore l'engagement (Chetouani, 2011; Morency, 2010).

En situation d'interaction humain-humain, reconnaître chez l'autre une émotion implique avoir la connaissance de différents paramètres. Des paramètres propres à la perception de l'expression de l'émotion, tout d'abord, comme par exemple les caractéristiques acoustiques ou issues de la transcription orale permettant une prédiction de l'état émotionnel (Arroyo, 2009; De Rosis *et al.*, 2006; Ochs *et al.*, 2006), ou encore une compréhension des signaux émotionnels transmis par le langage du corps et du visage (citons notamment (Ekman, 1982)). L'analyse du support langagier peut également venir en renfort de la compréhension du message émotionnel (Cosnier, 2003; Ipgrave, 2009), et plusieurs études utilisent le canal linguistique et paralinguistique (Batliner, 2006; Devillers et Vasilescu, 2004; Devillers et Vidrascu, 2006; Vaudable et Devillers, 2012).

Analyser correctement l'émotion perçue chez un autre humain est de même facilité lorsque nous connaissons certaines qualités de l'énonciateur. Par exemple, la relation sociale que l'on entretient avec lui peut tempérer l'expression de ses émotions ; le statut social, des expériences partagées ou encore l'affinité, exprimés par la voix, la posture, les gestes et le contenu du discours sont autant d'éléments qui contribuent à moduler les émotions transmises lors de l'acte de communication (Baron-Cohen, 2004; Collins et Miller, 1994; Isbister, 2006; Mazur, 1985). La personnalité présumée de l'énonciateur, ses « traits émotionnels » connus, permettent également de renforcer le diagnostic de l'émotion exprimée (Markel *et al.*, 1972; McCrae *et al.*, 1986; McCrae et John, 1992; Revelle et Scherer, 2009) : par exemple, nous serons moins interpellés par l'expression de colère d'une personne de notre entourage s'exprimant habituellement de façon exubérante et très emphatique, que dans le cas d'une personne que nous voyons usuellement calme et introvertie. Connaître la personne, avoir en mémoire ses tendances émotionnelles, est donc un élément important dans la compréhension du message émotionnel.

2.3. Évaluation des systèmes d'Interaction Humain-Machine

Dans le processus de conception d'un système d'Interaction Humain-Machine, l'évaluation est une étape primordiale pour valider le bon fonctionnement du système, et son adéquation avec la tâche finale visée. Nous nous penchons sur ce que l'on cherche à évaluer, plus précisément : cherchons-nous à évaluer la réussite du système ? La formalisation des caractéristiques utilisateurs ? Nous nous interrogeons alors sur les

différentes méthodes d'évaluation les plus pertinentes : une évaluation subjective et qualitative (via l'utilisation de questionnaires par exemple), ou bien une évaluation objective et quantitative (par analyse statistique). Pour résumer, nous cherchons des réponses à : que faut-il évaluer ? Et comment l'évaluer ?

Dans cette section, nous présentons différentes méthodes d'évaluation proposées par la communauté, tout d'abord dans l'évaluation des performances d'un système d'interaction – qu'il s'agisse d'interfaces utilisateurs ou plus précisément de robots sociaux –, puis sur l'évaluation du comportement émotionnel (au sens large) et attitudinal de l'utilisateur.

2.3.1. Évaluation des performances du système

Dans ses travaux de thèse, (Balbo, 1994) propose une méthode d'évaluation ergonomique et automatique des interfaces utilisateur, qui s'appuie sur la constatation que « l'évaluation est souvent négligée par les développeurs ou reste le produit d'une démarche artisanale ». Le système d'analyse automatique proposé, baptisé EMA, agit comme un détecteur d'anomalies en se penchant sur trois sources d'informations susceptibles de représenter l'adéquation d'un système au but recherché : des normes de comportement du système (des règles de profils de comportement), une évaluation de la tâche (une représentation formelle de l'espace des tâches réalisables par le système), et le comportement effectif du système (des données comportementales enregistrées au cours des sessions). Dans cette évaluation, chaque tâche est subdivisée en sous-tâches, toutes liées à des sous-buts et des procédures. La réussite est évaluée pour chacune de ces sous-parties.

Dans le domaine des systèmes de dialogue, le cadre théorique d'évaluation PARADISE (Walker, 1997) se propose de fournir, en se basant sur la théorie de la décision, un cadre d'évaluation des systèmes de dialogue en termes de performances nécessaires à la réalisation de l'objectif (à savoir accroître la satisfaction de l'utilisateur). Un traitement statistique sur ces unités minimales permet, à l'aide de mesures qualitatives et d'efficacité, de minimiser le coût pour atteindre l'objectif et d'accroître les chances de réussite de la tâche. L'évaluation d'un système selon le cadre théorique PARADISE requiert donc de spécifier les objectifs et les unités minimales constituant l'objectif à atteindre (nommées les « *performance measures* »). Ce paradigme a été utilisé pour comparer des systèmes de dialogue, par exemple avec différentes stratégies de génération (explicite, implicite) (Devillers *et al.*, 2002).

En Interaction Humain-Robot, (Dautenhahn et Werry, 2002) proposent des métriques pour évaluer l'interaction, en décrivant et analysant quantitativement des critères de comportement bas niveau (des « micro-comportements »). Cette approche a été proposée dans un contexte d'interaction où l'utilisateur évoluait de façon spontanée (non contrainte et non structurée) autour du robot, sans que le robot ait de tâche spécifique à effectuer, et sans la possibilité de réaliser un questionnaire de satisfaction auprès de l'utilisateur. Dans ces circonstances, l'évaluation méthodique de l'interaction est donc particulièrement ardue. Les micro-comportements sont présentés comme des unités minimales de comportement, facilement identifiables de façon automatique ou experte (position de la tête, regard, gestuelle, etc). En se focalisant sur ces sous-éléments de l'interaction, les chercheurs ont pu se livrer à une analyse comparative des comportements d'interaction dans différents contextes d'application.

Encore dans le cadre des métriques pour l'évaluation des Interaction Humain-Robot, (Steinfeld, 2006) suggèrent des ensembles de métriques exploitables dans une vaste majorité d'applications Humain-Robot,

permettant notamment d'évaluer l'impact de l'interaction sur la performance du système. Ils s'appuient sur l'étude d'un robot mobile dédié à des tâches d'orientation et de navigation. Pour évaluer l'application, il convient de déterminer les différents paramètres de réalisation de la tâche. Dans le cas de la navigation par exemple, les paramètres consistent en un pourcentage de sous-tâches de navigation réussites, de couverture de terrain, de déviation par rapport à la route prévue, de taux de réussite dans la gestion des obstacles. D'autres paramètres, traitant de la perception de l'environnement, peuvent inclure des mesures de la mobilité du robot, de la réussite de la détection et de la reconnaissance de l'environnement. La gestion de l'environnement, ensuite, peut être un autre ensemble de paramètres permettant d'évaluer la performance du système : le temps de réaction, l'autonomie. Les auteurs proposent également une évaluation de « l'efficacité sociale » du système, en termes de style d'interaction, du pouvoir de persuasion du robot (a-t-il réussi à influencer sur le comportement de l'humain ?), de confiance que l'humain porte au robot, d'engagement de l'utilisateur, et enfin de conformité du comportement social du système (apparence, cohésion avec les normes, etc.). Tous ces ensembles de mesures permettent d'estimer la performance du système, par des mesures quantitatives (efficacité), subjectives (qualité), ainsi que par une analyse de la coopération Humain-Robot (pourcentages de demandes d'assistance réalisées par le robot ou l'utilisateur). La performance du robot proprement dit peut également être évaluée via une estimation de ce que nous pourrions appeler la « conscience de soi » du robot (avoir conscience d'être en situation problématique, de savoir ou non répondre à un problème), de la conscience de l'humain (perception de l'humain et réaction idoine en conséquence), et du degré d'autonomie du robot (par exemple, par adaptation automatique (Benureau et Oudeyer, 2013)).

Dans un contexte applicatif plus spécifique, (Feil-Seifer *et al.*, 2007) proposent des standards pour l'évaluation de robots sociaux dédiés à l'assistance. L'évaluation peut être réalisée à partir d'études issues de la psychologie, de l'anthropologie, de la médecine, et de l'Interaction Humain-Robot, et vise à jauger 1) l'aspect technologique (sécurité, extensibilité), 2) l'interaction sociale (autonomie, imitation, intimité, compréhension du domaine applicatif, réussite sociale), 3) la technologie d'assistance (impact sur les soins portés à l'utilisateur, sur sa vie, sur le personnel soignant). Ces trois caractéristiques sont jugées déterminantes par les auteurs pour évaluer les performances du système en termes d'assistance et d'interaction sociale. Les auteurs estiment que le but à atteindre pour un robot social d'assistance consiste en trois dimensions : la sécurité, l'efficacité, et l'impact positif du robot sur l'utilisateur.

2.3.2. Évaluation psychologique et émotionnelle de l'utilisateur

Le psychométricien et psychologue de l'éducation Robert L. Thorndike propose en 1966 (Thorndike, 1966) une discussion sur la mesure de la croissance intellectuelle de jeunes sujets. Selon lui, une bonne évaluation d'un aspect psychologique est dépendante de trois critères : a) la fiabilité des mesures (éviter les erreurs de mesure), b) l'utilisation de bonnes unités de mesure, et c) l'équivalence entre les différentes mesures réalisées par différents instruments et à différents niveaux (dans le cas d'une étude comparative). Un test d'évaluation psychologique devrait donc reposer sur des critères de validité et de fiabilité.

En Interaction Humain-Machine, l'évaluation de la personnalité de l'utilisateur peut être réalisée de façon subjective, par l'utilisation de questionnaires notamment, ou objectivement par une analyse statistique.

Dans le cas de l'analyse subjective de la psychologie de l'individu, les professionnels de la psychologie peuvent mener une série d'évaluations psychologiques afin de déterminer, par exemple, l'aptitude d'un individu à l'adoption, ou sa dépendance psychologique aux produits stupéfiants, voire estimer une suspicion de maladies liées à la sénilité. L'évaluation réalisée dans ces contextes est basée notamment sur des entretiens avec le sujet, avec ses proches (propriétaires, employeur, etc.), ou encore, selon les cas, des tests de personnalité. Des questionnaires d'auto-évaluation sur la personnalité, validés par la communauté de chercheurs, peuvent être proposés au sujet. Le questionnaire Eysenck Personality Questionnaire (EPQ) (Eysenck et Eysenck, 1975) évalue l'individu en termes d'extraversion, de neuroticisme et de psychotisme (ou respectivement d'introversion, de stabilité et de socialisation). Le questionnaire Big Five Inventory (BFI) (John et Srivastava, 1999) évalue l'individu à l'aide de quarante-quatre assertions pour lesquelles l'individu choisit un degré d'accord (sur une échelle de Likert à cinq points). Les résultats permettent de définir la personnalité de l'individu en termes de niveaux d'extraversion, d'agréabilité, d'application, et de neuroticisme et d'ouverture à l'expérience. L'outil d'évaluation Myers-Briggs Type Indicator (Briggs, 1976), qui selon sa forme peut être soit proposé en auto-évaluation à l'individu, soit en estimation par un tiers, permet de déterminer quatre dimensions du sujet : l'orientation de l'énergie (introversion versus extraversion), la prise d'information (sensation versus intuition), la prise de décision (pensée versus sentiment), et le mode d'action (jugement versus perception). Cette évaluation permet de représenter seize grands types de personnalité.

Tous ces tests sont considérés de façon générale comme des évaluations valables de la personnalité de l'individu. Cependant, en Interaction Humain-Machine, l'estimation du profil de l'utilisateur en temps réel, et donc de façon autonome et dynamique, est l'un des verrous du domaine. Dans ce contexte, l'utilisation de questionnaires est alors difficilement envisageable. Pour lever ce verrou, la communauté en Interaction Humain-Machine se livre à des estimations automatiques de la personnalité en se basant sur des méthodes statistiques. La détection automatique de la personnalité est, par exemple, l'objet du challenge « INTERSPEECH 2012 Speaker Trait Challenge » (Schuller *et al.*, 2012). Le corpus d'étude, « Speaker Personality Corpus », propose 330 locuteurs dont le profil de personnalité a été réalisé par onze juges à l'aide notamment du questionnaire BFI-10. Ce Challenge a mené à toute une série de travaux sur la reconnaissance automatique du profil de personnalité de l'individu à partir d'indices acoustiques, comme par exemple les travaux de (Mohammadi et Vinciarelli, 2012). Ce principe d'estimation automatique de la personnalité est fiable du point de vue des méthodes d'évaluation. Tout d'abord, l'annotation des corpus met en œuvre la perception subjective de juges, dont le nombre important permet de rendre l'accord à la majorité valide. Ensuite, les méthodes de détection automatique s'appuient sur des scores de détection (en rapport à la perception subjective des juges) permettant une évaluation fiable de la bonne performance du système.

Il existe peu de moyens fiables pour déterminer de façon automatique l'émotion ressentie (par opposition à l'émotion exprimée) par l'humain. Des indices physiologiques tels que la sudation, le rythme cardiaque, sont des mesures objectives permettant de fournir un panel d'informations, malheureusement difficilement interprétables. Une auto-évaluation par l'humain fournit des informations plus fiables quant à l'émotion ressentie, mais elle comporte le biais 1) de la subjectivité du sujet par rapport à sa propre émotion, 2) que l'émotion décrite n'est qu'une réminiscence.

L'estimation de l'état émotionnel exprimé par l'utilisateur par un système d'Interaction Humain-Machine nécessite une approche subjective de caractérisation de l'individu similaire aux méthodes de

détermination de la personnalité. Elle nécessite une base de données annotées émotionnellement par différents juges, desquelles des paramètres acoustiques sont extraits et analysés à l'aide d'outil statistiques. Un système automatique de classification permet alors d'obtenir des scores de classification pour toute nouvelle occurrence de parole en termes d'émotion exprimée (Clavel, 2007; Devillers *et al.*, 2005a; Tahon, 2012; Vidrascu, 2007).

2.3.3. Interaction Humain-Robot émotionnelle : que faut-il évaluer, et comment ?

D'après les études que nous avons répertoriées précédemment, nous constatons que le but d'un robot social est de présenter des comportements socialement cohérents (c'est-à-dire compréhensibles par l'utilisateur, et représentatifs des codes de communication humaine), et de procurer de la satisfaction à l'utilisateur.

Pour l'évaluation de ces deux aspects, tout d'abord, nous notons l'importance de décliner l'interaction en une quantité de sous-modules plus faciles à identifier, analyser, et comparer. La définition des métriques à évaluer, ensuite, consisterait dans notre cas à la reconnaissance des émotions exprimées, du profil psychologique de l'utilisateur, et des comportements sociaux du robot.

Il convient également d'estimer la performance du robot. Dans notre cas, l'objectif est de satisfaire socialement l'utilisateur, par le biais de deux tâches – compagnon de jeu, et assistant. En reprenant (Feil-Seifer *et al.*, 2007), nous estimons que la sécurité, l'efficacité et l'impact positif des comportements du robot peuvent être évalués via les réactions émotionnelles exprimées par l'utilisateur, car la réalisation de ces trois paramètres entraînerait un état positif chez le locuteur.

2.4. Corpus d'interaction existants

La conception de modèles réalistes nécessite de s'appuyer sur des collectes de données en situation concrète d'application. Selon (Luzzati, 2007), par exemple, un corpus de dialogue oral spontané pour l'Interaction Humain-Machine doit refléter la vitalité de l'interaction. Pour être réellement riche et vivant, le corpus doit également présenter plusieurs types de représentations (d'annotations), sur tous les supports enregistrés (audio, vidéo).

Ces corpus permettent de constituer des modèles d'apprentissage, et de définir des règles « expert » sur des phénomènes spécifiques. L'interaction entre un utilisateur humain et une machine peut s'appuyer sur les codes de communication interpersonnelle, et donc nécessiter l'étude des relations Humain-Humain. Il est également essentiel d'étudier les interactions entre l'utilisateur et le système final (robot, interface, etc.), et à cette fin d'analyser les comportements d'utilisateur en situation réelle.

Il existe dans la communauté de nombreux corpus d'Interaction Humain-Humain, impliquant soit des interactions deux à deux ou en groupe. Ces corpus sont souvent la base d'études pour l'Interaction Humain-Système, qu'il s'agisse de systèmes de dialogue ou d'outils de détection automatique des émotions. Nous présentons quelques uns de ces corpus.

Nous proposons ensuite une sélection de corpus d'Interaction Humain-Robot constitués par la communauté. Chacun de ces corpus est dédié à une tâche spécifique, en interaction dyadique d'un utilisateur avec le robot, ou en interaction plus complexe du robot avec un groupe d'utilisateurs. Nous étudions la composition et la disponibilité de chacun de ces corpus.

2.4.1. Interaction Humain-Humain

Dans la catégorie des corpus d'interaction de groupe, nous pouvons citer par exemple le corpus ISL (Burger *et al.*, 2002) permettant l'étude des différences de styles de parole selon le type d'entrevue, par la mise en scène de scénarios réels ou scriptés de jeu, de discussion, de débat. Le corpus AMI (McCowan *et al.*, 2005) présente des interactions naturelles ou élicitées dans le cadre de présentations suivies de discussions. Les données sont annotées notamment en fonction de la domination exercée par les participants. Le corpus ELEA (Sanchez-Cortes *et al.*, 2011) met en scène des interactions traitant de la survie en conditions difficiles. Les annotations réalisées concernent la performance, la personnalité (sur le modèle OCEAN), et des données interpersonnelles telles que la domination, l'affinité, la compétence, et les capacités à diriger.

Dans l'étude des interactions un-à-un, nous pouvons citer par exemple le corpus CReST (Eberhard *et al.*, 2010), qui met en scène deux humains réalisant des tâches de coopération à distance. Ce corpus est composé de données audio annotées et transcrites, pour lesquelles sont repérés des indices dialogiques et syntaxiques.

Le corpus SEMAINE (McKeown *et al.*, 2010), disponible publiquement, propose des scénarios au cours desquels un opérateur et un participant, assis dans des pièces séparées, dialoguent via l'audio et la vidéo. L'opérateur adopte des comportements particuliers susceptibles de générer des réactions émotionnelles chez le participant. Cette étude s'appuie notamment sur les travaux réalisés dans le cadre du corpus SAL (Douglas-Cowie *et al.*, 2008), pour « Sensitive Artificial Listener », dans lequel des participants ont une conversation avec un expérimentateur qui prétend être un agent conversationnel automatisé. Le principe de l'expérimentation SAL vise à pouvoir mettre au point des systèmes susceptibles de reproduire de façon crédible cette action d'écoute dans l'interaction avec un utilisateur humain.

Le LIMSI a collecté différents corpus d'interaction Humain-Humain, parmi lesquels nous pouvons citer les corpus EmoVox, BOURSE et CEMO collectés en centres d'appel, et le corpus EmoTabou. Le corpus BOURSE collecté dans un centre d'appels de gestion de portefeuille boursier a été utilisé avec des indices acoustiques, lexicaux et dialogiques (Devilleers et Vasilescu, 2004). Le corpus EmoVox (Vaudable *et al.*, 2010), collecté dans un centre d'appel d'EDF, a mené à l'annotation émotionnelle et dialogique d'entretiens téléphoniques client-agent. Le corpus CEMO (Devilleers *et al.*, 2005b) collecté dans un centre d'appel d'urgence a permis l'étude de la peur et du stress. Le corpus EmoTabou (Devilleers et Martin, 2008) consiste en une interaction dyadique basée sur le jeu, où un des deux participants, complice de l'expérimentation, tente de faire réagir émotionnellement l'autre joueur.

2.4.2. Interaction Humain-Robot

Nous assistons depuis une quinzaine d'années à un important développement dans la création et l'utilisation de corpus en robotique. Si certains corpus robotiques n'impliquent pas d'interaction, comme par exemple (Zivkovic *et al.*, 2008) qui présentent une collecte destinée à l'aide à la représentation de l'espace et à la navigation, d'autres proposent des interactions extrêmement contraintes, telles qu'un appairage entre des ordres de l'humain (« Soulève la palette et pose-la sur le sol ») et les actions concordantes du robot pour l'apprentissage de systèmes de manipulation autonomes (Tellex *et al.*, 2011).

Dans le contexte de la robotique sociale et d'assistance, nous trouvons une quantité importante d'études menées en interaction avec l'humain, ayant mené à la constitution de corpus de données. Ces corpus traitent

de la relation dyadique entre un utilisateur et un robot, mais également d'interactions, plus complexes, à plusieurs. Nous présentons une sélection des corpus actuellement constitués par la communauté, résumée dans le Tableau 1.

Tout d'abord, le corpus BITT (Maas et Wrede, 2006) constitué à l'aide du robot Biron (ActivMedia PeopleBot) dans le cadre du projet COGNIRON met en scène un utilisateur ayant pour mission de décrire son environnement au robot. L'interaction consiste juste en des monologues de la part de l'utilisateur, le robot étant dans le rôle de l'auditeur. Ce corpus permet l'analyse de la communication multimodale, et propose une annotation sur l'audio (pauses, F0), sur la vidéo (références gestuelles) et sur la transcription (hésitations, structure de dialogue, références verbales, sujet de discussion). Une étude menée conjointement dans le cadre du projet COGNIRON (Green *et al.*, 2006) a mené à la constitution d'un corpus dans des conditions similaires, pour lequel les actes de communication dans la parole et la gestuelle sont annotés sur le schéma DAMSL (Core et Allen, 1997), ainsi que les gestes déictiques, la direction du regard, la posture et la distance au robot.

Le corpus d'interaction ludique Inter-ACT entre des enfants et le robot iCat est présenté dans (Castellano *et al.*, 2010b). Un enfant joue aux échecs avec le robot. Selon l'état du jeu, le robot met à jour son état affectif, et présente une expression faciale concordante. Ce corpus vise à l'étude de données expressives naturelles dans un scénario de jeu éducatif. Les données vidéo collectées sont annotées en termes d'émotions et d'intérêt pour l'iCat.

Nous trouvons également un corpus de retours non verbaux (Lang *et al.*, 2009) dans lequel le robot Biron interagit avec un utilisateur. Ce dernier montre un objet au robot, et doit jouer une émotion via une expression du visage selon que le robot a utilisé le bon terme pour désigner l'objet ou pas. Dans cette interaction, le robot parle afin de nommer l'objet, et bouge. L'annotation porte sur l'expression du locuteur, une transcription de la parole et un repérage des phases de l'interaction.

Le corpus Aibo Emotion (Batliner *et al.*, 2004) est constitué avec des enfants anglais et allemands interagissant avec le robot chien AIBO. Ce dernier, dirigé en Magicien d'Oz, doit exécuter les ordres de l'enfant dans un scénario de localisation d'objet et de parcours. Le robot est soit obéissant, soit désobéissant. Une annotation émotionnelle du corpus est réalisée sur l'audio.

La collecte H3R (Mohammad *et al.*, 2008) présente des interactions de type Humain-Humain, mais également de type Humain-Robot. Toutes sont basées sur un même principe : un instructeur humain explique une procédure à un auditeur (humain, ou robot). Le robot est Robovie II, et son seul comportement est de bouger les yeux au hasard. Cette étude permet de comparer les évaluations subjectives et physiologiques de différentes dimensions de l'attitude d'un auditeur. L'annotation porte sur des données physiologiques (conductivité de la peau, pouls, respiration), et sur le positionnement de la tête, du corps, et du regard.

Dans certaines études, lorsque la plateforme robotique est en cours de développement notamment, le robot est représenté par un avatar dans une interface sur ordinateur plutôt que physiquement. Nous trouvons par exemple les corpus constitués dans le cadre du projet ARMEN (Chastagnol et Devillers, 2012), qui ont pour objectif d'étudier l'interaction émotionnelle entre un robot et des personnes en perte d'autonomie. L'entité robotique est représentée soit par une voix de synthèse, soit par un Agent Conversationnel Animé

(ACA) dirigé en Magicien d'Oz, avec lequel l'utilisateur interagit au cours de scénarios de projection sur le thème de l'assistance. Le comportement du robot consiste, par exemple, à montrer de l'empathie, ne pas comprendre, se tromper. Les données audio sont étiquetées en termes d'émotions.

Dans la même idée d'un avatar graphique représentant le robot, la collecte de simulation de navigation présentée par (Koulouri et Lauria, 2009) présentent un avatar graphique contrôlé en Magicien d'Oz auquel l'utilisateur doit donner des indications de direction. Cette interaction se fait par l'écrit, et le retour fourni par le robot consiste en la compréhension (bonne ou mauvaise) des indications de direction saisies. Cette étude permet l'analyse du langage spatial dans l'Interaction Humain-Robot, en s'appuyant sur l'annotation des actes de dialogue effectuée.

Un autre corpus présentant un robot à représentation graphique est décrit dans (Chernova *et al.*, 2011). Cette interaction met en relation un joueur humain amené à coopérer avec un robot pour réaliser des tâches, dans le cadre du jeu en ligne « Mars Escape ». Le robot est dirigé par un second joueur, qui dispose d'un ensemble limité d'actions. Le joueur « humain » peut interagir avec l'environnement en manipulant des objets du jeu et en communiquant avec le joueur « robot » par l'écrit. La réussite du jeu ne peut être effectuée que si l'humain et le robot coopèrent. Cette étude montre la possibilité d'utiliser des jeux en ligne pour générer des corpus d'Interaction Humain-Robot.

Dans le cadre des interactions de groupe, c'est-à-dire impliquant plusieurs humains et un robot, nous trouvons par exemple le corpus Vernissage (Jayagopi *et al.*, 2013), qui met en scène le robot NAO chargé de jouer le guide artistique : il explique des peintures, puis interroge les participants sur l'art et la culture. Deux interlocuteurs humains sont présents par session. Une annotation verbale est réalisée (pour différencier parole, silence et rire), une transcription, ainsi qu'une indication du destinataire et un repérage vidéo de la position de la tête, des hochements de tête, du regard. Ce corpus d'interaction de groupe est le point d'appui, par exemple, d'une étude pour la détermination du destinataire dans l'interaction (Jayagopi et Odobez, 2013).

Le corpus RAVEL (Alameda-Pineda *et al.*, 2011) présente des interactions dyadiques avec le robot, et également de groupe. Des acteurs jouent des scénarios incluant ou pas le robot. Dans un cas, par exemple, l'acteur demande la direction des toilettes au robot. Celui-ci, dirigé en Magicien d'Oz, indique la direction en fonction du genre de l'acteur. Dans un autre scénario, un acteur demande la direction du bureau de monsieur Smith, qui est absent actuellement. Monsieur Smith vient ensuite s'informer auprès du robot de si quelqu'un est venu le demander. Le robot observe également des dialogues entre les acteurs sans interagir, ou en intervenant dans le dialogue. Cette étude riche et complexe est filmée et enregistrée par microphone sur la tête robotique POPEYE. Les actions sont annotées par fenêtre d'une seconde sur la vidéo.

Les corpus d'Interaction Humain-Robot proposés par la communauté présentent tous des spécificités différentes et particulièrement intéressantes. Qu'il s'agisse d'interaction un-à-un ou de groupe, ou que le robot soit présenté de façon physique ou virtuelle, chacun de ces corpus est consacré à une tâche dédiée, selon les besoins des équipes de chercheur. Nous retrouvons des grandes thématiques telles que le jeu, la coopération, ou encore l'enseignement, mais les protocoles d'acquisition et d'annotations rendent chaque corpus unique et difficilement exploitable pour d'autres études.

Il convient alors de s'appuyer sur des données collectées correspondant à trois critères : la plateforme avec laquelle l'utilisateur interagit (un avatar, un système de dialogue, un robot – et dans ce cas, quel robot), le

public concerné (adultes, enfants, adolescents, étudiants, personnes en perte d'autonomie...), puis enfin la tâche (enseignement, jeu, assistance...). Dans le cas du projet ROMEO, le contexte applicatif final est complexe, puisqu'il s'agit d'assister et d'accompagner des personnes dépendantes sur le long terme, dans des situations de la vie quotidienne. Nous entrevoyons toute la difficulté de l'application : le robot doit être un compagnon social (pouvant par exemple remonter le moral, discuter, etc.), et également réaliser des tâches d'expert (surveillance, assistance...). L'étude menée dans le cadre de ce travail de thèse est volontairement limité aux situations d'assistance lors du réveil du matin, et au jeu des enfants avec le robot. L'absence, à notre connaissance, de corpus disponibles pour notre contexte spécifique appuie notre décision de nous livrer à nos propres collectes pour évaluer l'interaction émotionnelle par la parole entre NAO et des enfants dans le cadre d'un jeu, tout d'abord, puis entre NAO et des utilisateurs en perte d'autonomie dans un contexte d'assistance à domicile.

Tableau 1. État de l'art des corpus d'Interaction Humain-Robot dyadique et en groupe: composition et disponibilité

Corpus	Robot	Interaction	Media	#locuteurs	Taille	Annotation	Disponibilité
BITT (Maas et Wrede, 2006)	BIRON (ActivMedia PeopleBot)	1 à 1	Vidéo Audio	29 adultes (24F, 5M)	~5 heures	Audio : Pauses, F0 Transcription : dialogue et références	Oui: Transcription et pitch
Corpus du projet Cogniron (Green <i>et al.</i> , 2006)	ActivMedia PeopleBot	1 à 1	Vidéo Audio	22 adultes (9F, 13M)	~5 heures	Actes de communication (parole et geste) ; regard ; posture	n/a
Inter-ACT (Castellano <i>et al.</i> , 2010b)	iCat	1 à 1	Vidéo	8 enfants	~30 minutes	Émotion (valence) ; Intérêt pour l'iCat	Non
Corpus de retours non verbaux (Lang <i>et al.</i> , 2009)	Biron	1 à 1	Vidéo	11 adultes (5F, 6M)	~2 heures	Expression faciale, transcription, phase du scénario	n/a
Aibo Emotion Corpus (Batliner <i>et al.</i> , 2004)	AIBO	1 à 1	Audio	81 enfants	~18h	État émotionnel	Oui
H ³ R (Mohammad <i>et al.</i> , 2008)	Robovie II	1 à 1	Vidéo Audio Physio	22 adultes	~7 heures	Physiologique Position tête et corps Regard	n/a
ARMEN_1 et ARMEN_2 (Chastagnol et Devillers, 2012)	Voix synthétisée et ACA	1 à 1	Audio Vidéo	52 adultes 25 adultes	17,3 heures 8,7 heures	Émotion dans l'audio	Non
Simulation de navigation (Koulouri et Lauria, 2009)	Avatar graphique	1 à 1	Écrit	16 adultes	96 dialogues	Acte de dialogue	n/a
Corpus sur Mars Escape (Chernova <i>et al.</i> , 2011)	Avatar graphique	1 à 1	Écrit	558 joueurs	~700 sessions	Information contextuelle (action, objet) Dialogue	n/a
Corpus Vernissage (Jayagopi <i>et al.</i> , 2013)	NAO	2 humains avec le robot	Vidéo Audio	26 adultes	~5 heures	Verbal : parole/silence/rire Transcription Destinataire Position tête et regard	Oui : Données + Annotation
RAVEL (Alameda-Pineda <i>et al.</i> , 2011)	POPEYE	1 à 1 et Plusieurs avec le robot	Vidéo Audio	n/a	n/a	Actions	Oui

2.5. Challenges actuels et pistes de recherche proposées

2.5.1. Challenges actuels de recherche

L'étude des travaux de la communauté dans le domaine de l'Interaction Homme-Machine nous permet de constater, tout d'abord, l'importance d'un traitement situé des informations perçues, que ce soit monomodal (linguistique et/ou paralinguistique) ou multimodal. En effet, fusionner des indices émanant de l'utilisateur (langagier, gestuel, paralinguistique) doit être fait en fonction des informations contextuelles (sujet de discussion, relation sociale, état du système, degré d'accomplissement de la tâche, etc.), ce qui permet de rendre le système plus robuste dans sa compréhension de la situation d'interaction, et d'améliorer son adaptation à l'utilisateur. La robustesse du traitement automatique des signaux perçus et leur fusion monomodale ou multimodale, dans un contexte, et en fonction de chaque individu, relèvent d'un premier défi de recherche.

Ensuite, nous notons l'intérêt crucial d'interpréter les indices sociaux ponctuels, afin de constituer un profil de l'utilisateur, susceptible de faciliter l'ajustement du comportement social du système et de favoriser une relation à long-terme. La question prépondérante est donc tout d'abord de déterminer les indices pertinents dans la modélisation de l'utilisateur. Viennent alors des interrogations supplémentaires : il s'agit de déterminer quelles informations nécessitent d'être traitées ponctuellement ou mémorisées, et la façon de les interpréter. Les objectifs principaux visés par la communauté en Interaction sociale Humain-Machine s'appuie sur une maximisation de l'engagement et de la satisfaction de l'utilisateur.

La représentation utilisateur, dynamique et sur le long-terme, est un facteur de sélection du comportement social du système. Ce comportement exprimé a inévitablement un impact sur l'utilisateur, et l'on souhaite naturellement que dans tous les cas cet impact soit positif, que la finalité soit la réussite de la tâche ou une satisfaction sociale. La sélection du comportement est donc un élément clé dans la conception d'un système d'Interaction Humain-Machine.

Notre analyse des travaux effectués dans la littérature nous a également éclairés sur l'importance de l'évaluation en Interaction Humain-Machine. Cette évaluation se doit d'être modulaire et de reposer sur des métriques susceptibles d'estimer la satisfaction de l'utilisateur et la réussite de la tâche (ou plus globalement, la performance du système). L'enjeu est donc de déterminer, pour un système d'interaction, les modules à évaluer, et de sélectionner des méthodes d'évaluation (subjectives ou objectives) en fonction des besoins.

Une dernière constatation consiste en l'importance de disposer de données spontanées (« real-life ») pour l'étude de l'interaction. Nous nous sommes penchés sur les différents corpus proposés en Interaction Humain-Humain et en Interaction Humain-Machine. Dans le second cas, plus proche de notre contexte, nous avons évalué la composition et la disponibilité des corpus, afin de déterminer la possibilité de s'appuyer sur ces données dans le cadre de notre étude.

2.5.2. Nos pistes de recherche

Il nous semble primordial d'adapter les comportements sociaux d'un robot à une représentation de l'utilisateur afin d'augmenter son niveau d'engagement et sa satisfaction dans l'interaction sociale. Nous estimons que pour ce faire, il convient de réaliser une interprétation haut-niveau d'éléments paralinguistiques

afin de doter le système d'une représentation dynamique de l'utilisateur, dans laquelle le système pourra venir puiser pour déterminer le meilleur comportement à adopter.

Nous nous sommes spécialisés sur l'analyse d'un type d'entrée du système, avant d'envisager une analyse multimodale. En effet, même si l'interprétation d'une entrée monomodale peut potentiellement induire des raisonnements incomplets par manque de contexte, l'analyse de cette modalité doit être approfondie afin de valider son impact sur l'interaction. Nous considérons donc cette étape comme essentielle, et nous nous attacherons donc à réaliser une analyse poussée de l'interprétation des indices paralinguistiques extraits du signal de voix, en vue de constituer la représentation de l'utilisateur.

Nous nous pencherons également sur la composition de ce profil de l'utilisateur, en fonction de deux critères : 1) les dimensions de représentation que nous estimerons pertinentes pour la gestion d'une interaction émotionnelle entre un humain et un robot, 2) les informations possiblement interprétables des seuls indices paralinguistiques extraits du signal de voix. Nous proposerons ensuite une formalisation et une implémentation permettant la constitution dynamique de ce profil.

Concernant le comportement social du robot, nous devons déterminer les comportements les plus désirables dans nos contextes d'applications, ainsi que leur impact sur l'interaction émotionnelle. Nous étudierons également l'impact possible du profil de l'utilisateur sur cette sélection, et proposerons une formalisation et une implémentation d'une sélection automatique.

Nous réaliserons des études en contexte d'Interaction Humain-Robot dans les deux cas applicatifs : le robot assistant des personnes en perte d'autonomie, et le robot jouant avec des enfants. Ces études nous permettront la collecte de données permettant notamment des analyses de différents constituants de la boucle d'interaction.

Nous nous attacherons également à évaluer notre système d'interaction, en nous appuyant sur nos données ou en proposant des protocoles expérimentaux. Cette évaluation visera à estimer la satisfaction de l'utilisateur, mais également la réussite de la mise à jour du profil et la sélection de comportement automatiques.

Chapitre 3

3. Profil émotionnel et interactionnel de l'utilisateur

3.1. Introduction

Dans une interaction sociale entre humains, reconnaître la personnalité de son interlocuteur, ou plus généralement ses tendances émotionnelles, permet d'adapter son propre discours et sa propre attitude émotionnelle, dans le but d'aménager un espace de communication sociale efficace. Cette étude intuitive du comportement émotionnel de l'interlocuteur s'effectue dès les premiers tours d'interaction, par une lecture d'indices pertinents, et résulte en l'établissement d'un profil sommaire de l'interlocuteur. S'adapter à ce profil, ou l'ignorer, sera dépendant de la personnalité du locuteur, de ses intentions de communication, et de son désir de sociabilité.

Doter un système de la capacité de constituer le profil « intuitif » de l'utilisateur avec lequel il interagit, et ce dès les premiers tours d'interaction, est un défi de recherche. Il existe de multiples façons de modéliser ce profil. Nous nous sommes donc interrogés sur les composantes pertinentes d'un tel profil, et les indices nécessaires à sa constitution. Pour répondre à cette interrogation, il est nécessaire tout d'abord de définir la notion de profil utilisateur et de déterminer quels sont les indices pertinents pour la représentation informatique du comportement humain. Il est certain qu'une machine ne pourra percevoir qu'une petite partie des comportements de la personne et que son interprétation sera limitée, simplifiée et subjective. Une étude des théories proposées dans la communauté en intelligence artificielle, ainsi qu'en psychologie et en sociologie, nous a permis ensuite d'élaborer une représentation émotionnelle et interactionnelle de l'utilisateur. Cette représentation est utilisée pour modéliser les réactions et attitudes affectives de ce dernier, et nous présentons la façon dont nous pouvons maximiser l'interprétation des sorties d'un module de détection des émotions afin de venir renseigner ce profil.

Nous fondons notre application sur une perception limitée du contexte d'interaction : en cela nous n'exploitons que les informations extraites du signal de voix, nous ne mettons pas en jeu de reconnaissance du lexique ou d'autres analyses de l'environnement. La perception du robot est donc réduite aux informations paralinguistiques. Notre enjeu est donc de savoir exploiter au maximum ces informations, tout en ayant conscience des limitations d'interprétation imposées par notre système. Nous avons donc mis au point une représentation émotionnelle et interactionnelle de l'utilisateur, dont nous décrivons les bases théoriques dans ce chapitre. Ce profil est constitué de quatre dimensions émotionnelles : l'Émotionnalité (« les émotions exprimées sont-elles fluctuantes ? »), l'Extraversion (« le locuteur est-il volubile ? »), l'Optimisme (« le locuteur s'exprime-t-il de façon plutôt positive ou négative ? »), la Confiance en soi (« est-il sûr de lui ? »), et de deux dimensions interactionnelles : l'Affinité (« est-ce qu'il apprécie l'interlocuteur ? ») et la Domination (« prend-il le contrôle de l'interaction ? »). Ces dimensions sont une adaptation des modèles

OCEAN, modèle définissant les traits de personnalité (McCrae et John, 1992), et du modèle de l'*interpersonal circumplex* définissant les relations sociales entre interlocuteurs, basées sur les axes de domination, d'affinité, d'introversion et d'engagement (Strong *et al.*, 1988).

Nous organisons ce chapitre, dans un premier temps, sur l'analyse des composantes nécessaires à la constitution d'un modèle utilisateur. Nous abordons ensuite les différentes théories sur la représentation du profil de personnalité et de l'attitude sociale chez l'être humain. Nous étudions plus particulièrement la relation entre ce profil et les éléments paralinguistiques reconnus comme pertinents. Nous nous appuyons pour cela sur une analyse des travaux dans la littérature.

Nous présentons ensuite notre propre apport, à savoir les dimensions composant le profil émotionnel et interactionnel de l'utilisateur du système, et le traitement multi-niveau des indices paralinguistiques extraits de l'audio permettant la constitution dynamique de ce profil.

3.2. Profil utilisateur et paralinguistique dans la littérature

3.2.1. Profil utilisateur

L'adaptation à l'utilisateur est le concept clé en Interaction Humain-Machine : selon les contextes applicatifs, ajuster les réactions du système permet de rendre l'interaction plus ergonomique, l'utilisation plus efficace, ou plus sécurisée. La personnalisation du système s'appuie tout d'abord sur une prise en compte du contexte d'application, c'est-à-dire la finalité du système, la tâche à effectuer. Selon la tâche considérée, le concepteur du système s'appliquera à déterminer les caractéristiques de l'utilisateur les plus pertinentes pour réaliser l'adaptation du comportement du système. La seconde étape de la personnalisation du système consiste donc à l'élaboration d'un modèle utilisateur.

Nous considérons donc dans un premier temps le domaine applicatif du système, qui conditionnera les caractéristiques de l'utilisateur à stocker. Si nous parlons par exemple d'un système industriel, l'interaction avec un humain pourra se limiter à assurer la sécurité de ce dernier, ou l'aider à mener à bien un process. La représentation de l'utilisateur pourra alors consister en la perception de l'activation de commandes, la traversée de faisceaux de sécurité, activant ou interrompant ainsi le fonctionnement du système. Dans le cas des machines à commandes numériques, l'utilisateur pourra aussi être représenté par un code saisi, figurant son niveau de compétences, et autorisant ainsi des actions plus ou moins pointues.

Dans le cas de l'apprentissage assisté par ordinateur, des caractéristiques différentes seront retenues par le système. Dans cette finalité, la prise en compte des caractéristiques individuelles de l'apprenant est primordiale pour maximiser l'enseignement, comme le rappellent (Vandewaetere *et al.*, 2011) dans leur étude sur les caractéristiques de l'apprenant pour l'adaptation des systèmes d'apprentissage. Par exemple, les connaissances initiales de l'apprenant, ou des données démographiques et socioculturelles, doivent être connues du système afin d'étalonner et adapter l'apprentissage. Ce type d'Interaction Humain-Machine met également en jeu des variables affectives de plus haut niveau, telles que la détection de la motivation, de l'intérêt de l'apprenant en temps réel. Selon les capteurs disponibles, cela pourra être réalisé en fonction, notamment, de l'intonation de voix de l'utilisateur, sa rapidité de réponse, sa posture, son taux d'erreur.

Nous voyons que toutes ces informations récoltées sur l'utilisateur humain servent de base à des interprétations de la part du système, à une estimation du comportement, de l'opinion, ou des capacités de

l'utilisateur. Le type d'informations nécessaires et l'interprétation réalisable dépendent fortement du contexte d'interaction. S'il n'est pas pertinent pour un robot industriel de savoir si l'utilisateur est heureux ou pas, cette information pourra être primordiale pour déterminer le comportement d'un robot social. En effet, comme nous l'avons déjà abordé dans la section 2.1 sur les systèmes d'Interaction Humain-Machine, un robot social se doit d'être sensible aux affects de l'utilisateur. La détection de l'expression d'une émotion ponctuelle est donc un élément clé pour déterminer le comportement utilisateur. Nous convenons donc naturellement qu'un robot dédié à l'assistance à domicile devra détecter si l'utilisateur est en situation de détresse, mais il devra savoir faire la part des choses : « Est-ce que cet utilisateur a l'habitude de s'exprimer si violemment, ou bien s'agit-il vraiment d'un cas d'urgence ? ». Selon le cas, l'interprétation à haut niveau et sur le plus long terme peut faire sens, et influencer sur la qualité du service rendu par le système.

Nous mettons donc, tout d'abord, en opposition les caractères statique et dynamique des informations à récolter. Les informations sur la personne qui ne varient pas pendant l'interaction, telles que l'âge ou le genre, peuvent être définies « une fois pour toutes », tandis que d'autres composantes nécessiteront d'être évaluées et réévaluées au cours de l'interaction. Dans le cas des affects tels que décrits dans (Scherer et Bänziger, 2004), la stabilité est variable : les émotions présentent la plus grande instabilité au cours d'une interaction, puisqu'elles sont susceptibles de varier rapidement (voir la Figure 4). Toutefois, l'émotion exprimée peut être identifiable de façon automatique (détection des émotions dans la voix, sur le visage, etc.), et peut être la base pour une interprétation sur du plus long terme. Cette interprétation peut mener à une connaissance des tendances émotionnelles de l'utilisateur (sa personnalité), et ses inclinations affectives (la relation entretenue avec l'autre), comme nous l'aborderons par la suite.

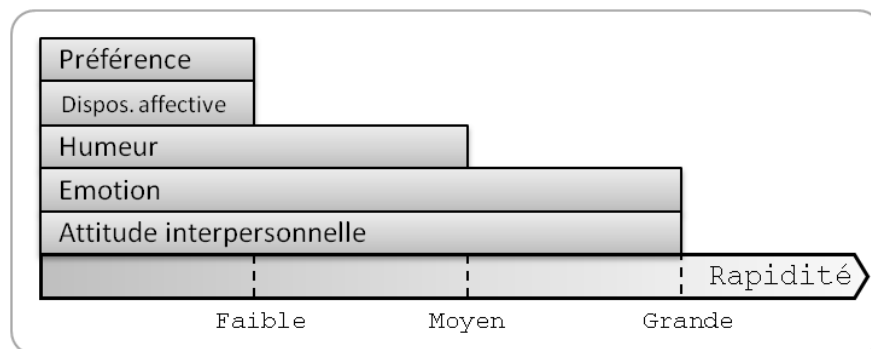


Figure 4. Rapidité de variation pour les différents types d'affect. D'après (Scherer et Bänziger, 2004).

La dynamique est une propriété qui permet d'opposer les descriptions d'états à celles d'événements (au sens étroit de ce terme). La dynamique de la collecte des informations est donc un premier point crucial pour améliorer l'adaptativité du système. La richesse des sources d'information, d'autre part, peut venir renforcer l'interprétation. En effet, la multimodalité rend l'interaction plus naturelle, et les informations complémentaires/redondantes renforcent la robustesse de l'interaction (Flickner *et al.*, 2004). Il serait alors possible d'apparier des informations issues de capteurs différents, tels que des caméras pour l'analyse visuelle, de micros pour l'analyse linguistique et paralinguistique, de capteurs de pression pour les études haptiques. Toutefois, l'étude réalisée au cours de cette thèse se concentrera principalement sur la paralinguistique extraite de l'audio, car l'interprétation de ces éléments extraits en termes de tendances émotionnelles et interactionnelles nécessite une première analyse approfondie, pour laquelle nous avons disposé d'outils de détection au cours de notre travail de thèse, et qui pourra par la suite être adaptée sur d'autres modalités.

Pour conclure, dans un système d'interaction entre un utilisateur humain et un système, il convient d'inventorier les éléments pertinents pour la représentation de l'humain. Cet inventaire dépend des modalités d'entrées (types d'inputs, nombre d'inputs, dynamique/en dur) et du but applicatif final du système (industriel/de service, plus ou moins intelligent, affectif/utilitaire).

Les informations paralinguistiques (liées aux émotions, à l'état de fatigue, à l'emphase du discours, etc.) sont une source d'informations pertinentes dans le cadre de l'interaction avec un système « affectif ». Ces informations sont des réponses conscientes ou inconscientes à différents événements (émotions, fatigues musculaire ou cognitive...) modifiant les caractéristiques de la voix. Nous nous interrogeons alors sur une représentation à plus long-terme que l'on peut réaliser à partir de ces informations émotionnelles : nous pouvons obtenir des informations de tendances émotionnelles de l'utilisateur (sa personnalité), et ses inclinations (la relation qu'il entretient avec l'autre).

3.2.2. Profil de personnalité et paralinguistique

Déterminer la personnalité d'un individu – à savoir son schéma de comportement émotionnel – dans la vie quotidienne est une tâche ardue et subjective ; en effet, s'il est possible de la discerner au travers du comportement de l'individu, ce comportement varie en réaction à son environnement. Pouvons-nous dire d'une personne qu'elle est impulsive sous prétexte qu'on la voit réagir vivement à une contrariété ? Nous ne pouvons probablement pas l'affirmer, sans une connaissance du contexte : est-ce la première fois que cette contrariété se présente ? Ou bien est-ce la dixième fois ? D'autres contrariétés ont-elles miné l'humeur de l'individu au préalable ? Pourtant, lors d'une première interaction avec cet individu, et ce même sans connaissance préalable de son environnement, nous établissons un premier profil, basé sur une « première impression ». Les interactions subséquentes feront évoluer ce profil, soit pour renforcer la première impression, soit pour le faire évoluer dans une autre direction.

Les recherches sur la personnalité en psychologie ont permis d'étiqueter certaines dimensions susceptibles de représenter l'individu, d'une façon indépendante de la culture. Le modèle prédominant dans la communauté sur l'*affective computing* est basé sur cinq dimensions, ou facteurs, qui recouvrent chacun un ensemble de termes (généralement adjectivaux) permettant de décrire l'individu de façon exhaustive, d'obtenir un phénotype de sa personnalité. Ces dimensions sont référencées usuellement sous les termes *Extraversion*, *Agreeableness*, *Conscientiousness*, *Emotional Stability* and *Openness*, généralement regroupés sous la dénomination de *Five-Factor Model*, et ont été adaptées pour différentes langues. Bien que gardant les mêmes concepts empiriques, ce modèle à cinq dimensions est exploité de façon différente selon les chercheurs, tant d'un point de vue terminologique que conceptuel.

D'une façon générale, ces différentes dimensions peuvent être comprises comme un continuum entre deux extrêmes, l'individu moyen se situant à un point quelconque entre ces deux polarités. Voici les définitions communément admises pour ces différentes dimensions :

- *Openness* (ouverture) : L'ouverture au monde de l'individu, c'est-à-dire son caractère imaginatif, adaptatif.

- *Conscientiousness* (application) : L'individu faisant preuve d'application tend à contrôler ses réactions et à ne pas s'éloigner de ses buts. L'individu appliqué est organisé et soucieux du détail.
- *Extraversion* (extraversion) : La propension de l'individu à s'exprimer, à affirmer ses points de vue, à communiquer aisément.
- *Agreeableness* (amabilité) : L'individu fait preuve d'un comportement prosocial. Il exprime de la gentillesse, de l'affection et s'intéresse aux autres.
- *Emotional stability* (stabilité émotionnelle) : L'individu stable émotionnellement tend à éprouver peu de fluctuations émotionnelles, et peu d'anxiété, de tristesse, de pessimisme.

Intéressons-nous à deux grandes approches : les *Big Five* et le modèle à cinq dimensions *OCEAN*. Les travaux de Goldberg, pour lesquels nous pouvons citer notamment (Hofstee *et al.*, 1992), présentent les dimensions *Surgency*, *Agreeableness*, *Conscientiousness*, *Emotional Stability* et *Intellect*, reprises sous la dénomination *Big Five*. Ces dimensions ont été regroupées en partant d'une hypothèse lexicale répandue dans la communauté du Five-Factor Model, selon laquelle les différences individuelles les plus pertinentes sont représentées dans la langue courante et peuvent être regroupées en différents facteurs propres à dresser le phénotype des traits de personnalité de tout individu. Goldberg présente ces dimensions en dix circomplexes reprenant les facteurs deux à deux, car ses études démontrent le grand nombre de chevauchement entre les éléments de différents facteurs, impliquant l'obligation de traiter les facteurs en corrélation les uns avec les autres. Ce modèle circulaire a été nommé *The Abridged Big-Five-Dimensional Circumplex* (connu sous l'acronyme AB5C). Quant au modèle à cinq facteurs de McCrae et John (McCrae et John, 1992), il présente les concepts d'*Openness to Experience*, *Conscientiousness*, *Extraversion*, *Agreeableness* et *Neuroticism* (communément regroupés sous l'acronyme OCEAN). Ce modèle se base sur l'analyse des différents facteurs testés dans les questionnaires de personnalité. Les équipes de McCrae et John ont analysé un grand nombre de questionnaires (citons par exemple le Hogan Personality Inventory, l'Adjective Check List, le Minnesota Multiphasic Personality Inventory, le Personality Research Form), afin de distinguer les facteurs communs à ces dimensions. Ces différents facteurs ont pu être regroupés en cinq dimensions susceptibles de représenter la personnalité d'un individu. (McCrae et John, 1992) présentent un tableau (voir Figure 5) exemplifiant la relation possible entre les cinq facteurs, les dimensions d'analyse des questionnaires de personnalité, et les traits adjectivaux associés.

En nous appuyant sur les cinq dimensions du *Five-Factor Model*, nous nous interrogeons maintenant sur la relation existant entre les émotions exprimées par l'individu, et ses traits de personnalité. Un individu présentant tel trait de personnalité tendra-t-il à exprimer plus fréquemment un certain type d'émotions ? À l'inverse, une émotion peut-elle être un indice de personnalité ? Dans leur ouvrage sur la personnalité et l'émotion, (Revelle et Scherer, 2009) décrivent cette relation de la sorte : « *[P]ersonality is to emotion as climate is to weather. That is, what one expects is personality, what one observes at any particular moment is emotion.* » (*La personnalité est à l'émotion ce que le climat est au temps qu'il fait. En d'autres termes, ce que nous présumons, c'est la personnalité, et ce que nous observons à un moment particulier, c'est une émotion*). L'émotion exprimée serait donc une des manifestations de la personnalité de l'individu.

Factor		Factor definers		
Name	Number	Adjectives ^a	Q-sort items ^b	Scales ^c
Extraversion (E)	I	Active	Talkative	Warmth
		Assertive	Skilled in play, humor	Gregariousness
		Energetic	Rapid personal tempo	Assertiveness
		Enthusiastic	Facially, gesturally expressive	Activity
		Outgoing	Behaves assertively	Excitement Seeking
Agreeableness (A)	II	Talkative	Gregarious	Positive Emotions
		Appreciative	Not critical, skeptical	Trust
		Forgiving	Behaves in giving way	Straightforwardness
		Generous	Sympathetic, considerate	Altruism
		Kind	Arouses liking	Compliance
Conscientiousness (C)	III	Sympathetic	Warm, compassionate	Modesty
		Trusting	Basically trustful	Tender-Mindedness
		Efficient	Dependable, responsible	Competence
		Organized	Productive	Order
		Planful	Able to delay gratification	Dutifulness
Neuroticism (N)	-IV	Reliable	Not self-indulgent	Achievement Striving
		Responsible	Behaves ethically	Self-Discipline
		Thorough	Has high aspiration level	Deliberation
		Anxious	Thin-skinned	Anxiety
		Self-pitying	Brittle ego defenses	Hostility
Openness (O)	V	Tense	Self-defeating	Depression
		Touchy	Basically anxious	Self-Consciousness
		Unstable	Concerned with adequacy	Impulsiveness
		Worrying	Fluctuating moods	Vulnerability
		Artistic	Wide range of interests	Fantasy
		Curious	Introspective	Aesthetics
		Imaginative	Unusual thought processes	Feelings
		Insightful	Values intellectual matters	Actions
		Original	Judges in unconventional terms	Ideas
		Wide interests	Aesthetically reactive	Values

Figure 5. Relation entre facteurs de personnalité, éléments de questionnaires et traits adjectivaux associés (McCrae et John, 1992)

Revelle et Scherer présentent des études menées à l'aide de cartographie cérébrale, permettant de corrélérer les traits de personnalité *Neuroticism* et *Extraversion* à des réactions à des stimuli à valence positive et négative. Par exemple, un individu neurotique aura tendance à réagir fortement à un stimulus à valence négative (l'étude ne sera pas probante dans le cas de stimuli positif), tandis qu'une personne extravertie réagira fortement à un stimulus à valence positive (non probant dans le cas de stimuli négatif). Revelle et Scherer décrivent également la notion de « trait émotionnel », ou « d'habitude émotionnelle », comme étant la tendance d'un individu à éprouver fréquemment une émotion donnée. Ils déterminent ainsi l'existence de différences individuelles susceptibles de constituer une prédisposition à éprouver certaines émotions. Les auteurs rapportent que la relation entre les autres traits de personnalité du Five-Factor Model et les affects n'a été que rarement examinée. De même, les études précises se limitent à l'analyse des traits émotionnels positifs, d'anxiété, et de colère.

Tableau 2. Sélection de termes qualificatifs pour un individu selon différentes facettes de personnalité.
Extrait et traduit de (Hofstee *et al.*, 1992).

Couple	Terme	Couple	Terme
Extraversion + Extraversion +	<i>bavard</i> <i>agressif</i> <i>affirmé</i> <i>spontané, direct</i>	Extraversion - Extraversion -	<i>calme</i> <i>silencieux</i> <i>taciturne</i>
Extraversion + Agreeableness +	<i>enthousiaste</i> <i>expansif</i> <i>énergique</i>	Extraversion - Agreeableness -	<i>réservé</i> <i>morose</i> <i>froid</i>
Extraversion + Agreeableness -	<i>énergique, déterminé</i>	Extraversion - Agreeableness +	<i>crainitif</i> <i>non agressif</i>
Extraversion + Conscientiousness +	<i>énergique, dynamique</i>	Extraversion - Conscientiousness -	<i>apathique</i> <i>lent, mou</i> <i>létargique</i>
Extraversion + Conscientiousness -	<i>chabuteur, bruyant</i> <i>démonstratif</i>	Extraversion - Conscientiousness +	<i>réservé</i> <i>discret</i>
Extraversion + Emotional Stability +	<i>désinhibé</i> <i>vigoureux</i>	Extraversion - Emotional Stability -	<i>pessimiste</i> <i>mélancolique</i> <i>mystérieux, secret</i>
Extraversion + Emotional Stability -	<i>impétueux</i> <i>verbeux</i>	Extraversion - Emotional Stability +	<i>paisible</i> <i>posé, calmé</i> <i>placide</i>

Cependant, les relations entre éléments lexicaux et traits de personnalité, tels que décrits notamment dans les travaux de (Hofstee *et al.*, 1992) sur le modèle AB5C, nous fournissent un grand nombre de concepts pouvant être mis en rapport avec l'expression d'affects, voire avec une attitude émotionnelle et communicationnelle globale. Le Tableau 2 présente une sélection des termes qualificatifs présentés dans les travaux de Hofstee, de Raad et Goldberg susceptibles d'illustrer la façon dont un individu s'exprime. Le modèle AB5C, complexe et riche, tend à représenter de façon exhaustive la personnalité d'un individu en couplant usuellement les dimensions des Big Five deux à deux (chaque couplage étant nommé « *facet* » dans leurs travaux, que nous traduisons ici simplement par « couple »). Le tableau présenté met en relation l'extraversion avec chacune des autres dimensions, montrant ainsi les manifestations de chacun des traits de personnalité selon que l'individu tendra à être communicatif ou non. Le critère de sélection de ces quelques termes parmi les quatre-vingt dix proposés repose sur un choix des concepts les plus simples, les plus clairement identifiables. Par exemple, nous éliminons les qualificatifs impliquant une compréhension du

contexte et des objectifs de l'individu (« rusé », « conservateur », « espiègle », « audacieux », etc.) ainsi que les termes trop généraux (« sociable », « introverti »), pour ne conserver que les termes décrivant une attitude facilement et objectivement identifiable.

3.2.3. Profil social et paralinguistique

La relation dyadique d'un individu avec un autre fait l'objet de la psychologie interactionnelle. La représentation de cette relation prend usuellement la forme d'un circomplexe interpersonnel, où l'attitude sociale d'un individu par rapport à l'autre s'exprime en termes de niveaux de *Domination* et d'*Amicalité* (« *Friendliness* » ou « *Liking* » – *Affinité* – selon les théories). Ces deux dimensions sont reconnues comme étant cartographiées sur le même principe que les dimensions *Surgeny* et *Agreeableness* du modèle des Big Five. Chaque dimension est un continuum, dont les extrêmes seront respectivement *Domination* versus *Soumission*, et *Amicalité*/*Affinité* versus *Hostilité* (voir Figure 6). Ces extrêmes peuvent être également compris comme « au-dessus de l'autre » versus « en dessous de l'autre », et « contre l'autre » versus « ensemble ». Dans la théorie interpersonnelle, la notion de complémentarité, élaborée initialement par (Carson, 1969), se base sur un principe de réciprocité (la domination appelle la soumission, au même titre que la soumission appelle la domination) et de correspondance (l'amicalité appelle l'amicalité, l'hostilité appelle l'hostilité). La position interactionnelle d'un individu dans sa relation avec l'autre variera donc en fonction de son interlocuteur, c'est-à-dire de leurs capacités respectives à s'adapter, de leur personnalité, de leurs intentions, ou encore de leur position dans la hiérarchie sociale.

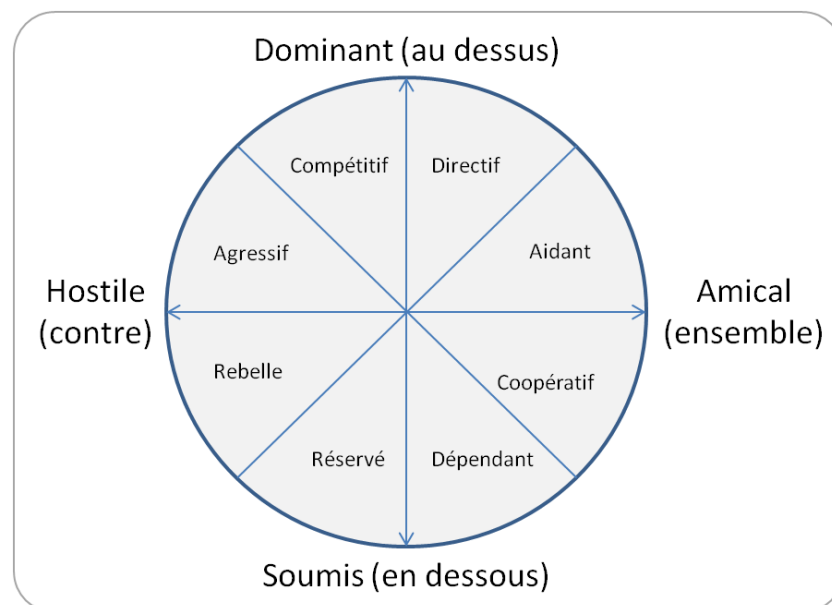


Figure 6 - Représentation des dimensions de Domination et d'Amicalité sur le Circomplexe Interpersonnel, et attitudes sociales dérivées de ces deux dimensions. Adapté de (Isbister, 2006).

Cette position interactionnelle se répercute évidemment sur l'attitude du locuteur par rapport à l'interlocuteur. (Leary, 1957) présente un circomplexe des comportements interpersonnels basé sur le principe de « en dessous » versus « au dessus », et « avec » versus « contre » (voir Figure 7). Le cercle central présente les tendances de comportement d'un individu présentant les traits définis en pourtour du cercle. Ces traits sont présentés à l'aide de deux termes, exprimant chacun un degré d'intensité : le premier cas présente

une intensité modérée pour ce trait, tandis que le deuxième cas s'avère pathologique (exemple : *Managerial-Autocratic*, « qui désire encadrer » jusqu'à « tyrannique »). D'après ce circomplexe, nous voyons qu'un individu compétitif/narcissique (donc « au dessus » et modérément « contre ») s'affirmera, et tendra à entrer en compétition. Le cercle intermédiaire indique les tendances de comportement que cela peut potentiellement déclencher chez l'interlocuteur. Dans le cas de notre exemple, il s'agirait d'une réaction de méfiance.

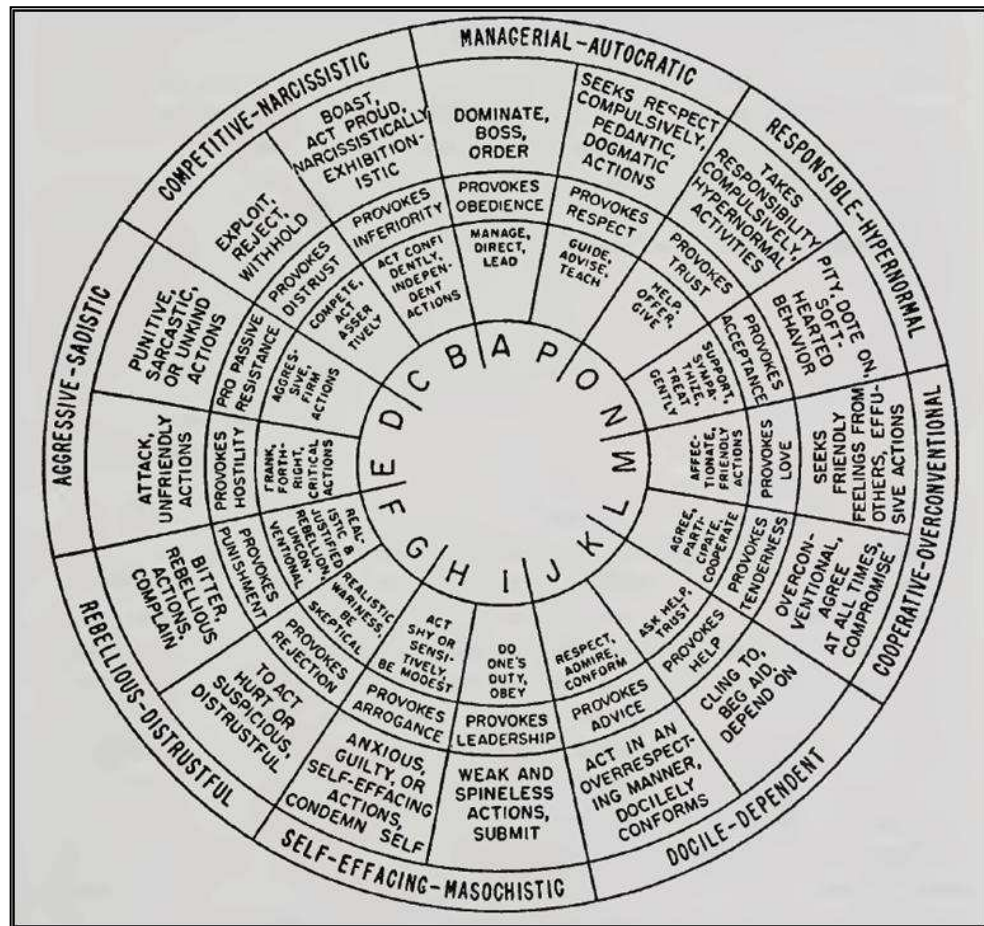


Figure 7 - Classification des comportements interpersonnels basés sur les notions de Domination et d'Amicalité. Extrait de (Leary, 1957), p. 65.

Le circomplexe interpersonnel peut également être représenté et illustré en termes d'attitudes interactionnelles, comme nous le voyons d'après la Figure 8 extraite de (Isbister, 2006), fondée sur les travaux de (Strong *et al.*, 1988). Cette représentation ajoute explicitement les vecteurs de « connexion avec l'autre » (*Connected* versus *Separated*), et d'« extraversion » (*Extraverted* versus *Introverted*), qui étaient sous-entendus dans les travaux de Leary présentés précédemment.

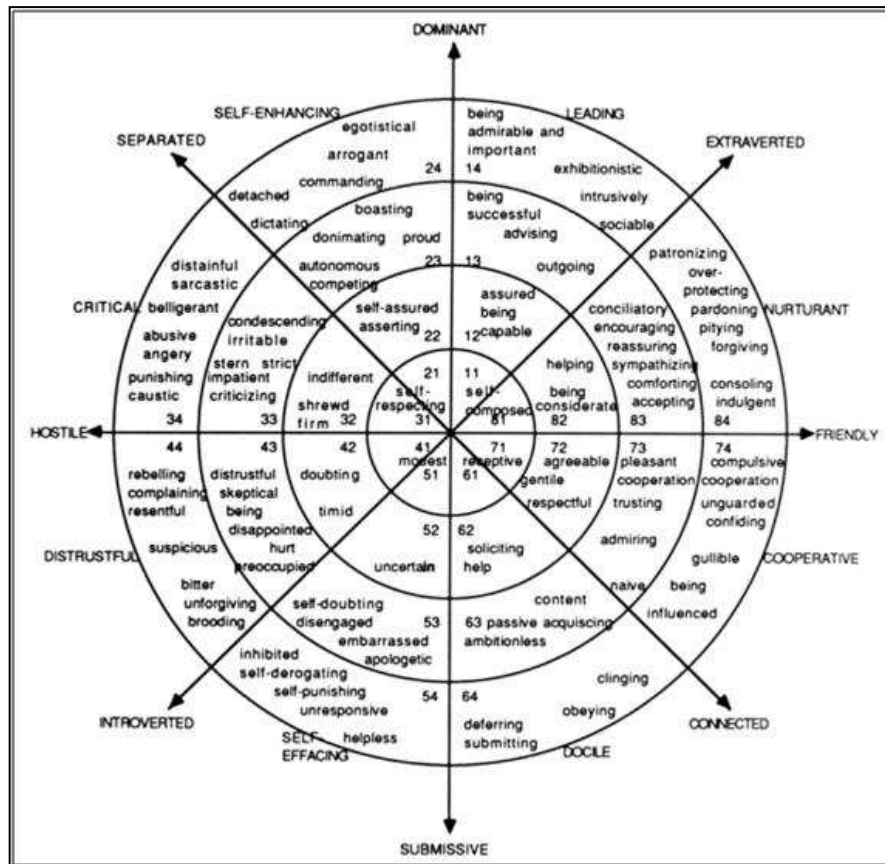


Figure 8 - Attitudes interactionnelles sur le Circomplexe Interpersonnel. Extrait de (Isbister, 2006), p. 26

La description d'une relation interpersonnelle, basée sur la Domination et l'Amicalité, peut être renforcée par l'adjonction d'une dimension de Familiarité (Svennevig, 1999). Dans la théorie de la pénétration sociale (Altman et Taylor, 1973), il est considéré que les relations se développent grâce à la divulgation progressive d'informations personnelles à l'interlocuteur. Si une relation naissante tendra à n'évoquer principalement que des sujets non personnels, une relation de plus longue durée évoluera vers un partage d'informations intimes et de connaissances communes. L'autodéclaration, ou autorévélation (*self-disclosure* en anglais), est donc un phénomène largement étudié dans la théorie de la pénétration sociale, car ses manifestations auraient un impact sur le niveau de familiarité de la relation.

La notion de Solidarité, également nommée distance sociale, a été présentée conjointement au concept de *power* (la capacité à influencer sur le comportement de l'autre) dans les travaux de (Brown et Gilman, 1960) sur l'emploi linguistique des formes familière et polie dans une relation sociale. La solidarité est considérée comme le fait de « penser pareil », ou de jouir de prédispositions comportementales communes (telles qu'une même profession, une même confession, être du même sexe). Comme il est rapporté dans (Bickmore et Cassell, 2001), la fréquence de contact et la solidarité peuvent avoir un lien, sans nécessairement avoir de relation causale.

Ces quatre dimensions peuvent former une représentation du profil social de l'individu dans sa relation avec l'autre. Nous nous interrogeons maintenant sur les manifestations observables d'un tel positionnement social. Nous trouvons des analyses intéressantes dans les travaux de (Isbister, 2006) et (Ochs *et al.*, 2009), mettant en relation la représentation du profil social d'agents virtuels et leur comportement effectif.

D'après ces études, l'Amicalité/Affinité se manifesterait alors par l'utilisation d'un ton chaud et énergique, au contraire d'une personne indifférente qui s'exprimerait de façon neutre, ou encore d'une personne hostile qui utiliserait un ton froid (énergique ou neutre). L'individu faisant preuve d'une affinité pour son interlocuteur présenterait également les signes caractéristiques de l'engagement, discernables via la posture, la chaleur du ton, la prise en compte de l'autre. L'Affinité du locuteur pour l'interlocuteur pourrait de même être accrue lorsque l'interlocuteur est à l'origine d'une émotion positive chez le locuteur.

Ces études mettent également en avant le fait que la Domination d'un individu sur l'autre peut être estimée à la façon dont celui-ci s'adresse à son interlocuteur. Une personne dominante parlerait plus fort, et prendrait volontiers le contrôle de la discussion par une prise de parole intensive, le choix du sujet et du ton de la conversation. Le déclenchement d'émotions morales (fierté, admiration, honte) peut être à l'origine d'un changement du niveau de domination que le locuteur pense exercer sur l'interlocuteur. Par exemple, le fait de faire éprouver à l'interlocuteur une sensation de honte ou d'admiration accroît le sentiment de domination du locuteur. Exprimer de la colère en réaction à l'attitude de l'interlocuteur peut être un indice de domination sur l'interlocuteur, au même titre que la peur serait un indice de soumission.

La Familiarité requerrait une connaissance du sujet de discussion, afin de pouvoir discerner s'il s'agit d'informations publiques ou privées. Si le sujet concerne la divulgation d'une information privée par le locuteur, on estimerait alors que la familiarité entre les interlocuteurs croît. (Ochs *et al.*, 2009) décrivent la diminution du sentiment de solidarité, si l'interlocuteur provoque une émotion négative chez le locuteur. La Solidarité n'augmenterait par contre pas si au contraire l'émotion est positive.

La notion de Solidarité peut être liée à la similarité de réaction à un même événement. En effet, la Solidarité est considérée comme le fait de « penser pareil ». Il est donc évident que si chaque individu exprime la même réaction comportementale ou émotionnelle face à un même événement, le degré de Solidarité peut être renforcé. L'on peut considérer alors qu'à l'inverse, des réactions dissemblables face à un même stimulus diminueraient la sensation de Solidarité entre les deux locuteurs. Nous trouvons également dans (Markel *et al.*, 1972) une relation entre la Solidarité et le ratio parole sur silence. Son étude s'appuie sur l'importance des pauses dans le discours, indiquant une différence statistiquement significative de ce ratio selon qu'il s'agit d'amis dialoguant, ou d'étrangers : des amis auraient moins recours aux silences que des étrangers.

En nous basant sur les travaux des communautés de recherche en interaction homme-robot, en psychologie et sociologie, nous pouvons donc déterminer les composantes du profil social d'un individu. Ce profil, à savoir sa relation avec l'autre, peut ainsi être représenté par quatre dimensions évoluant sur un continuum : l'Affinité, la Domination, la Familiarité et la Solidarité. La reconnaissance de ces dimensions nécessite une certaine compréhension du contexte d'interaction, comme la relation sociale entre les interactants, les événements ayant mené à la situation d'interaction, le sujet de discussion ou l'opinion des intervenants sur ce sujet. Si nous sommes conscients de la quantité d'informations nécessaires à l'établissement de ce profil social, il nous faudra néanmoins adapter la précision du profil en fonction des indices que l'on pourra détecter automatiquement dans notre système.

3.3. Nos travaux : Représentation émotionnelle et interactionnelle de l'utilisateur

Nos études bibliographiques préliminaires (détaillées dans les points 3.2) nous ont permis de déterminer les différentes dimensions susceptibles de représenter le profil interactionnel et de personnalité d'un individu. Nous avons repéré dans la littérature les différents indices à même d'être corrélés avec les dimensions de ce profil.

Ces études ont mené à l'établissement d'un profil émotionnel et interactionnel de l'utilisateur, susceptible de représenter dynamiquement les tendances de personnalité de celui-ci, et son attitude sociale face au système. Cette représentation utilisateur n'a pas valeur de vérité hors d'un contexte de communication Humain-Machine spécifique. En effet, la modélisation est invariablement limitée par les connaissances que l'on peut transmettre au système. Nous présentons donc dans cette section les différentes dimensions de profil émotionnel et interactionnel que nous avons jugées comme pertinentes pour une Interaction Humain-Machine, et les indices qu'il nous sera possible d'exploiter à différents niveaux pour constituer chacune de ces dimensions.

3.3.1. Le profil émotionnel et interactionnel : dimensions

3.3.1.1. Dimensions émotionnelles

Les dimensions proposées dans le Five-Factor Model sont présentées comme permettant de représenter l'individu de façon exhaustive. L'évaluation de ces dimensions est usuellement réalisées par l'utilisation de questionnaire d'auto-évaluation sur la personnalité (EPQ, ou NEO PI-R, etc.), ou par l'analyse des locuteurs par des experts à l'aide de grilles d'évaluation, tel qu'a été le cas pour le « Speaker Personality Corpus » analysé à l'aide du BFI-10 (Mohammadi *et al.*, 2010).

Il semble inconcevable de poser à l'utilisateur des questions directes sur sa personnalité afin de fournir les informations résultantes au robot. Il pourrait éventuellement être envisageable, lors de la première utilisation du système, de proposer à l'utilisateur un questionnaire ludique, voire un petit *serious game*, afin de renseigner un premier profil. Mais le vrai défi technologique, au cœur des préoccupations actuelles de la communauté, est la mise au point d'un système muni d'un modèle dynamique, dont les données seraient actualisées au fil des interactions, et qui serait donc susceptible de comprendre de façon autonome les tendances de l'utilisateur. Les données paralinguistiques et émotionnelles collectées fourniraient des indications permettant de définir une "intuition de la personnalité" du locuteur, avec probablement de plus en plus de précision au fil du temps.

La richesse du modèle des cinq facteurs de personnalité n'est pas pertinente dans notre contexte d'étude. En effet, s'agissant d'une application destinée à interagir naturellement avec l'utilisateur, et donc pas d'un outil de diagnostic psychologique, il convient de doter le système d'intuitions, d'une représentation globale et orientée sur la tâche. D'autre part, le modèle des cinq facteurs implique une compréhension très poussée de l'individu et de son contexte, compréhension que nous ne pouvons à l'heure actuelle détenir dans notre système informatique. Nous sélectionnons et adaptons alors les dimensions du Five-Factor Model propres à répondre à notre besoin.

La dimension d'**Optimisme**, tout d'abord, est l'attitude positive du locuteur par rapport aux remarques et aux comportements du robot, et par rapport à la situation en général. Elle se rapproche de la dimension

d'Agreeableness en ce qu'elle inclut une dimension d'adaptation positive au contexte, et de la notion de Neuroticism, puisque cette dernière oppose notamment des attributs de pessimisme et d'optimisme. Toutefois, ces deux dimensions incluent des paramètres beaucoup plus larges et difficilement modélisables. La dimension d'Optimisme est un continuum dont les pôles se situent entre *Peu optimiste* et *Très optimiste*. Nous notons à cette occasion l'absence volontaire de termes négatifs (tels que *pessimiste*) dans la conception que le système se fait de l'utilisateur. Ce choix terminologique est motivé par le souhait de ne pas élaborer un système propre à traiter des cas pathologiques (dépression, troubles de personnalité, etc.), et donc de se différencier de ces champs de recherche. Nous restons dans un contexte d'assistance robotique à domicile, de compagnon social.

Nous traitons également la notion de **Confiance en soi**, qui est un des traits bipolaires du modèle des Big Five permettant d'évaluer le couple « Emotional Stability + » et « Extraversion + ». Cet aspect de la personnalité de l'individu nous intéresse, car il permet au système de déterminer (conjointement à d'autres facteurs) si l'utilisateur est susceptible de prendre la situation en main, ou s'il désirerait plutôt être accompagné dans ses décisions. Les valeurs de cette dimension évoluent sur un continuum allant de *Peu confiant en lui-même* à *Très confiant en lui-même*.

Le degré d'**Émotionnalité** est représenté par les variations dans les émotions du locuteur. Un locuteur très émotionnel utilisera toute une gamme d'émotions, allant du négatif au positif. Un locuteur peu émotionnel sera beaucoup plus stable, invariant. Cette dimension se rapproche indubitablement de la notion d'Emotional Stability. Cependant, la notion du Five-Factor Model inclut également une propension à exprimer des émotions négatives telles que tristesse ou anxiété, que nous n'estimons pas inclure dans cette acception. Le continuum varie entre *Peu émotionnel* et *Très émotionnel*.

L'**Extraversion** caractérise la propension du locuteur à exprimer des émotions intenses, ou si au contraire il est réservé. Cette dimension recoupe la notion d'Extraversion du Five-Factor Model. Cette information peut permettre au système de déterminer si l'utilisateur aime l'interaction, et viendra engager l'interaction de son propre chef, ou si au contraire l'utilisateur a besoin d'être stimulé pour interagir. Le continuum varie entre *Peu extraverti* et *Très extraverti*.

Tableau 3. Dimensions émotionnelles du profil utilisateur

Dimensions	Pôles du continuum
Optimisme	<i>Peu optimiste à Très optimiste</i>
Confiance en soi	<i>Peu confiant en lui-même à Très confiant en lui-même</i>
Emotionnalité	<i>Peu émotionnel à Très émotionnel</i>
Extraversion	<i>Peu extraverti à Très extraverti</i>

3.3.1.2. Dimensions interactionnelles

Nous avons vu dans le point 3.2.3 quatre dimensions susceptibles de représenter la relation sociale entre deux interactants : l'Amicalité et la Domination, d'une part, représentées en un circomplexe, et d'autre part les notions de Solidarité et de Familiarité, évaluant le degré de proximité sociale entre deux individus.

Les concepts de Solidarité et de Familiarité sont bien sûr essentiels dans la compréhension du lien social entre deux individus. Cependant, elles font appel à des informations dont nous ne disposons pas dans le système que nous désirons mettre au point : en effet, nous désirons exploiter les éléments provenant du système de détection audio des émotions. Nous sommes quasiment aveugles au contexte ou à d'autres sources de perception, qui bien évidemment renforceraient le diagnostic quand à la personnalité et au lien social de l'individu. Nous nous limitons donc dans le cadre de cette étude aux dimensions Affinité et Domination.

Le degré d'**Affinité** caractérise le lien qui semble unir l'utilisateur au système intelligent, la vision que ce premier a du dernier. En d'autres termes, cette dimension évalue si l'utilisateur semble apprécier le système, s'il entretient avec lui une relation familière avec lui ou s'il est plutôt froid. Cette information peut être portée par les inflexions, la chaleur de la voix de l'utilisateur. Le continuum varie de *Peu amical* à *Très amical*.

La **Domination** exercée par l'utilisateur sur le robot indique si l'on ressent que l'utilisateur se laisse guider dans son interaction avec le robot. Par exemple, cette dimension nous permet de distinguer l'utilisateur qui attendra les consignes du robot pour engager ou poursuivre l'interaction, en écoute passive des commentaires du robot, ou à l'inverse si l'utilisateur semble « prendre les choses en main » et tenter de diriger le robot. Le fait de donner des ordres peut par exemple être un indice du niveau de domination. Les valeurs de cette dimension évoluent de *Peu dominant* à *Très dominant*.

Tableau 4. Dimensions interactionnelles du profil utilisateur

Dimensions	Pôles du continuum
Affinité	Peu amical à Très amical
Domination	Peu dominant à Très dominant

Nous considérons donc le profil de l'utilisateur comme une représentation à deux composantes : les tendances émotionnelles, et les tendances interactionnelles, toutes deux potentiellement interprétables à partir du signal audio. L'interprétation n'est évidemment pas entièrement fiable, puisqu'elle requerrait des informations contextuelles supplémentaires permettant de compléter l'estimation : c'est bien sûr l'intérêt d'un système multimodal. Cependant, un système multimodal est modulaire, et il convient de mener une analyse fine sur chacun des inputs. Dans le cadre de notre travail de thèse, nous nous concentrons sur l'interprétation des informations paralinguistiques que les inputs audio peuvent apporter. La section suivante aborde la façon dont nous comptons réaliser ce traitement.

3.3.2. Le traitement multi-niveau des indices paralinguistiques

Dans l'interaction humaine, les informations non-verbales telles que la gestuelle, l'expression faciale et les indices paralinguistiques viennent renforcer la compréhension du message transmis. Une interprétation haut-niveau de ces informations permet l'élaboration d'une représentation plus complexe de l'intervenant, par une gestion du long-terme, et de la répétition des informations. Par exemple, nous tirons naturellement des conclusions sur les traits émotionnels d'une personne si nous la voyons régulièrement sourire, ou au contraire régulièrement en colère. Par la gestion et la compréhension de ces signes, nous nous construisons une représentation instinctive de cet individu, représentation qui interviendra dans notre choix de comportement vis-à-vis de lui.

En Interaction Humain-Machine, il convient donc de discerner plusieurs étapes dans le traitement des indices bas-niveau récupérés par le système. Tout d'abord, une interprétation basée sur des méthodes statistiques ou expert permet de déterminer une liste des informations paralinguistiques reconnues. Par exemple, l'observation de la position de certains muscles du visage permet de reconnaître un sourire : les Unit Actions du système FACS (Facial Action Coding System) décrivent précisément les unités minimales intervenant dans la réalisation des signaux sociaux (Ekman et Rosenberg, 1997). A un plus haut niveau d'analyse, ces marqueurs peuvent être traités de façon à obtenir des informations sur les tendances émotionnelles et interactionnelles de l'individu : par exemple, s'il est sûr de lui, s'il apprécie son interlocuteur.

La Figure 9 résume la façon dont nous organisons notre interprétation multi-niveau des indices paralinguistiques extraits du signal de voix : tout d'abord, un traitement statistique des indices bas-niveau permet la récupération d'informations paralinguistiques. Ce premier niveau d'analyse est réalisé de façon ponctuelle, lorsque l'émotion vient d'être exprimée. Ce champ disciplinaire n'est pas l'objet de notre étude, nous nous pencherons principalement sur les indices paralinguistiques obtenus, et la façon dont nous les traiterons pour la réalisation du profil émotionnel et interactionnel de l'utilisateur. L'élaboration du profil s'appuie sur un historique regroupant les marqueurs paralinguistiques, constitué à l'aide de règles de logique floue. Des informations contextuelles, ainsi que la connaissance du comportement robot au tour précédent, peuvent venir renforcer ce système : nous traiterons ce point dans la suite de ces travaux.

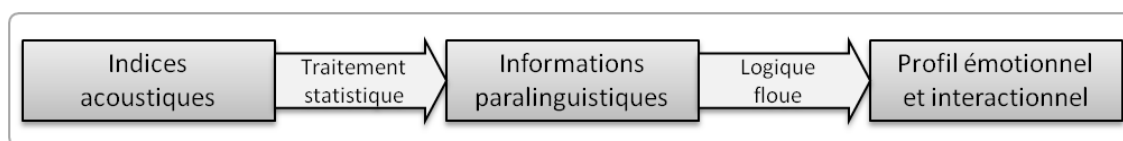


Figure 9. Traitement multi-niveau des indices paralinguistiques extraits de l'audio, contribuant à la constitution d'un profil émotionnel et interactionnel

Dans cette section de la thèse, nous présenterons tout d'abord les marqueurs moyen-niveau extraits du signal de voix disponibles dans notre système, et servant de base à l'élaboration de notre profil. Nous aborderons donc brièvement la façon dont ces marqueurs sont extraits par la détection automatique des émotions. Nous traiterons ensuite de la façon dont ces marqueurs peuvent être interprétés pour chaque dimension du profil.

3.3.2.1. Interprétation paralinguistique du signal audio

Dialoguer entre humains va au-delà de la simple compréhension linguistique. En effet, un message transmis n'est pas qu'une suite de mots organisés selon des règles grammaticales et sémantiques : il s'agit également d'un ensemble de marqueurs porteurs de sens pouvant être portés par la voix (rire, intonation, rythme, débit...), par les expressions faciales (sourire, grimace...), par le corps en général (posture, haussement d'épaule, mouvement de mains...). Ces marqueurs viennent ajouter un complément d'informations notamment sur l'humeur du locuteur, sa relation sociale, une émotion ponctuelle, ses intentions, ses dispositions affectives, son état d'esprit. L'humain perçoit et interprète ces informations bas-niveau multimodales, qui contribuent à une compréhension globale de l'interlocuteur.

Cette reconnaissance des expressions multimodales chez l'humain fait l'objet de la recherche en détection automatique de signaux sociaux. Notre étude prend appui essentiellement sur la reconnaissance automatique de l'état émotionnel de l'individu à partir du signal de parole (Clavel, 2007; Devillers *et al.*, 2005a; Tahon, 2012; Vidrascu, 2007). Nous retiendrons principalement que l'élaboration d'un module de détection consiste tout d'abord en une analyse automatique d'une base de données de parole orientées sur l'objet à reconnaître automatiquement (une base de données de service après-vente sera plus adaptée pour développer un module de détection de satisfaction client, par exemple).

De cette base de données sont extraits des descripteurs permettant de décrire des phénomènes (l'occurrence d'un mot ou d'un phonème, d'une émotion, la présence de parole ou pas, etc.). En traitement automatique de la parole, et plus particulièrement dans la détection audio des émotions, les indices bas-niveau calculés à partir du signal de parole peuvent être la fréquence fondamentale, l'énergie, des coefficients acoustiques tels que les MFCC, etc. Une sélection des paramètres pertinents et le choix d'un modèle d'apprentissage permettent d'élaborer un modèle acoustique du phénomène, et ainsi de mettre au point un système automatique de classification qui permettra de classer toute nouvelle occurrence de parole (dans la mesure où le contexte est similaire à l'objet étudié pour la conception du modèle).

L'équipe « Dimensions affectives et sociales dans les interactions parlées », dirigée par Laurence Devillers au LIMSI-CNRS, a développé un module de détection de l'état émotionnel du locuteur. Ce module fait l'objet d'une amélioration et d'une adaptation constantes. La conception de l'outil se base sur des corpus de données, ces données étant scindées en segments temporels et étiquetées en termes d'étiquette émotionnelle (telle que la Joie, la Tristesse, la Colère, la Peur, le Neutre), de valence (Positif, Négatif ou Neutre), d'activation (Actif ou Passif). De ces segments sont extraits des descripteurs bas niveau (à différentes échelles temporelles) qui permettent de constituer des modèles statistiques susceptibles de classer automatiquement un segment selon son étiquette émotionnelle, sa valence, ou son activation. Nous retrouvons une étude des descripteurs extractibles de nos corpus d'interaction humain-humain et humain-machine notamment dans le manuscrit de thèse de Marie Tahon (Tahon, 2012).

Comme le présente la Figure 10, des marqueurs paralinguistiques de différents niveaux peuvent donc être calculés sur le signal audio, et peuvent fournir à un système des informations telles que l'identité du locuteur, sa tranche d'âge, des informations émotionnelles (quelle émotion, avec quelle force, quelle polarité), des indices sur sa loquacité (un rapport parole sur silence dans le cas où il est attendu de la parole, ou juste une durée de temps de parole), ou encore sur les interruptions dans son discours (les disfluences).

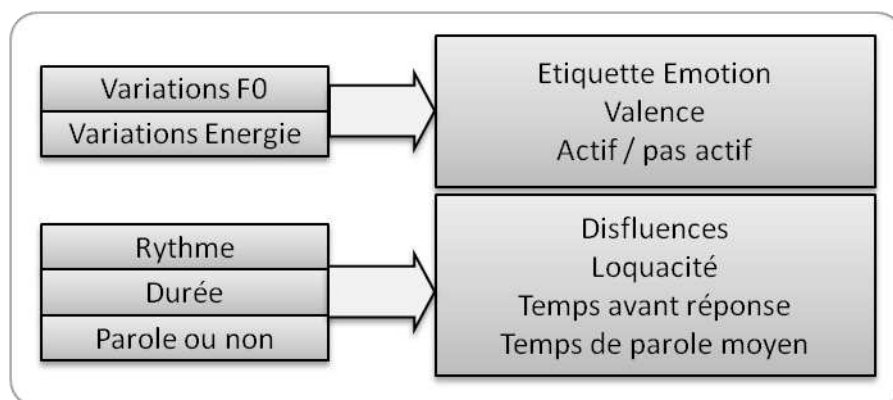


Figure 10. Quelques indices bas-niveau calculés sur le signal audio et interprétation en informations paralinguistiques

Le module de détection développé par le LIMSI et utilisé dans le cadre de notre étude nous permet d'obtenir les étiquettes paralinguistiques répertoriées dans le Tableau 5. La colonne « Informations dérivées » présente les informations pertinentes et significatives qu'il est possible de calculer à partir de ces étiquettes, telles que des moyennes, des totaux, etc. La colonne « Valeur » présente les différentes valeurs proposées par le système. Nous notons que l'étiquette émotion « Peur » n'est pas présente pour notre étude. En effet, les modèles sont construits à partir de données naturelles (émotions exprimées en contexte naturel), or la collecte d'un corpus de locuteurs exprimant de la peur spontanée est difficile à obtenir. Nous nous sommes donc limités aux étiquettes Colère, Joie, Tristesse et Neutre pour lesquels nous possédions suffisamment de données naturelles.

Il est à noter que les informations obtenues par le module de détection sont soumises aux scores de détection. Certains paramètres influent sur la réussite ou l'échec de la classification, que cela soit dû à un modèle peu adapté au signal traité, à des conditions sonores (bruit, échantillonnage), ou des caractéristiques dans la voix même du locuteur (citons par exemple l'étude de (Tahon *et al.*, 2010) sur l'influence de l'âge et du sexe sur les scores de détection).

C'est sur ce contexte perceptif limité que notre système élaborera le profil émotionnel et interactionnel du locuteur. Ces informations pourraient être complétées par une connaissance du contexte linguistique, gestuel, postural du locuteur, mais également une perception des événements survenant lors de l'interaction. Comme nous le détaillerons plus loin, les seuls éléments contextuels autres que nous exploitons sont le comportement actuel du robot et l'état de la tâche en cours (quand pertinent).

Tableau 5. Étiquettes paralinguistiques obtenues via le système de détection du LIMSI

Étiquette	Description	Valeurs	Informations dérivées
Parole	Lorsque de la parole spécifiquement est détectée	Booléen	Nombre de tours de parole Durée du tour de parole Moyenne de la durée des tours de parole
Silence	Lorsqu'un silence est détecté (une baisse d'énergie significative)	Booléen	
Fin de parole	La fin de parole a été détectée	Booléen	
Émotion	Une étiquette émotion parmi quatre	« Neutre », « Colère », « Tristesse », « Joie »	-
Valence	Calculée à partir d'un modèle de Valence, ou à partir de l'émotion reconnue	« Positif », « Négatif »	Moyenne et écart-type depuis le début de l'interaction Écart à la moyenne du dernier énoncé
Activation	L'activation reconnue	« Passif », « Actif »	
Locuteur	Le locuteur reconnu	« Locuteur x » parmi des locuteurs connus, ou « Inconnu »	-

3.3.2.2. Interprétation émotionnelle et interactionnelle des informations paralinguistiques

Étudions la relation entre les informations paralinguistiques récupérables depuis le système de détection, et les dimensions du profil. Nous répertorions tout d'abord les différentes informations exploitables selon les dimensions du profil, puis la façon de calculer dynamiquement le profil à l'aide de règles de logique floue.

En ce qui concerne le profil interactionnel de l'utilisateur, rappelons que nous nous baserons, dans notre contexte d'étude actuel, uniquement sur l'Affinité et la Domination. L'Affinité d'un individu A pour un individu B peut être caractérisée, de façon sommaire, par le ton chaud et énergique de A lorsqu'il s'adresse à B. Une façon de le décrire via nos étiquettes paralinguistiques sera de considérer les valeurs et les valeurs moyennes de l'activation et de la valence. Des déclarations positives et fortes de l'utilisateur, lors de son interaction avec le système, auront ainsi tendance à faire augmenter son niveau d'Affinité estimé pour le système. La Domination, ou la prise de contrôle de l'utilisateur sur son interaction avec le système, pourra être déterminée selon que l'utilisateur parle fort, longtemps, de façon sèche. Nous nous intéresserons donc à la présence de valence négative exprimée conjointement à une activation forte, une durée de parole longue, et un grand nombre de tours de parole. Le Tableau 6 référence toutes les informations paralinguistiques nécessaires au calcul des dimensions interactionnelles.

Tableau 6. Dimensions du profil interactionnel et informations paralinguistiques afférentes

Dimension	Informations paralinguistiques exploitées
Affinité	Activation
	Moyenne activation
	Valence
	Moyenne valence
Domination	Nombre de tours de parole
	Durée du tour de parole
	Moyenne des durées de tour de parole
	Activation
	Moyenne activation
	Valence
	Moyenne valence

Le profil émotionnel regroupe les dimensions d'Optimisme, de Confiance en soi, d'Émotionnalité, et d'Extraversion. Nous considérons l'Optimisme comme l'attitude positive du locuteur au cours de son interaction avec le robot. La valence est donc l'indice clé pour l'analyse de cette dimension. Dans le contexte du jeu, nous connaissons un élément contextuel susceptible de consolider cette estimation : si le joueur exprime des émotions positives alors qu'il est en train de perdre (une situation à valence négative, donc), on évalue l'Optimisme du joueur à la hausse. La Confiance en soi indique la propension du locuteur à se sentir à l'aise dans son interaction. Perceptivement, on notera une activation globalement élevée, et l'expression de colère spontanée peut être également un indice. Le degré d'Émotionnalité est représenté par les variations dans les émotions du locuteur. Un locuteur très émotionnel utilisera toute une gamme d'émotions, allant du négatif au positif (valence), du faible au fort (activation). Un locuteur peu émotionnel sera beaucoup plus stable, invariant. L'extraversion caractérise la propension du locuteur à être volubile et à exprimer des émotions intenses. Nous nous pencherons sur les indices portant sur la durée du temps de parole et l'activation. Le Tableau 7 présente toutes les informations paralinguistiques requises pour l'estimation des valeurs du profil émotionnel.

Tableau 7. Dimensions du profil émotionnel et informations paralinguistiques afférentes

Dimension	Informations paralinguistiques exploitées
Émotionnalité	Moyenne valence
	Écart-type valence
	Écart à la moyenne valence
	Moyenne activation
	Écart-type activation
	Écart à la moyenne activation
Optimisme	Valence
	Moyenne valence
Extraversion	Activation
	Moyenne activation
	Durée du tour de parole
	Moyenne des durées de tour de parole
Confiance en soi	Activation
	Moyenne activation

Comme nous l'avons déjà évoqué, la détection de l'état émotionnel du locuteur est une réalisation ponctuelle, qui ne saurait en aucun cas, seule, constituer une représentation dynamique de l'utilisateur sur du moyen à long terme. Gérer le profil dans le temps implique que le locuteur soit connu, c'est-à-dire que le système ait gardé trace des interactions précédentes : qu'il s'agisse des interactions venant immédiatement de se produire, comme celles datant des jours précédents. Comme nous l'avons vu dans le point 2.1.2 sur la mémorisation et la gestion de l'interaction à long terme, il existe différentes façons de traiter les « souvenirs » du système. Dans notre cas, cette mémorisation est gérée par l'agrégation des données en valeurs moyennes. Cette méthode simple nous semble particulièrement pertinente dans notre contexte. Dans une situation de communication humain-humain, nous connaissons tous le phénomène de « première impression », qui conditionne notre première interaction avec l'interlocuteur, plus ou moins fortement selon que l'on désire rester prudent face à cette première impression, ou si au contraire nous agissons de façon plus impulsive. Dans le cas de notre système, l'utilisateur est considéré initialement de façon « neutre » (i.e. représenté par

des valeurs moyennes : ni trop peu optimiste, ni trop optimiste, etc.). Une première déclaration de l'utilisateur aura de grandes chances de faire varier ce profil, mais puisque toutes les dimensions se basent également sur la moyenne de la totalité des déclarations de cet individu, et que cette moyenne variera par nature de moins en moins au fil des interactions, nous parviendrons à un système globalement stable au fil des interactions, qui ne sera plus soumis aux variations ponctuelles de l'état émotionnel de l'utilisateur. Il est donc possible, à partir d'une détection ponctuelle d'états émotionnels, de réaliser un traitement à plus long terme s'approchant d'une « opinion » que le système porte sur l'utilisateur, plus ou moins étayée par la quantité d'interactions selon que le système a beaucoup côtoyé l'utilisateur ou pas, et qui respecte les mécanismes d'agrégation de souvenirs propres à l'homme.

Dans notre système, il serait évidemment intéressant d'ajouter un historique inter-sessions afin de suivre le locuteur d'une session d'interaction à l'autre. C'est un mécanisme que nous n'avons pas eu l'occasion de mettre en œuvre et tester, pour évaluer l'interaction sur un terme bien plus long (de l'ordre de plusieurs jours), mais le fonctionnement et la mise à jour du profil auraient été similaires.

3.4. Conclusion

L'utilisation d'un système « social » implique que ce système soit capable de mener une interaction suivie et intelligente avec un utilisateur humain. Cette interaction requiert la compréhension du message affectif transmis par l'utilisateur, ce qui met en œuvre tout d'abord la notion de détection de l'émotion, puis l'interprétation de cette émotion.

Nous avons présenté dans cette section ce qui selon nous est crucial pour une interaction « intelligente » avec le système : une compréhension de la paralinguistique, et son interprétation à plus long terme en un profil émotionnel et interactionnel de l'utilisateur. Nous avons établi les relations possibles entre chaque indice récupérable en temps réel d'un système de détection des émotions extraits de l'audio, et les différentes dimensions du profil. Un traitement multiniveau des indices paralinguistiques extractibles du signal audio permet de renseigner le profil. Nous présenterons la formalisation en règles de logique floue dans le chapitre 6 de ce manuscrit.

Cette estimation du profil de l'utilisateur nécessite bien évidemment d'être pondérée. Tout d'abord, notons que les règles de constitution du profil ne sont pas absolues. Comme Revelle et Scherer le soulignent par exemple dans (Revelle et Scherer, 2009), un individu sensible aux affects positifs ne peut être automatiquement considéré comme extraverti : il ne s'agit pas d'une preuve formelle de son extraversion, mais plutôt d'un indice prouvant une tendance à l'extraversion, qu'il est nécessaire d'étudier de concert avec les composantes cognitives et comportementales de l'individu. Nous en venons donc à l'importance d'un traitement multimodal pour la constitution d'un profil utilisateur robuste. L'interprétation est le fruit de l'analyse d'un ensemble d'indices, afin de pouvoir confronter différents résultats. Certains types d'indices pourraient faire l'objet d'un travail subséquent à cette thèse, tel que l'ajout d'une compréhension linguistique ou visuelle du locuteur, ou d'autres facteurs inhérents à son environnement (état de fatigue, maladie, degré d'autonomie...).

Pour pouvoir parler d'adaptation à l'utilisateur, le profil émotionnel et interactionnel de l'utilisateur doit venir modifier le comportement du robot. Dans le chapitre suivant, notre réflexion portera sur la détermination des comportements sociaux envisageables dans un système d'Interaction Humain-Machine

sociale, et la relation de ces comportements avec le profil utilisateur : quel comportement adopter en fonction du profil utilisateur ? Quel peut être l'impact du comportement système sur l'interprétation du profil de l'utilisateur ? Nous présenterons les théories sur lesquelles nous appuierons notre réflexion, et proposerons une codification des comportements selon le contexte d'application.

Chapitre 4

4. Comportements sociaux du système

4.1. Introduction

La compréhension du profil social et de personnalité de l'utilisateur humain par la machine doit évidemment avoir un impact sur le comportement du système. En effet, lors d'une interaction sociale « classique », chaque interactant s'adapte à l'autre ; en fonction de son humeur, de ses buts et de ses prédispositions de personnalité, il va ajuster son comportement et ses propos. Cet ajustement doit alors être perceptible par l'utilisateur. Il convient donc de coder les comportements sociaux du système de telle sorte qu'ils soient tout d'abord reconnaissables par l'utilisateur (i.e. qu'ils répondent à certaines normes de comportement social humain), mais également admissibles. Ce dernier point est à notre sens le plus délicat en Interaction Humain-Robot : les attentes en termes de robotique varient selon les utilisateurs comme nous l'avons vu précédemment, et il est difficile de se représenter ce qu'un utilisateur pourra attendre d'une machine « pensante », ayant des intentions de communication émotionnelle, et susceptible de générer des liens et des sentiments, lorsque cette machine sera *véritablement* devenue son compagnon à domicile. Pour faire bref, une interaction qui sera jugée divertissante et adaptée en conditions d'expérimentation pourra être inadaptée, voire tout à fait dérangeante, une fois le compagnon installé à demeure.

S'il nous est encore à l'heure actuelle impossible de nous projeter entièrement dans ce que sera l'interaction à domicile, au quotidien, entre un utilisateur et son robot, il nous semble envisageable toutefois de déterminer et de typer les comportements sociaux pouvant potentiellement être acceptables par l'utilisateur. Il nous faut donc déterminer deux éléments : quelle est la finalité recherchée par le robot, et comment peut-il l'exprimer. Pour le premier élément, nous nous interrogeons alors sur les stratégies de comportement que le robot pourra adopter. Nous étudierons tout d'abord les théories de la communauté sur lesquelles nous appuyons notre choix en termes de comportements sociaux.

Nous présentons ensuite nos propres analyses quant au lien entre les comportements du robot et le profil, en fonction du contexte. Notre sélection des modes d'expressions possibles de ces comportements sera également proposée.

4.2. Stratégies de comportement du système dans la littérature

4.2.1. Détermination des comportements sociaux du système

Lors de la définition des comportements sociaux du système, et globalement des normes permettant l'interaction avec son environnement, il convient de discerner un premier grand axe : le système sera-t-il hostile (machines de combat), neutre (observation, tâches mécaniques), ou amical (compagnon, assistant) ? Nous nous situons dans le cadre d'une interaction avec un robot compagnon social, il est donc tout naturel de déterminer les intentions du système comme étant délibérément bénéfiques pour l'utilisateur. La finalité

recherchée par le robot, en termes méta, sera donc d'apporter de la satisfaction à l'utilisateur. Les moyens pour y parvenir, cependant, constituent évidemment un défi de recherche.

Un robot doté d'une personnalité, tel que décrit dans l'étude de Goetz et Kiesler (Goetz et Kiesler, 2002), a un impact sur la représentation mentale que l'utilisateur se fait du robot. Leur étude montre que l'apparence du robot, ainsi que son discours, affectent l'attitude de l'utilisateur, ainsi que son implication dans l'interaction avec le système. Goetz et Kiesler proposent des études mettant en scène des robots jouets (assemblage mécanique non anthropomorphe, ou grossièrement anthropomorphe), et un robot assistant social expressif. Ces robots sont présentés à l'utilisateur comme étant soit joueurs, soit sérieux. Les résultats de leur étude montrent qu'un robot joueur et extraverti engendrera chez l'utilisateur une forte représentation mentale positive, mais tendra moins à stimuler la coopération de l'utilisateur qu'un robot sérieux. Une autre étude intéressante sur l'impact du comportement du système, menée sur le robot Aibo (Batliner *et al.*, 2004), montre qu'un robot qui désobéit aux ordres de l'utilisateur aura, comme l'on peut s'y attendre, un impact sur les expressions émotionnelles de l'utilisateur. Ce comportement, s'il est intéressant pour l'expérimentation, n'est toutefois pas envisageable dans le cadre d'un robot compagnon au quotidien, pour lequel nous attendons l'accomplissement de tâches prédéfinies. Si ce comportement peut s'avérer ponctuellement amusant, voire probablement stimulant dans le cadre d'un robot à visées thérapeutiques, il serait tout à fait contreproductif sur le long-terme. Le comportement général du système a donc un impact direct sur l'attitude de l'utilisateur, et la sélection de ce comportement doit être effectuée avec soin, en fonction des intentions que l'on souhaite prêter au robot afin d'améliorer la relation de l'utilisateur (comme le note Breazeal dans (Breazeal, 1999)).

Nous avons abordé dans la section précédente les principes de réciprocité et de correspondance intervenant dans la théorie interpersonnelle (Carson, 1969). Il s'agit de mécanismes d'adaptation sociale, innés et dépendants des buts de l'interactant, ainsi que de son humeur et de sa personnalité. Le principe général consiste en ce que la domination exercée par un individu A sur un individu B entraînera potentiellement la soumission de B envers A, et que l'amicalité de A envers B attirera potentiellement l'amicalité de B envers A (de même pour l'hostilité). Nous resituons ces pôles de domination et d'amicalité sur le circomplexe interpersonnel dans la Figure 11, en nous focalisant sur les comportements amicaux. En effet, comme nous l'avons noté précédemment, nous travaillons sur un système pour lequel il est inconcevable de déterminer des comportements hostiles. Nous reconnaissons donc quatre îlots de comportements potentiels pour le système : directif, aidant, coopératif, dépendant.

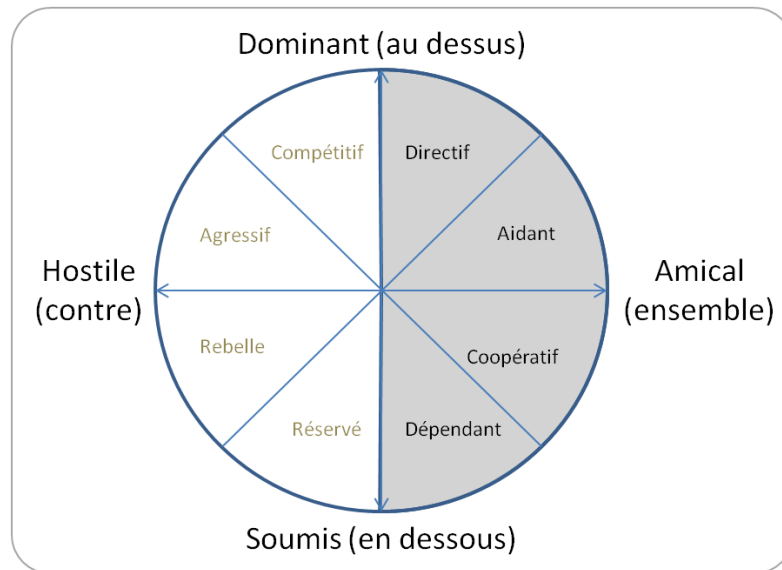


Figure 11. Comportements sociaux envisageables pour un système d'interaction Humain-Machine. Adapté de (Isbister, 2006).

Nous notons que, selon l'étude menée par (Ray *et al.*, 2008) sur les attentes des utilisateurs en matière de robotique, une des attentes principales est de pouvoir garder le contrôle sur le robot, qu'il n'ait pas une autonomie trop prononcée. Nous notons également que dans le cas de personnes souffrant d'une perte d'autonomie, il importerait de laisser au système un degré d'autonomie dans la réalisation des tâches (Dautenhahn et Werry, 2002; Tapus et Mataric, 2006). Un comportement non désirable pourrait donc être représenté par un robot qui impose sa façon de considérer les événements (Directif). Nous estimons toutefois que ce choix peut être dépendant du contexte applicatif dans lequel le système est invoqué, et des caractéristiques émotionnelles de l'utilisateur. Nous évaluerons ce point au cours de ce travail de thèse.

4.2.2. Relation entre comportement social du système et profil utilisateur

Nous nous interrogeons sur la relation entre le profil de l'utilisateur et la sélection de comportements. Pour cela, nous devons considérer deux aspects de l'interaction entre l'humain et le système. Tout d'abord, comme dans toute interaction classique, ce que nous disons, et la façon dont nous le disons, ont un impact sur le comportement de notre interlocuteur. D'autre part, l'attitude de l'interlocuteur peut venir avoir un impact sur notre propre attitude. Considérer cette double composante est essentiel dans la gestion d'une interaction entre un humain et un système.

L'adaptation du comportement du système en fonction des caractéristiques utilisateurs est une composante essentielle du champ de recherche en Interaction Humain-Robot. L'adaptation peut s'appuyer, par exemple, sur un processus d'imitation fondé sur les mécanismes de l'Empathie : nous pouvons citer notamment les travaux de (Tapus et Mataric, 2007) et (Duhaut, 2010). L'adaptation peut également être réalisée à partir d'une compréhension de la personnalité de l'utilisateur, comme dans le cas du système présenté par (Aly et Tapus, 2013). Dans ce système, une interprétation des traits de personnalité (introverti, extraverti) est effectuée d'après une reconnaissance lexicale de la part de l'utilisateur. En fonction des traits de personnalité reconnus, le système génère des phrases relatives à ces traits de personnalité, en partant de l'hypothèse que l'utilisateur préfère interagir avec un robot qui présente une personnalité similaire à la sienne.

Dans notre cas, le robot n'est pas doté d'une personnalité, mais d'une attitude sociale vis-à-vis de l'utilisateur. Nous partons plutôt de l'hypothèse que la théorie de la réciprocité détaillée dans les sections précédentes sera plus pertinente pour la sélection de comportements robotiques.

Vouloir considérer l'impact du comportement exprimé par le robot sur le profil de l'utilisateur revient à estimer que ce comportement a eu un effet sur les tendances émotionnelles du locuteur, qu'il peut venir le moduler. Cela pourrait éventuellement être envisageable si l'on considère la théorie des faces de (Goffman, 1974), selon laquelle l'individu réagit plus ou moins bien selon que son interlocuteur respecte l'image qu'il a de lui-même ou pas. Pour éviter de générer du mal-être dans une interaction, il convient donc de veiller à une bonne adaptation à l'autre, afin que ce dernier puisse « garder la face ».

Ainsi, si le comportement du robot face à l'utilisateur est en adéquation avec l'image que ce dernier a de lui-même, cela peut nous permettre de gérer son bien-être. L'adéquation (ou la non-adéquation) va impacter l'expression émotionnelle du locuteur, et peut-être son attitude sociale (affinité et dominance), mais pas sa personnalité. En effet, les traits de personnalité sont des constantes qui varient peu dans le temps et en fonction des circonstances. L'inadéquation du comportement pourrait alors juste faire fluctuer temporairement l'expression de cette personnalité, à savoir les émotions. Il faut donc se pencher sur l'impact des comportements sur les manifestations émotionnelles de l'utilisateur.

4.3. Nos travaux : Codification des comportements sociaux du système

Outre l'adaptation à l'utilisateur, la sélection de comportements sociaux doit donc être fonction du contexte. En effet, l'objectif recherché lors d'une interaction de type « assistance » n'est pas le même que lors d'une interaction de type « jeu ». Dans un premier cas, l'objectif à atteindre sera l'efficacité, tandis que dans l'autre ce sera prioritairement la distraction. Nous estimons donc que les comportements sociaux doivent varier également en fonction du contexte applicatif.

D'autre part, nous avons vu au cours de ces travaux que dans un système robotique émotionnel traditionnel, l'état interne du robot peut impacter la sélection du comportement à jouer. Cependant, dans le cadre du projet Romeo, NAO n'est pas doté d'un système émotionnel interne complexe : même s'il sait mimer des émotions par la voix ou par la posture (modules développés par Aldebaran Robotics), son système de gestion et d'actualisation de ses émotions internes est à l'heure de nos travaux toujours en développement. Il était pourtant impératif pour nous que les émotions exprimées par l'utilisateur, ainsi que son profil émotionnel et interactionnel, aient un impact visible sur le comportement système, afin de représenter l'adaptation du système à l'utilisateur. Nous nous interrogeons alors sur la façon dont ces comportements peuvent être représentés.

4.3.1. Comportements sociaux en fonction du contexte

En Communication Humain-Machine, on considère qu'un système est efficace et convivial si l'utilisation de ce système remplit plusieurs critères. Tout d'abord, le système doit accomplir la tâche qui lui est dévolue. Si nous parlons par exemple d'un système de réservation en ligne, on attend évidemment qu'il réalise la réservation en fin de manipulation. D'autre part, l'utilisation du système doit être intuitive et ergonomique. Le système doit « aider » l'utilisateur à l'utiliser, que ce soit par des conventions informatiques connues telles que des menus, ou encore par des sons, des voyants lumineux, des explications orales. Enfin,

si le système est destiné à réaliser une interaction sociale avec l'utilisateur, nous attendons qu'il partage les codes de communication interpersonnelle propres à l'homme pour fluidifier l'interaction.

Comme nous l'avons vu dans la section 4.2.1, les comportements du robot peuvent être exprimés sur un circomplexe « Amicalité – Domination ». Dans une situation d'assistance robotique à la personne plus précisément, le comportement du robot doit se situer dans un domaine exclusivement amical. Le niveau de domination exercé par le système sur l'utilisateur doit être fonction des tendances émotionnelles et interactionnelles exprimées par l'utilisateur. Nous avons localisé quatre zones pouvant potentiellement délimiter les comportements sociaux désirables de la part d'un système d'assistance (voir Figure 12, illustration de gauche) :

- 1) Directif : il s'agit pour le système de gérer l'interaction, de prendre des décisions. Cela peut être envisageable dans le cas d'un utilisateur en perte d'autonomie, dans le cas d'une urgence, mais également pour les individus préférant être guidés ;
- 2) Aidant (la stratégie « Encourageant » que nous utiliserons plus tard fait partie de ce domaine) : cette attitude consiste à prendre une position plus dominante sur l'autre, puisqu'elle peut faire intervenir une part de décision, ou du moins de conseils, tout en actant cependant dans un but amical ;
- 3) Coopération (« Empathique » fait partie de ce domaine) : il s'agit de s'adapter à l'utilisateur pour réaliser un but commun ;
- 4) Dépendant (« Soumis », « Hésitant » font partie de ce domaine) : cette attitude sociale peut être envisageable dans les cas où l'utilisateur ne désire pas du tout être dirigé, mais préférera garder la mainmise sur le système. Dans l'absolu, cette attitude pourrait également être souhaitable dans le cas d'un utilisateur déprimé, qui aurait alors satisfaction à venir apporter son aide, mais nous n'envisagerons pas ce cas de figure complexe.

Dans le cas du robot en tant que compagnon de jeu, la finalité est de distraire l'utilisateur. Dans ce contexte applicatif, nous estimons que l'utilisateur n'aura pas les mêmes attentes sociales. S'il pourra toujours potentiellement être aidé, il n'appréciera peut-être pas (ou n'aura pas besoin de) une prise de contrôle trop directe du système. Un système soumis, dépendant, n'est pas non plus envisageable. Nous distinguons donc deux aires potentiellement désirables :

- 1) Aidant (« Encourageant ») : le système peut venir aider l'utilisateur, lui prodiguer des conseils au cours du jeu ;
- 2) Coopératif (« Empathique ») : le système ne vient pas forcément aider le joueur en termes de réussite de jeu, mais il lui fournit un appui affectif au cours de l'activité.

Nous avons localisé ces deux aires sur le circomplexe (Figure 12, illustration de droite). Dans le cadre du jeu, nous avons jugé bon d'ajouter un comportement « Amusant », qu'il est difficile de situer sur le circomplexe. En effet, une situation humoristique est très souvent créée via un écart entre ce qui est attendu, et ce qui est réellement obtenu. De nombreuses études se sont penchées sur l'utilisation de l'humour dans les systèmes (notons par exemple les travaux de la communauté HUMAINE (Strapparava *et al.*, 2011), et les ateliers sur l'humour computationnel (Nijholt, 2012), desquelles ressort principalement le fait que l'humour

est un procédé de communication qui joue sur les croyances et attentes de l'individu. Ces études nous renseignent également sur le fait que l'humour est générateur de réactions émotionnelles et d'implication dans l'interaction. Bien que la gestion de l'humour soit difficile, et puisse produire des réactions contraires à celles escomptées (de l'Ironie, ou de l'Agacement, de l'incompréhension), l'humour doit être une composante de l'interaction ludique.

Nous n'avons cependant pas opté pour l'utilisation de la stratégie « Amusant » dans le cas des personnes en situation d'assistance, car, bien que cette stratégie nous semble tout à fait pertinente, sous bon contrôle, avec des personnes accessibles à l'amusement, ce n'est à notre sens pas sur ce type de stratégies que nous devons prioritairement mettre le focus dans le cadre d'une assistance à personnes en perte d'autonomie.

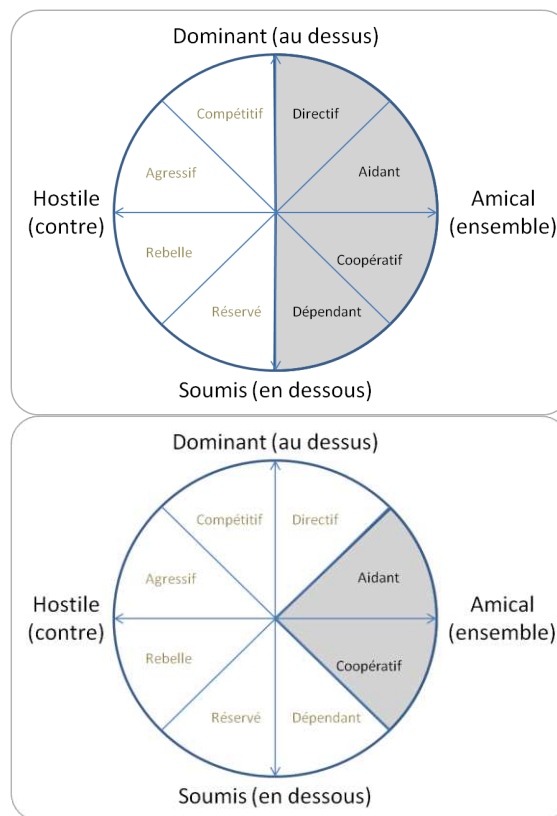


Figure 12. Délimitation des comportements sociaux « désirables » (gris foncé) en fonction du contexte applicatif (assistance à l'utilisateur en haut, compagnon de jeu en bas). Adapté de (Isbister, 2006).

4.3.2. Comportements sociaux en fonction du profil utilisateur

Notre considération de l'attitude sociale tant chez l'utilisateur, que chez le système, étant basée sur les théories du circomplexe Domination-Amicalité, nous estimons alors que la sélection du comportement social du robot doit s'appuyer sur une adaptation à l'utilisateur mettant en œuvre le principe de réciprocité basé sur la domination : plus l'utilisateur sera soumis, i.e. montrera les signes caractéristiques et détectables de la soumission, plus le système prendra le contrôle de l'interaction. La Figure 13 présente cette relation entre le niveau de soumission/domination de l'utilisateur, et les comportements robotiques afférents.

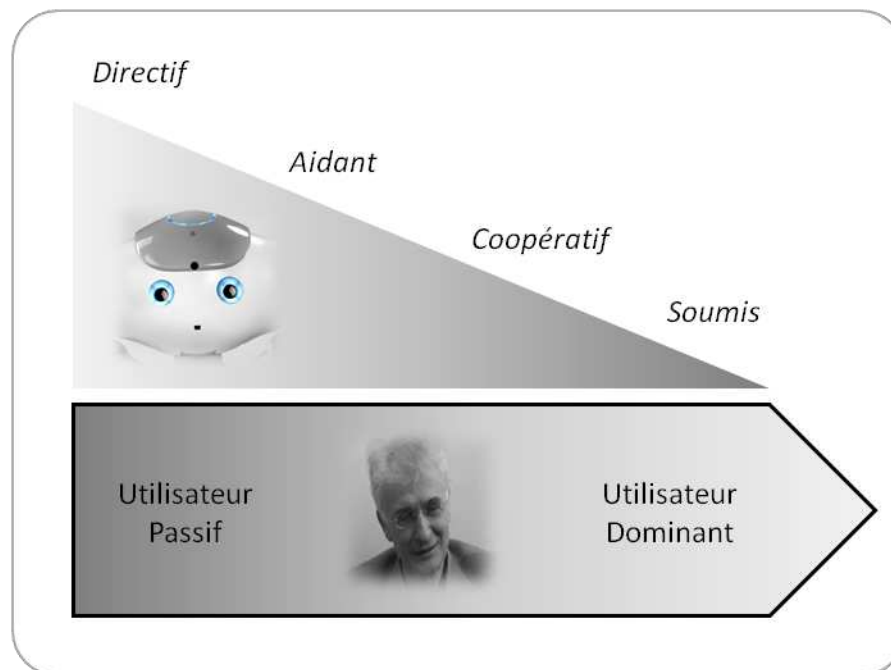


Figure 13. Comportements sociaux du système sélectionnés (du plus dominant au moins dominant) en fonction du niveau de domination exprimé par l'utilisateur

Le degré de soumission et de domination de l'utilisateur, dans notre étude, ne sera pas réduit à la notion telle que décrite dans la section précédente. Comme nous l'avons vu en effet, le degré de domination s'exprime par des phénomènes d'expression qui sont également liés à la personnalité de l'utilisateur. Notons par exemple que l'extraversion (parler fort, longtemps), la stabilité émotionnelle (la capacité de l'individu à contrôler ses émotions) seront autant d'indices permettant de déterminer le degré de domination que l'individu veut, et peut, instaurer dans sa relation avec l'entité robotique.

Le Tableau 8 ci-dessous récapitule les tendances de sélection des différents comportements sociaux du système, en fonction des valeurs du profil émotionnel et interactionnel de l'utilisateur. Ces tendances ont été déterminées à l'aide de références bibliographiques consacrées à la relation entre personnalité de l'utilisateur et l'attitude sociale du système, et sur les théories sociales humain-humain, mais également par les observations réalisées dans le cadre d'expérimentations en contexte (décrites dans le chapitre 5).

Ces tendances de sélection sont présentées en termes de « plus cette composante est forte (+ ou ++), plus ce comportement aura tendance à être sélectionné », et identiquement dans le cas « plus cette composante est faible (- ou - -), plus ce comportement aura tendance à être sélectionné ».

Tableau 8. Récapitulatif de l'impact du profil émotionnel et interactionnel de l'utilisateur sur la sélection du comportement social du système.

Utilisateur		Robot
Profil émotionnel	Profil interactionnel	Comportement social
Confiance en soi - - Emotionnalité +	Domination - - Affinité +	Directif
Confiance en soi - Optimisme -	Domination -	Encourageant
Confiance en soi + Optimisme + Extraversion +	/	Amical/Aimable
Confiance en soi + Emotionnalité +	Domination +	Empathie
Confiance en soi + + Extraversion + Emotionnalité +	Domination + + Affinité -	Soumis/Hésitant

Ces tendances sont très liées à notre contexte d'application et à notre échantillon d'étude, et ne peuvent avoir valeur de généralité, de par la complexité du fonctionnement humain. Elles représentent néanmoins une base de réflexion qui pourrait être, selon le contexte d'application et la finalité recherchée, améliorée par des experts en psychologie pour adapter un système d'interaction émotionnelle à différents contextes.

4.3.3. Expression du comportement social

Comme nous l'avons vu dans la section 4.2.1, le comportement du système (ce qui inclut l'expression de sa personnalité ou de ses désirs, ou son comportement social en général) a un impact sur le comportement et l'opinion de l'utilisateur humain. D'autre part, les préférences de l'utilisateur au sujet du comportement de la machine sont également liées à ses caractéristiques émotionnelles et de personnalité. Enfin, l'impact des comportements du système peut dépendre du contexte d'application. Les comportements que l'interface (robot, agent conversationnel, etc.) doit adopter sont donc étroitement liés à la tâche à accomplir, ainsi qu'aux caractéristiques propres à l'utilisateur.

En termes de communication Humain-Robot, au-delà des attentes fonctionnelles de l'utilisateur, c'est-à-dire que le robot réalise la tâche qui lui est dévolue (assistance quotidienne, jeu, enseignement...) et que son

utilisation puisse se faire de façon instinctive et réponde aux critères d'ergonomie, un robot social se doit de partager certains des codes de communication interpersonnelle humaine (Duhaut et Pesty, 2011; Feil-Seifer *et al.*, 2007) afin de répondre efficacement aux messages de l'utilisateur. Ainsi, le robot pourra être plus à-même d'instaurer et d'entretenir une relation naturelle et socialement acceptable.

Par ailleurs, il est essentiel de prendre en compte que, afin que l'interaction ne soit pas rédhitoire (comme cela pourrait être le cas dans le cadre de l'utilisation de nouvelles technologies par des personnes âgées), il est essentiel d'analyser les préférences en termes de modalités de communication entre l'interface et l'utilisateur (Granata *et al.*, 2010). Outre la sélection du comportement social en lui-même, il convient alors de s'intéresser à la façon dont ce comportement sera exprimé.

En conséquence, il semble essentiel dans notre présente étude de sélectionner soigneusement la stratégie de comportement du robot comme nous l'avons vu en 4.2, mais également de choisir soigneusement la modalité d'expression de ce comportement.

4.3.3.1. En situation d'assistance à personnes en perte d'autonomie

Dans le cadre du robot assistant à domicile, dans le projet ROMEO, nous avons travaillé avec des personnes souffrant d'une perte d'autonomie. Ces personnes étaient déficientes visuelles, de totale à partielle. Si certaines étaient malvoyantes depuis leur naissance, d'autres avaient connu une dégénérescence plus tardive due à l'âge, la maladie ou un accident. Les impératifs liés à ce public ont donc conditionné notre approche de la codification de l'expression des comportements sociaux du système : il est évident qu'un système de communication basé sur la gestuelle du robot ou sur des codes lumineux serait inefficace. Le robot NAO avec lequel nous avons travaillé n'était pas doué de parole expressive à l'heure de nos premières études, c'est pourquoi nous ne nous sommes pas attardés sur la gestion des émotions exprimées par le système, mais avons concentré nos efforts sur les marqueurs lexicaux propres à communiquer l'attitude sociale du robot par rapport à l'utilisateur.

Nous nous sommes appuyés sur les théories linguistiques de l'analyste P. Charaudeau. Ses théories onomasiologiques, présentées dans l'ouvrage *Grammaire du Sens et de l'Expression* (Charaudeau, 1992), sont encore aujourd'hui largement exploitées, notamment en Traitement Automatique des Langues, comme par exemple pour la détection de l'opinion (Clavel *et al.*, 2012), ou la détection de la subjectivité dans les textes (Vernier *et al.*, 2009). La classification de Charaudeau propose cinq catégories de modalités évaluatives du langage : l'opinion, le jugement, l'appréciation, l'accord (ou désaccord), et l'acceptation (ou refus). Chacune de ses modalités est liée à un positionnement du locuteur par rapport au contexte ou à l'interlocuteur, et aux différents marqueurs linguistiques associés.

Charaudeau distingue trois types d'actes locutifs (« un certain nombre d'actes énonciatifs de base qui correspondent à une position particulière [...] du locuteur dans son acte de locution ») : l'acte allocutif, qui implique le locuteur et l'interlocuteur, l'acte élocutif, qui se limite au rapport du locuteur à son propre propos, et l'acte délocutif, qui précise la manière dont le propos existe en tant que tel et s'impose aux interlocuteurs. (Charaudeau, 1992) propose des modalités allocutives (« sous-catégories ») spécifiant chacun de ces actes : par exemple l'injonction, l'interrogation, l'interdiction sont considérés comme des actes allocutifs ; l'opinion, le constat, l'appréciation, etc. sont élocutifs ; l'assertion, le discours rapporté tiennent de l'acte délocutif.

Nous distinguons donc dans nos stratégies de comportement (Encouragement, Empathie, Amabilité, Neutre) si celles-ci impliquent que l'on se réfère : uniquement au locuteur, à ses pensées et opinions (élocutif), ou bien que l'on inclue également l'interlocuteur, afin de l'amener à réaliser un but commun par exemple, ou qu'on le juge (allocutif), ou bien si les qualités (i.e. caractéristiques) du locuteur et de l'interlocuteur n'ont aucune importance dans l'acte de communication (délocutif).

L'Encouragement suppose que le locuteur implique lui-même et l'interlocuteur dans son discours, en cela qu'il tente de convaincre l'interlocuteur que son avis est celui à suivre pour atteindre un but considéré comme positif. Le fait que le locuteur soit ici un robot nécessite que la stratégie d'encouragement soit progressive : les participants n'attribueront pas tous au robot un même degré d'humanité dans son raisonnement et ses connaissances des préoccupations humaines, c'est pourquoi un avis très tranché de la part du robot a de fortes chances d'être ignoré ou méprisé.

Dans le cas d'un contexte d'assistance à domicile, le but à atteindre concerne directement l'utilisateur, puisqu'il s'agit pour lui d'aller mieux, ou de faire perdurer son bien-être. Nous décomposons l'énonciation en deux parties, en renforçant graduellement le lien entre le robot et l'humain : le robot donne un conseil pour que la situation soit plus favorable au participant (élocutif), puis exhortera ce dernier à prendre courage (allocutif) dans le cas de situations négatives (une urgence, un mal-être). Dans le cas d'une situation déjà profitable à l'utilisateur (bonne nouvelle, en forme), le robot incite l'utilisateur à profiter de cette situation favorable (élocutif, puis allocutif).

Le robot peut prodiguer ses conseils sous forme d'avis, d'exhortation indirecte. Dans ce cas, il exprime une Opinion de type « Certitude forte » (« Je pense que ») plutôt qu'une Proposition (« Tu devrais »). La locution qui doit suivre cette proposition montre la progression dans la relation humain-robot initiée par le robot : d'une part, par la prise en compte de l'interlocuteur dans son discours (dans la première phrase, même si la référence est l'humain, le robot ne fait qu'exprimer son opinion), et d'autre part le robot enjoint plus directement, plus impérativement le participant à se lier à son avis.

Dans le cas de l'Empathie, nous attendrions un acte allocutif, en cela que l'acte même d'empathie implique une prise en considération forte des qualités visibles de l'interlocuteur. Nous adoptons plutôt la notion d'empathie cognitive, plus réaliste de la part d'un robot qui n'est pas doué d'émotions naturelles, mais régi par des règles. L'interlocuteur est pris en considération dans le sens où son état est la cause du sentiment du locuteur, mais c'est cet état en lui-même qui est jugé, la relation à l'interlocuteur est donc indirecte (je ne suis pas content parce que c'est toi spécifiquement, je suis content parce que ta situation « en pleine forme » évoque pour moi des choses positives). Nous nous orientons donc vers un lexique se rapportant à un acte élocutif.

Lorsque le robot se comporte de façon simplement Aimable, il ne prend pas de position marquée en termes de domination, et s'attache simplement à se montrer amical et respectueux. Afin de montrer son intention amicale, il ne peut pas faire que constater l'état de santé du participant, nous choisissons donc de nuancer le Constat (acte élocutif, indiqué par le verbe) avec une marque d'Opinion (acte élocutif, indiqué ici par un adverbe ou un mot fortement connoté émotionnellement, e.g. « en détresse »). Cette évaluation permet de montrer l'intérêt porté par le robot à son interlocuteur, sans toutefois être trop intrusive.

Tableau 9. Relations entre actes locutifs et rôles sociaux du robot en fonction des différentes stratégies de comportement

Comportement sélectionné	Faut-il mettre en avant/établir la relation entre le locuteur (robot) et l'interlocuteur (humain) ?	Acte locutif relatif à la relation établie	Modalité énonciative	Relation sociale du robot à l'humain (circomplexe)
Encouragement	OUI « Je formule une opinion sur la façon de profiter de/ améliorer cet état puis je formule une suggestion »	Élocutif puis Allocutif	Opinion puis Suggestion	Protecteur (Dominant +, Amical ++)
Empathie	OUI « J'ai des sentiments, et je comprends les tiens »	Élocutif	Appréciation	Coopératif (Dominant -, Amical ++)
Amabilité	OUI / NON « Je suis agréable avec toi, mais je ne cherche pas à décider pour toi. »	Élocutif et Allocutif	Constat- Opinion	Aimable (Dominant 0, Amical +)
Neutre	NON « Je ne prends pas en compte le degré d'humanité de chacun d'entre nous deux. »	Délocutif	Assertion	Neutre (Dominant 0, Amical 0)
Directif	OUI « Je considère avoir de l'autorité sur toi, tu n'as pas le choix. »	Allocutif	Injonction	Directif (Dominant ++, Amical 0)
Hésitant	OUI « Je te prends en compte, mais j'insiste sur mon incertitude »	Élocutif	Opinion (faible)	Soumis (Dominant --, Amical 0)

Dans le cas d'une stratégie de comportement Neutre, le robot ne se positionne ni par rapport à son interlocuteur, ni par rapport à son propre discours. Il ne fait qu'énoncer un état de fait, un accomplissement de la tâche, une observation. Cette stratégie ne suppose aucune implication émotionnelle ou sociale, et s'exprimera à l'aide d'un lexique propre à un acte délocutif, tel qu'une Assertion.

Si l'on désire que le robot apparaisse comme Directif, nous pouvons choisir une modalité allocutive d'Injonction : (Charaudeau, 1992) définit que dans ce type d'énoncés, le locuteur considère son interlocuteur comment n'ayant pas d'alternative pour l'exécution de la tâche demandée ; il se donne un « statut de pouvoir (autorité absolue) ».

Le comportement Hésitant peut être considéré comme une mise en avant d'une Opinion faible de la part du robot, introduite par « Je crois », « Il me semble », et renforcée par une description de son hésitation via des adverbes et adjectifs (« pas sûr », « peut-être »).

Le Tableau 9 résume la relation entre les actes locutifs et les rôles sociaux du robot en fonction des différentes stratégies de comportement. Nous pouvons trouver un exemple d'application de cette sélection linguistique en Annexe I, dans la section Expérimentation IDV-HR : énonciations du robot NAO.

4.3.3.2. En situation de jeu (sujets autonomes)

L'expression de l'attitude sociale d'un robot est réalisable selon les modalités disponibles. Comme nous l'avons vu dans la section précédente, dans le cas d'un public malvoyant, il est possible de diffuser le comportement social de ce robot par le biais d'un fort support lexical. Si le sujet ne présente pas de déficiences visuelles, un robot doué d'un visage expressif, comme nous le voyons sur les robots Reeti et Nexi dans la Figure 14, peut permettre de présenter une large gamme d'expressions émotionnelles, voire d'intentions et d'attitudes sociales, de par le grand nombre de degrés de liberté pour chacun des éléments constituant leur visage (mâchoire, paupières, globes oculaires, etc.).



Figure 14. Exemples de robots à visages expressifs : expressions émotionnelles du robot Reeti (Robopec) et du robot Nexi⁷ (MIT).

Le robot NAO sur lequel nous avons basé notre étude ne possède pas de visage expressif. Si nous considérons le contexte du robot jouant avec un public d'enfants (ou plus largement de sujets non déficients visuels), il est toutefois possible d'exprimer le comportement social via la posture et les gestes du robot, en plus de la codification lexicale présentée dans la section précédente.

Des études menées sur le robot NAO par les équipes d'Aldebaran Robotics (France) et de la School of Computer Science de l'université d'Hertfordshire (UK) (Beck *et al.*, 2010) ont permis de définir un « *Affect Space* » d'après les études de (Breazeal, 2002) définissant une large gamme d'expressions émotionnelles basées

⁷ Illustrations tirées de "*Reeti and his expressions*", de la Robopec (<http://www.reeti.fr>), et l'article « *Head and Face* » à propos de Nexi sur le site du MIT (<http://robotic.media.mit.edu/projects/robots/mds/headface/headface.html>).

sur un continuum d'*Arousal*, *Valence* et *Stance*. Cet *Affect Space*, sur le robot NAO, lui permet d'exprimer un panel étendu d'expressions émotionnelles complexes via la posture corporelle.

La posture de NAO peut être décrite à l'aide de trois continuums :

- Joie à Tristesse
- Excitation à Colère
- Fierté à Peur

Une valeur médiane sur ces trois axes est interprétée comme une posture « Neutre ».

Comme le notent (Beck *et al.*, 2010), réaliser l'animation d'une entité robotique de façon fluide et convaincante est un véritable challenge, et ils limitent leur étude de l'*Affect Space* aux simples postures statiques du robot. Cette constatation est renforcée par les théories de (Mori, 1970) sur l'*Uncanny Valley*, qui indiquent que le risque qu'une entité artificielle inspire le malaise augmente si l'entité est en mouvement.

Nous avons décrit les différents comportements sociaux de NAO en fonction de ces trois continuums, en partant de règles simples. Tout d'abord, pour l'Empathie, nous partons du postulat que le but du joueur est de réussir le jeu. En cas de réussite, nous nous imaginons donc qu'il éprouvera une émotion positive, et une émotion négative dans le cas inverse (Ochs *et al.*, 2012). L'Empathie pour une situation d'échec sera matérialisée par une grande proportion de Tristesse, tandis que la réussite sera associée à de la Joie.

Pour l'expression de l'Encouragement, nous partons d'une étude menée lors de l'annotation perceptive par notre équipe d'une collecte de données actées (Rollet *et al.*, 2009). Pour l'analyse de ces données, le jeu d'étiquettes d'annotation disponible ne permettait pas de renseigner l'Encouragement, qui était un phénomène récurrent dans le corpus. Les annotateurs s'étaient alors penchés sur le mélange « Positif – Énervement », arguant que l'intention de communication était positive, mais que le dynamisme requis pour encourager une personne pouvait, au mieux, s'apparenter à un énervement. Les dimensions proposées pour la description de l'état affectif de NAO proposant un continuum d'excitation positive (Excitation) à négative (Colère), nous associons donc l'Encouragement à une expression d'Excitation et de Joie.

La posture du robot, par défaut debout, bras le long du corps un peu écartés, est plutôt amicale qu'hostile. Le comportement Amical, qui est un comportement social relativement neutre, pourra être exprimé par une faible proportion de Joie de la part du robot, voire simplement par une posture Neutre.

Dans le cas de l'attitude sociale Soumis, nous estimons qu'il est possible d'envisager la Peur comme une manifestation de la Soumission. Le côté excessif et amusant de cette posture indique de façon extrêmement claire l'infériorité du robot face à l'utilisateur. L'attitude Directif, au contraire, jouera sur les valeurs de Fierté et de Colère afin de représenter la détermination et la puissance sociale du robot.

Notons que cette sélection d'attributs émotionnels pour représenter le comportement social est très dépendante du support. En effet, ces mêmes associations, sur un autre robot, ne renverront peut-être pas la même intention communicative.

4.4. Conclusion

Nous avons présenté dans ce chapitre la relation entre le comportement social exprimé par le système, et les tendances émotionnelles et interactionnelles de l'utilisateur. Dans cette relation, nous avons noté que la sélection du comportement social du système en fonction du profil de l'utilisateur est l'une des techniques en Interaction Humain-Machine permettant de simuler une adaptation affective et sociale de la part du système. Nous présenterons l'implémentation de ces règles en logique floue dans le chapitre 0.

Nous avons présenté une codification des comportements selon le contexte d'application tout d'abord, qui conditionne la délimitation des comportements désirables sur le circomplexe interpersonnel : nous partons de l'hypothèse qu'un utilisateur, en situation robotique quelconque, n'appréciera que des comportements sociaux positifs. Dans le cas plus précis de l'assistance à domicile, nous estimons que le système pourra présenter des comportements allant sur le continuum Soumis à Directif, mais que dans le cas du jeu nous pourrions nous limiter à des comportements plus neutres du point de vue de la domination.

La sélection de comportement dépend également du profil de l'utilisateur, et nous nous sommes intéressés principalement à la notion de réciprocité dans l'interaction, et la façon dont les différentes composantes du profil émotionnel et interactionnel pouvaient déterminer la sélection de comportement.

Nous avons remarqué que l'impact du comportement social sur le profil, à l'inverse, pouvait être un phénomène très pertinent à prendre en compte, mais que ces mécanismes mettent en jeu des facteurs sociaux et psychologiques très délicats à gérer. Nous nous bornerons à étudier l'impact du comportement social sur les expressions émotionnelles des utilisateurs.

Les travaux réalisés au cours de cette thèse portent sur deux contextes applicatifs très spécifiques, au cours desquels nous avons élaboré et testé notre proposition de détermination des comportements sociaux désirables chez le robot : le robot dans un rôle d'assistance à domicile de personnes en perte d'autonomie (déficiences visuelles), et le robot comme compagnon de jeu pour des sujets autonomes. L'expression des comportements sociaux du robot a donc dû être soigneusement travaillée, pour assurer le bon impact du comportement social du système. Nous avons proposé une codification via le support linguistique basé sur les théories de (Charaudeau, 1992), et une représentation du comportement social via les postures émotionnelles développées par (Beck *et al.*, 2010).

Chapitre 5

5. Expérimentations et corpus

5.1. Introduction

Le système d'interaction robotique sur lequel nous nous appuyons à l'occasion de nos travaux de thèse est constitué de plusieurs modules (voir Figure 15) :

- un module de perception gérant la détection des émotions et leur interprétation en indices émotionnels et interactionnels d'après la parole expressive de l'utilisateur ;
- un modèle utilisateur qui réalise la représentation émotionnelle et interactionnelle de l'utilisateur et met à jour l'historique affectif avec cet utilisateur ;
- un module comportemental qui sélectionne le comportement en fonction du profil utilisateur ;
- des effecteurs qui réalisent le comportement social du système.

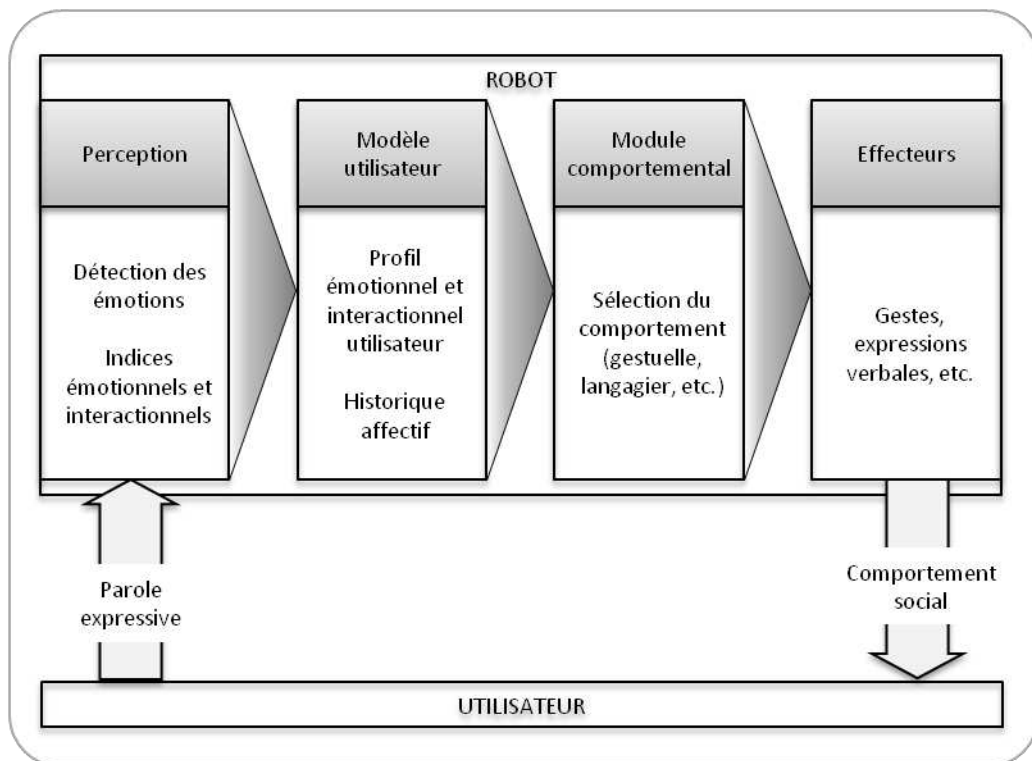


Figure 15. Boucle d'interaction émotionnelle Utilisateur-Robot : perception de la parole expressive utilisateur, système émotionnel, module comportemental et génération de comportement social.

Notre recherche est focalisée sur la constitution du profil émotionnel et interactionnel de l'utilisateur à partir des indices extraits du signal de voix expressive, ainsi que la sélection de comportement en fonction de ce profil. Afin de mener à bien cette étude, nous avons considéré certains paramètres de la boucle

d'interaction nous permettant de mieux comprendre les mécanismes d'interaction entre un humain et le système.

Tout d'abord, il est intéressant d'observer et de comprendre les expressions émotionnelles des utilisateurs dans un contexte d'Interaction Humain-Robot. Ceci est impératif afin d'orienter le développement du module de détection des émotions automatique, par la constitution d'un corpus de données pour l'apprentissage statistique, mais également pour l'étude perceptive des réactions émotionnelles potentielles qui permettent d'affiner la sélection des indices paralinguistiques pertinents pour gérer un système d'interaction émotionnelle.

Puisque notre étude se concentre également sur la sélection des comportements sociaux robotiques, il convient d'examiner l'impact des comportements du robot dans l'interaction. Nous mettons donc en œuvre deux études au cours desquelles le robot adopte des comportements sociaux spécifiques, et étudions sur les corpus obtenus les réactions émotionnelles collectées. Cette étude nous permet de confirmer nos hypothèses sur la détermination des comportements sociaux désirables selon le contexte d'interaction robotique.

L'évaluation de la relation entre indices paralinguistiques et profil utilisateur permettrait de renforcer et d'appuyer nos hypothèses. En effet, outre les études bibliographiques et empiriques nécessaires pour jeter les bases théoriques nécessaires à notre étude, il est intéressant de recourir à des analyses statistiques sur les données collectées afin de déterminer la pertinence des indices en relation avec le profil.

Le lien entre la sélection du comportement social robot et le profil utilisateur constitue également un élément de réflexion. Nous considérons ici des dialogues sans tâches, principalement du *small talk*. Cela nécessite une reconnaissance (automatique ou perceptive) du profil de l'utilisateur en temps réel, une interprétation en fonction du contexte (ce que dit le robot et la tâche d'interaction s'il y en a une) et une adaptation du comportement social du robot en fonction. Cet aspect de l'étude de la boucle d'Interaction Humain-Robot sera abordé dans le chapitre 0 de ce manuscrit, lorsque nous présenterons nos premiers tests d'implémentation du système semi-automatique.

Afin de mener à bien ces études, nous avons réalisé différentes expérimentations qui nous ont permis de constituer des corpus d'étude. Ces différents corpus ont permis tant la collecte de données émotionnelles en contexte pour les études acoustiques, que les études sur l'interaction. Ces travaux ont été menés en collaboration avec les membres du thème « Dimensions affectives et sociales dans les interactions parlées » (laboratoire LIMSI-CNRS, Orsay) dirigé par Laurence Devillers : Marie Tahon (analyse des indices acoustiques pertinents dans la voix émotionnelle en Interaction Humain-Robot), Mariette Soury (étude de l'expression du stress dans l'interaction), et Clément Chastagnol (reconnaissance automatique des dimensions affectives dans l'interaction orale Humain-Machine pour des personnes dépendantes). Les annotations de ces corpus ont été réalisées par Julieta Lencina, Caroline Benoit, Virginie Mouilleron et Nicolas Rollet, formés sur l'analyse perceptive des données émotionnelles.

Au cours de ce chapitre, nous présentons les différents protocoles d'acquisition et d'annotation de ces corpus, et analysons les résultats obtenus en fonction de ces données. Tous les corpus constitués au cours de ces travaux respectent des protocoles éthiques pour la collecte et la diffusion, et ne sont pas disponibles pour la communauté. Nous proposons en Annexe V un récapitulatif des corpus et des annotations réalisées.

5.2. Étude des réactions émotionnelles utilisateur

Nous avons déterminé dans la section 3.2 l'ensemble des théories sur lesquelles nous basons la conception du profil interactionnel et émotionnel de l'utilisateur, ainsi qu'une première approche des mécanismes de sélection du comportement du robot en fonction de ce profil.

Une première recherche expérimentale a été effectuée afin de déterminer, tout d'abord, si les indices qui seront utilisés dans le profil sont effectivement produits par les utilisateurs cible en Interaction Humain-Machine, et également s'il est possible de les annoter de façon fiable afin de les exploiter dans le module de reconnaissance automatique. Une expérimentation a ainsi été menée afin d'obtenir une première approche des réactions émotionnelles d'utilisateurs en interaction avec un système sensible à leurs émotions.

5.2.1. Interaction Adulte-Système – Corpus IDV-HS

Dans l'étude nommée IDV-HS (HS : Humain-Système) menée à l'Institut de la Vision avec les partenaires du projet ROMEO, notre but est de confronter des personnes en perte d'autonomie à un système de détection d'émotions déclaré comme étant « en test » : les sujets se projettent dans une situation d'assistance robotique à l'aide de scénarios, où l'état affectif exprimé doit être reconnu par le système dirigé en Magicien d'Oz (Tahon *et al.*, 2010). Le bon fonctionnement du système de détection, ou au contraire son mauvais fonctionnement, génèrent des expressions d'émotions chez les participants. Cette étude nous a permis de prendre un premier contact avec le public cible de notre application, d'identifier leurs réactions émotionnelles et leur attitude générale face à des dispositifs techniques, avant la mise en place d'une expérimentation en interaction homme-robot réelle.

5.2.1.1. Protocole d'acquisition

Les séances d'enregistrement réalisées dans le cadre de l'expérimentation IDV-HS ont été menées dans un appartement mis à notre disposition par l'Institut de la Vision, au sein de la résidence Saint-Louis (structure d'hébergement et d'accompagnement des personnes déficientes visuelles gérée par le Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, Paris). Les volontaires pour l'expérimentation, adultes et souffrants de déficience visuelle, étaient invités à contribuer à la mise au point d'un robot assistant personnel sensible aux émotions, en nous permettant de tester notre système de reconnaissance des émotions grâce à leur concours.

Dans la salle de séjour de l'appartement, deux expérimentateurs accueillent le participant. L'un est en charge de réaliser les enregistrements audio, tandis que l'autre gère le système de détection des émotions en magicien d'Oz. Assis de part et d'autre d'une table, l'expérimentateur système et le participant dialoguent tout d'abord au sujet de l'expérimentation, ses objectifs, et ce qu'il sera attendu du participant. Nous avons estimé essentiel, dans le cadre de cette étude, de communiquer avec le participant sur les enjeux de l'étude, afin de les engager dans la réussite de l'expérimentation. En effet, l'assistance à domicile peut présenter pour eux un besoin réel. L'expérimentateur présente alors son travail de recherche afin que le sujet se sente responsabilisé. Il est également nécessaire que ce dernier soit bien disposé face aux questions qui lui seront posées, et face aux efforts d'imagination, d'expression, voire d'attention qui lui seront demandés. Contrairement au travail avec les enfants, il semble impératif d'insister sur la dimension scientifique et sur les répercussions d'un tel projet.

Le projet est alors présenté comme un travail de recherche sur la création d'un robot domestique, apportant de l'aide aux personnes souffrant de problèmes d'autonomie dans leur intérieur. Les objectifs sont que le robot soit capable d'effectuer des tâches ménagères simples, de lire le courrier électronique, de jouer avec les enfants, de surveiller la maison, de s'assurer que l'utilisateur va bien. Pour être capable de réagir avec plus de précision, le robot a besoin de reconnaître les émotions humaines. Cette étude se concentre sur la détection des émotions dans la voix. Le robot devra réagir différemment en fonction de l'humeur de la personne qu'il assiste. Le robot pourra venir en aide lorsqu'une personne est en difficulté chez elle. Lorsque le robot aura détecté dans la voix de l'humain que la situation est urgente (de la peur), il pourra mettre en œuvre des procédures pour aider la personne. Et ainsi, pour lui apprendre à détecter correctement la peur, il faut que le robot ait enregistré une multitude d'échantillons de voix, de personnes exprimant cette émotion.

Le participant est équipé d'un micro-cravate, et l'expérience débute. Tout d'abord, le participant est invité à répéter quelques consignes simples, présentées dans le Tableau 10. Ces instructions nous permettent d'obtenir un échantillon de voix neutre et sont également destinées au partenaire du projet travaillant sur la reconnaissance vocale. Il est bien précisé au participant que ces instructions ne sont présentées que dans le cadre de cette expérimentation : les capacités technologiques du robot ne lui permettent pas d'effectuer ces ordres à l'heure de l'expérimentation.

Tableau 10. Expérimentation IDV-HS – Instructions fictives jouées par le participant

« Viens par ici »	« Ramasse ça »	« Mets le plat au four »
« Arrête-toi »	« Donne »	« Descends la poubelle »
« Stop »	« Roméo réveille-toi »	« Va chercher le courrier »
« Écoute-moi »	« Approche »	« Va chercher à boire »
« Va-t-en »	« Aide-moi à marcher »	« Aide-moi à me lever »

Cinq scénarios de projection sont proposés ensuite au participant. Pour chaque scénario, il est demandé au participant d'écouter la mise en situation présentée par l'expérimentateur, ainsi que les indications sur les états affectifs à jouer, et de s'exprimer de la façon attendue. Le contenu lexical choisi par le participant est libre. Ces scénarios reprennent les thèmes de l'assistance robotique à domicile : la détection d'une alerte (médicale, intrusion), échanges informels sur l'état de santé, définition d'un planning. Il est alors attendu du participant qu'il s'exprime tel qu'il le ferait si le robot était dans la même pièce que lui. Le Tableau 11 reprend les cinq scénarios présentés au participant par l'expérimentateur, ainsi que les différents états affectifs attendus.

L'expérimentateur système évalue perceptivement si le participant a réalisé l'émotion attendue, et choisi sur l'interface Magicien d'Oz l'émotion prétendument reconnue. Les **Erreur ! Source du renvoi introuvable.** et **Erreur ! Source du renvoi introuvable.** présentent cette interface : l'expérimentateur lance prétendument l'analyse audio à chaque énonciation du locuteur. Il choisit en réalité un des paramètres (noms fictifs) dissimulant un des états affectifs potentiellement attendus (*bien-être, peur, inquiétude, douleur, joie,*

agacement, urgence, tristesse, pas d'état émotionnel), ce qui déclenche une simulation de traitement du signal audio et d'annonce de résultat (par le biais d'un fichier son préenregistré annonçant : « Émotion reconnue : ... »).

Si l'état affectif exprimé par le participant n'est pas conforme à celui attendu pour ce scénario, l'expérimentateur l'encourage à recommencer. Nous entendons par « non conformes », les expressions non identiques à l'état attendu (*joie* au lieu de *stress*, par exemple, ou encore *tristesse* au lieu de *stress*), mais également les émotions exprimées trop faiblement. L'expérimentateur fait recommencer le participant un nombre limité de fois, en fonction de l'implication et du bien-être du participant. Dans le cas des participants à l'aise, il n'était pas rare que nous sélectionnions volontairement un état affectif faux, même si le participant avait manifestement exprimé une émotion correcte. Ces différentes stratégies permettaient d'éliciter des réactions d'agacement et de satisfaction naturelles de la part des participants.

Ces premières « règles de comportement », adoptées de façon intuitive de manière à générer des réactions émotionnelles, ont jeté les premières bases vers l'élaboration de nos comportements robotiques, par la représentation d'un système parfois accommodant ou fonctionnel, et parfois dans l'incompréhension ou dans l'ignorance des attentes utilisateur.



Figure 16. Figure 16. Expérimentation IDV-HH : Interface Magicien d'Oz pour la sélection de l'état affectif du locuteur

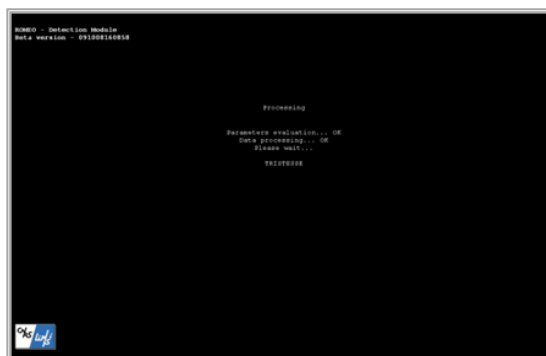


Figure 17. Figure 17. Expérimentation IDV-HH : Interface Magicien d'Oz pour les phases de « traitement » et annonce orale de l'état affectif détecté

Tableau 11. Expérimentation IDV-HS – Scénarios et état affectifs attendus

Scenario	Expressions d'états affectifs attendues	Mise en situation
Urgence médicale	DOULEUR, STRESS	« Vous êtes dans un appartement qui a été équipé pour recevoir votre robot domestique. Vous êtes dans votre salon, le robot est dans la cuisine en train de mettre de l'ordre. Vous êtes installé dans votre fauteuil, mais quand vous vous relevez votre tête heurte violemment la fenêtre de la salle à manger, et vous vous sentez mal. Vous voulez que le robot vienne de toute urgence depuis la cuisine pour vous mettre en relation avec un médecin. Que dites-vous pour qu'il vienne à vous ? »
Réveil du matin (en forme)	SATISFACTION, BIEN-ÊTRE	« C'est le matin, vous vous réveillez. Le robot entend que vous vous levez, il vient vous interroger sur votre état de santé. Vous êtes dans une santé excellente, vous vous sentez extrêmement dynamique, faites-le lui comprendre. »
Réveil du matin (mauvaise santé)	DOULEUR, AGACEMENT	« C'est le matin, vous vous réveillez. Le robot entend que vous vous levez, il vient vous interroger sur votre état de santé. Vous avez la migraine, vous êtes barbouillé, faites-le lui comprendre. »
Bruits suspects	PEUR, ANGOISSE, INQUIETUDE	« Vous êtes chez vous, la nuit, couché dans votre lit. Vous entendez un bruit suspect en provenance de la salle à manger, que vous êtes incapable d'identifier. Romeo est présent dans la chambre, en veille. Vous avez peur, vous voulez qu'il aille voir ce qui a causé ce bruit. »
Réveil du matin (déprime)	TRISTESSE, AGACEMENT	« C'est le matin, vous vous réveillez. Le robot entend que vous vous levez, il vient vous interroger sur votre état de santé. Vous ne vous sentez pas en forme, sans trop savoir pourquoi... Votre moral n'est pas au beau fixe. »
Venue des enfants, d'amis	SATISFACTION, JOIE	« Vos (petits-)enfants/des amis viennent vous rendre visite. Cette visite vous est agréable. Faites comprendre au robot que tout va bien pour vous, il peut vaquer à d'autres occupations. »

5.2.1.2. Annotation et scores

Suite aux enregistrements pratiqués dans le cadre de l'expérimentation IDV-HS, nous avons obtenu des fichiers audio qui ont été segmentés puis annotés par des codeurs formés pour l'annotation émotionnelle. La segmentation n'est réalisée que sur la piste du participant, et lorsque la voix de celui-ci ne se chevauche pas avec tout bruit extérieur (bruit ambiant, autre personne qui parle). Chaque segment obtenu présente une émotion homogène, fixe sur toute la durée du segment (Devillers, 2006; Devillers et Martin, 2008). Nous renvoyons à l'Annexe III de ce présent document pour une description détaillée du protocole de segmentation et d'annotation utilisé, et présentons ci-après les étiquettes et dimensions d'annotation utilisées pour ce protocole.

Pour chaque segment émotionnel, trois étiquettes décrivent l'émotion exprimée. Ces étiquettes sont choisies par l'annotateur sur une liste prédéfinie d'émotions et d'états émotionnels, en fonction de sa

perception des expressions du locuteur. L'annotation doit au maximum être perceptive, et non basée sur le contenu linguistique de l'énoncé. Par exemple, si le locuteur dit « je suis triste » sur un ton joyeux non équivoque, nous attendons de l'annotateur qu'il annote « Joie ». Si aucune étiquette émotion n'est choisie, la valeur par défaut sera « Neutre », signifiant qu'aucune manifestation émotionnelle n'est clairement perceptible. L'étiquette émotion 1 (ou émotion majeure) est l'émotion exprimée la plus prégnante dans le segment, celle que l'on perçoit immédiatement, dès la première écoute. Les étiquettes émotion 2 (appelée également émotion mineure) et 3 caractérisent plus précisément l'émotion 1, lui ajoutent une nuance. Toutes les combinaisons sont possibles (exemple : amusement-stress, satisfaction-énervement). Il est possible de décrire une émotion avec une étiquette positive liée à une négative. Il est possible de ne pas annoter d'émotions (étiquette « Neutre »), si aucune émotion n'est perçue dans le segment.

Le Tableau 12 présente les différentes valeurs des étiquettes d'annotation proposées aux annotateurs, et un regroupement en macro-classes (Joie, Colère, Peur, Tristesse, Neutre). Nous proposons également des étiquettes d'annotation supplémentaires (catégorie Autres), selon les manifestations émotionnelles et sociales que nous avons pu repérer lors des enregistrements, ou selon nos attentes au vu d'expérimentations en interaction humain-robot de la communauté. Tel est le cas par exemple pour le phénomène de maternage (*motherese* en anglais, on parle également de « parler bébé »), utilisé notamment dans les travaux de (Batliner *et al.*, 2004) où des enfants sont en interaction avec le robot chien Aibo. Le maternage consiste à s'adresser, initialement, à un enfant ou à un bébé tel que le ferait un parent, en adaptant ses intonations et son rythme. Les indices perceptibles sont notamment une tonalité plus élevée qu'en temps normal, et une extra-articulation. Ce phénomène a été repéré dans la communauté de recherche sur l'interaction humain-robot lorsque l'utilisateur s'adresse au compagnon robotique, c'est pourquoi nous désirions proposer ce choix aux annotateurs afin de déterminer si cet usage était employé dans nos contextes. Parmi les étiquettes catégorisées dans « Autres », citons également le cas de l'Empathie et la Compassion : il ne s'agit pas à proprement parler d'émotions, mais plutôt de mécanismes d'adaptation sociale, qui mène à une prise de distance entre l'émotion réellement ressentie, et l'émotion délibérément exprimée. Ce phénomène a été jugé intéressant à repérer pour étudier les mécanismes d'adaptation interactionnelle potentiellement utilisés par les locuteurs dans une situation d'Interaction Humain-Machine.

Nous pouvons retrouver, dans l'Annexe IV de ce document, les définitions pour chacune des étiquettes présentées dans ce tableau. L'annotateur estime également la **valence** qu'il prête au segment, à savoir sa polarité émotionnelle : le segment est-il « positif », « négatif », « ambigu », « positif et négatif à la fois », ou « neutre » ? L'annotateur ne corrèle pas impérativement cette étiquette à ce qu'il aura sélectionné comme étiquettes émotions, mais cette association peut être interprétée comme un ensemble par la suite : l'association d'une étiquette « surprise », par exemple, avec une valence « négative » pourra permettre de décrire une réaction à un événement imprévu jugé comme étant néfaste au locuteur.

Tableau 12. Corpus IDV-HS : Étiquettes d'annotation et regroupement en macro-classes.

Macro-classe de l'émotion	Valeur d'annotation	Macro-classe de l'émotion	Valeur d'annotation
JOIE	Joie	PEUR	Peur
	Amusement		Inquiétude
	Satisfaction		Stress
Gêne			
COLERE	Colère	NEUTRE	Neutre
	Agacement		Positif
TRISTESSE	Tristesse		Négatif
	Déception		Surprise
AUTRES	Intérêt	Empathie	Compassion
	Mépris	Ironie	Souffrance
	Maternage	Ennui	

Une fois les étiquettes sélectionnées, l'annotateur doit typer plus précisément l'émotion en estimant des valeurs dimensionnelles. L'**intensité** (la force de l'émotion), l'**activation** (la quantité de moyens mis en œuvre pour l'expression de l'émotion), le **contrôle** (la faculté perceptible de dominer l'émotion) sont annotés sur des échelles allant de 1 (très faible) à 5 (très fort). Ces dimensions, exploitées dans les autres corpus de cette thèse, sont détaillées dans l'Annexe IV. L'annotateur détermine également si l'expression est jouée dans le cadre du scénario, ou si elle provient d'une réaction spontanée du locuteur (par exemple discussion avec l'expérimentateur, réaction au Magicien).

Voici quelques exemples de tours de parole exprimés par les locuteurs du corpus IDV-HS, et annotés de façon consensuelle par les annotateurs. Tout d'abord, une femme de 23 ans dans un scénario de colère : « *Va-t-en, je suis en colère, je suis pas bien. Tu m'énerves.* » Ici, trois segments ont été repérés par les annotateurs : « *va-t-en, je suis en colère* » (annoté colère), « *je suis pas bien* » (annoté déception), « *tu m'énerves* » (annoté agacement). Dans un deuxième exemple, nous retrouvons un homme de 57 ans jouant le scénario de déprime : « *Tu... tu restes tranquille parce que je suis pas trop en forme, c'est pas... non je sais pas, c'est pas la pêche aujourd'hui hein, c'est pas...* ». Les annotateurs ont distingué deux segments émotionnellement homogènes, « *tu... tu restes tranquille parce que je suis pas trop en forme, c'est pas...* » (annoté déception), « *non je sais pas, c'est pas la pêche aujourd'hui hein, c'est pas...* » (annoté tristesse).

Scores d'annotation

Comme expliqué précédemment, cette première étude vise à identifier les expressions émotionnelles du public cible du projet. Cette acquisition de données, initialement basée sur la génération de comportements

émotionnels expressifs chez les locuteurs par le biais de scénarios, visait également à induire différentes réactions émotionnelles. Nous avons donc intégré à notre protocole d'annotation certaines émotions, ou plus généralement certains phénomènes interactionnels, recensés soit dans la littérature, soit dans nos constatations sur le terrain, et susceptibles d'apparaître dans notre contexte d'interaction. Il est donc intéressant de recenser les étiquettes utilisées par les annotateurs.

Le Tableau 13 présente la matrice de confusion des annotations émotionnelles réalisées par les deux annotateurs. Un premier regroupement en termes de macro-classe correspond à la liste du Tableau 12, complétées selon les annotations de la valence réalisées (et détaillées ci-après). Il s'agit de catégoriser chaque étiquette selon sa macro-catégorie. Nous obtenons ainsi :

- Positif : « joie », « amusement », « satisfaction », « positif », « empathie », « excitation »
- Colère : « colère », « agacement »
- Tristesse : « tristesse », « déception »
- Peur : « peur », « inquiétude », « stress », « gêne »
- Neutre : « neutre », « intérêt »
- Négatif : « négatif », « mépris »
- Autres : « compassion », « surprise », « maternage », « ironie », « ennui », « souffrance »

Chaque segment portant plusieurs étiquettes annotations, la macro-classe d'un segment émotionnel nécessite d'être calculée. Nous présentons les règles de détermination de la macro-classe dans l'Annexe III. Huit macro-classes permettent de décrire l'ensemble des 3066 segments du corpus IDV-HS : Positif, Colère, Peur, Tristesse, Neutre, Positif-Négatif, Autres, et Négatif. Parmi ces segments, nous trouvons environ pour moitié de segments actés (segments prononcés dans le cadre des scénarios), et moitié de segments spontanés (réactions des locuteurs). Nous obtenons 1562 segments pour lesquels les annotateurs s'accordent sur la macro-classe.

Nous nous appuyons sur la mesure Kappa (Cohen, 1960) pour déterminer le taux d'accord entre les annotateurs, selon la formule :

$$K = \frac{P_o - P_e}{n1 - P_o}$$

où P_o correspond à la proportion d'accord observé, et P_e la proportion d'accord aléatoire, calculés à partir de la matrice de confusion représentant les différentes évaluations des juges. La valeur K peut prendre des valeurs décimales entre 0 et 1 ; 0 signifiant une absence complète d'accord entre les juges, et 1 un accord total. De faibles scores d'accord entre les juges (inférieurs à 0,5), outre un désaccord réel, peuvent également être dus à une non-homogénéité des données, dans le cas où une catégorie spécifique serait sur-représentée ou au contraire sous-représentée.

Tableau 13. Corpus IDV-HS : Matrice de confusion des annotations émotionnelles. Étiquettes regroupées en macro-classe.

	Positif	Colère	Peur	Tristesse	Neutre	Pos-Nég	Autres	Négatif	Total
Positif	409	32	39	9	67	46	12	9	623
Colère	3	192	45	23	13	3	7	12	298
Peur	26	47	376	22	41	25	7	11	555
Tristesse	10	17	48	103	45	1	8	7	239
Neutre	138	31	225	42	431	17	54	22	960
Pos-Nég	64	14	10	2	6	10	6	2	114
Autres	17	16	37	10	15	5	31	12	143
Négatif	2	27	43	19	21	1	11	10	134
Total	669	376	823	230	639	108	136	85	3066

L'analyse des scores d'agrément entre les annotateurs n'a pu être réalisée sur un tel schéma d'annotation, du fait de sa finesse. Nous avons donc opéré des regroupements plus basiques de macro-classes et écarté les classes peu représentées ou à l'interprétation difficiles, pour ne conserver que Positif, Colère, Peur, Tristesse et Neutre, permettant d'obtenir un score d'agrément Kappa de $K = 0.5$. La distinction Neutre versus autre étiquette est le point ayant été le plus délicat pour l'accord inter-annotateurs, et induisant ce faible taux d'accord. La distinction entre Positif, Colère, Peur et Tristesse entraîne un Kappa de $K = 0.67$, et une distinction entre positif (Positif) et négatif (Colère, Peur, Tristesse) présente un taux d'agrément de $K = 0.79$.

Nous remarquons des disparités pour l'annotation de la valence (cinq dimensions possibles). Le désaccord réside essentiellement dans la discrimination Ambigu contre les autres valeurs, et Positif-Négatif contre les autres valeurs. Sur cinq dimensions, les annotateurs obtiennent un Kappa de $K = 0.4$. Nous obtenons un taux d'accord de $K = 0.6$ si nous négligeons les segments pour lesquels au moins un des annotateurs a perçu une valence ambiguë ou positive-négative.

L'annotation de l'activation et de l'intensité a suscité très peu d'accord de la part des annotateurs, les scores Kappa étant très faibles. La complexité de cette annotation réside dans la finesse de la distinction entre les notions, mais également sur la difficulté de présenter des valeurs étalon aux annotateurs afin de niveler les échelles.

5.2.1.3. Résultats

L'expérimentation IDV-HS visait à analyser les réactions émotionnelles susceptibles de se produire de la part de notre public cible dans un contexte proche du contexte final d'application. A cette fin, nous avons mis au point un protocole d'annotation encourageant les annotateurs à analyser le plus finement possible les manifestations émotionnelles des locuteurs. Nous proposons donc de renseigner trois étiquettes afin de décrire chaque segment émotionnel.

L'analyse de l'utilisation de la troisième étiquette nous montre de façon non équivoque qu'aucun des deux annotateurs n'est parvenu à décrire aussi finement les segments émotionnels que nous l'avions attendu. La Figure 18 présente la répartition moyenne des étiquettes utilisées : 96% des étiquettes choisies pour l'émotion 3 a été Neutre, ce qui signifie que les annotateurs laissaient la valeur par défaut. Pour les 4% restants, nous trouvons par ordre décroissant les étiquettes Négatif (0,7%), Stress (0,6%), Positif (0,5%), Inquiétude (0,2%), puis d'autres étiquettes présentes à moins de 0,2% du corpus.

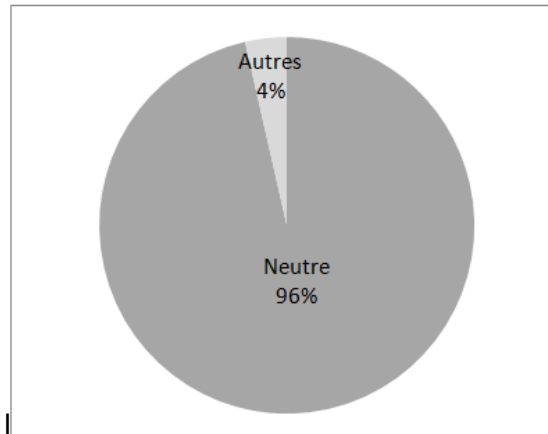


Figure 18. Corpus IDV-HS : Utilisation moyenne de la troisième étiquette d'annotation par les annotateurs.

Ces résultats montrent, étant donné la proportion de « neutre » en Émotion 3, que les étiquettes majeures et mineures ont été jugées par les annotateurs suffisantes pour décrire les émotions exprimées par les locuteurs.

Nous nous penchons ensuite sur le nombre de fois où l'annotateur a utilisé certaines étiquettes émotion (que ce soit en tant qu'émotion majeure, ou émotion mineure, ou émotion 3), et sur la valence choisie conjointement. Le Tableau 14 présente la répartition moyenne des valences pour les segments sur lesquels les annotateurs s'accordent sur la présence d'expression de Surprise, Intérêt, Excitation, Ironie ou Stress (que ce soit en tant qu'émotion majeure, mineure, ou en troisième émotion). Ces répartitions nous permettent d'analyser plus finement ces manifestations.

Nous nous demandons tout d'abord si l'excitation perçue par les annotateurs est assimilable à un état de stress. C'est la raison pour laquelle nous avons volontairement dissocié ces deux étiquettes, afin d'étudier la corrélation avec la valence sélectionnée par les annotateurs. Les résultats d'annotation indiquent que sur 294 segments de Stress, un peu plus de la moitié sont liés à un état émotionnel négatif, 17% sont difficiles à identifier, 12% sont liés à un état positif, et 8% ne sont pas reliés à un état émotionnel spécifique. L'étiquette Excitation quant à elle est distinctement liée à un état émotionnel positif : sur les 37 segments étiquetés Excitation, 80% sont jugés positifs. Si l'excitation exprimée par les participants est donc nettement une excitation positive, le stress exprimé par les locuteurs est plus difficile à identifier comme une manifestation négative, puisque ce stress est parfois exprimé conjointement à des émotions positives ou neutres. Détecter la présence de stress dans une interaction Humain-Machine peut être un élément impératif, surtout lorsqu'il s'agit d'un robot assistant à domicile. Cependant, cette première approche nous montre, de façon simple, que le stress est une manifestation ambiguë, qu'il n'est pas forcément évident à catégoriser.

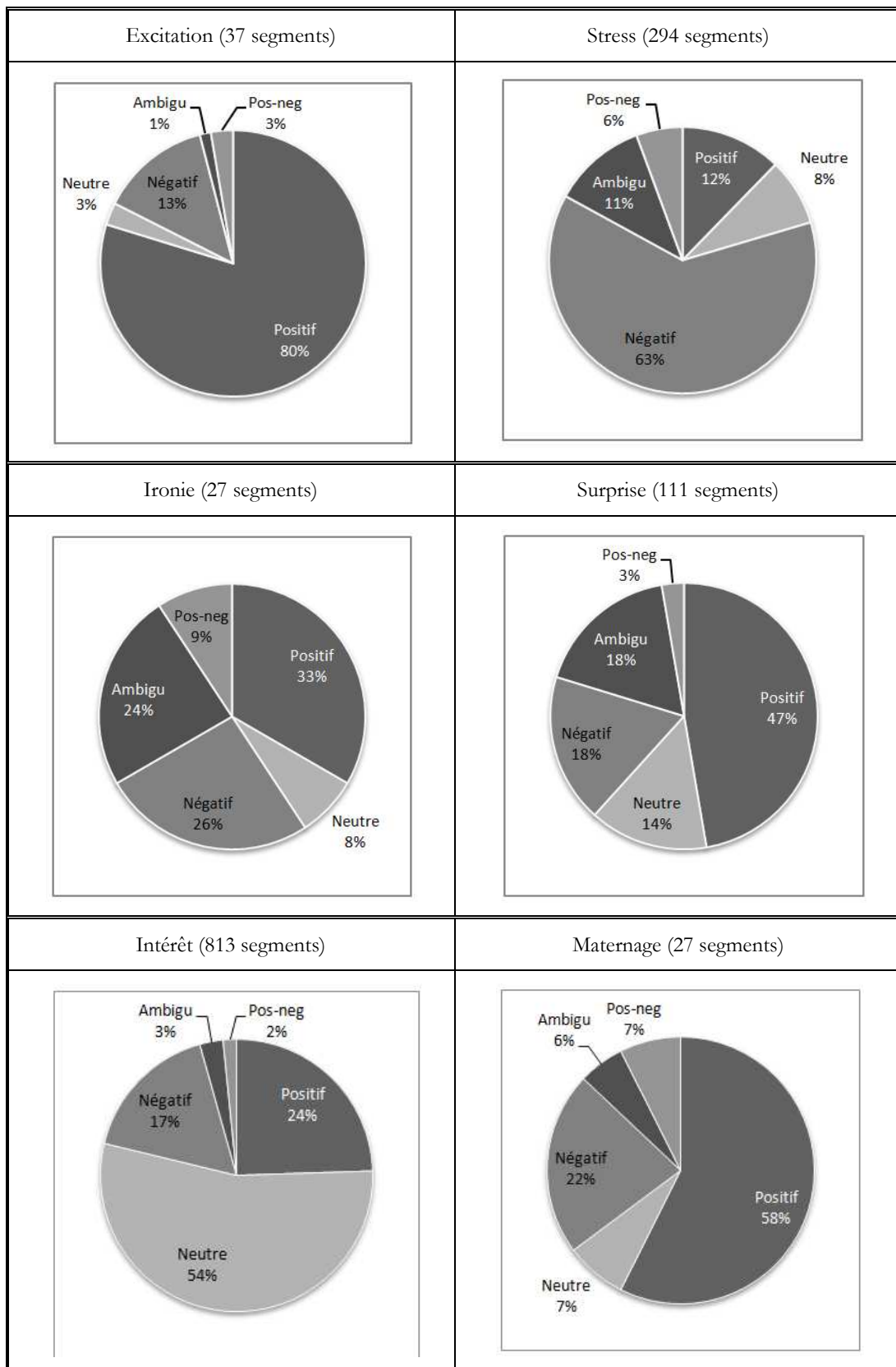
A titre exploratoire encore, nous nous interrogeons sur l'utilisation qu'il serait faite des étiquettes Surprise et Ironie. Utiliser l'ironie pourrait impliquer que l'utilisateur estime que le système est susceptible de comprendre le double sens humoristique, mais une autre possibilité serait que l'utilisateur se moque du système. Il nous semble bien sûr important que l'utilisateur n'éprouve pas le besoin ni l'envie d'exprimer de sentiments condescendants par rapport au système, c'est pourquoi il est important de repérer les utilisations d'ironie dans nos corpus. De façon basique, nous considérons qu'une ironie liée à une attitude émotionnelle positive peut être souhaitable dans l'interaction, tandis qu'une ironie teintée de négatif pourrait sans doute être de la critique. Sur les 27 segments pour lesquels les annotateurs s'accordent quant à la présence d'Ironie, 33% ont été repérés dans un contexte émotionnel positif, 26% en contexte négatif, et 24% ambigu. La répartition en termes de valence n'est donc pas probante : quelques manifestations sont non désirables, mais tout autant de manifestations sont potentiellement admissibles (les positives) ou à étudier plus avant (les ambiguës). D'autre part, la très faible proportion de perception d'ironie dans le corpus (moins de 1% du corpus total) nous amène à négliger ce phénomène dans le contexte du corpus IDV-HS.

Les 111 segments étiquetés Surprise par les annotateurs sont présentés pour la moitié en contexte émotionnel positif (47%). Pour l'autre moitié, 18% des segments sont négatifs, 18% ambigus, 14% neutres, et 3% positifs-négatifs. Si l'expression de Surprise est donc principalement liée à une expression positive, son contexte émotionnel d'apparition reste variable.

La moitié des 813 segments annotés comme présentant de l'Intérêt est reliée à une absence de manifestation émotionnelle. Proposer cette étiquette aux annotateurs avait pour but de leur imposer une réflexion supplémentaire, plutôt qu'utiliser l'étiquette « Neutre » lorsqu'ils ne ressentaient pas d'émotion. Nous voulions que, même si aucune émotion n'était perceptible, ils évaluent l'implication du locuteur dans la situation communicationnelle. Cela a permis de réduire le nombre de segments étiquetés comme étant de type Neutre (c'est-à-dire avec pour étiquette majeure Neutre) ; 431 segments du corpus sont purement Neutre. Même si le résultat final de cette démarche n'entraîne pas de différence notable – les segments annotés Intérêt n'apportent pas plus d'information pour la compréhension de l'état émotionnel qu'une étiquette Neutre –, cette approche a enjoint les annotateurs à adopter plus de finesse dans leur réflexion.

Nous voulions repérer d'éventuels phénomènes de maternage de la part des locuteurs. Nous obtenons 27 segments exploitant cette étiquette, dont 58% sont exprimés en contexte positif. La présence de 22% de segments négatifs attire principalement notre attention, car il ne s'agit non pas d'une expression de maternage (action positive), mais d'infantilisation (volonté de rabaisser, acte de communication avec un « inférieur »). Ces manifestations sont donc à repérer lors des expérimentations suivantes, car il s'agit d'une expression non désirable de la part de l'utilisateur.

Tableau 14. Corpus IDV-HS : Valences moyennes utilisées conjointement aux étiquettes d'annotation



Étant donné la quantité infime d'utilisation de l'étiquette Empathie (les annotateurs s'accordent sur seulement deux segments comportant cette étiquette), nous ne la faisons pas apparaître dans notre tableau.

Nous notons toutefois que nous avons catégorisé cette étiquette sous la macro-classe Positif puisque dans notre corpus, ces deux occurrences étaient liées à une valence positive.

Par le biais de cette analyse des annotations, nous désirions appréhender la présence ou l'absence de certaines manifestations émotionnelles spécifiques exprimées par le public cible de notre étude, dans un contexte conceptuellement proche du contexte d'application final. L'expérimentation se déroulait de façon à stimuler la production d'émotions de la part des participants, afin d'obtenir un corpus exploitable pour les études acoustiques de l'équipe, et pour une première piste d'analyse haut niveau des réactions émotionnelles.

La première constatation réalisée, pour la suite de nos expérimentations, est que l'annotation du Stress seule n'est pas suffisante pour déterminer le type d'excitation exprimé par le locuteur. Si l'on désire repérer le stress négatif, le malaise, chez l'utilisateur du système, il convient de prêter également attention à la valence, ou de renforcer auprès des annotateurs la notion de sentiment négatif lié au stress.

L'Ironie est une manifestation primordiale à repérer dans l'interaction Humain-Machine, du point de vue de sa valence : une ironie négative pour être le signe d'un ressentiment, d'une non-adéquation entre les prestations attendues du système.

Nous n'avons pas constaté beaucoup de phénomènes de Maternage de la part des utilisateurs adultes. Nous attirons néanmoins l'attention sur le maternage à valence négative, interprétable comme une moquerie, une infantilisation.

L'Intérêt aurait pu être annoté à part, plutôt qu'en alternative à l'étiquette Neutre. En plus de l'annotation émotionnelle, la présence d'Intérêt ou pas, évaluée dans l'attitude générale du locuteur, aurait pu renseigner sur le degré d'implication du locuteur dans son interaction avec le système. Cette notion pourrait être alors reliée à d'autres indices afin de déterminer automatiquement l'engagement du locuteur.

L'expérimentation IDV-HS nous a fourni un premier contact avec le public cible, nous a sensibilisé à ses besoins et attentes, et à ses réactions émotionnelles. Ces premières analyses nous ont ouvert des pistes pour l'analyse des émotions dans l'interaction entre un utilisateur adulte et le système en situation d'assistance. Cette expérimentation a également posé les premières bases pour l'élaboration et le test de comportements robotiques, « désirables » versus « non-désirables », qui ont été par la suite bien plus approfondis.

5.2.2. Interaction Enfant-Robot – Corpus NAO-HR1

Dans le cadre de l'expérimentation NAO-HR1 (HR : Humain-Robot), le robot Nao est présenté aux enfants comme un compagnon de jeu. Il propose de jouer à trois jeux différents : questions-réponses, le jeu des chansons, et le jeu des émotions (Delaborde et Devillers, 2010; Delaborde *et al.*, 2009; Delaborde *et al.*, 2010; Tahon *et al.*, 2011).

5.2.2.1. Protocole d'acquisition

Nous enregistrons des enfants qui se connaissent (amis, frères-sœurs), afin de stimuler l'interaction, en nous basant sur l'étude de (Shahid *et al.*, 2009) qui tend à montrer que la présence d'un ami peut booster les réactions positives pendant une expérimentation. D'autre part, la présence d'enfants qui se connaissent est

représentative d'un contexte réel où le robot serait à domicile, en train de jouer avec les enfants : ces enfants seraient soit de la même famille, soit amis.

Un expérimentateur tient le rôle du « maître du jeu ». Le maître du jeu a pour charge de présenter les règles des jeux et distribuer les cartes. Il veille également à laisser s'écouler un temps raisonnable entre chaque nouvelle phase du jeu (changement de question, changement de jeu) afin de permettre le chargement des comportements de NAO. D'autre part, il veille au bon déroulement du jeu, et au respect des consignes d'installation audio/vidéo pour optimiser les conditions d'enregistrement. Il peut occasionnellement encourager les enfants à interagir avec le robot. Son rôle dans ce cas est très cadré, afin que ses interventions entravent au minimum le bon déroulement de l'interaction entre les enfants et le robot. Nous spécifions que si un comportement du robot codé afin de faire réagir les enfants échoue (c'est-à-dire que les enfants semblent ne pas le remarquer), alors le maître du jeu peut intervenir et attirer l'attention des enfants sur le comportement exprimé par le robot. Toutefois, le maître du jeu n'intervient pas si l'enfant a exprimé une réaction émotionnelle (qu'elle corresponde à celle attendue ou non). Cette légère ingérence du maître du jeu a été jugée nécessaire pour pallier une éventuelle timidité ou passivité des enfants dues aux conditions d'expérimentation.

Chaque enfant porte un micro-cravate, et est assis sur un coussin à même le sol (l'un face à Nao, l'autre à sa droite). Les paroles des enfants sont enregistrées depuis le début du jeu, jusqu'à la fin de l'expérimentation. Nao est posé debout à même le sol. La Figure 19 illustre cette installation. La scène est filmée par deux caméras, permettant de voir l'expression faciale et corporelle de chaque enfant. Ces enregistrements vidéo n'ont qu'un but illustratif pour l'expérimentation, et de contrôle pour le manipulateur du Magicien d'Oz (décrit plus loin), et n'ont pas servi de support pour l'annotation émotionnelle basée uniquement sur l'audio.

Chacun des trois participants (Nao et les deux enfants) pose à tour de rôle une question inscrite sur une carte, et les deux autres doivent essayer de trouver la réponse. Bien que Nao soit lui-même un joueur, il est plus stimulé par le désir d'accompagner les enfants dans leur jeu que par le désir de gagner. Il fournit donc des indices à son concurrent si celui-ci ne connaît pas la réponse. Toutefois, pour stimuler la compétitivité des enfants, et ainsi enrichir leur expressivité, Nao ne favorisera parfois qu'un seul des deux enfants.



Figure 19 - Expérimentation NAO-HR1 : deux garçons (10 et 12 ans) jouent avec le robot NAO.

Dans le jeu, le robot ne montre pas de reconnaissance particulière de l'individualité de l'enfant (il ne focalise que rarement son regard sur un enfant précis, ne s'adresse pas à eux par leurs prénoms). La prise en compte de leur individualité ne se manifeste que lorsqu'il donne spécifiquement des indices à l'un mais pas à l'autre, ou qu'il demande des explications à un enfant en particulier. D'une façon globale, le discours de NAO et des stratégies s'adresse aux deux enfants à la fois. Cependant, dans les faits, les réactions de NAO peuvent être déclenchées après qu'un enfant ait pris la parole. L'enfant prend donc dans ces cas la réponse directement pour lui.

Le robot Nao est contrôlé par un Magicien d'Oz. Les stratégies utilisées par le Magicien au cours du jeu de questions-réponses ont pour but d'élucider et de collecter des expressions émotionnelles chez l'enfant, dans un cadre ludique. Le jeu des chansons permet la collecte de chants d'enfants pour un partenaire du projet, et nous permet de tester également quelques stratégies de comportement pour Nao ; par exemple, Nao ne reconnaîtra pas tout de suite la chanson, et le maître du jeu pourra proposer à l'autre enfant de chanter la chanson jusqu'à ce que Nao la reconnaisse. Le but du jeu des émotions permet la collecte d'expressions émotionnelles actées prototypiques, et nous en profitons pour appliquer des stratégies de comportement de la part de Nao. Nao indique une émotion à jouer, et les enfants doivent la jouer, jusqu'à ce que Nao parvienne à la détecter correctement. Nao ne détectera parfois pas correctement l'émotion exprimée par un des enfants, même si l'expression est manifestement correcte, afin de stimuler la compétition entre les deux enfants.

Le Tableau 15 reprend les différentes stratégies comportementales appliquées par le robot Nao lors de l'expérimentation. Nous reconnaissons certaines stratégies telles que décrites dans le circomplexe des attitudes sociales présenté en Figure 6 : « Aidant », « Coopératif », « Compétitif », « Rebelle ». Une stratégie n'entre par contre pas complètement dans le cadre du circomplexe : nous voulions faire rire les enfants. A cette fin, le robot devait réaliser une action en inadéquation avec les attentes qu'un utilisateur peut avoir d'un système robotique. Cette action n'a pas de vocation sociale à proprement parler, mais joue simplement sur le décalage entre une attente fonctionnelle, et une défaillance du système. Nous avons catégorisé les comportements exprimés par le robot en « comportements socialement désirables » et « indésirables ». Nous nous y référons plus loin également en tant que « comportements positifs » et « comportements négatifs ». Certaines actions du robot étaient considérées comme « comportements neutres », comme le fait de saluer les

enfants au début du jeu, se présenter, ou des phrases liées au déroulement du jeu (« D'accord. », « Donne-moi la carte. », etc.).

Tableau 15. Expérimentation NAO-HR1 : Comportements sociaux désirables et non désirables du robot Nao, réactions émotionnelles attendues chez l'enfant, et exemples

	Comportement social robot	Stratégie associée	Réaction émotionnelle attendue	Exemple
Comportements socialement désirables (comportements « positifs »)	Coopératif	Encouragement, félicitations	Joie	Nao félicite le joueur pour une bonne réponse
	Aidant	Aide à la réussite	Joie	Nao fournit un indice pertinent pour une question difficile
Comportements socialement indésirables (comportements « négatifs »)	Non coopératif	Non fonctionnel persistant	Colère	Nao fait répéter des explications Nao ne reconnaît pas une émotion manifestement bien jouée / une chanson bien chantée
	Humour	Non fonctionnel	Joie	Nao donne la réponse au lieu de fournir un indice
	Compétitif	Mise en compétition	Colère, Tristesse	Nao favorise un joueur en refusant de donner un indice à l'autre
	Rebelle	Non fonctionnel	Colère	Nao s'éteint au milieu du jeu sans vouloir se rallumer
Comportements neutres	Amical	Politesse, cours du jeu	Joie, Neutre	Nao se présente


5.2.2.2. Annotation et scores

Le protocole de segmentation et d'annotation de ces données est similaire au protocole d'IDV-HS. Deux annotateurs formés à l'annotation émotionnelle ont annoté chaque segment émotionnel à l'aide de **trois étiquettes émotion**, suivant la même liste qu'IDV-HS. Ils ont typé la **valence**, l'**intensité**, l'**activation**, le **contrôle**, mais également l'**état mental** supposé du locuteur (voir Annexe IV pour la

définition et les valeurs possibles). Pour chaque segment, ils ont également indiqué à **qui le locuteur s'adresse** (à Nao, au maître du jeu, à l'autre enfant, à lui-même, ou autre), et **quel a été le déclencheur** de la réaction émotionnelle de l'enfant le plus probable (un comportement de Nao, une action du maître du jeu, etc.).

Le Tableau 16 présente un extrait transcrit d'une phase du jeu questions-réponses mettant en scène une fille de 9 ans et son frère de 11 ans. Au cours de cette phase, Nao se présente comme non fonctionnel : il ne comprend pas les réponses, et insiste. Nous constatons l'expression d'attitudes émotionnelles variées chez les enfants, allant de l'Amusement à l'Irritation.

Tableau 16. Expérimentation NAO-HR1 : Extrait transcrit du jeu des questions-réponses. Une fille de 9 ans (Loc. 1) et son frère de 11 ans (Loc. 2) jouent avec le robot.

		
Locuteur	Émotion perçue	Transcription
Loc. 1 (Fille)	Neutre	[Lit la carte] Quel est le nom du film dans lequel un extraterrestre dit "Téléphone maison" ?
Loc. 2 (Garçon)	Satisfaction	E.T.
Loc. 1	Joie / Déception	Oh mais c'est facile aussi !
Nao	n/a	C'est quoi "E.T." ?
Loc. 2	Neutre	Euh... E.T. c'est un film avec un extra-terrestre qui dit "Téléphone maison".
Loc. 1	Neutre	Allo. Téléphone. Maison.
Nao	n/a	Mais pourquoi il téléphone à la maison ?
Loc. 2	Surprise	Euh parce qu'il est sur la Terre.
	Amusement	Et il veut parler avec ceux qui sont chez lui, sur une autre planète.
Loc. 1	Neutre	E.T. Téléphone. Maison. E.T. Téléphone. Maison.
Nao	n/a	Un extraterrestre n'a rien à faire avec un téléphone. Je ne comprends pas.
Loc. 2	Amusement	Non, mais en fait il parle par la tête, par la pensée.
Nao	n/a	Mais pourquoi il téléphone à la maison ?
Loc. 2	Amusement	Ben euh parce qu'il s'ennuie tout seul.
	Amusement/Maternage	Et peut-être qu'il demande à sa maman s'il peut rester plus longtemps sur la Terre.

Nao	n/a	C'est quoi "E.T." ?
Loc. 2	Neutre	Euh... E.T.... Euh...
Loc. 1	Amusement	C'est un extra-terrestre.
Loc. 2	Irritation	C'est un extra-terrestre, mais "E.T." ça s'écrit en anglais
	Amusement	donc ben comme je fais pas d'anglais, ben je sais pas
Loc. 1	Amusement	[rire]
Loc. 2	Neutre	Extraterrestre. Terre.
Nao	n/a	Mais pourquoi il téléphone à la maison ?
Loc. 1	Ennui/Irritation	[soupir]
Loc. 2	Amusement	Euh... Bon, je sais pas.

Scores d'annotation

Nous obtenons 1287 segments au total, dont 560 présentant un accord des annotateurs quant à la macro-classe (voir matrice de confusion Tableau 17).

Nous constatons un désaccord important sur la distinction entre l'annotation Neutre versus autres étiquettes, et un désaccord quant aux positifs-négatifs et aux ambigus. Le désaccord sur les neutres s'explique par le fait que les annotateurs ne réalisent pas la même distinction entre une émotion faible et une absence d'émotion ; quant aux désaccords sur Positif-Négatif et Ambigu, cela s'explique bien sûr par la subtilité de ces émotions mélangées, qui engendre naturellement un désaccord d'interprétation.

La distinction entre positif et négatif pour le corpus NAO-HR1 est plus sujette à désaccord qu'IDV-HS. En effet, le score Kappa sur la distinction « Positif », « Colère », « Peur », « Tristesse » et « Neutre » est de $K = 0.4$, et l'accord « Positif » versus « Négatif » seulement de $K = 0.5$. Le choix des étiquettes émotion, dans le contexte de ce corpus, semble avoir posé problème aux annotateurs, puisque si nous nous penchons sur leur score d'agrément sur l'annotation de la valence (en restreignant sur « Positif » et « Négatif »), le score est $K = 0.7$. Les annotateurs semblent donc raisonnablement d'accord sur la polarité émotionnelle des expressions des enfants, mais la variété des étiquettes à leur disposition ne permet pas d'obtenir un score d'agrément correct.

Tableau 17. Corpus NAO-HR1 : Matrice de confusion des annotations émotionnelles. Étiquettes regroupées en macro-classe.

	Positif	Colère	Peur	Tristesse	Neutre	Pos-Nég	Ambigu	Négatif	Total
Positif	297	21	31	5	31	53	16	2	456
Colère	1	31	1	2	0	11	0	1	47
Peur	16	6	34	6	8	13	1	1	85
Tristesse	1	12	10	45	2	7	3	0	80
Neutre	89	27	139	9	112	32	14	1	423
Pos-Nég	27	11	10	6	1	27	0	0	82
Ambigu	24	4	14	3	17	3	14	2	81
Négatif	4	7	8	6	2	4	2	0	33
Total	459	119	247	82	173	150	50	7	1287

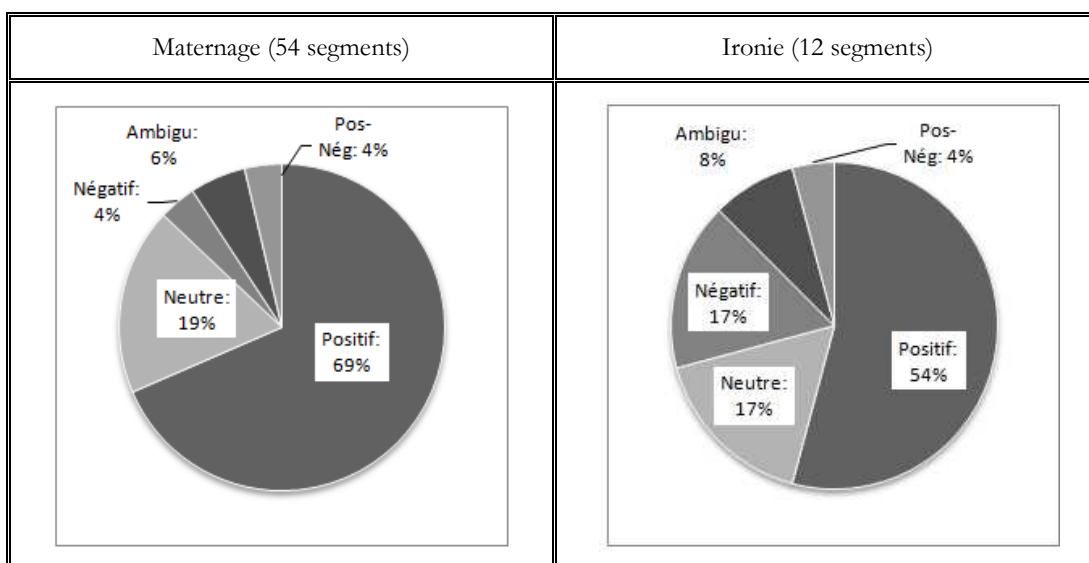
5.2.2.3. Résultats

Nous nous intéressons, comme pour l'expérimentation IDV-HS, à l'usage que les locuteurs ont fait des étiquettes maternage et de l'ironie au cours des enregistrements. Dans la même optique que précédemment, nous souhaitons repérer les marques d'infantilisation et d'ironie négative de la part des locuteurs.

Une première étude a été menée sur les annotations de l'Ironie, du Maternage et du Mépris (Delaborde et Devillers, 2010) afin de repérer ces phénomènes sur les segments consensuels (i.e. pour lesquels les annotateurs s'accordent quant à la macro-classe). Cette étude a été réalisée au cours du processus d'annotation du corpus, sur les dix premiers locuteurs. Nous remarquons que, dans l'ensemble, les enfants expriment rarement une Ironie négative ou du Mépris au cours de l'expérience (seules 7 occurrences ont été annotées Ironie et Mépris). Nous notons que les filles n'expriment jamais d'ironie envers le robot, tandis que les garçons l'expriment parfois lorsque le robot fait une erreur (4 fois), or quand le robot les stimule à interagir (4 fois). Nous ne constatons pas de relation particulière entre l'âge du locuteur, ou la relation entre les deux enfants (fratrie ou amis), et l'expression d'Ironie ou de Mépris. Nous estimons que l'expression de telles attitudes relève plutôt du tempérament du locuteur. Cependant, le sexe de l'enfant peut être une indication des tendances de comportement que nous pouvons retrouver chez l'enfant quand ce dernier n'est pas d'accord avec le comportement du robot.

Nous remarquons en effet des expressions de Maternage à la fois chez les garçons et les filles lors de l'expérimentation. Toutefois, parmi les 42 segments ayant été annotés par l'étiquette Maternage, 8 présentent une annotation double (Agacement, Négatif, ou Ironie). Nous associons ce phénomène d'annotation (Maternage + étiquette négative) à l'expression d'infantilisation, de moquerie. Dans notre petit corpus d'étude, nous retrouvons cette annotation uniquement lorsque les couples de joueurs sont masculins.

Tableau 18. Corpus NAO-HR1 : Valences moyennes utilisées conjointement aux étiquettes d'annotation.



Une fois le corpus total annoté, nous avons étudié à nouveau les annotations du Maternage et de l'Ironie (Tableau 18). Nous nous sommes penchés sur les segments où les annotateurs s'accordent pour trouver l'étiquette (que ce soit en tant que majeure, mineure ou étiquette 3). Nous avons cette fois étudié la relation entre l'étiquette et la valence moyenne choisie par les annotateurs (Positif, Négatif, Neutre, Ambigu ou Positif-Négatif). Lors de l'expression du Maternage, la valence perçue par les annotateurs est négative seulement pour 4% des segments, ambiguë pour 6%, et mixte pour 4%. La proportion de manifestations négatives liées au Maternage est donc faible. Nous remarquons quelques expressions localisées, que l'on pourrait perceptivement associer à du Mépris, lorsque les enfants jouent avec le robot. Dans le cas de l'Ironie, sur les 12 segments où les annotateurs s'accordent à percevoir la présence de cette expression, 54% sont positifs. En proportion, l'emploi d'Ironie semble donc se faire plutôt de façon négative dans notre contexte. Toutefois, la faible quantité d'expression d'Ironie sur notre corpus total permet de dire que cette manifestation n'est également que ponctuelle.

5.2.3. Conclusion

Les expérimentations IDV-HS et NAO-HR1 visaient à prendre contact avec le public cible, dans des conditions proches des contextes applicatifs finaux. Nous avons donc travaillé avec des personnes adultes déficientes visuelles dans le cadre de projection de situations d'assistance, et avec des enfants dans le cadre de jeu avec le robot. Dans les deux cas, nous nous sommes penchés sur l'analyse des expressions émotionnelles des locuteurs. Nous avons présenté notre étude des annotations perceptives réalisées sur ces deux corpus.

Ces premières études nous ont permis de discerner la présence d'expressions émotionnelles et affectives pertinentes pour comprendre la qualité de l'interaction entre l'humain et le robot. Tout d'abord, l'Ironie et le Mépris – même si ce sont des expressions affectives difficiles à détecter automatiquement – sont des marques très pertinentes de la prise de distance du participant vis-à-vis du robot. Les manifestations repérées d'Ironie et de Mépris nous ont donc fourni quelques pistes sur les comportements du système à éviter.

Nous n'avons pas constaté beaucoup de phénomènes de Maternage de la part des utilisateurs adultes, mais plus de la part des enfants. Au même titre que l'Ironie et le Mépris, le Maternage à valence négative peut être un signe de moquerie, ou d'infantilisation. Ces expressions indiquent potentiellement une inadéquation entre les capacités du système, et les attentes effectives de l'utilisateur, qu'il s'agisse d'attentes sociales ou techniques.

Ces expérimentations nous ont fourni un premier contact avec les publics cibles, et nous ont sensibilisés à leurs besoins et attentes, et à leurs réactions émotionnelles. Ces premières analyses nous ont ouvert des pistes pour l'analyse des émotions dans l'interaction entre les utilisateurs et le système, qu'il s'agisse d'un contexte d'assistance ou de jeu. Ces expérimentations ont également posé les premières bases pour l'élaboration et le test de comportements robotiques « désirables » versus « non-désirables », que nous avons approfondis dans ce document de thèse. Les corpus IDV et NAO ont également été utilisés pour la détection automatique des émotions dans la thèse de M. Tahon (Tahon, 2012).

5.3. Expressions émotionnelles en réaction aux comportements robotiques

Nous avons mené l'expérimentation NAO-HR1 (Nao Humain-Robot), détaillée dans la section précédente, qui vise à étudier les réactions émotionnelles des enfants en situation de jeu. Des enfants sont invités à venir jouer, deux par deux, avec le robot NAO. Les deux enfants jouent à des activités ludiques avec le robot, qui exprime des comportements sociaux codés selon les théories présentées en 4.2.2. Ces interactions enfant-robot nous ont fourni une base d'étude quant aux réactions émotionnelles de jeunes utilisateurs en situation cible.

Nous avons également réalisé l'expérimentation IDV-HR, au cours de laquelle des personnes adultes souffrant de déficience visuelle interagissent avec le robot NAO. Les comportements sociaux du robot sont codés selon les mêmes théories. Cette expérimentation vise à analyser les réactions émotionnelles de ce public dans des situations d'assistance à domicile.

Nous retrouvons certaines de ces analyses dans (Delaborde, 2012a) et (Delaborde, 2012b).

5.3.1. Interaction Enfant-Robot – Corpus NAO-HR1

Cette expérimentation, dont le protocole de collecte et d'annotation ont été détaillés en 5.2.2, vise à mettre en interaction le robot NAO et deux enfants, dans une situation ludique. Les enfants jouent à des jeux de questions-réponses, ainsi qu'au jeu des émotions et au jeu des chansons. Les stratégies comportementales du robot sont codées selon les théories évoquées dans 4.2.2, et sont présentées dans le Tableau 15.

5.3.1.1. Résultats

Analyse de réactions aux stratégies comportementales du robot

Nous désirons, dans le cadre de cette expérimentation, étudier les réactions émotionnelles des enfants face aux comportements du robot. Cependant, et bien que nous ayons essayé de les réduire au minimum, les interventions du maître du jeu ne sont pas à négliger. En effet, les annotateurs ont dénombré que sur la totalité des sessions d'enregistrement :

- 39,2% des réactions des enfants proviennent d'une intervention du Maître du jeu (cours du jeu, motivation à l'interaction et renforts des stratégies du robot) ;
- 38,9% des réactions sont jugées comme étant déclenchées par une action de Nao ;
- 13,9% sont déclenchées par une intervention de l'autre enfant ;
- 8% des réactions présentes dans le corpus sont issues d'autres déclencheurs (évènement extérieur, fil de pensées, etc.)

Nous nous intéressons principalement ici à ce que nous dénommons les « renforts de stratégie ». Il s'agit des interventions du maître du jeu propres à attirer l'attention sur les comportements du robot. Par exemple lorsque le robot se trompe (il lit la question, et enchaîne en lisant également la réponse), les enfants, portés par le fil de l'expérimentation, ne réagissent parfois pas à l'incongruité de l'action. Le maître du jeu encourage les enfants à réagir pour faire comprendre au robot son erreur. Autre exemple, lorsque NAO s'éteint de façon inattendue, le maître du jeu indique la marche à suivre aux enfants : « Utilisez la commande que j'ai utilisée tout à l'heure pour le remettre en marche : 'Nao, réveille-toi !' ».

Les interventions de « renfort » du maître de jeu consistent donc à tourner les enfants vers le robot ; nous retrouvons en effet à chaque fois que les segments émotionnels qui suivent sont adressés au robot, et non au maître du jeu. Nous avons donc assimilé ces interventions du maître du jeu aux comportements robot associés, ce qui nous permet d'obtenir :

- 9% des réactions émotionnelles des enfants sont dues à une intervention du maître du jeu
- 69% proviennent des stratégies du robot (+ renforts maître du jeu)
- 14% déclenchées par l'autre enfant
- 8% autres

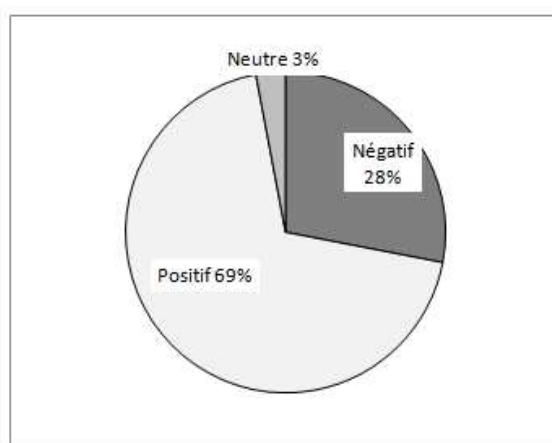


Figure 20. Distribution moyenne des stratégies comportementales du robot (comportement positif, négatif ou neutre) présentées à chaque enfant pour la totalité de sa session d'enregistrement.

Comme nous le voyons Figure 20, pour chaque session d'enregistrement, le robot se comporte avec chaque enfant, en moyenne, de façon positive 69% du temps, 28% de façon négative, et 3% de façon neutre.

Nous retrouvons la définition des catégories « comportement positif », « comportement négatif » et « comportement neutre » au Chapitre 4.

Nous désirons maintenant estimer les réactions émotionnelles des enfants en fonction des comportements du robot. Nous nous sommes basés sur les segments consensuels entre les annotateurs sur la macro-classe. La Figure 21 présente les réactions émotionnelles parmi les macro-classes Positif, Colère, Tristesse, Peur, Neutre, Ambigu et Positif-Négatif. Nous pouvons constater de façon notable que les comportements négatifs du robot n'ont pas nécessairement entraîné de réactions émotionnelles négatives de la part des enfants. Malgré le biais dû au taux d'accord de l'annotation (une proportion d'accord sur les segments positifs plus élevée que sur les segments négatifs), l'observation des séances d'enregistrement nous fournit globalement une impression similaire. Nous constatons que les comportements négatifs du robot entraînent beaucoup de réactions d'amusement.

Les comportements positifs exprimés par le robot ont entraîné naturellement des réactions positives de la part des enfants, mais également une proportion intéressante de réactions « neutres ». Ces réactions ne signifient pas nécessairement que le locuteur n'a pas exprimé d'émotions, mais aussi qu'il a exprimé des émotions très faiblement marquées. L'observation rejoint la constatation que les comportements positifs du robot n'ont pas impérativement entraîné de réactions très marquées de la part des enfants.

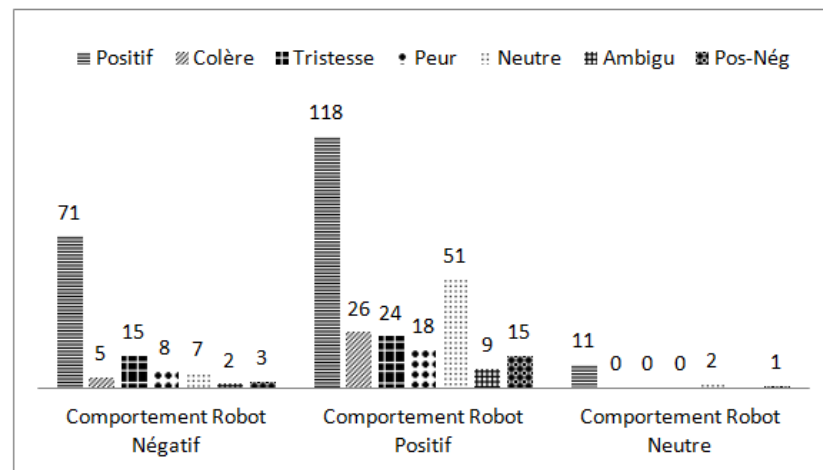


Figure 21. Expérimentation NAO-HR1 : Réactions émotionnelles des enfants en fonction des comportements Robot et #segments

Nous nous sommes également penchés sur les différences de réaction liées au sexe. Comme nous le disions précédemment, le phénomène de Maternage négatif (que l'on peut associer à du mépris, à une moquerie) n'a été constaté que chez les garçons dans notre corpus. D'une façon plus générale, nous avons étudié les différences de réactions émotionnelles (Positif, Négatif, Neutre ou Autres) en fonction des comportements du robot (Positif, Négatif ou Neutre). La Figure 22 présente les proportions de réactions émotionnelles selon le comportement robot et le sexe locuteur.

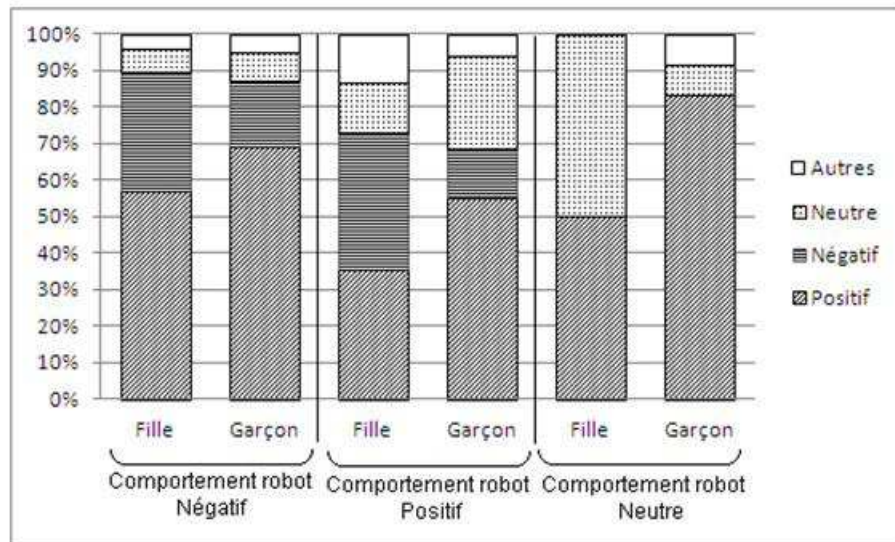


Figure 22. Expérimentation NAO-HR1 : Réactions émotionnelles des enfants en fonction des comportements Robot. Distinction Garçon/Fille. (« Négatif » : Colère, Tristesse, Peur ; « Autres » : Ambigu et Positif-Négatif)

Les comportements robotiques neutres ont engendré des manifestations différentes chez les garçons et les filles : les réactions des garçons semblent plus positives, tandis que les filles présentent pour moitié des réactions positives, et pour moitié des réactions neutres ou très faibles. Toutefois, la proportion de comportements robotiques neutres étant très faible dans le corpus (3%), nous ne disposons que de très peu de segments pour cette analyse. Dans le cas des réactions aux comportements robot négatifs et positifs, nous constatons que les garçons semblent réagir, en proportion, plus fréquemment positivement que les filles. Les réactions émotionnelles négatives des filles étaient composées de comportements de gêne, entraînant soit des expressions neutres ou négatives.

Nous nous sommes intéressés à une éventuelle répartition des réactions émotionnelles selon la tranche d'âge. Nous estimons qu'un jeune enfant n'aura pas la même attitude qu'un préadolescent face à un dispositif robotique. Il nous suffit par exemple d'étudier les résultats de l'étude de (Okita et Schwartz, 2006) sur la représentation mentale des enfants de 3 à 5 ans sur les robots, qui indique que les enfants sont particulièrement sujets à une représentation animiste des systèmes robotiques, mais que la façon dont ils perçoivent le robot semble évoluer rapidement avec l'âge. Les résultats que nous avons obtenus dans le cadre de l'étude de notre corpus ne sont pas probants. Nous avons réparti les enregistrements des enfants, âgés de 8 à 12 ans, en deux groupes autour de la moyenne d'âge (11 ans). La Figure 23 présente les réactions des enfants en fonction de leur tranche d'âge, selon le comportement du robot.

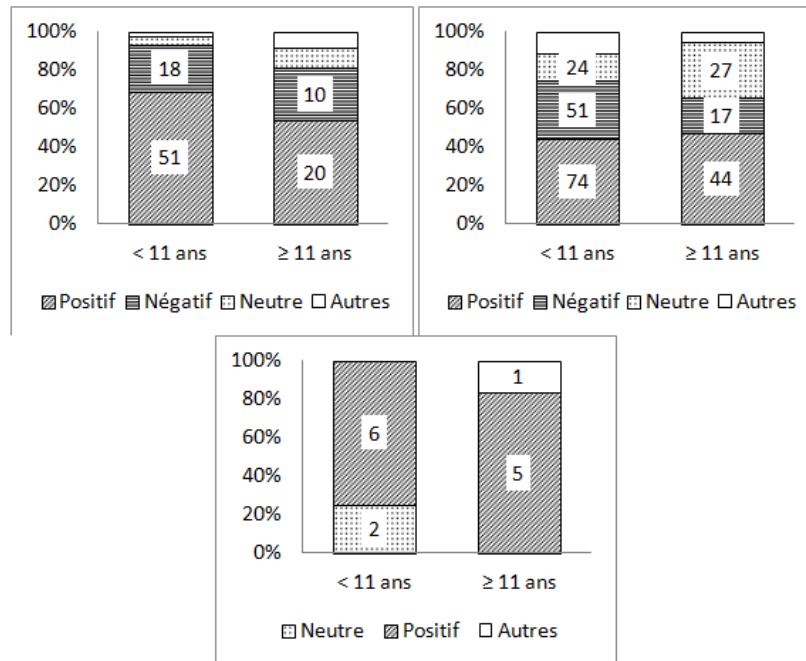


Figure 23. Expérimentation NAO-HR1 : Réactions émotionnelles des enfants (et #segments) en fonction des comportements du robot et de leur tranche d'âge. De gauche à droite, première ligne : comportement négatif, positif ; seconde ligne : comportement neutre.

En proportion, les enfants plus jeunes semblent réagir de façon légèrement plus positive que leurs aînés lorsque le robot se comporte de façon inappropriée. Dans le cas de comportements positifs de la part du robot cependant, nous ne constatons aucune différence notable, et nous ne pouvons rien conclure concernant le peu de segments obtenus pour les comportements robotiques neutres.

Nous avons analysé également l'impact des stratégies sociales du robot sur l'activation (la force de l'émotion exprimée) exprimée par les enfants. Nous nous sommes basés sur les 746 segments sur lesquels les annotateurs s'accordent quant à l'activation. Nous remarquons (Figure 24) que lors des comportements négatifs du robot, les enfants ont exprimé des émotions légèrement plus fortes qu'en réaction aux comportements positifs. Nous pouvons opérer un rapprochement entre cette constatation et le fait que les émotions exprimées lors des comportements négatifs étaient plus positives, et souvent mêlées de rire.

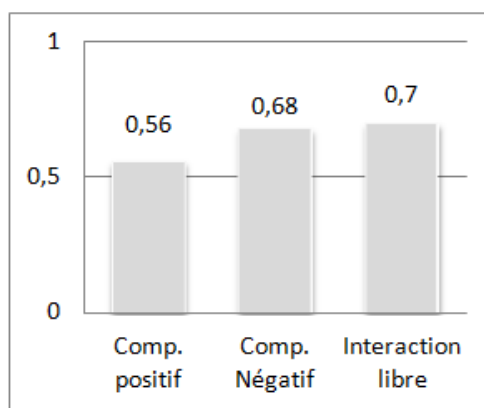


Figure 24. Expérimentation NAO-HR1 : Activation moyenne exprimée par les enfants (de 0/faible à 1/fort) en fonction des comportements positifs ou négatifs du robot, et comparaison avec l'activation exprimée en interaction libre

Cette première expérimentation mettant en relation des enfants jouant avec le robot Nao nous a permis d'étudier certaines tendances de comportement émotionnel des enfants. Tout d'abord par l'observation, puis par l'analyse des annotations émotionnelles sur le corpus. Ces résultats nous indiquent que ce que nous considérons comme « comportement négatif » n'est pas forcément réhibitoire dans un contexte de jeu. Il y aura probablement certainement une plus grande tolérance chez les enfants qui jouent avec le robot, qu'avec un adulte en situation d'assistance (ce que l'expérience IDV-HR, que nous décrivons plus loin dans ce document, tend à nous montrer).

Si nous remarquons quelques différences selon le sexe des enfants (l'attitude par rapport au maternage et la moquerie notamment), la différence n'est pas forcément très marquée en ce qui concerne l'âge, car les enfants ne sont pas dans des catégories d'âge radicalement différentes (enfants en bas âge versus adolescents, par exemple) : notre public ne se prêtait pas forcément à une comparaison sur l'âge.

5.3.2. Interaction Adulte-Robot – Corpus IDV-HR

L'expérimentation IDV-HR met en scène des personnes adultes souffrant de déficience visuelle, enregistrées lors de leurs interactions avec le robot NAO. Elles jouent pendant environ quarante-cinq minutes des scénarios de la vie quotidienne, au cours desquels le robot adopte différentes attitudes sociales (Delaborde, 2012b; Tahon *et al.*, 2011).

L'expérimentation vise à remplir plusieurs objectifs, tant du point de vue de l'analyse émotionnelle et interactionnelle entre un utilisateur adulte déficient visuel et le robot, que du point de vue des analyses acoustiques du signal audio émotionnel. Tout d'abord – et ce manuscrit de thèse se focalise sur cet aspect de l'expérimentation –, nous visons l'acquisition de données audio de test sur les interactions émotionnelles entre le robot NAO et des personnes malvoyantes, en réaction à des stratégies sociales de communication du robot. L'expérimentation vise également à obtenir des échantillons audio d'émotions actées et spontanées en vue de la reconnaissance des émotions et de l'identification du locuteur. Une dernière étude acoustique porte sur l'analyse de la normalisation au locuteur, relatée dans les travaux de thèse de Marie Tahon (Tahon, 2012).

5.3.2.1. Protocole d'acquisition

Au cours de l'expérimentation IDV-HR, nous accueillons un à un des participants adultes souffrant de déficience visuelle dans l'appartement témoin de l'Institut de la Vision mis à notre disposition (voir Figure 25). Le participant est installé assis sur le canapé (que nous voyons au premier plan de la photographie), face à NAO posé sur la table basse à sa hauteur. Un expérimentateur présent dans la salle se charge des captures audio (micro-cravate) et vidéo du participant, ainsi que d'expliquer les objectifs de notre recherche. Il fait passer les questionnaires au participant, mais n'intervient pas du tout lors des scénarios.



Figure 25. Plateforme de recherche de l'Institut de la Vision visant à améliorer l'habitat pour les personnes atteintes de déficience visuelle: l'appartement témoin Homelab. (photographie <http://www.institut-vision.org>)

Un second expérimentateur est dissimulé dans la pièce attenante, et est en charge de la manipulation du système en Magicien d'Oz qui gère le comportement du robot. Il bénéficie d'un retour audio et vidéo de la scène, lui permettant de sélectionner les réponses au moment opportun.

La prise de contact entre l'utilisateur et le robot NAO est identique pour tous les locuteurs : le robot demande le prénom du participant, puis il se présente lui-même et lui demande s'il aime sa voix. A chaque fois qu'une réponse est attendue de la part du locuteur, l'automate gérant les réponses du robot se met en pause, et l'expérimentateur Magicien d'Oz relance la suite du scénario une fois la réponse formulée. NAO demande ensuite au participant de lui toucher le sommet de la tête afin de « démarrer l'enregistrement ». Cette manipulation vise seulement à faire prendre conscience au participant de la présence physique du robot : il ne s'agit pas que d'un système informatique, mais d'une entité physique susceptible d'avoir une efficacité réelle dans les scénarios que le participant doit jouer ensuite. La Figure 26 présente cet automate sous Choregraphe, le logiciel développé par Aldebaran Robotics permettant la gestion des mouvements et comportements du robot.

Pour la suite de l'expérimentation, l'expérimentateur Magicien d'Oz présélectionne entre deux et trois attitudes sociales différentes pour le robot (en veillant à une bonne répartition attitude sociale robot/sexes du locuteur parmi l'ensemble des participants reçus). Pour chaque comportement, le participant doit jouer cinq scénarios : le thème et les sujets abordés dans les scénarios ne varient pas d'une attitude à l'autre, seul le choix du lexique dans les énonciations du robot varie.

Chaque suite de scénarios est organisée de la sorte : un scénario « positif » (qui demande de générer des émotions positives) débute et conclut la série. Le robot propose donc de jouer les scénarios « en forme », puis « mauvaise santé », « déprime », « urgence », et « content ». Chaque scénario est organisé autour du thème du réveil du matin. Nous retrouvons dans le Tableau 19 les mises en situation initiales proposées au participant.

Chaque participant est amené à inventer sa propre suite au scénario, en étant toutefois fortement influencé par les indications et questions du robot. Une fois l'attitude sociale du robot sélectionnée, la seule influence du Magicien d'Oz sur le cours du scénario consiste à sélectionner des réponses en fonction de la sémantique des énonciations du participant. Nous voyons par exemple dans la Figure 27 les remarques possibles du robot en fonction de la façon dont le participant répond (nous retrouvons en Annexe I la totalité des énonciations du robot). Dans le cas où aucune réponse du robot ne correspond parfaitement, l'expérimentateur choisit la réponse la plus logique parmi celles qui lui sont permises.

L'attitude sociale du robot est fixée et exprimée selon les théories évoquées dans 4.3. Nous avons sélectionné des attitudes « socialement positives » telles que « Aimable », « Empathique », « Encourageant », et des attitudes que nous avons estimées a priori « négatives » : « Directif », « Hésitant », et « Neutre ». Nous estimons qu'un robot social d'assistance à domicile peut être encourageant pour l'utilisateur, ou compatir à ses soucis et à ses joies, ou bien être tout simplement amical, sans marque de domination particulière. Le comportement « Hésitant » fait référence à un robot qui ne serait pas sûr de lui, ni de ses décisions. Ceci nous semble évidemment incompatible avec la fonction d'aide à domicile. D'autre part, un robot directif serait excessivement dominateur, et ne laisserait pas d'autonomie à l'utilisateur. Nous supposons que cette attitude sociale ne serait pas dans tous les cas négative : la prise en main d'une situation de crise peut être envisageable. Certains utilisateurs, en fonction de leur tempérament, pourraient également souhaiter que le robot prenne constamment en main les décisions, par commodité ou par passivité, par exemple. Nous estimons néanmoins que d'une façon générale, un robot se doit de préserver l'autonomie de l'utilisateur, c'est pourquoi nous le catégorisons comme « négatif ». Dans le cas de l'attitude sociale neutre, nous estimons qu'un robot social ne peut pas être « neutre » : il doit avoir un impact affectif sur l'utilisateur, une volonté marquée de sociabilisation, de penchant affectif.

Dans le déroulement de l'expérimentation, nous proposons toujours au minimum une attitude sociale positive, et une attitude sociale négative. Pour chaque attitude, le participant joue les cinq scénarios. Nous rajoutons parfois une attitude supplémentaire (donc une nouvelle série de scénarios) en fonction du temps disponible et de la motivation du participant.

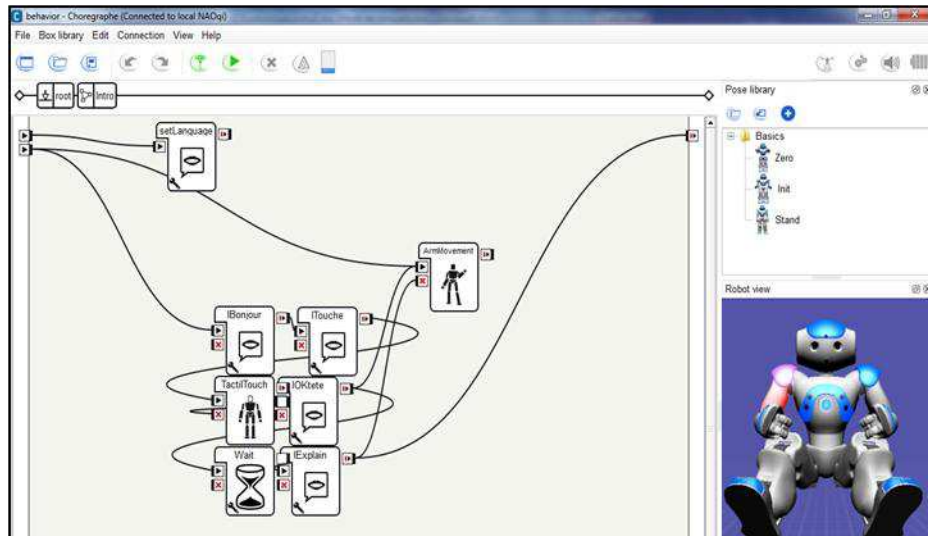


Figure 26. Expérimentation IDV-HR : automate Choregraphe gérant l'initiation de l'interaction avec l'utilisateur. (logiciel Choregraphe par Aldebaran Robotics)

Tableau 19. Expérimentation IDV-HR : Scénarios et états affectifs attendus.

Scenario	Expressions d'états affectifs attendus	Mise en situation
Réveil du matin (en forme)	SATISFACTION, BIEN-ÊTRE	« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin? »
Réveil du matin (mauvaise santé)	DOULEUR, AGACEMENT	« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin? »
Réveil du matin (déprime)	TRISTESSE, AGACEMENT	« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas au beau fixe. Bonjour, comment vas-tu ce matin? »
Réveil du matin (urgence)	STRESS, DOULEUR	« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence. Bonjour, comment vas-tu ce matin ? »
Réveil du matin (content)	SATISFACTION, JOIE	« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin? »

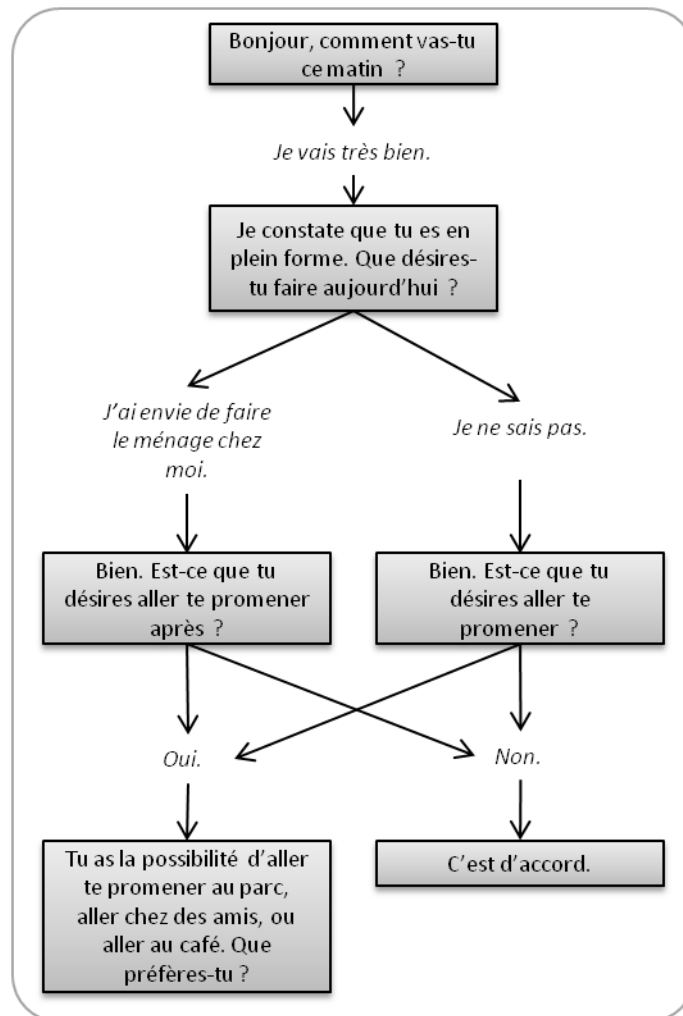


Figure 27. Expérimentation IDV-HR : Extrait du scénario du réveil du matin (en forme). Énonciations prédéterminées du robot, et réponses possibles du participant.

Le cours de l'expérimentation est ponctué de phases de questionnaires, proposés à l'oral par l'expérimentateur en charge de l'accueil des participants. Un premier questionnaire porte sur des informations sociologiques telles que l'âge, le sexe, la profession. Puis, après chaque série de cinq scénarios, l'expérimentateur demande au participant d'indiquer l'attitude sociale qu'il aura reconnue chez le robot. Cinq questions lui sont donc posées, pour lesquelles le participant répond librement : « Avez-vous trouvé le robot (compassant | encourageant | sûr de lui | aimable | directif) ? ».

Le questionnaire final s'oriente sur des informations plus générales, d'un point de vue émotionnel tout d'abord, puis interactionnel avec le robot. Nous interrogeons le participant sur les émotions qui sont pour lui les plus fréquentes, et d'autres émotions qu'il serait amené à exprimer le matin et qui n'ont pas été évoquées dans le scénario. Ces questions ont pour but d'évaluer l'adéquation de notre sélection de scénarios avec ce que des utilisateurs finaux pourraient nécessiter.

Afin d'évaluer l'interaction avec le robot, nous demandons ensuite s'il estime que Nao l'a bien compris, et s'il répondait d'une façon adaptée aux situations. Nous demandons également au participant si en cas d'urgence, il estime que le robot doit prendre les choses en main. Afin de comprendre plus précisément la

relation pouvait s'établir entre le robot et l'utilisateur, nous avons également ajouté des questions portant sur des composantes interactionnelles de plus haut niveau :

- « Est-ce que vous avez aimé qu'il vous tutoie ? »
- « Est-ce que vous trouvez que le robot répondait d'une façon adaptée aux situations ? »
- « Imaginez que ce robot soit à vous, est-ce que vous lui donneriez un nom ? Lequel ? »
- « En cas d'urgence, est-ce que vous estimez que le robot doit prendre les choses en main ? »
- « Aimerez-vous avoir un robot pour vous aider dans la vie de tous les jours ? »
- « Dans quelles situations est-ce que le robot serait le plus approprié dans votre vie quotidienne ? »

Ces questions, très brèves et au panel de réponse possible extrêmement large, ont simplement pour but d'illustrer la relation pouvant s'établir entre l'utilisateur et le robot tel que nous leur avons présenté au cours de cette expérimentation. Cette partie du questionnaire permet également au participant de se détendre suite à l'expérimentation, et de dialoguer avec l'expérimentateur.

Une première passe d'enregistrement (les huit locuteurs de la première journée d'expérimentation) nous a permis de tester les différentes stratégies du robot, et d'affiner le discours du robot afin qu'il soit plus impactant.

5.3.2.2. Annotation et scores

Au cours de l'expérimentation IDV-HR, nous avons recueilli les enregistrements de 22 participants, parmi lesquels 11 femmes et 11 hommes. La moyenne d'âge est de 58 ans (maximum 80 ans et minimum 28 ans). Le corpus final est composé de 6071 segments, segmentés à partir des enregistrements audio selon le protocole de segmentation présenté en Annexe III. L'annotation du corpus est réalisée par deux annotateurs, en deux étapes. Tout d'abord, nous avons utilisé les enregistrements des huit premiers locuteurs afin de déterminer les critères d'annotation les plus pertinents. La seconde partie du corpus a été annotée à l'aide d'un schéma plus réduit.

Les huit premiers enregistrements ont été annotés à l'aide de :

- trois étiquettes **émotion** (voir Tableau 20),
- une **macro-classe** parmi « positif », « négatif », « colère », « tristesse », « peur », « neutre », « positif négatif »,
- une **valence** (« positive », « négative », « neutre », « ambiguë », « positif négatif »),
- un niveau d'**activation** de 1 à 5,
- un niveau de **contrôle** de 1 à 5,
- le **déclencheur** de l'expression émotionnelle (« robot », « expérimentateur », « soi-même », « autre »),
- la **direction** de l'expression émotionnelle (« robot », « expérimentateur », « soi-même », « autre »),

- la **phase** (« scénario » ou « libre ») afin de déterminer si le discours du participant est contraint ou pas.

Tableau 20. Corpus IDV-HR : Étiquettes d'annotation et regroupement en macro-classes.

Macro-classe de l'émotion	Valeur d'annotation	Macro-classe de l'émotion	Valeur d'annotation
JOIE	Joie	PEUR	Peur
	Amusement		Inquiétude
	Satisfaction		Stress
Gêne			
COLERE	Colère	NEUTRE	Neutre
	Agacement		Positif
TRISTESSE	Tristesse		Négatif
	Déception		Surprise
AUTRES	Excitation	Surprise positive	Surprise négative
	Mépris	Compassion	Empathie
	Ironie	Maternage	Ennui

L'annotation des enregistrements des locuteurs 9 à 22 a été réalisée par deux annotateurs à l'aide de ces étiquettes et dimensions :

- une **macro-classe** parmi « joie », « colère », « tristesse », « peur », « neutre »,
- un niveau d'**activation** de 1 à 5,
- le **déclencheur** de l'expression émotionnelle (« robot », « expérimentateur », « soi-même », « autre »),
- la **direction** de l'expression émotionnelle (« robot », « expérimentateur », « soi-même », « autre »),
- la **phase** (« scénario » ou « libre »).

Scores d'annotation

Comme nous l'avons vu pour les corpus précédents, l'annotation des émotions est sujette à un faible taux d'accord entre les annotateurs dès lors que nous considérons les étiquettes fines. Nous étudions alors les annotations de la macro-classe. Le Tableau 21 présente la matrice de confusion de l'annotation de la macro-classe pour chaque segment des enregistrements 1 à 8. À partir de cette matrice, nous avons calculé le score Kappa représentant l'agrément entre les annotateurs, $K = 0,7$. Nous trouvons dans le Tableau 22 les résultats d'annotation de la macro-classe pour les locuteurs 9 à 22. Le Kappa calculé sur cette matrice est de $K = 0,25$. Cette grande différence entre les scores d'agrément pour les deux parties du corpus s'explique par

L'application apportée à l'annotation de la première partie, pour laquelle les annotateurs ont procédé à des ré-annotations. Si nous réduisons les annotations de la seconde partie du corpus à une distinction entre « positif » (Joie) et « négatif » (Tristesse, Colère, Peur), nous obtenons un score d'agrément de $K = 0,8$.

Tableau 21. Corpus IDV-HR : Matrice de confusion de l'annotation de la macro-classe pour les enregistrements des huit premiers locuteurs.

	Positif	Négatif	Colère	Tristesse	Peur	Neutre	Pos-Nég	Total
Positif	280	0	5	1	0	157	0	443
Négatif	0	339	5	1	1	26	0	372
Colère	1	3	79	1	0	57	0	141
Tristesse	3	4	1	114	0	44	0	166
Peur	13	0	6	2	19	95	0	135
Neutre	17	8	2	0	0	862	0	889
Pos-Nég	0	0	0	0	0	0	52	52
Total	314	354	98	119	20	1241	52	2198

Tableau 22. Corpus IDV-HR : Matrice de confusion de l'annotation de la macro-classe pour les enregistrements des locuteurs 9 à 22.

	Joie	Tristesse	Colère	Peur	Neutre	Total
Joie	418	6	29	0	595	1048
Tristesse	7	120	55	0	328	510
Colère	27	16	189	0	363	595
Peur	32	41	110	48	549	780
Neutre	40	3	25	0	856	924
Total	524	186	408	48	2691	3857

L'annotation de l'activation présente, comme pour les corpus précédemment décrits, un faible taux d'accord brut, de $K = 0,2$. La réduction de l'annotation à des pôles « très faible », « moyen », et « très fort » permet d'obtenir un score Kappa de $K = 0,45$. Notons toutefois que la quantité de segments annotés comme présentant une activation « moyenne », et pour lesquels les annotateurs s'accordent, représente plus de 90% du corpus (3711 segments), ce qui constitue un déséquilibre dans l'homogénéité du corpus et peut grandement expliquer ce score si bas.

5.3.2.3. Résultats

L'analyse des données obtenues (questionnaire et annotations audio) nous ont permis de déterminer les attentes des participants déficients visuels en termes de robotique, et leurs réactions émotionnelles en fonction de l'attitude sociale du système.

Analyse des questionnaires

Cohérence émotionnelle des scénarios. A la question « Quelles autres émotions pourriez-vous exprimer au lever ? », quelques états émotionnels ont été cités par les participants (parmi lesquels nous trouvons l'agacement, l'inquiétude), mais tous ces états ont déjà été pris en compte dans les scénarios. Nous noterons juste l'état physique de « Fatigue » proposé par 3 participants sur les 22 interrogés, mais qui ne nous intéresse pas dans le cadre de notre étude actuelle sur les émotions.

Nous retrouvons dans le Tableau 23 les résultats des autres questions posées aux participants.

Crédibilité du robot. Nous désirions savoir si, globalement, l'attitude et les énonciations du robot semblaient cohérentes pour le participant, et si le robot leur semblait comprendre les situations jouées. En d'autres termes, si le participant se doutait qu'il s'agissait d'un système Magicien d'Oz ou non. Si le participant se doute que le robot n'est qu'une boîte vide lors de l'expérience, il est alors peu probable qu'il s'implique dans le dialogue avec le robot, et soit sensible à ses différences de comportement.

A la question « Durant l'expérience, pensez-vous que Nao vous a bien compris ? », 77% des participants trouvaient que le robot comprenait correctement, voire très bien, et 18% ont déclaré ne pas savoir. A la question « Estimiez-vous que le robot répondait d'une façon adaptée aux situations ? », 59% ont estimé que les réponses du robot étaient bien ou très bien adaptées. Seuls 14% ont estimé que les réponses n'étaient pas du tout adaptées, et 27% n'avaient pas d'avis. Nous estimons donc que globalement, les participants ont cru à la compréhension du robot, et à son adaptation lors des scénarios joués.

Affinité avec le robot. Savoir si l'utilisateur refuse catégoriquement le tutoiement de la part du robot nous indique quelle relation sociale il souhaite instaurer. Refuser le tutoiement indique une volonté claire d'établir des distances, d'instaurer une relation de supériorité sociale. Parmi les réponses obtenues à notre questionnaire, 32% des participants souhaitaient explicitement que le robot les tutoie. Pour les 68 autres pourcents, qui ont choisi l'option « peu importe », nous considérons que ce choix indique une certaine neutralité affective par rapport au robot.

Proposer aux participants de donner un nom au robot nous permet de discerner également le degré d'affinité qu'ils établissent avec le robot suite à leurs interactions, mais également savoir comment ils se projettent dans une éventuelle relation. Nous avons catégorisé les noms obtenus en quatre catégories thématiques : « amical », « imaginaire », « technique », et « assistance ». La catégorie « technique » est prédominante (36%). Dans cette catégorie, nous incluons des réponses telles que tout simplement « NAO », ou bien « un numéro de code », « pas un nom humain ». Ce choix indique une prise de distance affective avec le robot. Vient ensuite la catégorie « amical », dans laquelle nous trouvons des prénoms humains attribués par les participants au robot (« Marcel », « Nicolas », « Gérard », « Mario », « Robert », « Jacques », « Siméon »). Ces noms, même si parfois attribués avec dérision, montrent une démarche de personnification du robot et peut amener à l'établissement d'un lien social. Les surnoms de la catégorie « imaginaire », proposés dans 18%

des cas, regroupent par exemple « E.T. », « Ouistiti », ou « Peluche ». Nous considérons ces appellations comme des marques d'humour de la part des participants, pour lesquelles il est difficile d'établir s'il y a lien affectif ou pas. Dans le cas de la catégorie « assistance » (9%), les dénominations du robot sont de type « L'aigle volant », « Mon bras droit », ou « Mes yeux ». Ces termes mettent en avant le désir fort des participants d'avoir un robot efficace et sachant répondre à leurs demandes concrètes, ce qui n'exclut bien sûr pas une éventuelle relation affective, mais ne l'implique pas nécessairement.

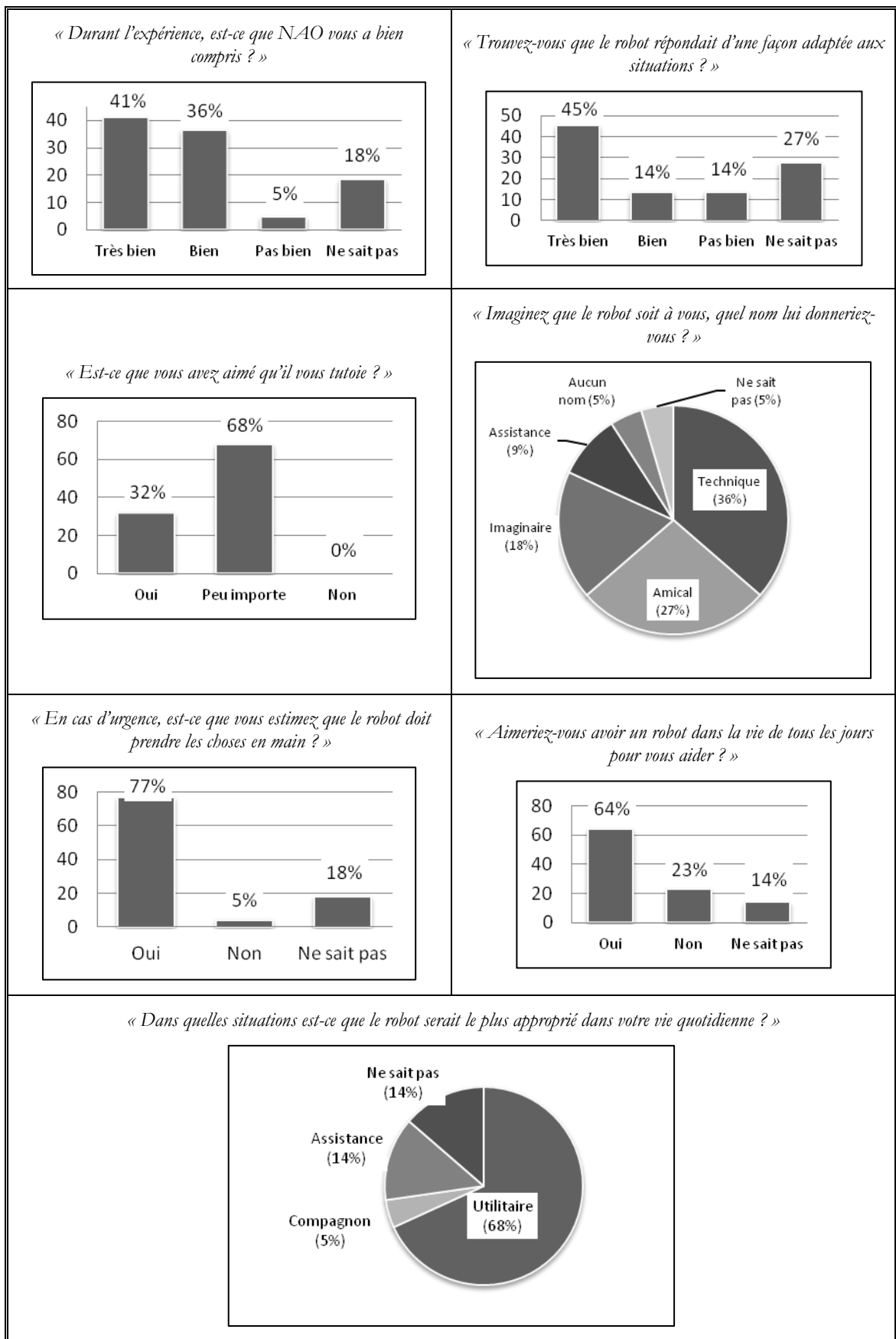
Fonctionnalités du robot. Les trois dernières questions du questionnaire visent à caractériser les fonctionnalités attendues de la part du participant. 77% des participants s'accordent sur le fait qu'en cas d'urgence, le robot doit pouvoir « prendre les choses en main ». Cette question confirme l'intuition que les participants peuvent accepter – a priori – de renoncer à une part d'autonomie dans un cas d'urgence, dans la mesure où ils sont face à un dispositif dont la vocation est d'assister.

Le désir d'avoir un robot à domicile pour les aider est présent chez 64% des participants interrogés, alors que 23% n'en souhaitent pas, et 14% n'ont pas d'avis défini. Parmi notre panel de participants, dont la moyenne d'âge est 58 ans, nous constatons que 80% des moins de 58 ans désireraient un robot à domicile, contre 50% parmi les plus de 58 ans.

Les fonctionnalités proposées par les participants peuvent être regroupées en trois catégories :

- Utilitaire (68%) : le participant attend tout d'abord du robot qu'il réalise des tâches d'aide à la vie quotidienne en relation avec son handicap, comme « retrouver un objet », « lire (manuscrit, courrier, journal...) », mais également des activités plus générales telles que « pense-bête », « faire le ménage »,
- Assistance (14%) : le robot pourrait réaliser les tâches que l'âge ou le handicap rendent difficiles – voire dangereuses –, telles que mettre un plat au four. Le robot serait également susceptible de détecter rapidement un malaise, chez les personnes vivant seules, et d'intervenir en cas d'urgence,
- Compagnon (5%) : seul un participant a évoqué l'idée que le robot puisse faire office de compagnon à domicile.

Tableau 23. Expérimentation IDV-HR : Résultat des questionnaires participant



Reconnaissance des comportements robotiques. Nous avons expérimenté différents comportements sociaux pour le robot : « Directif », « Hésitant », « Neutre », « Amical », « Empathique » et « Encourageant ». Il est important de déterminer si le participant a su différencier ces différents comportements, et s'il a été capable de les identifier. D'après les questions posées à la fin de chaque série de scénarios, la distinction fine – à savoir, identifier précisément ce qui était par exemple « Empathique » ou « Amical » – semble délicate, et présente un faible taux d'identification. Cependant, 64,3% des participants ont correctement identifié que les comportements positifs du robot étaient en effet parmi ceux que nous estimions socialement désirables, et 72% des participants ont bien reconnu les comportements négatifs comme tels. La distinction entre macro-catégorie « socialement désirable » et « socialement indésirable » a donc bien été réalisée par les participants.

Analyse des attitudes émotionnelles en réaction aux comportements du robot

Nous avons évalué l'impact des comportements sociaux positifs et négatifs du robot sur la valence (émotion positive versus négative) et l'activation (force de l'émotion) exprimées par les participants. Les scénarios sont divisés en deux types : les situations positives (« content » et « en pleine forme ») et les situations négatives (« déprimé », « malade », « urgence »). Nous nous attendons donc à ce que les participants expriment plus d'émotions négatives au cours des scénarios négatifs, et des émotions positives dans les scénarios positifs. Néanmoins, nous nous attendons à observer également une différence notable selon que le robot adopte un comportement socialement désirable, ou indésirable.

Il est important de noter que les locuteurs que nous avons enregistrés n'étaient pas des acteurs professionnels. Ces personnes doivent jouer un rôle certes, mais elles ne contrôlent pas nécessairement leurs expressions d'émotions, tel qu'un acteur le ferait. Nous partons donc du principe que, bien que les sujets aient des consignes quant à l'émotion à jouer, les comportements du robot and ses types de réponses auront un impact notable sur leurs expressions d'émotions. Par exemple, nous avons remarqué qu'un participant pourra avoir du mal à jouer un scénario déprimé de façon convaincante, s'il trouve le comportement du robot amusant, ou s'il est gêné.

La Figure 28 présente la valence moyenne exprimée par les participants au cours des sessions d'enregistrement. La valence a été ramenée à une valeur de 0 (négatif) à 1 (positif), en passant par une valeur neutre de 0,5. Nous nous basons sur les segments pour lesquels les annotateurs s'accordent quant à la distinction « positif »/ « négatif ». Nous remarquons une différence en fonction du comportement du robot : lorsque le robot adopte un comportement jugé socialement indésirable, les participants tendent à exprimer des émotions plus neutres, ou plus négatives. Durant les scénarios positifs, la valence décroît de 9%, et de 5% pour les scénarios négatifs.

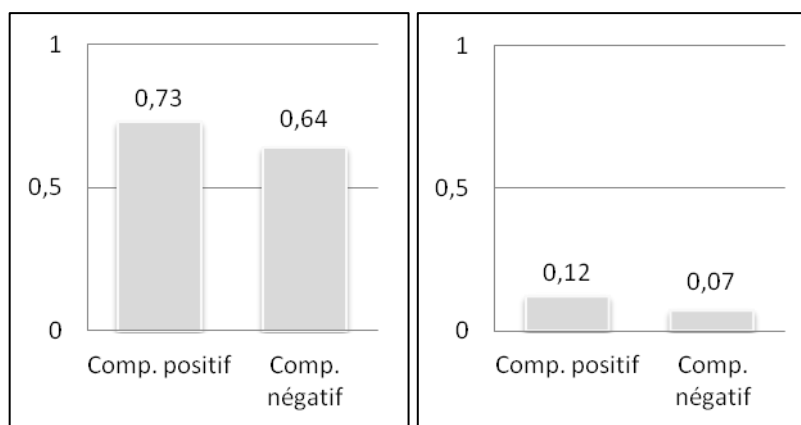


Figure 28. Expérimentation IDV-HR : Valence moyenne exprimée par les participants (de 0/négative à 1/positive) en fonction des comportements positifs ou négatifs du robot. Figure de gauche : scénarios positifs ; figure de droite : scénarios négatifs.

L'analyse de l'activation moyenne exprimée par les participants (voir Figure 29), basée sur les 2313 segments sur lesquels les annotateurs s'accordent quant au niveau d'activation, nous permet de constater que l'activation croît légèrement, en moyenne, lorsque le robot présente un comportement négatif, et ce quelque soit le type de scénario joué.

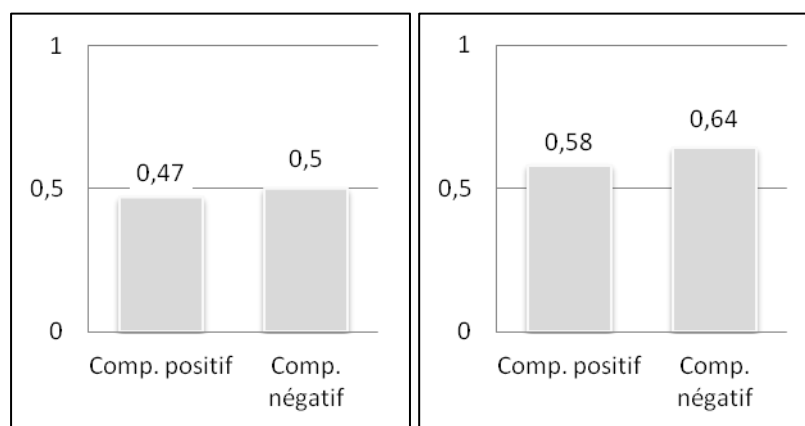


Figure 29. Expérimentation IDV-HR : Activation moyenne exprimée par les participants (de 0/faible à 1/fort) en fonction des comportements positifs ou négatifs du robot. Figure de gauche : scénarios positifs ; figure de droite : scénarios négatifs.

Nous observons donc que les comportements indésirables du robot, dans ce contexte, ont un impact sur le type et la force des émotions exprimées par les participants : la valence tend à être plus négative, et l'activation augmente. Nous nous attendions à ce que les stratégies sociales négatives du robot aient un impact sur le comportement émotionnel des participants. Toutefois, les écarts sont mineurs (6% en moyenne). Il apparaît évident que le contexte in vitro (pas de tâche concrète à effectuer, et conditions expérimentales) atténue les réactions des participants.

5.3.3. Conclusion

Nous avons dans cette section étudié les réactions émotionnelles des sujets face aux différents comportements sociaux du robot. Dans les études sur lesquelles nous nous sommes appuyés, les sujets

interagissent avec le robot dans deux contextes applicatifs différents : les enfants jouent à des jeux avec le robot, et les adultes déficients visuels jouent des scénarios de la vie quotidienne où le robot pourrait venir les assister. La codification des comportements sociaux du robot était similaire dans les deux cas, avec toutefois – évidemment – des manifestations adaptées aux situations d'interaction.

L'analyse des enregistrements nous a permis, malgré le faible nombre de locuteurs étudiés, de constater que les comportements indésirables du robot, dans un contexte d'assistance, ont un impact négatif sur l'expression d'émotion des adultes : la valence de leurs émotions tend à être plus négative, même s'ils jouent un scénario dans lequel nous attendions d'eux des émotions positives. Chez les enfants, quant à eux dans un contexte de jeu, les comportements que nous jugions initialement indésirables n'ont pas déclenché nécessairement d'émotions négatives, mais étaient plutôt perçus comme drôles par les enfants. Bien que l'impact soit modéré dans des conditions expérimentales, cela révèle très certainement les tendances émotionnelles des utilisateurs in vivo. Ces résultats motivent la prise en compte, dans l'élaboration du profil émotionnel et interactionnel de l'utilisateur, de l'impact du comportement du robot sur l'expression d'émotion de l'utilisateur.

L'augmentation de l'activation, dans les deux contextes d'interaction, pourrait être interprétée comme une augmentation du niveau d'engagement du locuteur envers le robot. Cependant, le type d'engagement que nous constatons dans nos corpus n'est pas celui que nous souhaiterions dans une Interaction Homme-robot : soit le participant se moquait du robot, ou bien essayait de parler plus clairement et fortement afin de se faire comprendre du robot, ou exprimait de l'irritation.

Cette expérimentation a permis de mettre en avant l'importance de la tâche et probablement de la tranche d'âge du public pour la sélection de ce qui serait un comportement socialement « désirable » ou « indésirable » pour un robot. En effet, l'utilisateur n'attendra pas les mêmes performances dans l'accomplissement d'une tâche de jeu, ni la même attitude sociale chez le robot, que lorsqu'il sera en situation d'assistance, où le robot devra proposer une réalisation concrète de la tâche, et une compréhension efficace de l'humain.

Notons enfin, en dernier point, l'importance des questionnaires d'évaluation des participants. Dans IDV-HR, le questionnaire rempli par les participants nous indique qu'il existe une bonne corrélation entre le contenu lexical choisi et le type de comportements reconnu par le participant, ce qui valide notre choix de modalité (contenu lexical fortement marqué) dans l'élaboration de comportements robotiques pour un public déficient visuel.

5.4. Relation entre indices paralinguistiques et profil émotionnel et interactionnel

Nous analysons maintenant, sur corpus, la relation entre indices paralinguistiques (activation, valence, durée de parole) annotés et le profil émotionnel et interactionnel des locuteurs. Les données utilisées étant complexes et en petit nombre, il convient de comprendre les résultats obtenus en termes de tendances. L'annotation des émotions est en soi un challenge : il est en effet difficile de trouver un consensus quant à la perception et à l'étiquetage des émotions. La difficulté de cette étude est encore renforcée par le défi représenté par l'analyse perceptive du profil émotionnel et interactionnel, qui consiste à réaliser une interprétation encore plus poussée du comportement social et émotionnel de l'individu. Il est donc impératif d'établir un lien clair et étayé entre les différents indices et le profil.

Pour réaliser cette analyse, il convient bien sûr de disposer d'un nombre convenable de données. Un regroupement de nos corpus peut être envisageable sous certaines conditions. Le profil émotionnel de l'individu n'est pas dépendant de la personne à laquelle il s'adresse, puisqu'il s'agit de tendances émotionnelles observables sur du plus long-terme (rappelons que les dimensions du profil émotionnel se rapprochent du concept de personnalité). Le profil interactionnel, au contraire, est totalement lié aux qualités de l'interlocuteur, puisqu'il s'agit d'une estimation de « Est-ce que je l'apprécie ou pas ? », et « Est-ce que je le considère comme socialement supérieur ou inférieur à moi ? ».

Nous estimons donc que l'étude du profil émotionnel est possible sur nos deux corpus IDV-HS et IDV-HR, car ils présentent tous deux un même public – des adultes déficients visuels. Nous obtenons donc un échantillon d'étude de 48 locuteurs. Toutefois, nous réalisons l'étude des dimensions d'Affinité et de Domination (le profil interactionnel) uniquement sur le protocole d'expérimentation le plus abouti en termes de codification de l'interaction Humain-Robot : le protocole IDV-HR (22 sujets), qui met en scène des adultes déficients visuels interagissant avec le robot NAO dans le cadre de scénarios de la vie quotidienne.

Nous présentons dans cette section une analyse des annotations réalisées par deux annotateurs, et du corpus de travail obtenu. Nous proposons ensuite un protocole d'analyse statistique des données, qui permettrait d'établir la pertinence des indices choisis pour la description des valeurs des dimensions du profil.

5.4.1. Interaction Adulte-Robot et Adulte-Système

Le protocole d'acquisition du corpus IDV-HR, détaillé dans la section 5.3.2, consiste à mettre en interaction des adultes déficients visuels avec le robot NAO. Les participants se mettent dans des situations de la vie quotidienne dont la thématique leur est donnée, et le robot adopte différents comportements sociaux pour venir les assister. Le corpus IDV-HS (décrit dans la section 5.2.1) est constitué d'enregistrements audio de personnes en perte d'autonomie confrontés à un système de détection d'émotions déclaré comme étant "en test". Dans ce cadre, les sujets se projettent dans une situation d'assistance robotique donnée, où l'état affectif exprimé doit être reconnu par le système (dirigé en Magicien d'Oz).

L'annotation des corpus IDV-HR et IDV-HS est détaillée dans les sections citées ci-dessus. Nous nous appuyons sur les annotations de la Valence et de l'Activation, et sur la durée du temps de parole déterminée grâce à la segmentation des pistes audio des locuteurs. Nous décrivons ci-après le protocole d'annotation perceptive du profil émotionnel et interactionnel.

5.4.1.1. Annotation perceptive du profil émotionnel et interactionnel

Nous avons mentionné précédemment le fait que, lors de l'annotation des émotions sur des enregistrements de voix, l'annotateur s'efforce de faire abstraction du contenu linguistique afin de se concentrer sur les caractéristiques de voix ayant conduit au ressenti de l'émotion exprimée. Dans le cas de l'annotation du profil émotionnel et interactionnel, au contraire, nous nous servons de l'intégralité des informations à notre disposition, qu'il s'agisse de ce que le sujet a dit, ou de comment il l'a dit. Nous pourrions ensuite tenter d'établir un lien entre cette interprétation haut niveau, et les indices paralinguistiques annotés. L'annotation des valeurs du profil émotionnel et interactionnel des sujets se fonde sur une série d'indices perceptifs qu'il est possible de repérer à l'écoute des enregistrements audio.

L'annotation du profil émotionnel a été réalisée sur 48 locuteurs. Le degré d'Optimisme, tout d'abord, est reconnu grâce au fait que le sujet exprime beaucoup d'émotions positives, même dans les situations négatives. Le sujet utilise un ton enjoué une grande partie du temps, et rit. Nous reconnaissons la Confiance en soi au fait que le sujet s'exprime de façon déterminée, directe. Il exprime peu d'hésitations, n'emploie pas de rires gênés. Il peut, par exemple, imposer ses points de vue. Le degré d'Émotionnalité est évalué en fonction des variations dans l'expression du sujet : si les émotions de ce dernier varient fortement ou, au contraire, peu. Dans un cas où l'Émotionnalité est forte, on peut avoir l'impression que le locuteur se laisse submerger par ses émotions. L'Extraversion, enfin, peut être évaluée en fonction du niveau de volubilité du sujet, et sa propension à exprimer des émotions fortes. Le sujet extraverti a tendance à ne pas hésiter à exprimer ses émotions.

Pour chacune des dimensions du profil émotionnel, les valeurs d'annotation possibles sont :

- 1 : peu optimiste, peu sûr de lui, très stable émotionnellement, peu extraverti ;
- 2 : moyennement optimiste, etc. ;
- 3 : très optimiste, etc.

L'annotation du profil interactionnel a été réalisée sur les 22 locuteurs du corpus IDV-HR. Nous limitons notre étude aux phases durant lesquelles le participant interagissait avec le robot dans le cadre des scénarios. Comme nous l'avons mentionné précédemment, l'Affinité et la Dominance varient en fonction de l'interlocuteur. Nous notons en effet que les participants ne parlent pas de la même façon à l'expérimentateur qu'au robot. Nous remarquons par exemple, chez certains sujets, une hyperarticulation lorsqu'ils s'adressent au robot, ou encore un ton plus maternalisant.

Les indices perceptifs nous permettant d'évaluer le degré d'Affinité du locuteur pour son interlocuteur sont les suivants : une voix « chaude », amicale, un ton enjoué, sont autant de signes que le sujet apprécie le robot (ou, à tout le moins, qu'il part d'un a priori positif à son sujet). Au contraire, un ton froid, monocorde, indique une absence d'Affinité. Nous ne parlons pas dans ce cas-là « d'hostilité », – puisque le cas ne s'est pas présenté lors de nos expérimentations –, mais plutôt juste de « froideur ».

La Domination peut être reconnue grâce à un ton direct de la part du sujet, une absence d'hésitations lorsqu'il s'adresse au robot pour lui donner une consigne. Le fait de donner des ordres pourra également constituer un indice, au même titre que parler d'une voix forte et claire.

Voici les différentes valeurs numériques d'annotation pour ces deux dimensions, et leur signification :

- Affinité : Froid (1), Indifférent (2), Amical (3)
- Domination : Passif (1), Neutre (2), Dominant (3)

5.4.1.2. Scores d'annotation

Les scores Kappa permettant d'évaluer l'accord entre les deux annotateurs sont très bas, et ce pour chaque dimension du profil. Nous remarquons un premier point : les confusions se trouvent généralement sur les classes 2 et 3. Les annotateurs s'accordent sur le fait qu'il est facile de déterminer qu'une personne est **très peu** optimiste, **très peu** émotionnelle, etc., mais que la distinction entre **moyennement** et **fort** est

beaucoup plus difficile. Nous avons donc opéré un regroupement des valeurs 2 et 3, en opposition à la valeur 1. L'évaluation du profil consiste donc en deux valeurs : par exemple « peu optimiste » versus « optimiste ».

Nous remarquons d'autre part que ce regroupement n'est pas suffisant pour calculer le score Kappa d'accord entre les annotateurs. Pour chaque dimension, si la proportion d'accord observé est correctement élevée (en moyenne de 70%), le Kappa reste toutefois faible (inférieur à 0,5). Les données ne sont pas équilibrées pour toutes les dimensions étudiées : les locuteurs présentant des valeurs faibles pour les dimensions du profil sont sous-représentés dans notre corpus (de l'ordre d'un locuteur « faible » pour 10 locuteurs « forts », soit environ 4 locuteurs), ce qui crée un déséquilibre dans le traitement statistique de nos données et exclut un rééquilibrage par retrait aléatoire d'échantillons dans la classe sur-représentée. Nous nous limiterons donc dans ce cadre à l'analyse de la proportion d'accord observé, lorsque le Kappa n'est pas pertinent. Nous présentons les scores d'agrément dans le Tableau 24.

Nous remarquons que la proportion d'accord observé est en moyenne de 80% pour toutes les dimensions du profil, avec un minimum de 64% pour l'Affinité, et 79% pour l'Émotionnalité et l'Optimisme. Le score Kappa est en moyenne de 0,4. Seule la Domination présente un bon score Kappa et un accord observé très élevé.

Tableau 24. Scores d'agrément pour l'annotation du profil émotionnel et interactionnel sur les corpus IDV1 et IDV2 : proportion d'accord observé et Kappa.

Dimension du profil	Proportion d'accord observé
Émotionnalité	0,79
Confiance en soi	0,75
Optimisme	0,79
Extraversion	0,71
Affinité	0,64
Domination	0,91

Comme nous l'avons signalé, la classe « faible » pour chaque dimension est sous-représentée dans notre corpus d'étude. Si, pour chaque dimension, nous considérons uniquement les locuteurs pour lesquels les annotateurs s'accordent sur la valeur de la dimension, et que nous veillons à une homogénéité dans la distribution des classes, notre base de travail se limiterait à 9 échantillons pour l'Émotionnalité, 2 pour la Confiance en soi, 6 pour l'Optimisme, 4 pour l'Extraversion, 2 pour l'Affinité et 4 pour la Domination (voir Tableau 25).

Tableau 25. Corpus IDV-HS et IDV-HR : Nombre d'échantillons pour les classes faibles et fortes des différentes dimensions du profil émotionnel et interactionnel. Restriction aux échantillons avec accord annotateurs.

Dimension	#échantillons classe « faible »	#échantillons classe « fort »	Corpus
Émotionnalité	9	29	IDV-HS + IDV-HR
Confiance en soi	2	34	IDV-HS + IDV-HR
Optimisme	6	32	IDV-HS + IDV-HR
Extraversion	4	30	IDV-HS + IDV-HR
Affinité	2	12	IDV-HR
Domination	4	16	IDV-HR

D'un point de vue perceptif, nous pouvons signaler quelques points ayant rendu l'annotation du profil émotionnel et interactionnel délicate. Tout d'abord, le fait qu'il faille « moyenner » l'intégralité d'une session d'interaction, même si cela était pertinent afin d'avoir une représentation globale du sujet, a engendré des prises de décision complexe pour les annotateurs.

D'autre part, par exemple, l'annotation de l'Affinité, en s'appuyant sur les critères prédéfinis, s'est avérée compliquée dans le cas d'un participant masculin qui n'exprimait un haut degré Affinité que dans le choix de ses appellations du robot (par exemple, « mon grand », « mon pote »), c'est-à-dire dans le lexique, alors que son ton neutre tendait à transmettre une information contraire. Dans un cas où une ironie éventuelle de la part du sujet est possible, le consensus entre les annotateurs est malaisé.

Dans le cas encore de l'Affinité, nous avons remarqué que cette dimension pouvait évoluer au fil de l'interaction : un participant gêné face à NAO lors de ses premières prises de parole peut se mettre à dialoguer plus chaleureusement avec le robot lorsque celui-ci a posé quelques questions. Cette évolution rend l'évaluation perceptive globale difficile.

5.4.1.3. Protocole pour une analyse statistique

Différentes analyses statistiques nous permettraient d'évaluer la pertinence des indices choisis pour la détermination du profil émotionnel et interactionnel. Notre échantillon d'étude, cependant, ne nous permet pas de mettre en œuvre ces analyses, car comme nous l'avons décrit dans la section précédente, la sous-représentation de la classe « faible » ne nous permet pas de disposer d'un échantillon homogène. Les

évaluations ne sont donc pas pertinentes, voire impossibles à effectuer dans les cas où la classe ne possède que deux échantillons. Nous proposons cependant un protocole qu'il est possible de mettre en œuvre pour analyser la pertinence de la relation entre indices paralinguistiques et dimensions du profil.

Tout d'abord, une analyse de nos données d'annotation des indices paralinguistiques nous permet d'identifier des caractéristiques pour chaque locuteur. Un sujet pourra donc être représenté par :

- une durée du temps de parole : totale, moyenne et écart-type ;
- une activation : moyenne, écart-type ;
- une valence : moyenne, écart-type.

Ces 7 valeurs peuvent être organisées pour représenter le profil du sujet en se basant sur la corrélation proposée dans la section 3.3.2.2 (Interprétation émotionnelle et interactionnelle des informations paralinguistiques). Par exemple, l'Extraversion moyenne d'un sujet peut être évaluée en fonction de la moyenne des activations, et la moyenne des durées de tour de parole.

Chaque sujet peut donc être considéré comme un échantillon statistique, pour lequel nous avons repéré une variable continue (la dimension du profil) que l'on suppose décrite en fonction de plusieurs autres variables explicatives continues ou catégorielles (les indices paralinguistiques). L'étude de la variation d'une variable en fonction de plusieurs facteurs, et de la pertinence de chacun de ces facteurs, peut être réalisée à l'aide d'une analyse de covariance ANCOVA.

Cette analyse nous permettrait de comparer :

- que les indices choisis sont pertinents, via le F-Test notamment, qui renseigne sur le fait que les indices apportent une quantité d'information significative ;
- que certains, parmi les indices choisis, sont plus pertinents que les autres dans le modèle, via l'analyse Type I SS ;
- que ces indices, plutôt que d'autres, sont pertinents, en lançant l'analyse ANCOVA sur différents sets d'indices paralinguistiques (par exemple, est-ce que l'Affinité pourrait encore mieux être décrite par la durée du temps de parole que par les variations de la valence ?).

Des tests de classification seraient également susceptibles de valider la relation entre les indices paralinguistiques et la variation des valeurs du profil. Cependant, dans ce contexte plus encore qu'avec l'analyse ANCOVA, le manque d'échantillons pose problème puisqu'il faudrait diviser nos données en des bases d'apprentissage et de test.

Il serait possible par exemple de réaliser une classification supervisée, dans le cadre de laquelle nous pourrions étudier les probabilités que telle série d'indices permette de prédire la classe à laquelle appartient le sujet. Cette évaluation, réalisée à l'aide de classifieurs Bayésiens par exemple, se fonderait sur les mêmes procédés d'analyse que dans le cadre de l'analyse ANCOVA :

- évaluation de la pertinence des indices ;
- évaluation de la pertinence de certains de ces indices plutôt que les autres ;

- évaluation de la pertinence de ce set d'indices plutôt qu'un autre set.

Obtenir un bon score de classification signifierait donc que nous avons trouvé un modèle valide pour décrire les valeurs des dimensions du profil en fonction de sets d'indices spécifiques. A partir de cette constatation, nous pourrions inférer que les indices choisis sont les bons, qu'ils sont les plus pertinents.

5.4.2. Conclusion

L'annotation du profil émotionnel a représenté pour les annotateurs un travail délicat. En effet, les dimensions à annoter se basent sur une grande quantité de critères, – tant linguistiques, qu'émotionnels, ou psychologiques –, qu'il est difficile d'analyser de concert sur toute la durée d'une session.

Nous pourrions certainement envisager d'avoir recours à un protocole plus approfondi pour évaluer le profil émotionnel et interactionnel du sujet, via les procédés courants en psychologie : une auto-évaluation via questionnaire, ou un questionnaire renseigné par l'expérimentateur expert basé sur quelques critères extrêmement précis et représentatifs. L'annotation perceptive du profil émotionnel et interactionnel demeure cependant un challenge intéressant sur lequel nous allons continuer à nous pencher.

Nous avons proposé quelques pistes d'analyses statistiques susceptibles de valider la relation entre les indices et les profils : une étude statistique de la covariance, et la recherche d'un modèle de classification automatique. Cependant, l'étude de la relation entre indices paralinguistiques et profil nécessiterait d'acquérir un plus grand corpus avec plus de locuteurs. Nous notons enfin que, comme pour toute annotation perceptive et subjective, il serait intéressant de confronter les annotations de plus de codeurs (naïfs et experts).

5.5. Conclusion

Au cours de ces travaux de thèse, nous avons mené une série d'expérimentations pour étudier les différents modules de la boucle d'interaction entre l'utilisateur et le robot. Ces études ont mené à la constitution de corpus, dont les protocoles d'annotation et les récapitulatifs sont présentés en détail en Annexe III, Annexe IV et Annexe V. Nous nous sommes intéressés à deux contextes applicatifs, et deux publics : les enfants jouant avec un robot compagnon, et les adultes en perte d'autonomie en interaction avec un robot social d'assistance. Un point intéressant et original de cette thèse est la constitution de corpus avec des participants susceptibles d'être de réels *end-users*. En effet, les personnes déficientes visuelles ayant participé à nos enregistrements sont des volontaires recrutés par l'Institut de la Vision intéressés par les technologies de demain liées à leur handicap.

Notre première étude visait à analyser les réactions émotionnelles de l'utilisateur en situation cible. Les données collectées dans le cadre d'IDV-HS et NAO-HR1 ont été annotées émotionnellement et analysées afin d'étudier le panel d'expressions affectives susceptibles d'apparaître dans une situation d'interaction. L'étude nous a permis d'évaluer les informations émotionnelles potentiellement annotables, et d'étudier le niveau de finesse d'annotation requis. Des expressions affectives négatives telles que l'Ironie, le Mépris et le Maternage à valence négative ont également été repérées, ce qui nous a offert de premières pistes pour déterminer les comportements souhaitables et non souhaitables du système.

Nous avons également procédé à l'étude des expressions émotionnelles utilisateur en réaction aux comportements sociaux robotiques. L'analyse des enregistrements d'IDV-HR et NAO-HR1 nous a amenés à constater que les comportements que nous jugeons indésirables chez le robot ont un impact négatif sur l'expression d'émotion des adultes dans un contexte d'assistance. Chez les enfants jouant avec le robot, les comportements que nous jugions initialement indésirables n'ont pas déclenché nécessairement d'émotions négatives, mais étaient plutôt perçus comme drôles par les enfants. Ceci nous permet de redéfinir la distinction entre socialement « indésirable » et « désirable » en fonction du contexte.

Nous avons ensuite étudié la relation entre les indices paralinguistiques et les dimensions du profil. Nous avons réalisé un test d'annotation du profil émotionnel chez l'adulte en perte d'autonomie dans une situation d'assistance (interaction avec un système IDV-HS, et interaction avec un robot IDV-HR). Nous avons proposé des analyses statistiques permettant de déterminer la pertinence de ces indices, que nous pourrions envisager sur un plus grand éventail de données.

L'analyse de la pertinence de la sélection du comportement en fonction du profil nécessite une évaluation du profil de l'utilisateur au cours de l'expérience. Cela demanderait une reconnaissance (automatique ou perceptive) du profil de l'utilisateur en temps réel, et une adaptation du comportement social du robot en fonction de l'évolution de la compréhension de ce profil. Nous avons vu au cours de l'expérimentation sur l'annotation du profil présentée, que la reconnaissance perceptive du profil nécessite, pour être plus fiable, une multi-annotation, ce qui n'est pas envisageable en temps réel. Une détection automatique du profil requiert la mise en place d'un système autonome (ou même semi-autonome), que nous présentons dans le chapitre suivant de ce manuscrit.

Chapitre 6

6. Implémentation et évaluation

6.1. Introduction

Nous avons présenté les théories sur lesquelles nous nous appuyons pour définir les règles de constitution du profil émotionnel et interactionnel de l'utilisateur, et exposé plusieurs expérimentations que nous avons menées afin de déterminer certains paramètres de l'interaction. Ces études nous ont menés à l'élaboration d'un système permettant la gestion de l'interaction entre l'utilisateur humain et le robot, en se basant sur la constitution et la mise à jour du profil émotionnel et interactionnel de l'utilisateur, et la sélection de comportements sociaux pour le robot.

Nous présentons dans ce chapitre la façon dont a été réalisée l'implémentation de la détection du profil émotionnel et interactionnel de l'utilisateur à partir de son signal de voix, et l'adaptation automatique du comportement du système. Nous présentons dans un premier temps l'architecture globale du système, constituée à partir d'une application C++ dont la base a été développée par l'entreprise SpirOps AI, ainsi que de leur système d'IA. Dans ce dernier, nous avons pu représenter nos règles de constitution automatique du profil de l'utilisateur et de sélection du comportement du robot. Nous détaillons ensuite la représentation formelle par des règles en logique floue de constitution du profil et de sélection de comportements. Nous présentons également la réalisation concrète, sur NAO, des comportements sociaux du robot d'après les théories détaillées dans le Chapitre 4, à l'aide des outils fournis par Aldebaran Robotics.

Nos expériences ont été menées avec des systèmes à différents degrés d'autonomie avec utilisateurs potentiels adultes et enfants pour différentes tâches. Des expérimentations ont été menées auprès de sujets déficients visuels susceptibles d'utiliser un robot dans la vie de tous les jours et ont été relatées dans la section 5. Il est rare de tester les systèmes avec de véritables *end-users*, cet aspect est un des points originaux de cette thèse.

Une expérimentation avec système semi-autonome a été réalisée dans un contexte de jeu avec des enfants. L'architecture s'appuie sur le modèle présenté en section 6.2, adapté à la situation d'interaction de jeu. Le système est considéré comme « semi-autonome » du fait de la saisie manuelle des informations émotionnelles par un manipulateur en Magicien d'Oz, à la place d'un système de détection des émotions automatique. Nous présentons la mise en place de cette expérimentation et les données obtenues en section 6.3.

Les différentes étapes du processus d'implémentation du système de gestion du profil utilisateur et de sélection du comportement ont permis la réalisation de systèmes en Magicien d'Oz à plusieurs degrés d'autonomie. Cette approche modulaire est susceptible de permettre l'évaluation des différents modules d'un système d'Interaction Humain-Robot ; nous la présentons en section 6.4. Enfin, nous concluons en 6.5

6.2. Implémentation

6.2.1. Architecture du système d'interaction

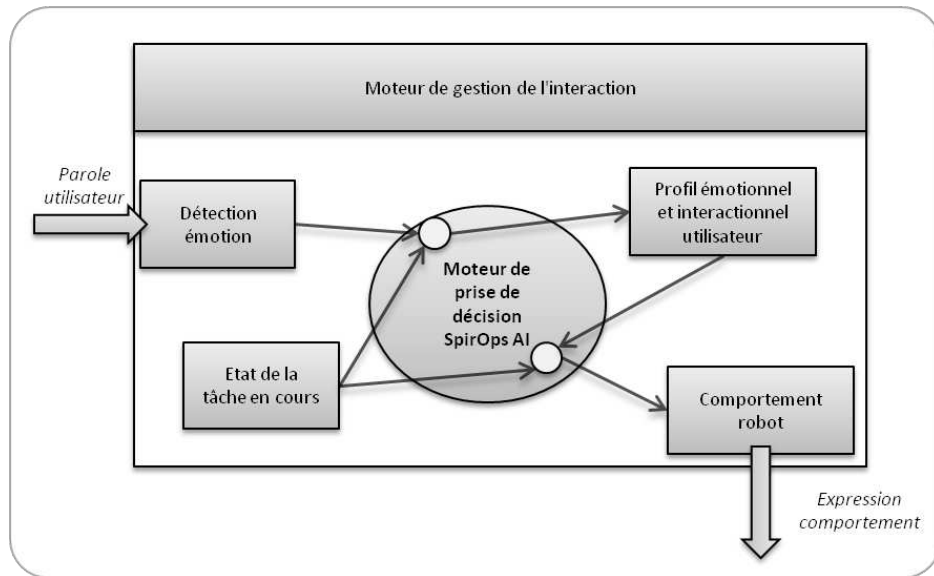


Figure 30. Architecture du système d'interaction entre le robot et l'utilisateur pour une mise à jour du profil émotionnel et interactionnel à partir du signal de voix de l'utilisateur, et une sélection automatique du comportement robot.

Selon le type de plateforme et la finalité recherchée, un système d'intelligence artificielle peut être basé sur des modèles statistiques ou des règles expert. Nous avons vu dans le cadre de l'expérimentation décrite dans la section 5.4, que la constitution d'un modèle d'apprentissage nécessite une quantité importante de données issues d'un même contexte applicatif, mettant en scène des publics comparables. Or, nous ne disposons pas dans notre étude de suffisamment de données audio pour entraîner un modèle.

Dans le cas du projet ROMEO, le moteur de décision régissant les comportements du robot sont implémentés en logique floue, sur l'éditeur graphique SpirOps Editor mis au point par l'entreprise SpirOps AI. Ce logiciel permet la description de comportements d'intelligence artificielle. Les décisions sont prises par le « cerveau » SpirOps, en évaluant les désirs de sélectionner tel ou tel comportement.

La Figure 30 schématise les différentes briques intervenant dans la gestion de l'interaction entre l'utilisateur humain et le robot, en vue de la constitution du profil émotionnel et interactionnel et la sélection de comportement. Un système général développé en C++ en collaboration avec SpirOps AI permet la gestion de l'interaction. Développée initialement dans le cadre du jeu des Histoires Interactives (voir section 6.3), cette interface permet le dialogue entre les différents modules :

- les informations paralinguistiques issues du module de détection des émotions du LIMSI (inscrites dans la mémoire du robot NAO AL Memory par le module autonome, et récupérées à cet emplacement par le système) ;
- les valeurs du profil de l'utilisateur (stockées et récupérées dans l'AL Memory), à savoir son profil émotionnel et interactionnel, son identité, son âge, son sexe ;

- une gestion simple du dialogue (dialogue linéaire orienté soit autour du jeu des Histoires Interactives, soit sur une discussion informelle avec l'utilisateur) dont découle une connaissance de la tâche en cours (en attente d'une réponse de l'utilisateur, échec ou réussite au jeu) ;
- le comportement du robot sélectionné, information transmise à Choregraphe afin de donner l'ordre au robot de réaliser ce comportement (via la synthèse de parole et la gestuelle selon le cas applicatif) ;
- le moteur de prise de décision SpirOps, autour duquel s'organisent tous les modules, et permettant de définir le cours du jeu, la mise à jour du profil, et la sélection de comportement.

Sur les versions semi-autonomes de notre système, le module de détection des émotions du LIMSI est remplacé par une interface graphique développée par SpirOps AI permettant le suivi des énonciations prononcées par le robot, le suivi des valeurs du profil émotionnel et interactionnel et le comportement sélectionné, et de saisir les informations émotionnelles (étiquette émotion, temps de parole et intensité) au clavier. Cette interface est utilisée par le manipulateur en Magicien d'Oz, et permet de simuler les entrées d'un module de détection émotionnelle.

Nous présentons à la Figure 31 un exemple de description d'une règle pour la diminution de la valeur de Confiance en soi, dans un contexte de jeu avec le robot. Cet exemple de règle est implémenté dans l'éditeur SpirOps Editor, qui génère un code C++ directement exploitable par le système. Dans cet exemple, toutes les boîtes de gauche reliées à la boîte de droite sont les paramètres intervenant dans la prise de décision. Ces paramètres sont directement liés à des variables instanciées dans le système. Nous voyons par exemple l'état de la tâche en cours (IsActedPhase) nous indiquant si l'on se trouve dans une phase de jeu ou pas, l'identité de l'enfant (Child.ID), le pas de décrémentation de la Confiance en soi (représenté par sa valeur flottante, ici -0.1f). Dans cet exemple, nous prenons en compte les informations paralinguistiques (l'intensité de l'émotion exprimée par l'enfant Activation, et le temps avant qu'il donne une réponse au robot TimeBeforeSpeech) et les informations sur l'état de la tâche en cours (le nombre d'échecs de l'enfant FailsCount). Désélectionné dans cet exemple, le paramètre IsDirective prend en compte le comportement précédent du robot.

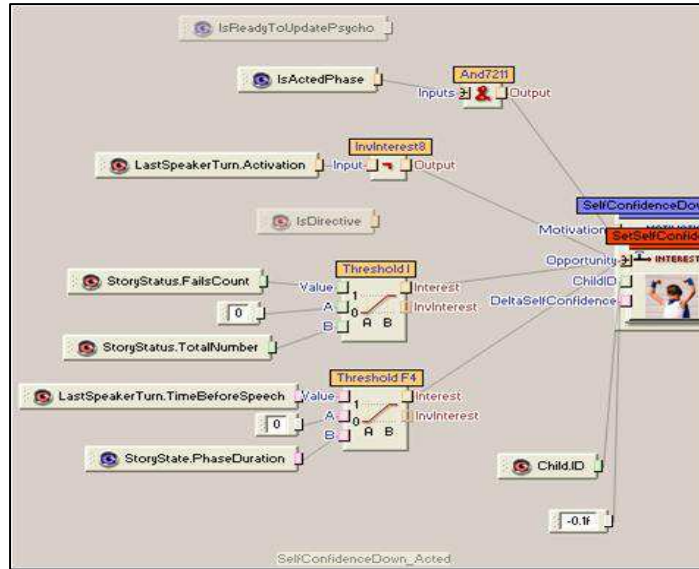


Figure 31. Exemple de description d'une règle de logique floue sur l'interface graphique SpirOps Editor. A gauche, représentation des paramètres contribuant à la prise de décision, et transmission (à droite) des valeurs de ces paramètres nécessaires pour le traitement.

6.2.1. Représentation formelle des règles de logique floue

Nous présentons dans cette section les représentations formelles des règles de mise à jour du profil émotionnel et interactionnel de l'utilisateur et de sélection du comportement robot en fonction du profil. Les décisions prises par le « cerveau » SpirOps dans ce contexte, sont de deux types :

- l'évaluation du désir d'augmenter ou de diminuer la valeur d'une dimension du profil (pour chaque dimension du profil),
- et l'évaluation du désir de sélectionner un comportement robotique plutôt qu'un autre.

Pour chacun de ces types, un ensemble de règles permet l'évaluation de chaque désir de réaliser un comportement. En nous appuyant sur la théorie des ensembles flous, nous pouvons donc dire que réaliser un désir revient à comparer les degrés d'appartenance de chaque entrée pertinente du système, par le biais de fonctions renvoyant des valeurs entre 0 et 1.

Nous prenons l'exemple de l'évaluation du désir d'augmenter le niveau de Confiance en soi du locuteur. Plusieurs entrées contribuent à l'évaluation de ce désir. Le degré d'appartenance de chacune de ces entrées est calculé à l'aide soit de règles de logique floues, soit de décision booléenne. Un OU logique appliqué entre tous ces degrés permet au système de prise de décision d'obtenir le degré d'appartenance maximal.

Augmenter le niveau de Confiance en soi du locuteur signifie alors qu'il y a au moins un élément de l'évaluation $x \rightarrow \{success, fastness, emotion, activation, emotion, activation, activation_mean, behaviour\}$ telle que :

$$0 < \max (m_a(success), m_b(fastness), m_c(emotion), m_d(activation), m_e(activation_{mean}), m_f(behaviour)) \leq 1.$$

Cela signifie que le maximum obtenu par les fonctions propres à chaque entrée résulte en une valeur entre 0 et 1. Cette valeur sera donc mise en opposition avec la valeur maximale obtenue pour le désir de diminuer la valeur de Confiance en soi du locuteur. Trois cas de figure peuvent alors se présenter : le désir

d'augmenter le niveau de confiance en soi est le plus fort ; le désir de diminuer le niveau de confiance en soi est le plus fort ; aucun désir n'est activé (valeurs à 0).

Nous présentons dans le Tableau 26 le détail des fonctions pour chaque entrée permettant l'évaluation du désir d'augmenter le niveau de Confiance en soi. Nous avons catégorisé les entrées en trois types :

- la représentation de l'environnement. Dans notre cas d'étude, cela se limite à l'état de la tâche en cours. Plus précisément, dans le cadre du jeu, il s'agira de la connaissance de l'échec ou de la réussite de l'utilisateur, duquel peut être à terme calculé un ratio succès/échec. Dans le cas où une réponse est attendue de la part de l'utilisateur, le « temps avant réponse » est une variable susceptible de décrire l'état de la tâche en cours : le robot est en attente, et la durée de cette attente a une signification pertinente pour l'analyse. Cette mesure est commune pour les situations de jeu, d'assistance, ou même de *small talk* lorsque le robot attend une réponse à une question.
- la représentation de l'émotion. Il s'agit des entrées émotionnelles et paralinguistiques obtenues via le module de détection des émotions ou par une saisie manuelle. Nous nous intéressons dans ce type d'entrée à la valence de l'émotion exprimée, sa force, la durée du temps de parole. Dans certains cas, l'étiquette émotion en elle-même est intéressante. Nous pourrions ajouter d'autres informations paralinguistiques, telles que la présence d'hésitations, de rires, ou de pleurs par exemple, selon le type de détection fournie.
- la représentation de l'attitude sociale du robot. L'impact possible de l'attitude du robot sur le profil émotionnel a été discuté en 4.2.2. Nous notons toutefois que la prise en considération de l'attitude dans la gestion du profil est représentable de cette façon. Nous pourrions également faire intervenir ici des informations subjectives sur l'image du robot telles que son degré d'humanité, d'intelligence, d'émotionnalité, etc., en bref toutes les manifestations susceptibles de faire varier l'attitude de l'utilisateur en réaction au robot.

Les fonctions définissent pour chaque entrée une valeur de sortie entre 0 et 1. Dans le cas de l'entrée *succès*, cette valeur est en relation avec le ratio de succès et d'échecs de l'utilisateur dans le cadre du jeu avec le robot. On estimera que moins le locuteur échouera, plus il pourra se sentir sûr de lui. La *rapidité* est une relation entre le temps écoulé avant que l'utilisateur donne une réponse et le temps total alloué à l'utilisateur pour cette réponse. Plus il répondra vite, plus cela signifiera qu'il est sûr de lui dans le jeu. La présence d'un certain type d'émotion (ici la Colère) constitue l'entrée *émotion*. Celle-ci est calculée sur une valeur booléenne indiquant la présence (valeur renvoyée 1) ou l'absence de Colère (valeur renvoyée 0). L'*activation* est directement liée à la valeur de l'activation perçue ou détectée pour ce tour de parole, et la *moyenne_activation* représente la moyenne de toutes les activations exprimées depuis le début de l'interaction. La prise en compte de la moyenne de l'activation sur toute la durée de l'interaction, et pas seulement de l'activation sur un seul tour de parole spécifique, permet de tempérer l'interprétation, c'est-à-dire de ne pas accorder une importance trop grande à une modification éventuellement ponctuelle de l'activation du locuteur. L'entrée gérant le comportement courant du robot (abrégée « *comport.* » dans le tableau) est une relation booléenne renvoyant une valeur de 0 ou de 1 selon qu'un des comportements potentiellement pertinents a été exprimé par le robot.

Tableau 26. Évaluation du désir d'augmenter le niveau de Confiance en soi, à chaque tour de parole

Entrée du système	Fonction	Description
Type d'entrée : Représentation de l'environnement (état de la tâche)		
Ratio succès/échec	$m_a(succès) = 1 - \frac{échecs}{essais}$	Moins le locuteur échoue
Temps avant la réponse attendue	$m_b(rapidité) = 1 - \frac{temps_avant_reponse}{duree_phase}$	Le plus vite il donne une réponse
Type d'entrée : Représentation de l'émotion (module de détection ou saisie manuelle)		
Présence ou absence de Colère	$m_c(émotion) = (émotion == "colère")$	S'il exprime de la colère
Niveau d'activation	$m_d(activation) = activation$	Plus son émotion est exprimée intensément
Moyenne activation	$m_e(moyenne_activation) = \frac{\sum_{i=1}^n activation_i}{n}$	Plus ses émotions sont exprimées intensément sur toute la session
Type d'entrée : Représentation du comportement du robot		
Comportement courant	$m_f(comport.) = (comport. == "empathique" OU "encourageant")$	Si le robot est actuellement encourageant ou empathique

Nous avons appliqué ce type de formalisation pour toutes les règles de mise à jour du profil émotionnel (désir d'augmenter ou de diminuer chaque dimension) et pour la sélection de comportement, en suivant les règles décrites notamment dans le Tableau 6 (« Dimensions du profil interactionnel et informations paralinguistiques afférentes »), le Tableau 7 (« Dimensions du profil émotionnel et informations paralinguistiques afférentes ») et le Tableau 8 (« Récapitulatif de l'impact du profil émotionnel et interactionnel de l'utilisateur sur la sélection du comportement social du système. ») de ce présent document.

L'Annexe II présente des simulations d'évolution des valeurs du profil en fonction de la prise en compte de un ou plusieurs paramètres.

6.2.2. Expression des comportements sociaux robot

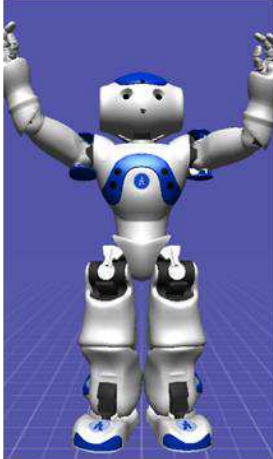
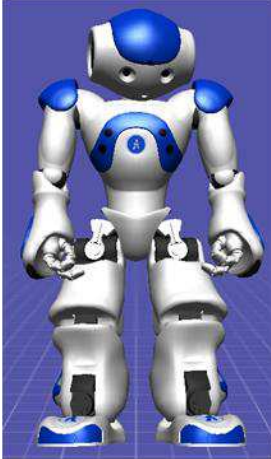
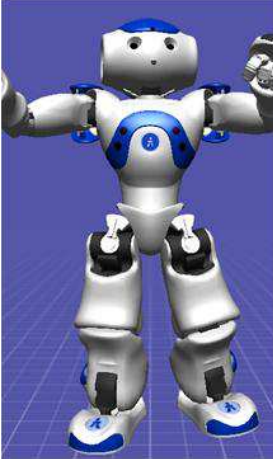
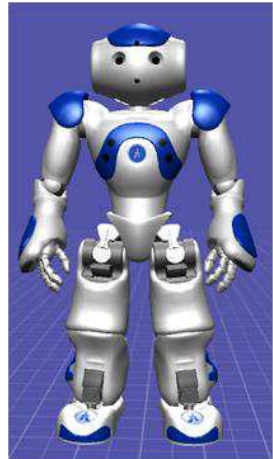
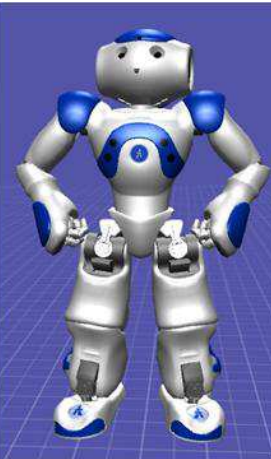
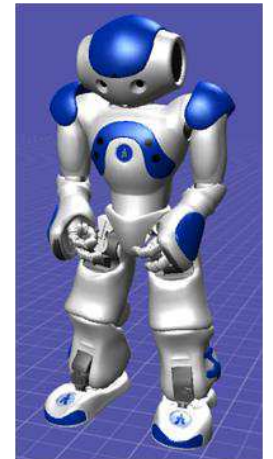
Nous avons proposé dans le chapitre 4 une codification des comportements sociaux du robot NAO. Nous nous sommes appuyés sur les théories sociales interpersonnelles, et avons déterminé un espace de comportements sociaux désirables pour un robot social. Nous avons sélectionné cinq comportements que nous jugeons socialement désirables et intéressants selon nos contextes d'applications : Encourageant, Empathique, Directif, Soumis et Amical. Nous avons détaillé la façon dont nous pouvons exprimer ces comportements sociaux via la sélection du lexique. La réalisation nécessite donc un système de synthèse de la parole par le texte, dont le robot NAO est doté. Nous avons pu aisément lui faire prononcer les énonciations voulues. Nous avons expliqué par ailleurs nos premières pistes de représentation posturale de ces comportements sur un robot humanoïde, en nous fondant sur les travaux menés par les architectes du comportement de l'entreprise Aldebaran Robotics. Nous présentons dans cette section la réalisation de ces comportements sociaux posturaux.

L'outil de création de postures émotionnelles fourni par Aldebaran Robotics se présente sous la forme d'un comportement Choregraphe (c'est-à-dire un script permettant de contrôler le robot) à trois curseurs émotionnels. Ces curseurs permettent de définir un point dans l'espace *Valence*, *Arousal*, et *Stance*, en déterminant un point sur les oppositions Joie-Tristesse, Excitation-Colère, Fierté-Peur. Ces trois valeurs sont définies sur une échelle de -1 (le pôle « positif » de l'opposition) à 1 (le pôle « négatif »). Par exemple, représenter une posture mélangée de Joie et de Fierté se présenterait sous la forme : Joie-Tristesse à -1, Excitation-Colère à 0, Fierté-Peur à -1. La Peur se présenterait comme Joie-Tristesse à 0, Excitation-Colère à 0, Fierté-Peur à 1.

Le logiciel Choregraphe permettant une gestion des comportements du robot tant par une interface graphique que par des scripts, nous pouvons également transmettre les codes de posture du robot depuis un programme extérieur sous la forme d'un tableau de 7 flottants : [Fierté, Joie, Excitation, Peur, Tristesse, Colère, Neutre]. Dans le cas du comportement Empathique en situation d'échec, la valeur transmise à Choregraphe sera [0, 0, 0, 0, 1, 0, 0], signifiant que seule la Tristesse sera marquée. Dans le cas de l'Empathie en situation de réussite, nous transmettons les valeurs [0, 1, 0, 0, 0, 0, 0] : seule la Joie est exprimée. Le comportement Amical est représenté par le tableau [0, 0, 0, 0, 0, 0, 0], puisque nous avons sélectionné l'attitude Neutre du robot NAO pour représenter l'amicalité du robot. Pour certains autres comportements, il nous a fallu jouer sur différents paramètres afin de représenter l'attitude souhaitée, les extrêmes ne nous semblant pas toujours pertinents. Tel est le cas de l'expression de l'Encouragement, représentée par le tableau de valeurs [0.5, 0.3, 0.7, 0, 0, 0, 0]. Afin de préserver la stabilité physique du robot, et de par les contraintes mécaniques liées à chaque attitude prototypique, le comportement est seulement teinté de Joie et d'Excitation, et non pas représenté avec des valeurs extrêmes de 1. L'expression de Fierté se justifiait par le besoin que le robot redresse la tête et les avant-bras afin « d'englober » l'interlocuteur, par une posture ouverte.

Comme nous l'avons déjà mentionné, l'attitude sociale du robot n'est pas exprimée via des mouvements, mais s'appuie sur des postures statiques. Le robot est donc en position Amical (Neutre) par défaut, et atteint ponctuellement les postures expressives pendant la durée de ses énonciations. A la fin d'une énonciation, il relâche la posture et revient à sa position par défaut.

Tableau 27. Postures sociales du robot NAO : Empathique, Encourageant, Directif, Soumis et Amical (captures d'écran Choregraphe).

<p>Empathique (réaction à évènement positif)</p>	<p>Empathique (réaction à évènement négatif)</p>	<p>Encouragement</p>
		
<p>Amical</p>	<p>Directif</p>	<p>Soumis</p>
		

Le Tableau 27 ci-dessus, présente les réalisations des différentes postures sociales sur le robot NAO, sur l'interface de visualisation 3D proposée par le logiciel Choregraphe. Nous pouvons voir à la Figure 32 la mise en place d'une réalisation du comportement Encouragement sur le robot, à l'occasion d'une démonstration au LIMSI, filmée par E. Fertil et montée dans le film « Vivre avec les robots » (Fertil, 2012) présenté en compétition des films scientifiques 2012. L'amplitude de la posture a été volontairement réduite par rapport au modèle prévu afin d'assurer la stabilité du robot en position debout.



Figure 32. Le robot NAO exprimant l'Encouragement lors de la vidéo documentaire *Vivre avec les robots* (Fertil, 2012).

Le comportement social du robot, au-delà des théories que nous avons sélectionnées et adaptées, est donc dépendant de la plateforme sur lequel il est réalisé. Tout d'abord, du point de vue de l'aspect du robot bien sûr (humanoïde ou non, expressif ou non, etc.), mais également en fonction de contraintes plus prosaïques telles que les degrés de liberté du robot, sa stabilité verticale, ou encore la fluidité de ses moteurs. Il est donc nécessaire de procéder à une adaptation pour minimiser l'écart entre les théories invoquées et les réalisations concrètes envisageables.

6.3. Mise en place d'un système semi-automatique avec sélection de comportement et mise à jour profil : expérimentation NAO-HR3

Les environnements de jeu sont propices pour la génération d'interactions spontanées. Nous avons imaginé un jeu original d'histoires interactives dans lequel des enfants sont en compétition lors d'une interaction avec le robot NAO. Pour gagner, il leur faut générer des expressions émotionnelles spécifiques (par des bruits, des mots) lorsque le robot prononce certains mots (au premier niveau de difficulté du jeu) ou concepts (à un niveau plus difficile). Si le robot reconnaît l'expression émotionnelle attendue, l'enfant marque un point. Le jeu est alors d'être le plus expressif et le plus rapide puisqu'il faut être attentif à l'histoire pour reconnaître quand il faut réagir. Ce jeu semblait intéressant pour stimuler l'attention, la compréhension, l'imagination et l'expressivité des enfants. De plus, le robot tentait de s'adapter aux enfants pour les aider à réaliser la tâche de jeu. Ce jeu va surtout nous permettre de collecter un corpus spontané de réactions émotionnelles de jeunes enfants. La création d'un jeu ludique, captivant nécessite de longues expérimentations. Notre but ici était principalement de collecter des données, nous n'avons pas cherché à optimiser le jeu.

Nous avons donc réalisé un premier test hors laboratoire de notre système à l'occasion d'une collecte de données auprès d'enfants de 6 à 11 ans en situation de jeu avec NAO (Tahon *et al.*, 2012). Le système fonctionnait en semi-automatique : une sélection manuelle des émotions exprimées par les enfants, réalisée par un expérimentateur en Magicien d'Oz, et une mise à jour automatique du profil enfant et de la sélection

du comportement robot via notre modèle en règles de logique floue implémenté sur SpirOps AI. Pour réaliser cette collecte, nous avons mis au point un scénario d'acquisition tout à fait nouveau.

6.3.1. Environnement de test

Deux enfants interagissent avec le robot NAO à une session de jeu, intitulé « Les histoires interactives ». Au cours de ce jeu, le robot raconte une histoire pour enfants (adaptée des Trois Petits Cochons), durant laquelle les enfants devront réagir en fonction d'une liste de règles indiquées sur un tableau. Ils devront par exemple exprimer de la tristesse lorsque le robot dira le mot – ou abordera le concept de – « Pauvreté » (selon le mode de difficulté). L'enfant gagne un point si l'émotion attendue a été détectée. Ce concept de jeu permettait à la fois d'obtenir des réactions émotionnelles de la part des enfants dans un contexte d'Interaction Humain-Robot pour la constitution d'un corpus riche d'Affect bursts pour une étude acoustique (Tahon *et al.*, 2012), et également d'observer l'interaction émotionnelle entre les enfants et NAO dirigée semi-automatiquement par notre système de constitution du profil et de sélection de comportement.

Les enfants jouent deux par deux, debout face au robot NAO (voir Figure 33). Le robot est assis et ne joue pas de comportements gestuels afin de ne pas saturer sa mémoire. Les enfants portent chacun un micro-cravate et sont enregistrés pendant toute la durée de l'expérimentation. La caméra Kinect posée aux pieds du robot permet de filmer les enfants pour une autre étude de l'équipe (en vidéo directe et en relevé de profondeur).



Figure 33. Expérimentation NAO-HR2 : Disposition du robot NAO devant le tableau des consignes du jeu (image de gauche). Deux filles jouant au jeu, debout face à NAO (image de droite).

Le système informatique C++ gérant le cours du jeu, la mise à jour automatique du profil émotionnel de chaque joueur, et la sélection automatique des comportements du robot, a été développé en partenariat avec la PME SpirOps (Paris). Ce système gère la progression de l'histoire et son énonciation par le robot. Lorsqu'une intervention est attendue de la part d'un enfant, le système se met en attente. Une fois que l'enfant a joué son tour (ou que le temps est écoulé), le robot interprète la réponse (ou l'absence de réponse) en fonction des mécanismes en logique floue décrits dans la section 6.2 (Implémentation), afin de mettre à jour le profil émotionnel de l'enfant. En fonction des valeurs de ce profil, le système sélectionne le comportement le plus approprié, selon les règles décrites dans ce document. La Figure 34 récapitule l'architecture du système semi-automatique.

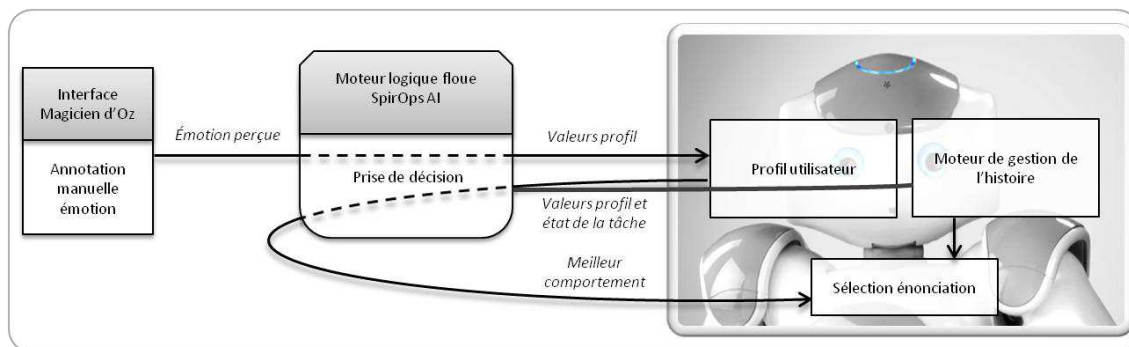


Figure 34. Architecture du système semi-automatique pour le jeu Histoires Interactives : annotation émotionnelle manuelle, et prise de décision automatique pour la mise à jour du profil utilisateur et la sélection comportement robot via les règles implémentées dans SpirOps AI.

Le comportement du robot s'adapte donc au fil du jeu. Les variations du comportement sont exprimées via le choix lexical (modalité détaillée dans la section 4.3), adapté en fonction du comportement social (Empathique, Amical, Encourageant). Lorsque le robot est Amical, par exemple, il dira simplement « *La réponse est correcte* » ou « *La réponse est incorrecte* ». S'il est passé en mode Empathique pour cet enfant, et que ce dernier échoue, le robot dira « *C'est pas facile, mon pauvre.* », etc. Au cours de cette expérimentation, nous avons choisi de ne pas faire réaliser de comportements gestuels au robot, mais de nous appuyer sur le contenu lexical, pour des contraintes de température des moteurs et de surcharge de la mémoire du robot, plusieurs sessions d'enregistrement étant réalisées successivement pendant plusieurs heures.

Le seul aspect non automatisé du système est la reconnaissance des émotions exprimées par les enfants. En effet, en lieu et place d'un module de détection censé transmettre automatiquement les informations sur « temps de parole », « étiquette émotion », « intensité » et « valence », un expérimentateur formé à l'annotation émotionnelle entre manuellement ces informations dans le système, en temps réel, après une évaluation perceptive des productions orales des enfants. Les moyennes et écart-type nécessaires à la mise à jour du profil sont ensuite calculés automatiquement par le système. Cette expérimentation ayant pour but de collecter des données supplémentaires pour renforcer la robustesse du système de détection de l'équipe, réaliser l'annotation manuellement permet de s'affranchir d'éventuelles erreurs de détection d'un modèle non stabilisé.

Une fois le jeu terminé, un expérimentateur pose une série de questions à chaque enfant afin d'évaluer :

- 1) Si le jeu était à sa portée :
 - « Est-ce que le jeu était intéressant ? »
 - « Est-ce qu'il était difficile ? Trop facile ? »
 - « Est-ce que c'était amusant ? »
 - « C'est un jeu auquel tu aurais envie de rejouer un jour ? »
- 2) Sa perception du robot (compréhension du robot, affinité pour le robot) :
 - « Est-ce que tu as l'habitude de jouer avec des robots ? »
 - « Comment est-ce que tu as trouvé le robot ? »

- « Est-ce que tu trouves qu'il comprenait bien ce que tu disais ? »
 - « Il arrivait à bien reconnaître si tu étais en colère, ou triste, ou content ... ? »
 - « Est-ce que tu l'aimes bien ? »
 - « Est-ce que tu as trouvé le robot ennuyeux ? »
 - « Est-ce que tu as trouvé le robot agaçant ? »
- 3) S'il a perçu une adaptation individuelle du comportement du robot :
- « Est-ce que tu as trouvé qu'il était plus gentil avec l'un d'entre vous deux ? »
 - « Est-ce que tu as l'impression qu'il t'aime bien ? »
 - « Est-ce que tu l'as trouvé gentil avec toi ? »

6.3.2. Analyse des données obtenues

Le corpus NAO-HR2 composé de voix émotionnelles d'enfants en interaction ludique avec le robot NAO contient, après segmentation selon le protocole détaillé en Annexe III, 460 segments émotionnels. Nous avons enregistré 12 enfants (7 filles et 5 garçons) d'une moyenne d'âge de 8 ans. La Figure 35 présente quatre enfants jouant deux par deux avec le robot.



Figure 35. Expérimentation NAO-HR2 : Quatre enfants jouant par paires lors du jeu Histoires Interactives (une fille de 9 ans et un garçon de 6 ans à gauche ; deux garçons de 7 ans à droite).

Voici ci-dessous un extrait du jeu transcrit, mettant en scène une fille de 9 ans (F) et le robot NAO (N). Le terme « grand méchant loup » devait amener comme réponse une expression de Peur de la part de l'enfant, qui a été effectivement reconnue comme de la Peur par le manipulateur du Magicien d'Oz. Le robot était passé en mode Encouragement vis-à-vis de ce joueur.

N — « Ils lui demandèrent, et voici la réponse qu'elle leur fit : 'Si vous n'êtes pas dans un endroit sûr, vous pourriez être mangés par le grand méchant loup'. »

F — « Ah, j'ai peur, ah non, j'ai peur. »

N — « Bravo, je savais que tu pouvais le faire. »

L'annotation de la voix expressive, réalisée par un annotateur formé à l'analyse perceptive de la voix expressive, a porté sur :

- **Émotion** : Joie, Colère, Tristesse, Peur, Gêne, Surprise, Neutre ;
- **Valence** : Positif, Négatif, Positif-Négatif, Neutre ;
- **Intensité** : de -2 (faible) à 2 (fort) ;
- **Activation** : de -2 (faible) à 2 (fort).

Selon les résultats d'annotation obtenus, le corpus était composé à 46% d'émotions de macro-classe négative (Colère, Peur, Tristesse et Gêne), 29% de Joie, et 25% de Neutre/Surprise. Nous rappelons que le jeu impliquait l'expression de Colère, Peur, Tristesse et Joie, et que ces scores comprennent autant les émotions spontanées qu'actées au cours du jeu. Si l'on ramène l'Intensité à une valeur comprise entre 0 (faible) et 1 (fort), l'Intensité moyenne exprimée par les enfants était de 0,6, donc globalement moyenne.

Comme le fait remarquer (Tahon *et al.*, 2012) à l'occasion de l'analyse des affects bursts de ce corpus, les segments audio sont globalement très courts : lorsque, dans le cours du jeu, on attendait une réponse de la part des enfants, ils s'exprimaient principalement par onomatopées brèves. Aucun support lexical ne leur ayant été imposé pour ne pas entraver leur spontanéité, les enfants ont majoritairement choisi de s'exprimer à l'aide de, par exemple, « grr » pour la colère, « aaah » pour la peur, ou encore des rires pour la joie.

Le questionnaire posé aux enfants avait pour but de déterminer, tout d'abord, si la difficulté du jeu proposé était à leur portée, puis d'évaluer leur vision du robot – leur niveau d'affinité –, et enfin s'ils s'imaginaient que le robot avait un avis sur eux – c'est-à-dire s'il s'adaptait à eux. Nous présentons ces résultats dans la Figure 36, la Figure 37 et la Figure 38 ci-dessous.

D'après le questionnaire, les enfants ont globalement apprécié le jeu, et l'ont trouvé assez facile. Les enfants ont apprécié le robot, et l'ont trouvé globalement gentil (nous n'avons par ailleurs programmé aucun comportement « indésirable » pour le robot). En ce qui concerne la perception que le robot différencie bien les deux joueurs l'un de l'autre (« Est-ce que tu as trouvé qu'il était plus gentil avec l'un d'entre vous deux ? »), les avis des enfants étaient partagés. Nous comprenons à l'écoute des réponses des enfants que le fait que l'un des deux joueurs réussisse mieux à jouer les émotions (et donc à obtenir des points) que l'autre pouvait influencer la réponse à cette question. Notre retour sur l'expérience nous fait dire que les enfants ont eu tout de même quelques difficultés à comprendre les règles du jeu, et ont sans doute eu du mal à s'impliquer.

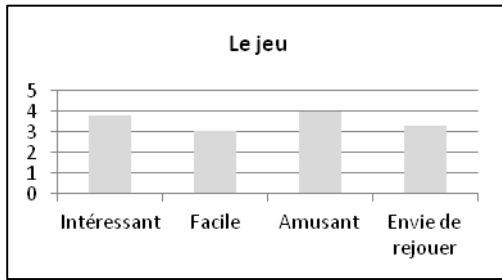


Figure 36. NAO-HR3 : Résultats moyens du questionnaire, évaluation des enfants quant au jeu (0 : pas du tout d'accord, 5 : tout à fait d'accord).

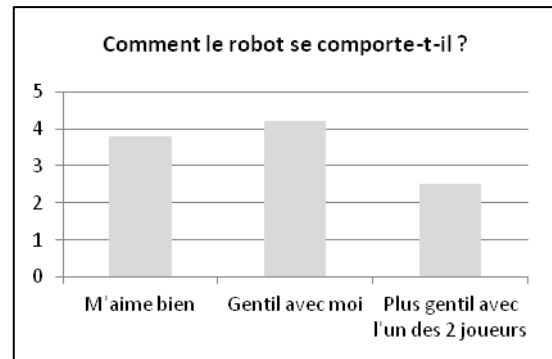


Figure 37. NAO-HR3 : Résultats moyens du questionnaire, évaluation du comportement du robot (0 : pas du tout d'accord, 5 : tout à fait d'accord).

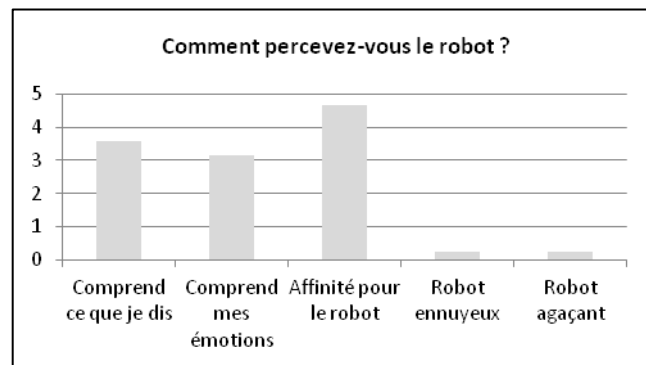


Figure 38. NAO-HR3 : Résultats moyens du questionnaire, évaluation des enfants quant au robot (0 : pas du tout d'accord, 5 : tout à fait d'accord).

Comme nous l'avons évoqué, nous avons profité de cette collecte de données émotionnelles pour tester notre système semi-automatique en conditions réelles d'application. Nous voyons d'après les informations collectées que les enfants présentent, d'après notre annotateur, 25% d'expressions Neutre, et 18% de Gêne. La Gêne est une expression affective subtile qui n'est pas prévue dans le système automatique (uniquement basé sur les émotions primaires). Nous l'avons principalement annotée comme Neutre dans le Magicien d'Oz, car la gêne des enfants s'exprimait par un manque d'énergie, une petite voix. Cette émotion était donc perceptivement la plus proche de l'état de Gêne.

La grande proportion de voix manquant d'énergie et d'expressions brèves a eu comme conséquence de niveler les profils émotionnels des enfants. Une analyse perceptive des profils émotionnels de tous les enfants, réalisée par le manipulateur du Magicien d'Oz en fin de chaque session, indique des valeurs moyennes pour chaque dimension du profil de 0,6 (0 : faible, 1 : fort), avec un écart-type moyen de 0,2, et des valeurs minimum de 0,2 pour certaines dimensions, et des maximums de 1.

Cette expérimentation, en plus d'avoir permis la collecte d'une grande quantité d'affect bursts de la part d'enfants en interaction avec un système d'interaction ludique semi-autonome, nous amène à envisager d'ores et déjà les futures améliorations de notre système. En effet, ne pas être à l'aise avec un système est une situation très fréquente et doit pouvoir être prise en compte automatiquement. La nouveauté, par exemple, peut être un facteur générant des manifestations de gêne de la part de l'utilisateur, et ce facteur pourrait être

pris en compte pour rendre l'interprétation du profil utilisateur encore plus robuste aux conditions d'interaction.

6.3.3. Conclusion

Nous avons réalisé un premier test hors laboratoire de notre système semi-automatique, avec sélection manuelle des émotions exprimées par les enfants, et mise à jour automatique du profil enfant et de la sélection du comportement robot. Nous avons réalisé ce test à l'occasion d'une collecte de données auprès d'enfants de 6 à 11 ans en situation de jeu avec NAO.

Nous avons constaté que les enfants étaient souvent très intimidés par notre scénario, ce qui s'est traduit par peu d'engagement dans le protocole, et beaucoup de gêne. Les enfants intimidés ont produit un grand nombre d'affect bursts en guise de réponse au cours du jeu, ce qui nous a permis de récolter un corpus extrêmement intéressant à étudier. L'expérimentation nous apporte également des pistes de réflexion intéressantes, puisque nous avons pu constater que la grande proportion d'émotions faibles et la durée moyenne très courte des productions des enfants entraînaient un nivelage des profils, et donc une sélection de comportements globalement toujours identique d'un enfant à l'autre. Cette constatation nous amène donc, pour des améliorations futures, à envisager d'autres critères d'estimation du contexte permettant une adaptation automatique encore plus robuste à l'utilisateur. Notamment, la gestion de la nouveauté chez l'utilisateur serait un indice permettant d'affiner la constitution du profil : si c'est la première fois que l'utilisateur joue à ce jeu, ou si c'est la première fois que l'utilisateur interagit avec le robot.

6.4. Évaluation d'un système d'Interaction Humain-Machine via un Magicien d'Oz à différents degrés d'autonomie

L'évaluation d'un système d'Interaction Humain-Machine est complexe. Évaluer un système signifie estimer l'adéquation des performances du système avec l'objectif désiré. Interrogeons-nous donc d'abord sur l'objectif que l'on souhaite atteindre avec notre système d'interaction émotionnelle. Dans le cas d'une interaction avec un robot social assistant à domicile, l'objectif le plus humble est d'abord que le robot ne soit pas rejeté ou ignoré par l'utilisateur, qu'il ne lui apparaisse ni comme un pantin tentant de mimer de façon non pertinente des interactions sociales (dont les mécanismes sembleront alors plus dirigés par le hasard que par l'intelligence), ni comme une machine dénuée totalement d'intelligence sociale (et avec laquelle on ne tentera même pas d'interagir intelligemment). Notre objectif principal est donc l'acceptation sociale du système robotique par l'utilisateur, cette acceptation étant conditionnée par le comportement exprimé par le robot (ce qu'il dit, ce qu'il fait).

De façon simpliste, il nous suffirait donc de demander aux utilisateurs : « Jugez-vous votre interaction avec le robot satisfaisante ? ». Cette approche soulève alors une autre interrogation chez l'expérimentateur : que signifie « être satisfait du système ? », quels sont les critères permettant de déterminer de façon scientifique la satisfaction de l'utilisateur ?, l'utilisateur est-il réellement satisfait de l'attitude du robot, ou bien d'autres facteurs interviennent-ils dans sa satisfaction ?

Nous voyons déjà les difficultés soulevées pour traiter l'évaluation de l'acceptation sociale de notre système : à partir d'une interrogation simple, se déroule un chapelet de sous-études à mener. Sans oublier que jusque là, nous n'aurions évalué que la réussite de l'objectif du système global. En effet, comme nous l'avons

présenté dans l'introduction du chapitre 5, notre système d'interaction émotionnelle Humain-Robot est composé de quatre modules interdépendants (rappelés dans la Figure 39 ci-dessous): un module de perception capture et réalise une première interprétation des indices paralinguistiques dans l'audio ; un système émotionnel interprète ces indices en termes de représentation émotionnelle et interactionnelle de l'utilisateur ; un module comportemental sélectionne le comportement du robot ; et enfin des effecteurs permettent de transposer ce choix de comportement en signes compréhensibles par l'humain (dans notre cas, sur les modalités gestuelles et linguistiques). Les informations transmises et les connexions établies entre ces modules nécessitent des évaluations distinctes, et parfois difficiles à chiffrer, dans les domaines de la psychosociologie et de la communication homme-machine.

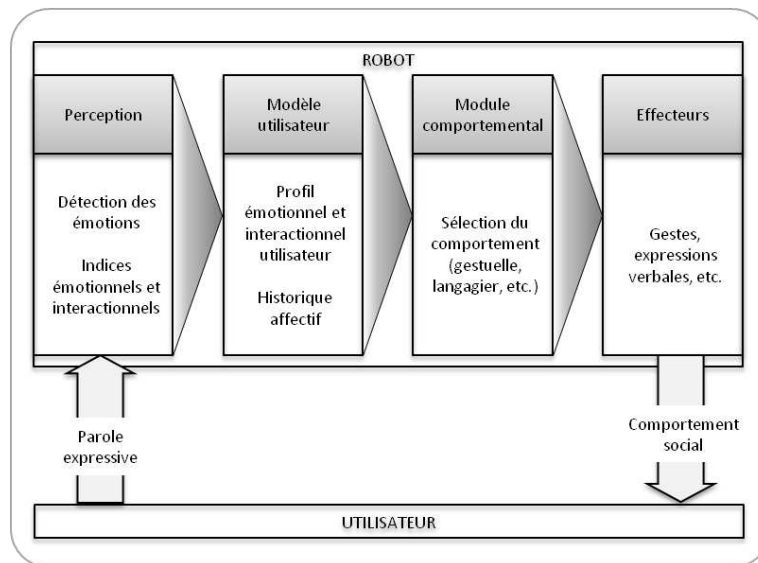


Figure 39. Boucle d'interaction émotionnelle Utilisateur-Robot : perception de la parole expressive utilisateur, système émotionnel, module comportemental et génération de comportement social.

Dans le cadre du projet Romeo, nous avons conçu et testé différents systèmes d'Interaction Humain-Robot présentant différents degrés d'autonomie, afin de maîtriser et d'étudier plusieurs paramètres de l'interaction. Selon les expérimentations menées, différents cas se présentent : le robot n'est pas sensible aux émotions ni au profil du locuteur, et suit un script de comportements prédéfinis ; dans un autre cas, un expérimentateur fournit au système, à la main, les informations paralinguistiques perçues (durée, émotion, force de l'émotion) et le système met à jour le profil et sélectionne le comportement du robot ; dans un dernier cas, le système est doté d'une perception automatique d'émotions, et met à jour son profil et sélectionne automatiquement le comportement robot d'après des règles de logique floue.

Le premier cas nous permet d'étudier les expressions d'émotions des utilisateurs de façon générale, et également en réaction à des comportements prédéfinis. Dans le second schéma, nous pouvons étudier les réactions émotionnelles de l'utilisateur face à des comportements sélectionnés automatiquement comme étant adaptés au contexte et à son profil, en faisant abstraction des contraintes techniques dues à la perception automatique des émotions. Et enfin le troisième cas nous met en situation concrète d'interaction émotionnelle homme-robot totalement automatisée.

Ces systèmes nous permettent donc de recueillir et d'analyser des données en contrôlant plusieurs paramètres d'élicitation des émotions au sein d'une interaction : nous pouvons choisir de faire varier, ou de

laisser en autonome la sélection du comportement robot (positif ou négatif vis-à-vis de l'utilisateur ?), la prise en compte ou non des émotions exprimées précédemment par le locuteur (gestion d'un historique ou non), et de l'état de la tâche (le robot coopérerait-il ou pas ?).

Dans cette thèse, nous avons mené différentes expériences avec des magiciens d'Oz à différents niveaux. Ces études nous ont permis d'étudier certaines composantes de l'interaction, et également de proposer d'autres protocoles pour évaluer les autres aspects du système, ainsi que le système totalement autonome. Ces évaluations seront menées dans la suite du projet ROMEO. Nous cherchons avant tout dans cette thèse à montrer le cheminement et les protocoles à mettre en œuvre pour la création d'un système d'interaction.

6.4.1. Définition et intérêt du Magicien d'Oz

L'expérience du Magicien d'Oz est une technique expérimentale d'évaluation d'un système d'Interaction Humain-Machine, qui consiste à faire croire à l'utilisateur qu'il interagit avec un véritable système. Le chercheur pourra, par exemple, observer l'utilisabilité de son système, tout en se dégageant des contraintes techniques dues à un système en cours d'élaboration. Nous retrouvons déjà le principe d'un opérateur humain manipulant un système à l'insu de l'utilisateur dans le cadre du test de Turing (Turing, 1950) visant à évaluer l'intelligence d'un système (si l'utilisateur croit qu'il interagit avec un vrai système autonome, le test est alors réussi).

(Dahlbäck *et al.*, 1993) proposent une discussion de l'utilité des systèmes en Magicien d'Oz dans le domaine des systèmes de dialogue, pouvant s'appliquer à d'autres types d'interfaces intelligentes. Les auteurs mettent en avant la complexité de la tâche de dialogue, qui fait intervenir des mécanismes complexes tant du point de vue de l'argumentation, que du contenu narratif, de la sémantique, etc. D'autre part, ils rappellent que le dialogue est influencé par des facteurs interpersonnels, c'est-à-dire une compréhension, et un impact de l'autre sur le locuteur.

Selon les auteurs, le système ne doit pas mimer l'humain, parce que les modalités sont différentes : on ne discute pas avec un humain, mais avec une machine. Cependant, les auteurs admettent que pour étudier correctement l'interaction d'utilisateurs finaux avec le système, il convient de proposer un système ayant su s'abstraire des contraintes technologiques propres au système en cours de développement (ressources, erreurs), et donc de proposer une version de bonne qualité et correctement contrôlée.

L'utilisation d'un système en Magicien d'Oz permet donc soit de donner au système une plus grande liberté d'expression pour obtenir des données d'interaction plus « idéale », soit au contraire de contraindre le système (éviter par exemple qu'un dysfonctionnement perturbe l'interaction, ou encore limiter les réponses possibles du système à un ensemble de réponses donnés).

Lorsque l'on débute l'implémentation d'un système, il convient de procéder à des tests pour évaluer les différents modules. L'évaluation modulaire, avec utilisation de Magiciens d'Oz, permet en premier lieu de réaliser des tests techniques des différentes composantes du système d'interaction. En effet, respecter la modularité du système, et donc ne proposer qu'un module automatisé à la fois, permet de ne pas créer un effet boule de neige sur le système en cas de défaillance.

Nous présentons tout d'abord l'architecture de ces systèmes en termes de « module contrôlé par l'humain » versus « module autonome », puis nous expliquons en quoi ces architectures permettent l'évaluation de certaines composantes précises de l'interaction.

6.4.2. Magicien d'Oz avec contrôle complet de l'expérimentateur

Au cours des expérimentations auxquelles nous avons participé durant cette thèse, nous avons mis en œuvre différents types de système nous permettant d'évaluer des interactions. Dans cette première partie, nous présentons les systèmes en Magicien d'Oz avec contrôle complet de l'expérimentateur. En d'autres termes, le comportement du système suit un script, et aucune décision automatique n'intervient dans la conduite de l'expérimentation. Dans les différents systèmes sur lesquels nous revenons dans cette section, les variations portent sur les différents modules que l'expérimentateur simule, et ceux qui sont juste ignorés.

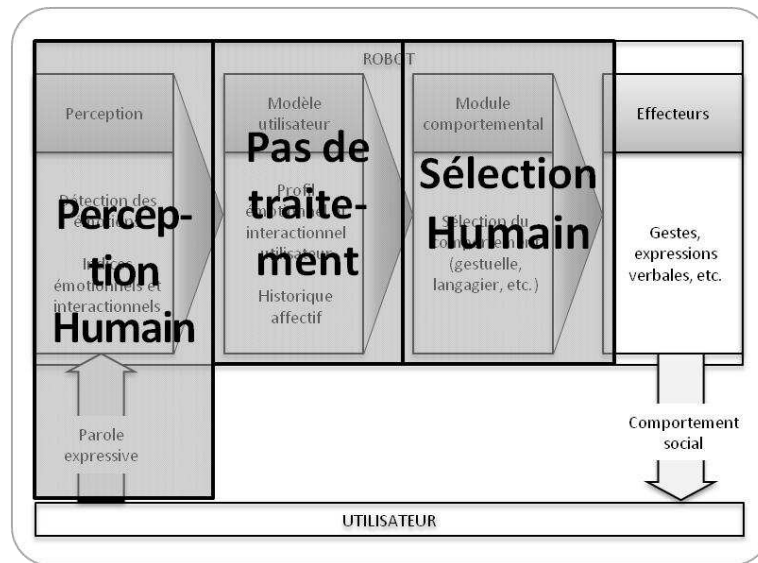


Figure 40. Boucle d'interaction émotionnelle Utilisateur-Robot avec perception de la parole expressive et sélection de comportement par l'expérimentateur.

Description de l'architecture

Dans le cas de l'expérimentation IDV-HS (voir paragraphe 5.2.1 pour une description du protocole), un expérimentateur était en charge de simuler le fonctionnement d'un système de détection des émotions.

Ici, l'interaction est limitée au maximum : détection d'émotion, énoncé oral de la détection. Les comportements possibles du « système » sont très limités : soit le système est fonctionnel et reconnaît bien l'émotion exprimée par le locuteur, soit il présente une défaillance et ne détecte pas l'émotion. Le seul comportement social exprimé par le système était un comportement « machine », neutre (« Émotion détectée : Colère »).

Comme le résume la Figure 40, le module de détection et la sélection du comportement (émotion reconnue ou non) sont remplacés par une décision humaine. D'une façon générale, l'expérimentateur tentait de s'abstraire de la personnalité du participant et de suivre son scénario visant à faire produire des émotions. Nous considérons donc qu'en théorie l'interprétation de la personnalité du locuteur est un module qui n'est pas pris en compte par l'expérimentateur. Toutefois, et là réside une des difficultés de maintenir la rigueur du

Magicien d'Oz, la personnalité et l'attitude du locuteur pouvait influencer la sélection du comportement (un utilisateur trop gêné, qui s'emporte, etc.) et pousser l'expérimentateur à choisir le comportement de réussite du système de détection. Nous avons cependant essayé de nous détacher au maximum de l'analyse de la personnalité du participant, c'est pourquoi nous marquons ce module comme « non traité ».

Dans le cas de cette expérimentation, nous pouvons noter que même les effecteurs (les éléments du système réalisant concrètement le comportement) sont simulés. En effet, la voix énonçant le résultat de la détection est un enregistrement de voix humaine diffusé par les haut-parleurs du système. Nous minimisons toutefois cette caractéristique dans notre étude des Magiciens d'Oz, puisque nous ne nous sommes pas intéressés à l'étude du caractère naturel des expressions de la machine.

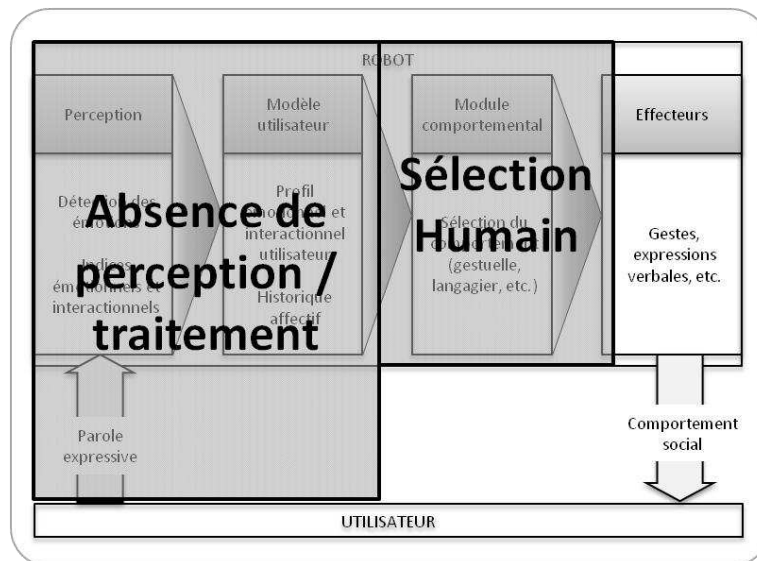


Figure 41. Boucle d'interaction émotionnelle Utilisateur-Robot avec seule sélection du comportement par l'expérimentateur.

Dans le cas de NAO-HR et IDV-HR (Figure 41), nous nous mettons en situation d'interaction que l'on pourrait croire au premier abord « idéale ». En effet, le robot comprend « tout », il sait mener le jeu ou le scénario, et fait simplement quelques erreurs (celles supposées éliciter de l'amusement ou de l'agacement). Ces versions intelligentes du système sont construites cependant sur une absence totale de prise en compte de ce que le locuteur a exprimé. L'expérimentateur réalise une reconnaissance basique du contenu dialogique, se limitant généralement à la reconnaissance du moment où l'individu a fini de parler et où l'énonciation suivante du système pourra être enclenchée. Les émotions exprimées sont ignorées, et l'expérimentateur ne s'adapte en aucune façon à la personnalité du locuteur ; dans les deux cas, les contraintes de l'automate Choregraphe régissant le fil du scénario ne permettent aucune adaptation.

Dans ces deux expérimentations, nous travaillons donc avec des systèmes insensibles à l'utilisateur, pour lesquels la sélection de comportement est réalisée par l'humain. Cette sélection de comportement ne fait que suivre un scénario préétabli et le participant n'a aucun moyen de faire varier cette sélection.

Pour quelle évaluation ?

Nous avons proposé deux systèmes mettant en œuvre la technique du Magicien d'Oz. La similarité des systèmes rappelés ci-dessus consiste dans le fait, d'une part, qu'aucun module n'est automatisé. D'autre part,

l'attitude du participant n'a pas (ou très peu) d'impact sur la sélection de comportement du système, puisque le manipulateur du Magicien d'Oz suit un script établi. Ces systèmes ne permettent donc pas d'étudier la boucle d'interaction à proprement parler, puisqu'une interaction suppose que chacun des interactants doit avoir un effet visible sur son interlocuteur.

Ces systèmes en Magicien d'Oz complets permettent de réaliser des tests utilisateurs en conditions d'application finale. Dans notre cas, cela permet de répertorier les manifestations émotionnelles des utilisateurs en contextes réels, c'est-à-dire en fonction d'une variété de comportements du système susceptibles de se produire en contexte d'application (ou dont nous voulons tester la pertinence).

D'une façon générale, cette architecture permet de tester la satisfaction utilisateur, puisque cela permet de présenter un système d'interaction presque « idéal ». En effet, même s'il n'existe aucune réelle compréhension ni traitement de l'énoncé de l'utilisateur, le système présente un contexte réel d'application dans lequel l'utilisateur peut se projeter. La tâche de Magicien d'Oz est ici complexe puisqu'il doit sélectionner beaucoup d'informations manuellement. La sélection de comportement peut également être réglée afin de ne présenter que des comportements souhaitables (contrairement à nos études qui visaient à étudier un panel de comportements possibles).

6.4.3. Magicien d'Oz semi-automatique

Si certains modules peuvent ne pas être du tout implémentés (et donc nécessiter une simulation par un expérimentateur humain), d'autres peuvent exister déjà mais ne pas être présents dans l'architecture à tester, afin de permettre d'effectuer le test des modules de façon logique.

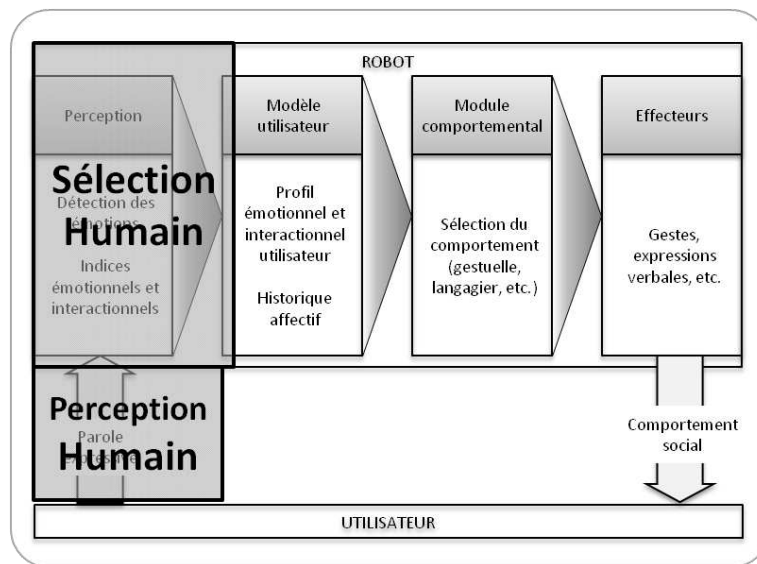


Figure 42. Boucle d'interaction Utilisateur-Système avec perception de la parole expressive par l'expérimentateur. La gestion des autres modules est automatique.

Description de l'architecture

La Figure 42 représente l'architecture du système que nous avons testé au cours de l'expérimentation NAO-HR3 (les « Histoires Interactives »). Cette expérimentation, menée auprès d'enfants en situation de jeu avec le robot, visait à collecter des données émotionnelles. Nous avons procédé également à un test de notre

système en conditions réelles, et fait passer aux enfants un questionnaire d'évaluation de la satisfaction et de l'adaptation robotique à la fin de l'expérimentation.

Dans ce système, la perception de l'émotion exprimée par l'utilisateur est prise en compte par l'expérimentateur, et transmise manuellement au système en charge d'en déduire automatiquement le profil utilisateur et de sélectionner le comportement robotique le plus approprié. Avec ce système, nous visions à valider les performances d'un module particulier. Dans ce contexte, il nous a semblé préférable de tester en même temps la constitution automatique du profil et la sélection de comportement. En effet, si nous avions voulu simuler la sélection de comportement robot par un manipulateur humain tout en réalisant la constitution automatique du profil, cette tâche d'interprétation à partir du profil se serait avérée fastidieuse à réaliser en temps réel par le manipulateur. Dans le cas inverse, avec définition manuelle du profil et sélection automatique du profil, la définition subjective des variations du profil à chaque tour de parole aurait été également fastidieuse, mais aussi sujette à une trop grande subjectivité : comme nous l'avons vu précédemment, l'annotation du profil est une tâche subjective, pour laquelle les annotateurs peinent à s'accorder lorsqu'il s'agit d'une évaluation globale de l'individu. L'annotation au niveau du tour de parole (en prenant en compte les énonciations précédentes, puisque notre théorie gère l'historique de l'interaction) nécessiterait un entraînement et une codification précise. Toutefois, cette manipulation peut être conceptuellement très intéressante afin d'évaluer la relation entre profil et sélection du comportement.

D'autres architectures en Magicien d'Oz pourraient permettre de tester les différents modules. Par exemple, il serait possible de tester l'impact des erreurs de détection émotionnelle sur la constitution du profil, en utilisant un module de détection automatique et une mise à jour du profil manuelle en fonction des inputs reçus. La sélection de comportements pourrait être réalisée ensuite automatiquement (si le système est stable et a été testé), ou manuellement en suivant les règles théoriques. L'évaluation de la satisfaction utilisateur serait alors très probablement motivée par la bonne reconnaissance de ses émotions en amont.

Pour quelle évaluation ?

Cette évaluation permet de tester l'impact d'un module automatisé spécifique sur la satisfaction de l'utilisateur (que ce soit une évaluation finale en fonction des émotions exprimées, ou d'un questionnaire). Ceci sous-entend que les autres modules ont fonctionné « parfaitement », puisqu'ils ne font pas intervenir de prises de décision autres qu'humaines (la détection de l'émotion en entrée), voire pas de prise de décision (l'expression du comportement est constituée de phrases scriptées).

6.4.4. Système entièrement autonome

Il va sans dire que concevoir un système d'Interaction Humain-Machine, l'implémenter puis le soumettre à des tests utilisateurs dans une même foulée peut mener à une analyse très difficile en cas d'erreur. En effet, si le but recherché pour ce système n'est pas atteint (réussite d'une tâche, satisfaction sociale, etc.), il serait impossible – ou du moins très délicat – de déterminer les causes. L'évaluation d'un système autonome nécessite donc obligatoirement d'avoir déjà validé tous les modules, un à un. Cette évaluation finale doit être réalisée en termes de réussite de l'objectif visé. Dans le cas d'un robot social sensible aux émotions, l'objectif est d'obtenir la satisfaction affective de l'utilisateur.

Description de l'architecture

Nous nous reportons à la Figure 39 (« Boucle d'interaction émotionnelle Utilisateur-Robot : perception de la parole expressive utilisateur, système émotionnel, module comportemental et génération de comportement social. ») représentant l'architecture de notre système d'interaction. Ici, tous les modules sont automatisés. La détection des émotions est réalisée par classification statistique, la mise à jour du profil et la sélection de comportement sont déterminés par prise de décision en logique floue, et les comportements sont générés par le système (synthèse à partir du texte, mouvements).

Pour quelle évaluation ?

Le système entièrement autonome nous permet de réaliser des tests utilisateur visant à déterminer la réussite de l'objectif global de notre système. La Magicien d'Oz complet, comme nous l'avons décrit précédemment, permet également de tester la satisfaction utilisateur, et de s'assurer que des théories sont valables ; toutefois, dans le cas d'un système entièrement autonomisé, nous nous assurons que ce sont nos théories et leur implémentation qui ont permis d'atteindre ce but.

Réaliser une évaluation du système final permettrait alors de :

- valider le respect des codes de communication interpersonnelle (le robot doit présenter une logique émotionnelle et avoir une attitude sociale cohérente),
- valider l'acceptation sociale du robot (via la satisfaction de l'utilisateur par rapport aux comportements du robot).

Nous travaillerions donc sur deux thèmes : la cohérence des comportements du robot en fonction de l'attitude de l'utilisateur, et la satisfaction de l'utilisateur.

Cette expérimentation pourrait être menée sur une tâche relative à notre contexte d'application finale, l'assistance lors du lever le matin, où le robot aide à la constitution du planning de la journée. Le locuteur dialoguerait avec le robot pendant un nombre limité de tours sur cette tâche. Le robot réagit à l'émotion exprimée par le locuteur, et répond d'une façon conforme au comportement sélectionné, puis lance une phrase de relance pour continuer à faire parler le locuteur.

Chaque utilisateur se verrait proposer deux passes d'interaction, selon la dimension testée. Dans le cas où nous souhaiterions tester la satisfaction utilisateur, le robot n'utiliserait pas de stratégie (i.e. serait neutre), puis dans une seconde passe nous lui soumettrions le système réel (adaptation du comportement en fonction du profil). Pour évaluer la cohérence, nous proposerions en première passe des stratégies de comportement aléatoires (pour chaque tour de parole, le robot sélectionne un comportement au hasard), puis en deuxième passe le vrai système.

Après chaque passe, le participant remplit un questionnaire sur des échelles de Likert à 5 points :

- Le robot reconnaissait bien mes émotions.
- Le robot s'adaptait à mon état émotionnel.
- J'ai trouvé l'attitude du robot avec moi plaisante.
- J'ai trouvé l'attitude du robot avec moi cohérente.

6.5. Conclusion

Ce chapitre a présenté la façon dont nous avons pu formaliser et implémenter le profil émotionnel et interactionnel de l'utilisateur, et la sélection automatique du comportement du système en fonction du profil. Cette implémentation a été réalisée à l'aide d'un système de prise de décision, sur des règles représentées en logique floue.

Nous avons présenté une expérimentation du système réalisée à l'aide d'un Magicien d'Oz semi-automatique simulant juste la détection des émotions, tandis que le profil et le comportement étaient traités automatiquement. L'analyse du comportement de notre système au cours de cette expérimentation nous a permis de collecter un riche corpus d'affect bursts produits par des enfants en interaction ludique avec un robot, et ouvert de nouvelles pistes d'améliorations de notre système.

Par ailleurs, cette expérimentation nous a amené à nous interroger sur l'intérêt de la technique du Magicien d'Oz dans l'évaluation d'un système d'Interaction Humain-Machine. Nous nous sommes donc livrés à un retour sur les différentes expérimentations réalisées au cours de ce travail de thèse, et la façon dont les différents degrés d'autonomie de nos systèmes nous ont permis de tester certaines caractéristiques précises de l'interaction. Nous avons également proposé d'autres expérimentations pouvant permettre de finaliser l'évaluation du système.

Le Magicien d'Oz complet est utile lors de la conception d'un système. En effet, cela permet d'évaluer les besoins de l'utilisateur en contexte, voire même de définir mieux le contexte, et de se représenter les différentes briques du système et leur éventuel impact.

Quant au système semi-automatisé, il peut permettre de procéder à des évaluations techniques des différents modules constituant le système.

La richesse et la difficulté de l'évaluation de ce type d'interaction réside également dans la connaissance de l'humain tout d'abord, mais aussi dans la compréhension de l'interaction, des liens entre les deux interactants. L'influence du caractère robotique de l'interactant (i.e. non-humain) peut également jouer un rôle, puisque l'utilisateur lui attribuera des caractéristiques spécifiques que nous ne contrôlerons pas nécessairement (intelligence présumée, affectivité, etc.). C'est pourquoi il est essentiel de se montrer particulièrement rigoureux dans l'évaluation d'un tel système, et de procéder tout d'abord à une analyse modulaire avant de songer à évaluer le système dans sa globalité.

Conclusion et perspectives

7.1. Résumé

Grâce aux robots sociaux complètement anthropomorphiques japonais (Gemini), aux robots humanoïdes expressifs (tels que Nexi, Kismet, Nao, Romeo, iCub), aux animaux (iCat, Paro), aux peluches (Keepon, Leonardo, Emotirob), la communauté mondiale des Interaction Humain-Robot dispose à l'heure actuelle d'une pléthore d'outils permettant l'étude des différentes composantes de l'interaction sociale. Pour l'instant, la majorité des travaux est cependant menée en laboratoire.

Dans les travaux de la communauté, une grande importance est accordée à l'analyse de l'émotion dans l'Interaction Humain-Robot. Qu'il s'agisse de l'analyse des émotions exprimées par l'utilisateur et de la constitution d'un profil utilisateur (citons par exemple les travaux de Telecom ParisTech ou du MIT), de la constitution d'un modèle émotionnel interne pour le robot (l'université Waseda ou le MIT, par exemple), ou encore des études sur le comportement social ou émotionnel du système (l'université de Carnegie Mellon), les approches sont très diversifiées.

Notre travail de thèse s'inscrit donc dans l'étude de l'interaction émotionnelle entre un utilisateur et un robot, et propose une étude approfondie des unités minimales de compréhension des émotions exprimées dans la voix par l'être humain, afin de se livrer à une interprétation du profil de l'utilisateur et permettre la sélection d'un comportement robotique jugé désirable.

Pour résumer, nous avons présenté dans ce manuscrit un état de l'art des différents systèmes d'Interaction Humain-Machine, et la façon dont ils gèrent des entrées à plus ou moins haut niveau. Nous avons étudié certains mécanismes de mémorisation proposés pour gérer l'interaction à long terme, et les différentes attitudes sociales et émotionnelles des systèmes. Nous avons circonscrit ensuite le domaine d'étude sur lequel nous nous appuyons dans le cadre de nos travaux, et ce dans deux domaines : le profil utilisateur et le comportement robotique. Nous avons proposé des méthodes de constitution d'un profil émotionnel et interactionnel de l'utilisateur, et de sélection et représentation de comportements sociaux pour le robot. Dans le cadre de notre travail de thèse, nous avons collaboré à des collectes de données émotionnelles qui ont permis l'étude de différentes composantes de l'interaction émotionnelle entre l'utilisateur et le système. Ces collectes et les résultats obtenus ont été détaillés dans ce manuscrit. Nous avons proposé ensuite une implémentation de nos règles de constitution du profil et de sélection automatique du comportement, ainsi qu'une expérimentation hors laboratoire d'un système d'interaction émotionnelle et un protocole d'évaluation de l'Interaction Humain-Robot via l'utilisation de systèmes à différents degrés d'autonomie.

7.2. Contributions de la thèse

Notre travail de thèse s'est déroulé dans un contexte de projet robotique impliquant un robot social susceptible d'assister à leur domicile des personnes en perte d'autonomie, et également de pouvoir jouer avec des enfants. Nous nous sommes appliqués à déterminer de quelle façon les expressions émotionnelles de ces deux types de public pouvaient être formalisées et traitées sur du plus long terme, afin d'obtenir une représentation émotionnelle dynamique et adaptative. Nous avons décidé d'utiliser les mêmes dimensions pour représenter le profil, quel que soit le public (adultes, enfants). Une fois ce profil utilisateur défini, il était important que la prise en compte du profil de la part du robot soit visible pour l'utilisateur. Nous nous sommes alors intéressés, au-delà des simples expressions d'émotions robotiques, aux expressions de comportements sociaux de la part du robot.

Ces études ont mené à, tout d'abord, différentes expérimentations et collectes de données en contexte cible, puis à l'implémentation d'un système en collaboration avec les membres du projet. Nous avons pu évaluer certaines composantes de l'interaction entre l'utilisateur et le robot, et définir des protocoles d'évaluation du système d'interaction en fonction du degré d'autonomie du système.

Nos contributions consistent donc en la constitution d'un profil émotionnel et interactionnel de l'utilisateur à partir du signal de voix expressive ; la sélection de comportements sociaux pour le robot ; la formalisation et l'implémentation du profil et de la sélection ; et enfin l'évaluation par systèmes à différents degrés d'autonomie.

7.2.1. Le profil émotionnel et interactionnel de l'utilisateur

La compréhension de l'être humain est au cœur du domaine de l'Interaction Humain-Machine. Le chercheur se doit de repérer les composantes essentielles permettant de décrire le comportement humain, puis de découvrir comment les modéliser et les exploiter dans un système à long-terme. Mais le comportement humain est complexe, son raisonnement fondé sur une analyse riche du monde, il est donc difficile d'interpréter ses expressions et de tenter d'en déterminer les causes. Il est donc important de subdiviser notre analyse de l'être humain en sous-tâches plus faciles à gérer et à interpréter.

Notre travail de thèse a été consacré à l'élaboration d'un profil de l'utilisateur, dans une interaction parlée entre un humain et un robot, susceptible de représenter ses tendances émotionnelles et interactionnelles sur du long terme. Nous nous sommes concentrés sur l'analyse d'un type d'entrée, les émotions exprimées dans la voix, et avons proposé une interprétation dynamique des indices récupérés depuis un module de détection automatique des émotions. Nous proposons donc un profil utilisateur à partir de la paralinguistique dans la voix, pour obtenir une représentation dynamique et à plus long terme que l'émotion ponctuelle. L'interprétation à différents niveaux des indices paralinguistiques extraits d'une analyse automatique du signal audio est un travail original.

La définition des dimensions du profil s'est appuyée sur une étude des travaux de la communauté en interaction. Les dimensions utilisées habituellement concernent la personnalité de l'interlocuteur : nous avons pu établir la relation entre les émotions exprimées et la personnalité, et convenir que la représentation de la personnalité pourrait potentiellement être déduite des émotions exprimées par l'utilisateur. De même, les émotions exprimées peuvent être des indices de la relation que l'individu entretient avec l'autre. Cependant, notre approche ne nous permet pas d'interpréter la personnalité de l'interlocuteur et son rapport à l'autre dans son ensemble : nous travaillons avec des unités minimales de compréhension de l'utilisateur (ses expressions paralinguistiques), qui ne suffisent pas à conclure sur des représentations aussi pointues de la personnalité. En effet, les dimensions de personnalité utilisées en psychologie touchent à des manifestations très précises et techniques de l'individu, qui dans l'absolu ne nous intéressent pas pour une interaction avec un robot social qui, dans notre cas, n'a pas vocation thérapeutique.

Les études que nous avons menées nous ont donc permis de sélectionner six grandes dimensions permettant de décrire, tout d'abord, les tendances de l'utilisateur du point de vue de ses attitudes affectives : Est-il sûr de lui ? Exprime-t-il de grandes variations émotionnelles ? S'exprime-t-il de façon optimiste ? Est-il extraverti ? D'autre part, ces dimensions décrivent également la relation affective qu'il entretient avec son interlocuteur robotique : Est-ce qu'il apprécie le robot ? Est-ce qu'il le domine ? Ces six dimensions constituent un profil émotionnel et interactionnel dynamique de l'utilisateur, sur lequel le système peut s'appuyer pour déterminer le meilleur comportement à adopter. L'apport d'informations multimodales pourrait permettre d'étoffer ce profil par l'ajout d'autres dimensions, notamment une gestion des notions de

Solidarité (nommée également distance sociale) et de Familiarité (divulgaration progressive d'informations personnelles à l'interlocuteur).

7.2.2. Sélection du comportement social robotique

Afin d'étudier la boucle d'interaction entre l'utilisateur humain et la machine, nous nous sommes penchés sur l'analyse du comportement robotique. Nous avons déterminé, grâce à la littérature et à nos propres expérimentations, les comportements sociaux que l'on pourrait juger adéquats selon le type d'interaction. Grâce aux études menées par la communauté, nous avons pu déterminer un espace dans le circomplexe interpersonnel susceptible de regrouper tous les comportements sociaux désirables de la part d'un robot social. Nos études des réactions émotionnelles des utilisateurs nous ont permis de distinguer ensuite une délimitation de cet espace en fonction du rôle du robot (robot assistant, ou robot compagnon de jeu).

Nous nous sommes intéressés, par ailleurs, à la relation entre le profil de l'utilisateur et les comportements robotiques, et avons déterminé les comportements pouvant potentiellement être le mieux acceptés en fonction du profil présenté par l'utilisateur. Notre analyse nous a amenés à déterminer que le comportement social du robot, évoluant principalement sur un axe soumission-domination, pouvait être relié au degré de passivité-domination de l'utilisateur. Dans le cas de l'utilisateur, ce degré est déterminé en fonction des différentes dimensions émotionnelles et interactionnelles de son profil.

Enfin, nous avons proposé une codification de l'expression des comportements sociaux du robot susceptibles de représenter son attitude sociale, selon deux modalités : la posture, et le lexique. Nous nous sommes intéressés à la relation entre postures émotionnelles et attitudes sociales, en nous appuyant sur des études existantes sur la posture émotionnelle du robot NAO. Concernant l'emploi du lexique, nous avons étudié les modalités évaluatives du langage proposées en grammaire onomasiologique permettant de représenter l'attitude sociale du robot.

7.2.3. Formalisation et implémentation de la gestion du profil et de la sélection du comportement

L'implémentation des mécanismes de constitution et de mise à jour automatique du profil, et de sélection des comportements robotiques a été réalisée au cours de ce travail de thèse. Nous avons collaboré avec un partenaire du projet ROMEO à l'élaboration d'un système de gestion de l'interaction entre l'utilisateur et le robot.

La formalisation des règles de mise à jour du profil émotionnel et interactionnel repose sur l'utilisation de règles expert en logique floue et booléennes, implémentées dans un moteur de décision sélectionnant automatiquement le meilleur choix possible (augmenter ou diminuer une valeur du profil, ou ne rien faire). Les valeurs d'entrée permettant cette prise de décision est effectuée soit via une interface graphique de saisie des données paralinguistiques (durée du temps de parole, émotion, intensité), soit via la détection automatique des émotions dans la voix de l'utilisateur.

La sélection du meilleur comportement du robot se présente de façon similaire, en règles de logique floue implémentée dans le moteur de décision. Le meilleur comportement est défini en fonction de l'état du profil émotionnel et interactionnel de l'utilisateur en cours, mais également de l'état de la tâche lorsque cette

information est pertinente (dans le cadre ludique, par exemple, il peut s'agir de si l'utilisateur a gagné ou perdu ce tour de jeu).

Cette implémentation a permis de réaliser différents tests et démonstrations, et de proposer des méthodes d'évaluation du système d'interaction.

7.2.4. Le Magicien d'Oz à différents degrés d'autonomie pour l'évaluation

La technique du Magicien d'Oz est une méthode très usitée pour réaliser l'évaluation des systèmes en Interaction Humain-Machine. Particulièrement utilisée dans les systèmes de dialogue, elle permet de simuler l'intelligence de la machine, face à un utilisateur naïf, et ainsi de tester notamment l'utilisabilité du système.

Dans le cas d'un système d'Interaction Humain-Machine où l'objectif principal est la satisfaction sociale de l'utilisateur, l'évaluation représente un défi complexe. La satisfaction de l'utilisateur tient à une multitude de paramètres qu'il convient de rendre modulaire afin de les tester séparément et logiquement. Parmi ces paramètres, nous pouvons citer par exemple l'aspect du robot (plus ou moins humain), qui va engendrer des représentations mentales différentes de la part des utilisateurs, et donc des attentes différentes. Un autre paramètre influant sur la satisfaction peut être la réussite de la tâche à accomplir (qu'il s'agisse d'une organisation de planning ou d'un jeu). Il peut s'agir également de l'attitude sociale du système, ou encore sa compréhension de l'utilisateur... Au final, la satisfaction utilisateur englobe tous ces paramètres, il est donc difficile de déterminer lequel intervient dans la réussite ou l'échec de l'objectif de satisfaction.

Au cours de notre travail de thèse, nous nous sommes livrés à une série d'expérimentations avec des utilisateurs finaux en contexte de scénarios d'application. Ces expérimentations ont permis la collecte de données émotionnelles, mais également l'étude de certains paramètres de la boucle d'interaction émotionnelle entre un utilisateur et un robot social. Nous avons proposé différentes architectures de systèmes, à différents degrés d'autonomie, et présenté la façon dont celles-ci permettent l'évaluation modulaire et finale du système d'interaction.

L'architecture des systèmes utilisés peut être classée selon leur autonomie : du système en Magicien d'Oz complet (i.e. sans décision automatique), au système totalement autonome, en passant par les systèmes semi-automatiques. Nous avons pu catégoriser nos expérimentations ainsi :

- Systèmes en Magicien d'Oz complet :
 - IDV-HS : Interaction entre des adultes et un système
 - NAO-HR et IDV-HR : Interaction entre des adultes et des enfants, et un robot social
- Système semi automatisé :
 - NAO-HR3 : Interaction entre des enfants et un robot social
- Système autonome :
 - Nous avons proposé une architecture pour un système totalement autonome, que nous testerons sur un ensemble représentatif d'utilisateurs.

Cette architecture modulaire, et cette approche selon l'autonomie du système, permettent de réaliser une évaluation efficace et organisée du système d'Interaction Humain-Machine. Dans le cas d'une interaction avec un robot social, nous avons montré que le système en Magicien d'Oz complet permet de réaliser des tests utilisateurs en condition d'interaction « idéale », ou au contraire d'évaluer des comportements non

désirables du système, mais parfaitement contrôlés. Dans les deux cas, l'analyse de la satisfaction de l'utilisateur porte bien sur le niveau d'intelligence que l'on espère atteindre avec notre système, et l'on teste ainsi uniquement les différentes présentations de notre système, en nous affranchissant d'éventuelles défaillances techniques.

L'utilisation de systèmes semi-automatisés est une étape intermédiaire dans l'évaluation du système d'Interaction Humain-Machine. Cette architecture consiste à ne faire fonctionner qu'un seul module en autonome, et de manipuler tous les autres en Magicien d'Oz afin de simuler un fonctionnement parfait. L'évaluation permet ainsi de tester l'impact de ce module spécifique sur la satisfaction de l'utilisateur.

Il est envisageable de tester le système complet qu'une fois tous ces tests effectués. L'utilisation d'un système autonome implique qu'il n'y a aucune interférence ni décision humaine de la part de l'expérimentateur. Cette architecture permet, au même titre que le Magicien d'Oz complet, de réaliser des tests utilisateurs en condition finale ; toutefois, l'utilisation d'un système autonome nous permet de tester nos théories de conception, et notre formalisation et implémentation des modèles. L'évaluation du produit final permet alors, dans le cas d'un système d'interaction sociale, de tester la cohérence sociale et la satisfaction de l'utilisateur.

7.3. Les perspectives de travail et ouvertures

7.3.1. L'annotation perceptive du profil émotionnel et interactionnel

Un des challenges rencontrés lors de notre travail de thèse consiste en l'annotation perceptive du profil émotionnel et interactionnel. Cette étape est incontournable, tout d'abord, pour l'évaluation de la constitution automatique du profil, car il conviendrait de comparer les résultats obtenus par le système à une interprétation réelle de l'humain. D'autre part, cette annotation permettrait de réaliser les traitements statistiques nous permettant de tester la classification automatique et ainsi de valider la bonne adéquation de nos indices pour la constitution du profil.

Lors de notre test d'annotation du profil, nous avons évoqué le besoin d'approfondir le protocole d'annotation, car les dimensions sont extrêmement subjectives, puisqu'elles se basent sur des interprétations des émotions et des propos exprimés par les locuteurs. Ce processus d'annotation, au même titre que l'annotation émotionnelle, requerrait donc une formation des annotateurs, et une formalisation perceptive des dimensions.

7.3.2. L'évaluation

Nous avons réalisé différentes évaluations, sur plusieurs aspects de l'interaction entre l'utilisateur final et le robot. Nous avons proposé des protocoles d'évaluation modulaire, et également un protocole d'évaluation du système autonome final. Certains aspects de l'évaluation modulaire n'ont pas pu être réalisés, au même titre que l'évaluation sur système final, car ces évaluations nécessitent du temps et la présence d'un panel important de participants. Notre objectif serait donc de réaliser ces évaluations afin de stabiliser l'intégralité du système d'interaction.

D'autre part, il nous semblerait très pertinent et intéressant d'évaluer notre système sur du plus long terme. En effet, notre profil se proposant comme une représentation dynamique et adaptative de l'utilisateur,

s'affinant et se stabilisant au fil du temps de par son principe, il serait intéressant d'évaluer la satisfaction de l'utilisateur à long terme. Il pourrait être très intéressant de tester l'interaction sur une durée de plusieurs jours avec des utilisateurs.

7.3.3. Vers une approche multimodale

Nos travaux se sont focalisés sur l'interprétation de la paralinguistique dans la voix, afin de déterminer une représentation de l'utilisateur.

Une suite logique de nos travaux serait de constituer le profil utilisateur à partir des émotions, en l'adaptant à un système gérant d'autres modalités, telle que la gestuelle, des expressions faciales, des émotions exprimées dans le lexique, etc. Deux volets d'étude nous semblent alors captivants : l'étude approfondie de chacun de ces constituants en suivant le même procédé que celui utilisé au cours de cette thèse, et également la fusion de ces modalités afin d'améliorer la robustesse du profil. Le traitement dialogique (par exemple, pour la compréhension du topic) pourrait notamment permettre une analyse des dimensions de Solidarité et de Familiarité afin de renforcer le profil social de l'utilisateur.

Il est important de prendre en compte les modalités dans un contexte multimodal situé. Par exemple, une personne qui s'exprime fortement, cela ne signifie pas obligatoirement qu'elle est sûre d'elle et s'exprime avec aplomb, et impose son point de vue. Cela peut être au contraire une personne peu sûre d'elle, qui ne parvient pas à se contrôler et à s'exprimer posément. Une analyse, par exemple, de sa posture, de sa gestuelle, ou des propos tenus, pourrait venir affiner notre sélection des traits émotionnels, et rendre notre modèle plus pertinent et robuste.

Nous avons envisagé d'autre part, au cours de notre expérimentation NAO-HR3, l'importance d'indiquer au système des informations sur le degré de nouveauté que la situation d'interaction peut présenter pour l'utilisateur. Nous estimons que ce paramètre peut avoir un impact sur l'attitude de l'utilisateur, et devrait donc être pris en compte pour l'estimation du profil. Il serait alors possible d'ajouter au profil utilisateur deux indications : s'agit-il de la première interaction (ou de premières interactions) de cet individu avec le robot ? Et s'agit-il de la première fois que le robot et l'utilisateur effectuent cette tâche ensemble (le jeu, la constitution d'un planning, etc.) ? Outre l'adaptation automatique des règles de constitution du profil, ces informations pourraient amener le robot à sélectionner automatiquement de nouvelles tâches de communication de type tutoriel (pour le jeu et le planning), ou une phase de discussion générale afin que l'utilisateur fasse connaissance avec le robot.

Le renforcement des indices, soit par l'ajout de nouvelles modalités ou de données contextuelles, permettra alors 1) de rendre le système de constitution du profil émotionnel et interactionnel encore plus robuste, dont découlera nécessairement 2) une meilleure sélection du comportement en fonction des besoins de l'utilisateur.

7.3.4. Vers une plus grande autonomie du système

La communauté en Interaction Humain-Machine s'inspire de plus en plus de modèles bio-inspirés dans la conception de ses systèmes informatiques. Qu'il s'agisse de l'utilisation de théories issues des neurosciences comme nous l'avons vu dans les travaux de (Buendia et Devillers, 2012), ou la reproduction des mécanismes cognitifs de l'être humain (notamment le principe de mémoire associative), toutes les composantes de l'être

humain sont intéressantes à modéliser, afin de rendre l'interaction avec le système plus naturelle et plus fluide.

En Interaction Humain-Machine, et plus particulièrement en robotique sociale, l'interaction à très long terme, c'est-à-dire à l'échelle de l'année, voire même « *lifelong* », est un centre d'intérêt qui inspire de nombreux chercheurs. De nombreuses conférences internationales (parmi lesquelles nous pouvons citer l'IEEE International Conference on Development and Learning and Epigenetic Robotics) s'attachent à réaliser un pont entre les différentes disciplines (psychologie comportementale, sociale et développementale, neurosciences, robotique, apprentissage automatique...) en vue de la constitution de modèles « à long terme ».

Le désir d'implanter un robot social à domicile implique une relation quotidienne à long terme avec l'utilisateur. Il importe de doter le système de mécanismes d'adaptation et d'une grande flexibilité, afin de lui permettre d'évoluer, de progresser de façon autonome. Il est intéressant alors d'envisager pour le robot les mêmes procédés cognitifs qu'un humain, notamment de pouvoir apprendre de ses erreurs, pour ainsi concevoir un robot qui pourrait apprendre tout seul.

Il serait alors fascinant de s'inspirer des méthodes d'apprentissage autonome en robotique (Benureau et Oudeyer, 2013), afin de proposer une représentation flexible de l'utilisateur permettant une adaptation automatique et autonome en fonction de nouveaux paramètres.

Publications et présentations associées à la thèse

Communications dans des conférences internationales avec actes

- Tahon, M., **Delaborde**, A., Devillers, L. *Corpus of Children Voices for Mid-level Markers and Affect Bursts Analysis*. In proc. 8th international conference on Language Resources and Evaluation (LREC). Istanbul, Turkey. Mai 2012.
- Tahon, M., **Delaborde**, A., Devillers, L. *Real-life Emotion Detection from Speech in Human-Robot Interaction: Experiments across Diverse Corpora with Child and Adult Voices*. In proc. 12th Annual Conference of the International Speech Communication Association (Interspeech). Florence, Italy. August, 2011.
- Delaborde**, A., Devillers, L. *Use of Nonverbal Speech Cues in Social Interaction between Human and Robot: Emotional and Interactional markers*. In proc. 3rd International Workshop on Affective Interaction in Natural Environments, ACM Multimedia, Firenze, Italy. October 2010.
- Tahon, M., **Delaborde**, A., Barras, C., Devillers, L. *A Corpus for Identification of Speakers and their Emotions*. In proc. of the International Workshop on Corpora for Research on Emotion and Affect, 7th LREC, Valletta, Malta. May 2010.
- Delaborde**, A., Tahon, M., Devillers, L. *Affective Links in a Child-Robot Interaction*. In proc. of the International Workshop on Corpora for Research on Emotion and Affect. LREC, Valletta, Malta. May 2010.
- Delaborde**, A., Tahon, M., Barras, C., Devillers, L. *A Wizard-of-Oz game for collecting emotional audio data in a children-robot interaction*. In proc. of the International Workshop on Affective-aware Virtual Agents and Social Robots, ICMI-MLMI, Boston, USA. November 2009.
- Rollet, N., **Delaborde**, A., Devillers, L. *Protocol CINEMO: The use of fiction for collecting emotional data in naturalistic controlled oriented context*. In proc. Affective Computing & Intelligent Interaction Conference (ACII), Amsterdam. Septembre 2009.

Communications dans des conférences nationales avec actes

- Delaborde**, A., Devillers, L. *Impact of the Social Behaviour of the Robot on the User's Emotions: Importance of the Task and the Subject's Age*. In proc of WACAI 2012: Workshop Affect, Compagnon Artificiel, Interaction. Grenoble, France. Novembre 2012.
- Delaborde**, A., Devillers, L. *Impact du Comportement Social d'un Robot sur les Émotions de l'Utilisateur : une Expérience Perceptive*. In proc. of Journées d'Études sur la Parole (JEP). Grenoble, France. Juin 2012.

Présentations scientifiques sans comité de lecture

- « Construction du Profil Émotionnel et Interactionnel dynamique de l'utilisateur pour piloter les comportements robotiques » groupe de travail "Affects, Compagnons Artificiels et Interactions" (ACAI) du GDR I3. 6 avril 2012, UPMC, Paris.
- « Modèle Émotionnel des Échanges pour l'Interaction avec un Robot ». Journée des Doctorants du projet ROMEO. 18 novembre 2010, Telecom Paris-Tech, Paris.

Bibliographie

- Aarsand, P. A. et Aronsson, K. (2009). Response cries and other gaming moves—Building intersubjectivity in gaming. *Journal of Pragmatics*, 41(8), 1557-1575.
- Alameda-Pineda, X., Sanchez-Riera, J., Franch, V., Wienke, J., Cech, J., Kulkarni, K., et al. (2011). *The RAVEL data set*. ICMI 2011 Workshop on Multimodal Corpora, Alicante, Spain.
- Altman, L. et Taylor, D. A. (1973). *Social Penetration: The Development of Interpersonal Relationships*. New York: Holt, Rinehart & Winston.
- Aly, A. et Tapus, A. (2013). *A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction*. 8th ACM/IEEE international conference on Human-robot interaction
- Arroyo, I., Cooper, D. G., Bursleson, W., Woolf, B. P., Muldner, K., Christopherson, R. (2009). *Emotion Sensors Go to School*. Artificial Intelligence in Education.
- Balbo, S. (1994). *Evaluation ergonomique des interfaces utilisateur: un pas vers l'automatisation*. (Thèse de doctorat), Université Joseph-Fourier-Grenoble I.
- Baron-Cohen, S. E., Tager-Flusberg, H. E. et Cohen, D. J. (2000). *Understanding other Minds: Perspectives from Developmental Cognitive Neuroscience*.: Oxford University Press.
- Baron-Cohen, S., Wheelwright, S. (2004). The Empathy Quotient: an Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. *Journal of autism and developmental disorders*, 34(2), 163-175.
- Batliner, A., Hacker, C., Steidl, S., Nöth, E., D'Arcy, S., Russel, M., et al. (2004). *"You Stupid Tin Box" - Children Interacting with the AIBO Robot: a Cross-linguistic Emotional Speech Corpus*. 4th Conference on Language Resources and Evaluation (LREC).
- Batliner, A., Steidl, S., Schuller, B., Seppi, D., Laskowski, K., Vogt, T., Devillers, L., Vidrascu, L., Amir, N., Kessous, L., Aharonson, V. (2006). *Combining Efforts for Improving Automatic Classification of Emotional User States*. IS-LTC.
- Beck, A., Cañamero, L., Damiano, L., Somavilla, G., Tesser, F., Cosi, P. (2011). *Children Interpretation of Emotional Body Language Displayed by a Robot*. International Conference on Social Robotics (ICSR).
- Beck, A., Hiolle, A., Mazel, A. et Cañamero, L. (2010). *Interpretation of Emotional Body Language Displayed by Robots*. 3rd International Workshop on Affective Interaction in Natural Environments.
- Benureau, F. et Oudeyer, P. Y. (2013). *Autonomous Reuse of Motor Exploration Trajectories*. International Conference on Development and Learning and on Epigenetic Robotics.

- Bickmore, T. W. et Cassell, J. (2001). *Relational agents: a model and implementation of building user trust*. Human-Computer Interaction CHI, Seattle, Washington.
- Bickmore, T. W. et Cassell, J. (2005). Social Dialogue with Embodied Conversational Agents. *Advances in Natural Multimodal Dialogue Systems*, 30, 23-54.
- Bosser, A. G., Leveux, G., Sehaba, K., Buendia, A., Corruble, V., De Fondaumière, G., et al. (2007). Dialogs taking into account experience, emotions and personality. *Entertainment Computing–ICEC 2007*, 356-362.
- Breazeal, C. (1999). *Robot in Society: Friend or Appliance?* Autonomous Agents Workshop on Emotion-Based Agent Architectures, Seattles, WA.
- Breazeal, C. (2002). Designing sociable robots. *Intelligent Robotics & Autonomous Agents*. MIT press.
- Briggs, K. C. (1976). *Myers-Briggs type indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Brown, P. et Levinson, S. (1987). *Politeness: Some universals in language usage*: Cambridge University Press.
- Brown, R. et Gilman, A. (1960). *The pronouns of power and solidarity* Cambridge, MA: MIT Press.
- Buendia, A. et Devillers, L. (2012). *From informative cooperative dialogues to long-term social relation with a robot*. International workshop series on spoken dialogue systems technology. IWSDS'12, Paris, France.
- Burger, S., MacLaren, V. et Yu, H. (2002). *The ISL meeting corpus: the impact of meeting type on speech style*. Interspeech 2002.
- Campbell, N. (2007). *Whom we laugh with affects how we laugh*. Interdisciplinary Workshop on The Phonetics of Laughter, Saarbrücken.
- Carson, R.C. (1969). *Interaction concepts of personality*. Chicago: Aldine.
- Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A. et McOwan, P. (2010a). Affect Recognition for Interactive Companions: Challenges and Design in Real World Scenarios. *Multimodal User Interfaces*, 3, 89-98.
- Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A. et McOwan, P. W. (2010b). *Inter-ACT: An affective and contextually rich multimodal video corpus for studying interaction with robots*. Proceedings of the international ACM conference on Multimedia.
- Charaudeau, P. (1992). *Grammaire du sens et de l'expression*: Hachette.
- Charaudeau, P. (2000). Une problématisation discursive de l'émotion. In C. Plantin, Doury, M., Traverso, V. (Ed.), *Les émotions dans les interactions* (pp. 126): Presses universitaires de Lyon.
- Chastagnol, C. et Devillers, L. . (2012). *Détection d'émotions dans la voix de patients en interaction avec un agent conversationnel animé*. JEP-TALN-RECITAL 2012.

- Chernova, S., DePalma, N., Morant, E. et Breazeal, C. (2011). *Crowdsourcing human-robot interaction: Application from virtual to physical worlds*. RO-MAN, 2011 IEEE.
- Chetouani, M. (2011). *Traitement du signal social et robotique personnelle : Signaux, actes de communication et comportements*. (Habilitation à diriger des recherches), Spécialité Sciences de l'ingénieur. Université Pierre et Marie Curie. Paris 6.
- Clavel, C. (2007). *Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales*. (Thèse de doctorat), École Nationale Supérieure des Télécommunications, Paris.
- Clavel, C., Adda, G., Cailliau, F., Garnier-Rizet, M., Cavet, A., Chapuis, G., et al. (2012). Spontaneous speech and opinion detection: mining call-centre transcripts. *Language Resources and Evaluation*, 1(37).
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measures*, 20(1), 37-46.
- Collins, N. L. et Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review. *Psychological bulletin*, 116(3), 457.
- Core, M. et Allen, J. (1997). *Coding dialogs with the DAMSL annotation scheme*. AAAI fall symposium on communicative action in humans and machines.
- Cosnier, J. (1987). *Expression et Régulation des Émotions*. Colloque international sur les émotions, Laboratoire européen de psychologie sociale des Sciences de l'Homme, Paris.
- Cosnier, J. (2003). Les deux voies de communication des émotions. In JM. Colletta & A. Tcherkassof (Eds.), *Perspectives actuelles sur les émotions. Cognition, langage et développement*: Mardaga, Hayen.
- Dahlbäck, N., Jönsson, A. et Ahrenberg, L. (1993). Wizard of Oz studies—why and how. *Knowledge-based systems*, 6(4), 258-266.
- Damm, O., Malchus, K., Hegel, F., Jaacks, P., Stenneken, P., Wrede, B., et al. (2011). *A computational model of emotional alignment*. 5th Workshop on Emotion and Computing.
- Dautenhahn, K. et Werry, I. (2002). *A Quantitative Technique for Analyzing Robot-Human Interactions*. International Conference on Intelligent Robots and Systems, Lausanne, Switzerland.
- Davidson, R. J., Scherer, K. R. et Goldsmith, H. H. (2003). *Handbook of Affective Sciences*. New York: Oxford University Press.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidisciplinary approach. *Journal of Personality and Social Psychology* 44, 113-126.
- De Rosis, F., Mazzotta, I., Miceli, M. et Poggi, I. (2006). Persuasion artifices to promote wellbeing. *Persuasive Technology*. Springer., 84-95.

- De Vignemont, F. et Singer, T. (2006). The empathic brain: how, when and why? *Trends in cognitive sciences*, 10(10), 435-441.
- Delaborde, A. et Devillers, L. (2010). *Use of Nonverbal Speech Cues in Social Interaction between Human and Robot: Emotional and Interactional Markers*. 3rd International Workshop on Affective Interaction in Natural Environments, ACM Multimedia, Firenze, Italy.
- Delaborde, A., Devillers, L. (2012a). *Impact du Comportement Social d'un Robot sur les Émotions de l'Utilisateur : une Expérience Perceptive*. Journées d'Études sur la Parole (JEP), Grenoble, France.
- Delaborde, A., Devillers, L. (2012b). *Impact of the Social Behaviour of the Robot on the User's Emotions: Importance of the Task and the Subject's Age*. WACAI 2012: Workshop Affect, Compagnon Artificiel, Interaction, Grenoble, France.
- Delaborde, A., Tahon, M., Barras, C. et Devillers, L. (2009). *A Wizard-of-Oz game for collecting emotional audio data in a children-robot interaction*. International Workshop on Affective-aware Virtual Agents and Social Robots, ICMI-MLMI, Boston, USA.
- Delaborde, A., Tahon, M. et Devillers, L. (2010). *Affective Links in a Child-Robot Interaction*. International Workshop on Corpora for Research on Emotion and Affect. LREC, Valetta, Malta.
- Devillers, L., Cowie, R., Martin, J., Douglas-Cowie, E., Abrilian, S., McRorie, M. (2006). *Real-life Emotions in French and English TV Video Clips: an Integrated Annotation Protocol Combining Continuous and Discrete Approaches*. 5th Int. Conf. on Language Resources and Evaluation (LREC), Genoa, Italy.
- Devillers, L. et Martin, J.-C. (2008). *Coding Emotional Events in Audiovisual Corpora*. 6th Int. Conf on Language Resources and Evaluation (LREC'2008), Marrakech, Morocco.
- Devillers, L., Maynard, H. et Paroubek, P. (2002). Méthodologies d'évaluation des systèmes de dialogue parlé: réflexions et expériences autour de la compréhension. *Traitement automatique des langues (TAL)*, 43(2), 155-184.
- Devillers, L. et Vasilescu, I. (2004). *Reliability of Lexical and Prosodic cues in two real-life spoken dialog corpora*. LREC, Lisbonne.
- Devillers, L. et Vidrascu, L. (2006). *Real-life emotions detection with lexical and paralinguistic cues on Human-Human call center dialogs*. Interspeech 2006.
- Devillers, L., Vidrascu, L. et Lamel, L. (2005a). Challenges in Real-life Emotion Annotation and Machine Learning Based Detection. *Journal of Neural Networks*, 18, 407-422.

- Devillers, L., Vidrascu, L. et Lamel, L. (2005b). Emotion detection in real-life spoken dialogs recorded in call center. *Journal of Neural Networks*, 18.
- Douglas-Cowie, E., Cowie, R., Cox, C., Amir, N. et Heylen, D. (2008). *The sensitive artificial listener: an induction technique for generating emotionally coloured conversation*. Workshop Corpora for Research on Emotion and Affect.
- Douglas-Cowie, E., Devillers, L., Martin, J.-C., Cowie, R., Savvidou, S., Abrilian, S., et al. (2005). *Multimodal databases of everyday emotion: content and labelling*. Interspeech.
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3), 177-190.
- Duhaut, D. (2010). *A way to put empathy in a Robot*. International Conference on Artificial Intelligence ICAI'10, Las Vegas, USA
- Duhaut, D. et Pesty, S. (2011). *Acceptability in Interaction: From Robots to Embodied Conversational Agents*. Computer graphics theory and applications, Algarve, Portugal.
- Eberhard, K. M., Nicholson, H., Kübler, S., Gundersen, S. et Scheutz, M. (2010). *The Indiana "Cooperative Remote Search Task" (CReST) Corpus*. LREC'10.
- Ekman, P. (1982). *Emotion in the Human Face*. New York: Cambridge University Press.
- Ekman, P. (1992a). Are There Basic Emotions? *Psychological Review*, 99, 550-553.
- Ekman, P. (1992b). An argument for basic emotions. *Cognition and Emotion*, 6, 169-200.
- Ekman, P. et Rosenberg, E. L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York: Oxford University Press.
- Elliott, C. D. (1992). *The Affective Reasoner: A Process model of emotions in a multi-agent system*. (Ph.D. thesis), The Institute for the Learning Sciences, Northwestern University, Evanston, Illinois.
- Eysenck, H. J. et Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.
- Fasola, J. et Mataric, M.J. (2012). Using Socially Assistive Human–Robot Interaction to Motivate Physical Exercise for Older Adults. *Proceedings of the IEEE*, 100(8), 2512-2526.
- Feil-Seifer, D. et Mataric, M.J. (2008). *Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders*. International Symposium on Experimental Robotics.
- Feil-Seifer, D., Skinner, K. et Matarić, M. J. (2007). Benchmarks for Evaluating Socially Assistive Robotics. *Interaction Studies: Psychological Benchmarks of Human-Robot Interaction*, 8(3), 423-429.

- Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391-428). New York: Oxford University Press.
- Fertil, E. (2012). *Vivre avec les robots (Living with Robots)*. (Documentaire, 52 minutes), Production Gédéon Programmes, France.
- Flickner, M., Darrell, T. et Oviatt, S. (2004). Multimodal Interfaces that Flex, Adapt, and Persist. *Communications of the ACM*, 47(1), 57.
- Fujita, M., Hasegawa, R., Costa, G., Takagi, T., Yokono, J. et Shimomura, H. (2001). Physically and Emotionally Grounded Symbol Acquisition for Autonomous Robots. *Association for the Advancement of Artificial Intelligence (AAAI)*.
- Goetz, J. et Kiesler, S. (2002). *Mental Models and Cooperation with Robotic Assistants*. Conference on Human Factors in Computing Systems.
- Goetz, J., Kiesler, S. et Powers, A. (2003). *Matching robot appearance and behavior to tasks to improve human-robot cooperation*. The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003.ROMAN 2003.
- Goffman, E. (1974). Perdre la Face ou Faire Bonne Figure? *Les rites d'interaction*.
- Goffman, E. (1983). *Forms of Talk*: University of Pennsylvania Press.
- Granata, C., Chetouani, M., Tapus, A., Bidaud, P. et Dupourque, V. (2010). *Voice and Graphical based Interfaces for Interaction with a Robot Dedicated to Elderly and People with Cognitive Disorders*. 19th IEEE International Symposium in Robot and Human Interactive Communication.
- Gratch, J. et Marsella, S. (2004). Domain-independent Framework for Modeling Emotion. *Journal of Cognitive Systems Research*, 5(4), 269-306.
- Green, A., Hüttenrauch, H., Topp, E. A. et Eklundh, K. S. (2006). *Developing a contextualized multimodal corpus for human-robot interaction*. Fifth international conference on Language Resources and Evaluation LREC.
- Hoffman, G. et Breazeal, C. (2006). Robotic partners' bodies and minds: An embodied approach to fluid human-robot collaboration. *Cognitive Robotics*.
- Hofstee, W. K., de Raad, B. et Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, 63(1), 146-163.
- Ipgrave, J. (2009). The Language of Friendship and Identity: Children's Communication Choices in an Interfaith Exchange. *British Journal of Religious Education*, 31, 213-225.

- Isbister, K. (2006). *Better Game Characters by Design: A Psychological Approach*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Itoh, K., Miwa, H., Takanobu, H. et al., Takanishi. (2005). Application of Neural Network to Humanoid Robots - Development of Co-associative Memory Model. *Journal of Neural Networks*, 18(5), 666-673.
- Jayagopi, D. B. et Odohez, J. M. (2013). *Given that, should i respond? Contextual addressee estimation in multi-party human-robot interactions*. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2013.
- Jayagopi, D. B., Sheiki, S., Klotz, D., Wienke, J., Odohez, J. M., Wrede, S., et al. (2013). *The vernissage corpus: a conversational human-robot-interaction dataset*. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2013.
- John, O.P. et Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102-138). New York: Guilford Press.
- Jung, M. F., Lee, J. J., DePalma, N., Adalgeirsson, S. O., Hinds, P. J. et Breazeal, C. (2013). *Engaging robots: easing complex human-robot teamwork using backchanneling*. 2013 conference on Computer supported cooperative work.
- Kanda, T., Hirano, T., Eaton, D. et Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1), 61-84.
- Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., et al. (2007). A humanoid robot that pretends to listen to route guidance from a human. *Autonomous Robots*, 22(1), 87-100.
- Kidd, C., Taggart, W. et Turkle, S. (2006). *A sociable robot to encourage social interaction among the elderly*. IEEE International Conference on Robotics and Automation, ICRA 2006, Orlando, Florida, USA.
- Koulouri, T. et Lauria, S. (2009). *A corpus-based analysis of route instructions in human-robot interaction*. TAROS'09.
- Lang, C., Hanheide, M., Lohse, M., Wersing, H. et Sagerer, G. (2009). *Feedback interpretation based on facial expressions in human-robot interaction*. The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009 (RO-MAN 2009).
- Lapeyre, M., Rouanet, P. et Oudeyer, P. Y. (2013). *Poppy: a New Bio-Inspired Humanoid Robot Platform for Biped Locomotion and Physical Human-Robot Interaction*. Proceedings of the 6th International Symposium on Adaptive Motion in Animals and Machines (AMAM).

- Lazarus, R.S., Kanner, A.D. et Kolkman, S. (1980). Emotions: a cognitive-phenomenological analysis. In Plutchick & Kellerman (Eds.), *Emotion: theory, research and experience* (Vol. 11). San Diego: Academic Press.
- Leary, T. (1957). *Interpersonal diagnosis of personality*. New York: Ronald.
- Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R. et Paiva, A. (2013). The Influence of Empathy in Human-Robot Relations. *International Journal of Human-Computer Studies*, 71(3), 250-260.
- Luzzati, D. (2007). Le dialogue oral spontané: quels objets pour quels corpora. *Revue d'Interaction Homme-Machine*, 8(2).
- Maas, J. F. et Wrede, B. (2006). *BITT: A corpus for topic tracking evaluation of multimodal human-robot-interaction*. Int. Conference on Language and Evaluation (LREC).
- Magnenat-Thalmann, N., Kasap, Z. et Ben Moussa, M. (2008). *Communicating with a virtual human or a skin-based robot head*. 8th IEEE International Conference on Automatic Face & Gesture Recognition. FG'08.
- Mahdhaoui, A. (2010). *Analyse de Signaux Sociaux pour la Modélisation de l'interaction face à face*. (Thèse de doctorat), Université Pierre et Marie Curie-Paris VI.
- Mariani, J., Besnier, J.-M., Bordé, J., Cornu, J.-M., Farge, M., Ganascia, J.-G., et al. (2009). Pour une éthique de la recherche en Sciences et Technologies de l'Information et de la Communication (STIC). Rapport validé par le Comité d'Éthique du CNRS (COMETS) sur proposition du groupe de réflexion sur l'éthique de la recherche en STIC.
- Markel, N. N., Phillis, J. A., Vargas, R. et Howard, K. (1972). Personality traits associated with voice types. *Journal of Psycholinguistic Research*, 1(3), 249-255.
- Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces*, 64(2), 377-402.
- McCowan, I., Carletta, J., Kraaij, W., Ashby, S., Bourban, S., Flynn, M., et al. (2005). *The AMI meeting corpus*. 5th International Conference on Methods and Techniques in Behavioral Research.
- McCrae, R. R., Costa, P. T. et Busch, C. M. (1986). Evaluating comprehensiveness in personality systems: The California Q-Set and the five-factor model. *Journal of Personality*, 54(2), 430-446.
- McCrae, R. R. et John, O.P. (1992). An Introduction to the Five-factor Model and its Applications. *Journal of Personality*, 60, 175-215.
- McKeown, G., Valstar, M. F., Cowie, R. et Pantic, M. (2010). *The SEMAINE corpus of emotionally coloured character interactions*. 2010 IEEE International Conference on Multimedia and Expo (ICME).

- Miwa, H., Okuchi, T., Itoh, K., Takanobu, H. et Takanishi, A. (2003). *A new mental model for humanoid robots for human friendly communication – Introduction of learning system, mood vector and second order equations of emotion*. ICRA 2003.
- Mohammad, Y. et Nishida, T. (2010). Modelling Interaction Dynamics during Face-to-Face Interactions. *Modeling Machine Emotions for Realizing Intelligence. Smart Innovation, Systems and Technologies, 1*, 53-87.
- Mohammad, Y., Xu, Y., Matsumura, K. et Nishida, T. (2008). *The H3R Explanation Corpus human-human and base human-robot interaction dataset*. International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2008 (ISSNIP 2008).
- Mohammadi, G., Mortillaro, M. et Vinciarelli, A. (2010). *The voice of personality: Mapping nonverbal vocal behavior into trait attributions*. International Workshop on Social Signal Processing, Firenze, Italy.
- Mohammadi, G. et Vinciarelli, A. (2012). Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features. *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, 3*(3), 273-284.
- Morency, L.-P. (2010). Modeling Human Communication Dynamics. *IEEE Signal Processing Magazine, 27*(5), 112-116.
- Mori, M. (1970). The Uncanny Valley (Bukimi no tani). *Energy, 7*(4), 33-35.
- Nalin, M., Baroni, I., Kruijff-Korbayová, I., Canamero, L., Lewis, M., Beck, A., et al. (2012). *Children's adaptation in multi-session interaction with a humanoid robot*. RO-MAN, 2012 IEEE.
- Nijholt, A. (2012). *Computational Humor 2012*. 3rd International Workshop on Computational Humor, Amsterdam, Universiteit Twente, Faculteit Elektrotechniek, Wiskunde en Informatica.
- Ochs, M., Niewiadomski, R. R., Pelachaud, C. et Sadek, D. (2006). Expressions Intelligentes des Emotions. *Revue Des Sciences et Technologies de l'Information, serie Revue d'Intelligence Artificielle, 20*(4-5).
- Ochs, M., Sabouret, N. et Corruble, V. (2009). Simulation of the Dynamics of Non-Player Characters' Emotions and Social Relations in Games. *IEEE Transactions on Computational Intelligence and AI in Games, 1*(4), 281-297.
- Ochs, M., Sadek, D. et Pelachaud, C. (2012). A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems, 24*(3), 410-440.
- Okita, S. Y. et Schwartz, D. L. (2006). Young children's understanding of animacy and entertainment robots. *International Journal of Humanoid Robotics, 3*(3), 393-412.
- Ortony, A., Clore, G. et Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

- Pal, P., Iyer, A. N. et Yantorno, R. E. (2006). *Emotion detection from infant facial expressions and cries*. International Conference on Acoustics, Speech and Signal Processing (ICASSP'06), Toulouse, France.
- Pedica, C. et Vilhjalmsson, H.H. (2009). *Spontaneous avatar behavior for human territoriality*. International Conference on Intelligent Virtual Agents.
- Pesty, S. et Duhaut, D. (2012). *Artificial companion and acceptability a problem of design?* IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom), 2012.
- Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. New York: Harper & Row.
- Poppe, R., Truong, K. P., Reidsma, D. et Heylen, D. (2010). Backchannel strategies for artificial listeners. *Intelligent Virtual Agents. Springer.*, 146-158.
- Ray, C., Mondada, F. et Siegart, R. (2008). *What do people expect from robots?* IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008), Nice, France.
- Revelle, W. et Scherer, K.R. (2009). Personality and Emotion *Oxford Companion to Emotion and the Affective Sciences* (pp. 304-306).
- Rivière, J., Adam, C. et Pesty, S. (2012). *Un aca sincère comme compagnon artificiel*. Workshop Affect, Compagnon Artificiel, Interaction (WACAI 2012), Grenoble, France.
- Rollet, N., Delaborde, A. et Devillers, L. (2009). *Protocol CINEMO: the Use of Fiction for Collecting Emotional Data in Naturalistic Controlled Oriented Context*. Affective Computing & Intelligent Interaction Conference, Amsterdam, NL.
- Russell, J. A. et Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3), 273-294.
- Sacks, H. (1992). *Lectures on conversation*. Blackwell Publishing.
- Sadek, D. (1991). *Attitudes mentales et interaction rationnelle: vers une théorie formelle de la communication*. (Thèse de doctorat), Informatique, Université de Rennes I.
- Sanchez-Cortes, D., Aran, O. et Gatica-Perez, D. (2011). *An audio visual corpus for emergent leader analysis*. Workshop on multimodal corpora for machine learning: taking stock and road mapping the future, ICMi-MLMI.
- Sander, D. et Scherer, K. R. (2009). *Traité de psychologie des émotions*. Paris: Dunod.
- Sawada, T., Takagi, T. et Fujita, M. (2004). *Behavior selection and motion modulation in emotionally grounded architecture for QRIO SDR-4XII*. IEEE Intelligent Robots and Systems, 2004.(IROS 2004).
- Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), *Handbook of Psychophysiology: Emotion and Social Behavior* (pp. 165-197). London.

- Scherer, K. R. (1994). Affect bursts. In S. H. M. van Goozen, N. E. van de Poll & J. A. Sergeant (Eds.), *Emotions: Essays on Emotion Theory* (pp. 161-196): Lawrence Erlbaum Associates, Inc.
- Scherer, K. R. (2005a). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion*. Chichester, UK: John Wiley & Sons, Ltd.
- Scherer, K. R. (2005b). What are emotions? And how can they be measured? *Social science information*, 44(4), 695-729.
- Scherer, K. R. et Bänziger, T. (2004). *Emotional expression in prosody: A review and an agenda for future research*. Speech Prosody.
- Scherer, K. R. et Peper, M. (2001). Psychological theories of emotion and neuropsychological research. *Handbook of Neuropsychology*, 5, 17-48.
- Scherer, K. R., Schorr, A. et Johnstone, T. (2001). *Appraisal considered as a process of multilevel sequential checking*: Oxford University Press US.
- Schröder, M. (2003). Experimental study of affect burst. *Speech Communication*, 40(1-2), 99-116.
- Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., et al. (2013). Paralinguistics in Speech and Language : State-of-the-Art and the Challenge. *Computer Speech & Language*, 27(1), 4-39.
- Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., et al. (2012). *The INTERSPEECH 2012 Speaker Trait Challenge*. Interspeech'12.
- Shahid, S., Krahmer, E., Swerts, M., Melder, W. A. et Neerinx, M. A. (2009). *You make me happy: Using an adaptive affective interface to investigate the effect of social presence on positive emotion induction*. Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009.
- Shiwa, T., Kanda, T., Imai, M., Ishiguro, H. et Hagita, N. (2008). *How quickly should communication robots respond?* 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI).
- Spencer-Oatey, H. (1996). Reconsidering power and distance. *Journal of Pragmatics*, 20, 1-24.
- SpirOps. (2005). SpirOps Artificial Intelligence - General presentation. Retrieved May 4th, 2012, from <http://www.spirops.com/SpirOpsAbstractQ2-2005.pdf>
- Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., Goodrich, M. (2006). *Common Metrics for Human-Robot Interaction*. HRI'06, Salt Lake City, Utah, USA.
- Sterlin, P. (2007). *Gestion dynamique des dialogues d'un Personnage Non Joueur*. (Mémoire de master), Université Pierre et Marie Curie Paris VI et Spir.Ops.

- Stiefelhagen, R., Fugen, C., Gieselmann, R., Holzapfel, H., Nickel, K. et Waibel, A. (2004). *Natural human-robot interaction using speech, head pose and gestures*. International Conference on Intelligent Robots and Systems, 2004.(IROS 2004).
- Stiehl, W. D., Lee, J. K., Breazeal, C., Nalin, M., Morandi, A. et Sanna, A. (2009). *The buggable: a platform for research in robotic companions for pediatric care*. 8th International Conference on interaction Design and Children.
- Strapparava, C., Stock, O. et Mihalcea, R. (2011). Computational humour. In R. Cowie, Pelachaud, C., Petta, P. (Ed.), *Emotion-oriented systems The HUMAINE Handbook* (pp. 609-634): Springer Berlin Heidelberg.
- Strong, S. R., Hills, H. I. et Kilmartin, C. T. (1988). The dynamic relations among interpersonal behaviors: A test of complementarity and anticomplementarity. *Journal of Personality and Social Psychology*, 54(5), 798-810.
- Svennevig, J. (1999). *Getting Acquainted in Conversation: A Study of Initial Interactions*: John Benjamins Publishing Co.
- Tahon, M. (2012). *Analyse acoustique de la voix émotionnelle de locuteurs lors d'une interaction humain-robot*. (Thèse de Doctorat), Université Paris-Sud XI, Orsay.
- Tahon, M., Delaborde, A., Barras, C. et Devillers, L. (2010). *A Corpus for Identification of Speakers and their Emotions*. International Workshop on Corpora for Research on Emotion and Affect, 7th LREC, Valletta, Malta.
- Tahon, M., Delaborde, A. et Devillers, L. (2011). *Real-life Emotion Detection from Speech in Human-Robot Interaction: Experiments across Diverse Corpora with Child and Adult Voices*. 12th Annual Conference of the International Speech Communication Association (Interspeech), Florence, Italy.
- Tahon, M., Delaborde, A. et Devillers, L. (2012). *Corpus of children voices for mid-level social markers and affect bursts analysis*. LREC, Istanbul, Turkey.
- Tapus, A. et Mataric, M. J. (2006). Towards socially assistive robotics. *Journal of robotics - Society of Japan*, 24(5), 14.
- Tapus, A. et Mataric, M. J. (2007). *Emulating Empathy in Socially Assistive Robotics*. AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, Palo Alto, CA.
- Tapus, A. et Mataric, M. J. (2008). *Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance*. AAAI Spring Symposium on Emotion, Personality, and Social Behavior.

- Tellex, S., Kollar, T., Dickerson, S., Walter, M. R., Banerjee, A. G., Teller, S. J., et al. (2011). *Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation*. AAAI Conference on Artificial Intelligence.
- Thomaz, A. L., Berling, M. et Breazeal, C. (2005). *An Embodied Computational Model of Social Referencing*. the IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005., Nashville, TN, USA.
- Thorndike, R. L. (1966). Intellectual Status and Intellectual Growth. *Journal of Educational Psychology*, 57(8), 131-127.
- Trager, G. L. (1958). Paralanguage: a first approximation. *Studies in Linguistics*, 13, 1-12.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind* 49, 443-460.
- Vandewaetere, M., Desmet, P. et Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27(1), 118-130.
- Vaudable, C. et Devillers, L. (2012). *Negative emotions detection as an indicator of dialogs quality in call centers*. IEEE icassp 2012.
- Vaudable, C., N., Rollet et L., Devillers. (2010). *Annotation of Affective Interaction in Real-life Dialogs Collected in a Call-center*. LREC'10, Valetta, Malta.
- Vernier, M., Monceaux, L. et Daille, B. (2009). *DEFT'09: détection de la subjectivité et catégorisation de textes subjectifs par une approche mixte symbolique et statistique*. Actes de l'atelier de clôture du 5ème Défi Fouille de Textes (DEFT'09).
- Vidrascu, L. (2007). *Analyse et détection des émotions verbales dans les interactions orales*. (Thèse de doctorat), Université Paris Sud - Paris XI.
- Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing*, 3(1), 69-87.
- Von der Pütten, A. M. et Krämer, N. C. (2012). *A survey on robot appearances*. Seventh annual ACM/IEEE international conference on Human-Robot Interaction.
- Wada, K., Shibata, T., Musha, T. et Kimura, S. (2005). *Effects of robot therapy for demented patients evaluated by EEG*. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005.(IROS 2005).

- Wainer, J., Feil-Seifer, D. J., Shell, D. A. et Mataric, M. J. (2006). *The role of physical embodiment in human-robot interaction*. 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006.
- Walker, M.A., Litman, D.J., Kamm, C.A, Abella, A. (1997). *PARADISE: a Framework for Evaluating Spoken Dialogue Agents*. 8th conference on European chapter of the Association for Computational Linguistics (EACL '97), Madrid, Spain.
- Zara, A., Maffiolo, V., J-C., Martin et L., Devillers. (2007). *Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics*. 2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007), Lisbon, Portugal.
- Zivkovic, Z., Booij, O., Krose, B., Topp, E. A. et Christensen, H. I. (2008). From sensors to human spatial concepts: An annotated data set. *IEEE Transactions on Robotics*, 24(2), 501-505.

Liste des Tableaux

Tableau 1. État de l'art des corpus d'Interaction Humain-Robot dyadique et en groupe: composition et disponibilité.....	40
Tableau 2. Sélection de termes qualificatifs pour un individu selon différentes facettes de personnalité. Extrait et traduit de (Hofstee <i>et al.</i> , 1992).....	49
Tableau 3. Dimensions émotionnelles du profil utilisateur.....	55
Tableau 4. Dimensions interactionnelles du profil utilisateur	56
Tableau 5. Étiquettes paralinguistiques obtenues via le système de détection du LIMSI.....	60
Tableau 6. Dimensions du profil interactionnel et informations paralinguistiques afférentes	61
Tableau 7. Dimensions du profil émotionnel et informations paralinguistiques afférentes.....	62
Tableau 8. Récapitulatif de l'impact du profil émotionnel et interactionnel de l'utilisateur sur la sélection du comportement social du système.	72
Tableau 9. Relations entre actes locutifs et rôles sociaux du robot en fonction des différentes stratégies de comportement.....	75
Tableau 10. Expérimentation IDV-HS – Instructions fictives jouées par le participant	82
Tableau 11. Expérimentation IDV-HS – Scénarios et état affectifs attendus	84
Tableau 12. Corpus IDV-HS : Étiquettes d'annotation et regroupement en macro-classes.....	86
Tableau 13. Corpus IDV-HS : Matrice de confusion des annotations émotionnelles. Étiquettes regroupées en macro-classe.	88
Tableau 14. Corpus IDV-HS : Valences moyennes utilisées conjointement aux étiquettes d'annotation.....	91
Tableau 15. Expérimentation NAO-HR1 : Comportements sociaux désirables et non désirables du robot Nao, réactions émotionnelles attendues chez l'enfant, et exemples.....	95
Tableau 16. Expérimentation NAO-HR1 : Extrait transcrit du jeu des questions-réponses. Une fille de 9 ans (Loc. 1) et son frère de 11 ans (Loc. 2) jouent avec le robot.	96
Tableau 17. Corpus NAO-HR1 : Matrice de confusion des annotations émotionnelles. Étiquettes regroupées en macro-classe.	98
Tableau 18. Corpus NAO-HR1 : Valences moyennes utilisées conjointement aux étiquettes d'annotation.	99
Tableau 19. Expérimentation IDV-HR : Scénarios et états affectifs attendus.	108
Tableau 20. Corpus IDV-HR : Étiquettes d'annotation et regroupement en macro-classes.	111

Tableau 21. Corpus IDV-HR : Matrice de confusion de l'annotation de la macro-classe pour les enregistrements des huit premiers locuteurs.....	112
Tableau 22. Corpus IDV-HR : Matrice de confusion de l'annotation de la macro-classe pour les enregistrements des locuteurs 9 à 22.....	112
Tableau 23. Expérimentation IDV-HR : Résultat des questionnaires participant	115
Tableau 24. Scores d'agrément pour l'annotation du profil émotionnel et interactionnel sur les corpus IDV1 et IDV2 : proportion d'accord observé et Kappa.....	121
Tableau 25. Corpus IDV-HS et IDV-HR : Nombre d'échantillons pour les classes faibles et fortes des différentes dimensions du profil émotionnel et interactionnel. Restriction aux échantillons avec accord annotateurs.....	122
Tableau 26. Évaluation du désir d'augmenter le niveau de Confiance en soi, à chaque tour de parole...	131
Tableau 27. Postures sociales du robot NAO : Empathique, Encourageant, Directif, Soumis et Amical (captures d'écran Choregraphe).....	133

Table des Figures

Figure 1. Roue des émotions de (Plutchik, 1980), composée d'émotions prototypiques desquelles sont dérivés les états mixtes.....	27
Figure 2. Les émotions morales selon (Davidson <i>et al.</i> , 2003), réparties sur deux critères : l'intérêt de l'individu, et le caractère prosocial.....	28
Figure 3. Différence de durée des différents types d'états affectifs. D'après (Scherer et Bänziger, 2004).	29
Figure 4. Rapidité de variation pour les différents types d'affect. D'après (Scherer et Bänziger, 2004).....	45
Figure 5. Relation entre facteurs de personnalité, éléments de questionnaires et traits adjectivaux associés (McCrae et John, 1992).....	48
Figure 6 - Représentation des dimensions de Domination et d'Amicalité sur le Circomplexe Interpersonnel, et attitudes sociales dérivées de ces deux dimensions. Adapté de (Isbister, 2006).....	50
Figure 7 - Classification des comportements interpersonnels basés sur les notions de Domination et d'Amicalité. Extrait de (Leary, 1957), p. 65.....	51
Figure 8 - Attitudes interactionnelles sur le Circomplexe Interpersonnel. Extrait de (Isbister, 2006), p. 26.....	52
Figure 9. Traitement multi-niveau des indices paralinguistiques extraits de l'audio, contribuant à la constitution d'un profil émotionnel et interactionnel	57
Figure 10. Quelques indices bas-niveau calculés sur le signal audio et interprétation en informations paralinguistiques.....	59
Figure 11. Comportements sociaux envisageables pour un système d'interaction Humain-Machine. Adapté de (Isbister, 2006).	67
Figure 12. Délimitation des comportements sociaux « désirables » (gris foncé) en fonction du contexte applicatif (assistance à l'utilisateur à gauche, compagnon de jeu à droite). Adapté de (Isbister, 2006).	70
Figure 13. Comportements sociaux du système sélectionnés (du plus dominant au moins dominant) en fonction du niveau de domination exprimé par l'utilisateur	71
Figure 14. Exemples de robots à visages expressifs : expressions émotionnelles du robot Reeti (Robopec) et du robot Nexi (MIT).	76
Figure 15. Boucle d'interaction émotionnelle Utilisateur-Robot : perception de la parole expressive utilisateur, système émotionnel, module comportemental et génération de comportement social.	79
Figure 16. Figure 16. Expérimentation IDV-HH : Interface Magicien d'Oz pour la sélection de l'état affectif du locuteur	83
Figure 17. Figure 17. Expérimentation IDV-HH : Interface Magicien d'Oz pour les phases de « traitement » et annonce orale de l'état affectif détecté.....	83

Figure 18. Corpus IDV-HS : Utilisation moyenne de la troisième étiquette d'annotation par les annotateurs.....	89
Figure 19 - Expérimentation NAO-HR1 : deux garçons (10 et 12 ans) jouent avec le robot NAO.....	94
Figure 20. Distribution moyenne des stratégies comportementales du robot (comportement positif, négatif ou neutre) présentées à chaque enfant pour la totalité de sa session d'enregistrement.	101
Figure 21. Expérimentation NAO-HR1 : Réactions émotionnelles des enfants en fonction des comportements Robot et #segments.....	102
Figure 22. Expérimentation NAO-HR1 : Réactions émotionnelles des enfants en fonction des comportements Robot. Distinction Garçon/Fille. (« Négatif » : Colère, Tristesse, Peur ; « Autres » : Ambigu et Positif-Négatif)	103
Figure 23. Expérimentation NAO-HR1 : Réactions émotionnelles des enfants (et #segments) en fonction des comportements du robot et de leur tranche d'âge. De gauche à droite, première ligne : comportement négatif, positif ; seconde ligne : comportement neutre.	104
Figure 24. Expérimentation NAO-HR1 : Activation moyenne exprimée par les enfants (de 0/faible à 1/fort) en fonction des comportements positifs ou négatifs du robot, et comparaison avec l'activation exprimée en interaction libre	105
Figure 25. Plateforme de recherche de l'Institut de la Vision visant à améliorer l'habitat pour les personnes atteintes de déficience visuelle: l'appartement témoin Homelab. (photographie http://www.institut-vision.org)	106
Figure 26. Expérimentation IDV-HR : automate Choregraphe gérant l'initiation de l'interaction avec l'utilisateur. (logiciel Choregraphe par Aldebaran Robotics).....	108
Figure 27. Expérimentation IDV-HR : Extrait du scénario du réveil du matin (en forme). Énonciations prédéterminées du robot, et réponses possibles du participant.	109
Figure 28. Expérimentation IDV-HR : Valence moyenne exprimée par les participants (de 0/négative à 1/positive) en fonction des comportements positifs ou négatifs du robot. Figure de gauche : scénarios positifs ; figure de droite : scénarios négatifs.	117
Figure 29. Expérimentation IDV-HR : Activation moyenne exprimée par les participants (de 0/faible à 1/fort) en fonction des comportements positifs ou négatifs du robot. Figure de gauche : scénarios positifs ; figure de droite : scénarios négatifs.	117
Figure 30. Architecture du système d'interaction entre le robot et l'utilisateur pour une mise à jour du profil émotionnel et interactionnel à partir du signal de voix de l'utilisateur, et une sélection automatique du comportement robot.....	127
Figure 31. Exemple de description d'une règle de logique floue sur l'interface graphique SpirOps Editor. A gauche, représentation des paramètres contribuant à la prise de décision, et transmission (à droite) des valeurs de ces paramètres nécessaires pour le traitement.	129

Figure 32. Le robot NAO exprimant l'Encouragement lors de la vidéo documentaire Vivre avec les robots (Fertil, 2012).....	134
Figure 33. Expérimentation NAO-HR2 : Disposition du robot NAO devant le tableau des consignes du jeu (image de gauche). Deux filles jouant au jeu, debout face à NAO (image de droite).	135
Figure 34. Architecture du système semi-automatique pour le jeu Histoires Interactives : annotation émotionnelle manuelle, et prise de décision automatique pour la mise à jour du profil utilisateur et la sélection comportement robot via les règles implémentées dans SpirOps AI.	136
Figure 35. Expérimentation NAO-HR2 : Quatre enfants jouant par paires lors du jeu Histoires Interactives (une fille de 9 ans et un garçon de 6 ans à gauche ; deux garçons de 7 ans à droite).	137
Figure 36. NAO-HR3 : Résultats moyens du questionnaire, évaluation des enfants quant au jeu (0 : pas du tout d'accord, 5 : tout à fait d'accord).....	139
Figure 37. NAO-HR3 : Résultats moyens du questionnaire, évaluation du comportement du robot (0 : pas du tout d'accord, 5 : tout à fait d'accord).....	139
Figure 38. NAO-HR3 : Résultats moyens du questionnaire, évaluation des enfants quant au robot (0 : pas du tout d'accord, 5 : tout à fait d'accord).....	139
Figure 39. Boucle d'interaction émotionnelle Utilisateur-Robot : perception de la parole expressive utilisateur, système émotionnel, module comportemental et génération de comportement social.	141
Figure 40. Boucle d'interaction émotionnelle Utilisateur-Robot avec perception de la parole expressive et sélection de comportement par l'expérimentateur.	143
Figure 41. Boucle d'interaction émotionnelle Utilisateur-Robot avec seule sélection du comportement par l'expérimentateur.	144
Figure 42. Boucle d'interaction Utilisateur-Système avec perception de la parole expressive par l'expérimentateur. La gestion des autres modules est automatique.	145
Figure 43. Évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart à la moyenne de l'activation.....	189
Figure 44. Simulation de l'évolution de l'émotionnalité au fil de la session, en fonction des valeurs d'activation pour chaque tour de parole	190
Figure 45. Évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart à la moyenne de la valence.....	191
Figure 46. Simulation d'une succession de tours de parole pour l'évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur des écarts de l'Activation et de la Valence.....	191
Figure 47. Simulation d'une suite de tours de paroles pour l'évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart de l'Activation et de la Valence.	192
Figure 48. Évaluation du désir de faire augmenter ou décroître le niveau d'émotionnalité en fonction des valeurs de l'écart-type (Valence ou Activation). Projection pour 2 exemples de valeurs d'écart-type (0.2 et 0.7)	193

Figure 49. Simulation d'une suite de tours de paroles pour l'évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart de l'Activation et de la Valence, et les écart-types Valence et Activation	194
Figure 50. Simulation de tours de parole et évolution de l'extraversion en fonction de la durée et la moyenne des durées.	195
Figure 51. Interface graphique du logiciel de segmentation, d'annotation et de transcription Transcriber.	196

Annexes

Annexe I

Énonciations prononcées par le système dans le cadre des expérimentations

Expérimentation IDV-HR : énonciations du robot NAO

Liste des énonciations possibles du robot NAO, pour tous les scénarios, en fonction de ses différentes attitudes sociales. Les « → » indiquent les différentes réponses possibles, pour chaque niveau de l'arbre de dialogue, selon la réponse que le participant aura fournie.

Comportement Directif

<p>Scénario – Réveil du matin : en forme</p> <p>« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin? »</p> <p>« Tu es en pleine forme, donc dis-moi ce que tu veux faire aujourd'hui. »</p> <p>→ « J'inscris dans ton planning que tu iras également te promener aujourd'hui. Indique-moi la destination. »</p> <p>→ « J'inscris dans ton planning que tu iras te promener aujourd'hui. Indique-moi la destination. »</p> <p>→→ « Compris. »</p> <p>→→ « Tu iras au café ou chez des amis. »</p> <p>→→ « J'inscris dans ton planning que tu iras te promener aujourd'hui. Dis-moi si tu valides ce choix. »</p> <p>« J'ajoute à ton planning de la journée d'appeler des amis pour qu'ils viennent te rendre visite. Indique-moi lesquels »</p> <p>→ « Bien compris. »</p>
<p>Scénario – Réveil du matin : mauvaise santé</p> <p>« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin? »</p> <p>« Comme tu ne vas pas bien, j'inscris au planning d'appeler un médecin. Indique-moi à quelle heure tu souhaites qu'il vienne. »</p> <p>→ « Bien compris »</p> <p>« Tu vas aller prendre ton petit déjeuner maintenant. Indique-moi si tu veux de l'aide. »</p> <p>→ « Bien compris. »</p>
<p>Scénario – Réveil du matin : déprime</p> <p>« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas</p>

au beau fixe. Bonjour, comment vas-tu ce matin ? »

« Tu es déprimé, donc prends sur toi. Il faut que tu fasses venir tes amis, dis-moi à quelle heure il faut les faire venir. »

→ « On les contactera une fois que tu seras debout. »

→ « Bien. »

« Je vais te lire un livre en attendant. Indique-moi quel genre de livre tu veux. »

« Je vais te lire un livre. Indique-moi quel genre de livre tu veux. »

→ « Bien compris. »

Scénario – Réveil du matin : content

« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin ? »

« Comme tu es de bonne humeur, il faudra que tu fasses plein de choses aujourd'hui. Dis-moi ce qui te rend si content. »

→ « Bien compris. »

« On va faire du rangement dans l'appartement ce matin. Indique-moi par quelle pièce de l'appartement on doit commencer. »

→ « Bien compris. »

Scénario – Réveil du matin : très mauvaise santé, urgence

« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence.

Bonjour, comment vas-tu ce matin ? »

« Tu as eu un accident, donc tu restes calme. Dis-moi la gravité de ton état. »

→ « Bien. Dis-moi où se situe la douleur précisément. »

« Je vais t'aider soit à appeler un médecin, soit appeler un proche, soit t'aider à trouver un médicament. Indique-moi ton choix. »

→ « Compris. Je vais engager la relation téléphonique. »

→ « Compris. Je vais t'assister. »

Comportement Aimable

Scénario – Réveil du matin : en forme

« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin ? »

« Je constate que tu es en pleine forme. Que désires-tu faire aujourd'hui ? »

→ « Bien. Est-ce que tu désires aller te promener après ? »

→ « Bien. Est-ce que tu désires aller te promener ? »

→→ « C'est d'accord. »

→→ « Tu as la possibilité d'aller te promener au parc, aller chez des amis, ou aller au café. Que préfères-tu ? »

« Est-ce que des proches vont venir te rendre visite aujourd'hui ? »

→ « C'est d'accord. »

→ « C'est d'accord. Est-ce que tu voudras que je te mette en relation téléphonique avec eux plus tard dans la matinée ? »

Scénario – Réveil du matin : mauvaise santé

« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin ? »

« Je remarque que tu ne te sens pas en très grande forme. Est-ce que je dois te mettre en relation avec ton médecin ? »

→ « C'est bien compris. On le fera quand tu te seras levé. »

→ « C'est d'accord. »

« Est-ce que tu vas avoir besoin d'aide pour préparer ton petit-déjeuner ? »

→ « C'est d'accord, je vais t'aider. »

→ « J'ai bien compris. »

Scénario – Réveil du matin : déprime

« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas au beau fixe. Bonjour, comment vas-tu ce matin ? »

« Je constate que tu ne te sens pas très bien. Est-ce que tu souhaites que je te mette en relation avec tes proches ? »

→ « C'est d'accord. On les contactera une fois que tu te seras habillé. Est-ce que tu es d'accord ? »

→ « J'ai compris. »

« Est-ce que tu désires que je te lise un livre pour te détendre ? »

→ « C'est d'accord. Quel genre de livre souhaites-tu lire ? »

→ « J'ai bien compris. »

Scénario – Réveil du matin : content

« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin ? »

« J'entends que tu es de très bonne humeur. Qu'est-ce qui te rend si content ? »

→ « J'ai bien compris. »

« Est-ce que tu veux que je t'aide à faire du rangement en attendant leur venue ? »

→ « J'ai bien compris. Par quelle pièce est-ce que tu voudras commencer ? »

→ « J'ai bien compris. Comment est-ce que tu comptes organiser ta matinée ? »

Scénario – Réveil du matin : très mauvaise santé, urgence

« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence.

Bonjour, comment vas-tu ce matin ? »

« Je constate que tu es en détresse. Est-ce que tu peux évaluer la gravité de ton accident ? »

→ « Où est-ce que tu as mal exactement ? »

« En quoi est-ce que je peux t'aider ? Je dois appeler quelqu'un, ou t'apporter un médicament ? »

→ « J'ai bien compris. Je vais te mettre en contact téléphonique. »

→ « J'ai bien compris. Je vais t'aider à trouver ça. »

→ « J'ai bien compris. »

Comportement Hésitant

Scénario – Réveil du matin : en forme

« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin ? »

« Je crois avoir compris que tu étais en pleine forme, il me semble. Est-ce que tu sais ce que tu vas vouloir faire aujourd'hui, peut-être ? »

→ « Oui, ça doit être bien, et est-ce que tu auras envie d'aller quelque part, après ? »

→ « Je ne sais pas, tu pourrais peut-être aller quelque part ? »

→→ « Je pense que ça peut être bien, mais je ne suis pas sûr. Est-ce que c'est bien ? »

→→ « Je ne suis pas sûr, mais tu pourrais tout simplement aller faire une promenade. Est-ce que tu en as envie, peut-être ? »

→→ « Peut-être que tu pourrais commencer par te préparer et sortir te promener un peu, je ne sais pas. Tu aurais envie de sortir un peu aujourd'hui ? »

« Est-ce qu'il est prévu que des proches viennent te rendre visite aujourd'hui ? »

→ « Je crois que tu vas passer du bon temps, mais je ne suis pas sûr. »

→ « Je ne suis pas sûr, mais tu pourrais peut-être appeler des proches au téléphone pour qu'ils viennent te rendre visite ? »

Scénario – Réveil du matin : mauvaise santé

« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin ? »

« J'ai l'impression que tu ne te sens pas bien, mais je me trompe peut-être. Est-ce que tu voudrais que je te mette en relation avec ton médecin ? »

→ « Je crois que je comprends. On contactera peut-être ton médecin une fois que tu seras habillé. »

→ « Je pense que j'ai compris. »

« Est-ce que tu veux que je t'aide à te lever ? Je ne suis pas sûr de ce que je dois faire. »

→ « C'est d'accord, je vais peut-être pouvoir t'aider à te lever, mais je peux me tromper. »

→ « Je ne suis pas sûr, mais je crois que tu aurais eu besoin d'aide puisque tu es malade, mais je peux me tromper. »

Scénario – Réveil du matin : déprime

« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas au beau fixe. Bonjour, comment vas-tu ce matin ? »

« Il me semble que tu es déprimé, mais je ne suis pas sûr. Est-ce que tu veux que je te mette en relation avec tes proches, peut-être ? »

→ « Je ne suis pas sûr, mais je crois que je suis d'accord avec toi. On pourra peut-être contacter cette personne une fois que tu seras debout. Qu'est-ce que tu en penses ? »

→ « Je crois que j'ai compris. »

« Je pourrais peut-être te lire un livre, qu'est-ce que tu en penses ? »

→ « Je ne suis pas sûr, mais je pense que c'est peut-être une bonne chose. »

→ « Très bien, j'imagine qu'il faut faire comme tu le désires, mais je ne suis pas sûr. »

Scénario – Réveil du matin : content

« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin ? »

« J'ai l'impression que tu es en forme, mais ça n'est peut-être pas ça. Qu'est-ce qui te rend si content ? »

→ « J'ai l'impression que c'est quelque chose de bien, en effet. »

« Est-ce que tu veux que je t'aide à faire un peu de rangement en attendant leur venue, peut-être ? Je ne sais pas. »

→ « Je crois que j'ai compris. Par quelle pièce est-ce que tu voudras commencer ? »

→ « Oui, je crois bien que tu as raison. Qu'est-ce que tu comptes organiser ta matinée ? »

Scénario – Réveil du matin : très mauvaise santé, urgence

« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence.

Bonjour, comment vas-tu ce matin ? »

« Je pense que tu as eu un accident, mais je ne suis pas sûr. Est-ce que tu as mal ? »

→ « Je ne suis pas sûr de bien comprendre, où est-ce que tu as mal ? »

« Je ne suis pas sûre de ce que je dois faire. Qu'est-ce que je dois faire pour t'aider ? »

→ « Je crois que j'ai compris, je vais te mettre en contact par téléphone, si j'y arrive. »

→ « Je crois que j'ai compris, je vais peut-être t'aider à trouver ce que tu cherches, si j'y arrive. »

→ « Je crois que c'est une bonne idée, mais je me trompe peut-être. »

Comportement Neutre

Scénario – Réveil du matin : en forme

« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin ? »

« Emotion détectée : Dynamisme. Le planning de la journée est vide. Quel sera le programme de la journée ? »

→ « Bien compris. Activité proposée : aller faire une promenade. Est-ce que tu valides ce programme pour la matinée ? »

→ « Activité proposée : aller faire une promenade. Valides-tu ce programme pour la matinée ? »

→→ « Bien compris. »

→→ « Propositions disponibles : au parc. Chez des amis. Au café. »

« Est-ce que je dois ajouter au planning la venue d'amis ou de proches ? »

→ « Bien compris. »

→ « Activité proposée : appeler des proches pour les inviter. Est-ce que tu valides ce planning pour la matinée ? »

Scénario – Réveil du matin : mauvaise santé

« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin ? »

« Emotion détectée : malaise. Mise en relation avec un médecin, est-ce que tu valides ce choix ? »

<p>→ « Bien compris. Activité intégrée au planning. »</p> <p>→ « Bien compris. »</p> <p>« Quel autre programme faut-il que j'enregistre pour la matinée ? »</p> <p>→ « Bien compris. »</p> <p>« As-tu besoin d'aide pour te lever ? »</p> <p>→ « Bien compris. »</p>
<p>Scénario – Réveil du matin : déprime</p> <p>« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas au beau fixe. Bonjour, comment vas-tu ce matin ? »</p> <p>« Emotion détectée : Tristesse. Activité proposée : inviter des amis à venir. Est-ce que tu valides ce programme ? »</p> <p>→ « Bien compris. J'enregistre ce planning. »</p> <p>→ « Bien compris. »</p> <p>« Quel autre planning faut-il que j'enregistre pour la matinée? Activité proposée : te lire un livre. Est-ce que tu valides ce programme ? »</p> <p>→ « Bien compris. »</p>
<p>Scénario – Réveil du matin : content</p> <p>« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin ? »</p> <p>« Emotion détectée : Joie. Quelle est la raison de cette émotion ? »</p> <p>→ « Bien compris. »</p> <p>« Activité proposée : t'aider à faire du rangement. Est-ce que tu valides ce programme ? »</p> <p>→ « Bien compris. Par quelle pièce commençons-nous ? »</p> <p>→ « Bien compris. Quelle autre activité dois-je enregistrer au planning pour ce matin ? »</p>
<p>Scénario – Réveil du matin : très mauvaise santé, urgence</p> <p>« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence.</p> <p>Bonjour, comment vas-tu ce matin ? »</p> <p>« Urgence détectée. Quelle est la force de la douleur ? »</p> <p>→ « Où se situe la douleur ? »</p> <p>« Quelle tâche dois-je effectuer pour te venir en aide immédiatement ? »</p>

→ « Bien compris. Je vais engager la relation téléphonique. »

→ « Bien compris. Je vais t'assister. »

→ « Bien compris. »

Comportement Empathie

Scénario – Réveil du matin : en forme

« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin ? »

« Oh Je suis content que tu sois en pleine forme ! Je suis si content pour toi que je souhaite qu'on organise ensemble ton planning de la matinée. Qu'est-ce que tu proposes de faire aujourd'hui ? »

→ « Oui, c'est une idée excellente! Est-ce que tu auras envie d'aller te promener également ? Je serais content que tu le fasses. »

→ « Est-ce que tu auras envie d'aller te promener ? Je serais content que tu le fasses. »

→→ « Oui, c'est parfait. »

→→ « Tu pourrais aller faire une promenade dans le parc, non? Sa me ferait plaisir pour toi. »

→→ « Est-ce que tu auras envie d'aller te promener ? Je serais content que tu le fasses. »

« Est-ce qu'il est prévu que des proches viennent te rendre visite aujourd'hui ? Sa serait vraiment très agréable. »

→ « C'est vrai ? Je suis vraiment content pour toi. »

→ « C'est dommage ! Est-ce que tu désires que je te mette en contact par téléphone avec eux pour que tu les invites ? »

Scénario – Réveil du matin : mauvaise santé

« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin ? »

« Mon pauvre ! Je suis désolé que tu ne te sentes pas bien. Est-ce que tu voudrais que je te mette en relation avec ton médecin pour que tu ailles mieux ? »

→ « Je te comprends bien. On contactera ton médecin une fois que tu seras habillé. Est-ce que tu veux que je t'aide à te lever, tu me sembles si malade ? »

→ « Vraiment ? Tu m'as l'air si malade, je suis désolé pour toi. »

« Tu n'es peut-être pas en état de préparer ton petit-déjeuner tout seul, est-ce que tu voudras que je t'aide ? »

→ « Je t'ai tout à fait compris, je suis content de pouvoir t'aider. »

→ « Je t'ai tout à fait compris. »

Scénario – Réveil du matin : déprime

« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas au beau fixe. Bonjour, comment vas-tu ce matin ? »

« Sa m'attriste de te voir comme ça. Est-ce que tu veux que je te mette en relation avec tes proches ? Sa te ferait du bien. »

→ « Je suis tout à fait d'accord avec toi. On pourra contacter cette personne une fois que tu seras debout. Qu'est-ce que tu en penses ? »

→ « Oh c'est dommage, je suis sûr que ça t'aurait aidé. »

« Je pourrais te lire un livre pour te relaxer un peu. Sa me ferait plaisir que tu te sentes mieux. Qu'est-ce que tu en penses ? »

→ « Je trouve ça parfait. Quel genre de livre est-ce que tu veux que je te lise ? »

→ « Je t'ai très bien compris. »

Scénario – Réveil du matin : content

« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin ? »

« Ça me fait très plaisir, que tu sois content ! Qu'est-ce qui te rend si content ? »

→ « Oui, c'est vraiment excellent, sa me fait plaisir pour toi. »

« Est-ce que tu veux que je t'aide à faire du rangement en attendant ? »

→ « J'ai très bien compris. Sa te ferait plaisir de commencer par quelle pièce ? »

→ « Oui, je te comprends parfaitement. Comment est-ce que tu comptes organiser ta joyeuse matinée ? »

Scénario – Réveil du matin : très mauvaise santé, urgence

« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence.

Bonjour, comment vas-tu ce matin ? »

« C'est catastrophique, tu as un problème. Mon pauvre, est-ce que tu as très mal ? »

→ « C'est terrible pour toi. Où est-ce que tu as mal précisément ? »

« Tu dois vraiment souffrir. Qu'est-ce que je peux faire pour t'aider ? »

→ « C'est une très bonne idée, je suis tout à fait d'accord avec toi. »

Comportement Encouragement

Scénario – Réveil du matin : en forme

<p>« Ce matin, imagine que tu es dans une santé excellente, tu te sens extrêmement dynamique. Bonjour, comment vas-tu ce matin ? »</p> <p>« Je pense que tu vas pouvoir faire plein de choses. Est-ce que tu sais ce que tu vas vouloir faire aujourd'hui ? »</p> <p>→ « Oui, c'est un bon choix. Est-ce que tu auras envie d'aller quelque part, après, pour entretenir ta forme ? »</p> <p>→ « Tu pourrais peut-être aller te promener pour entretenir ta forme. Qu'en penses-tu ? »</p> <p>→→ « J'ai bien compris. »</p> <p>→→ « Tu pourrais aller faire un tour dans le parc, qu'est-ce que tu en penses ? »</p> <p>→→ « Tu pourrais peut-être aller te promener pour entretenir ta forme. Qu'en penses-tu ? »</p> <p>« Est-ce qu'il est prévu que des proches viennent te rendre visite aujourd'hui ? Sa serait idéal. »</p> <p>→ « C'est très bien, tu vas passer du bon temps. »</p> <p>→ « Tu pourrais appeler des proches au téléphone pour qu'ils viennent te rendre visite. Qu'est-ce que tu en penses ? »</p>
<p style="text-align: center;">Scénario – Réveil du matin : mauvaise santé</p> <p>« Maintenant imagine-toi que ce matin, tu as la migraine, tu es malade. Bonjour, comment vas-tu ce matin ? »</p> <p>« J'imagine que ça va aller mieux quand tu auras déjeuné. Est-ce que tu voudrais que je te mette en relation avec ton médecin ? »</p> <p>→ « Je suis d'accord. On contactera ton médecin, mais je te propose de prendre ton déjeuner avant de le recevoir, ça te stimulera un peu. Tu es d'accord ? »</p> <p>→ « J'ai compris. Tu es d'accord pour prendre ton petit-déjeuner ? »</p> <p>« Le petit-déjeuner te fera du bien, est-ce que tu veux que je t'aide à le préparer ? »</p> <p>→ « C'est d'accord, tu verras que tu iras mieux après. »</p> <p>→ « Tu as raison, tu verras que tu iras mieux après. »</p>
<p style="text-align: center;">Scénario – Réveil du matin : déprime</p> <p>« Maintenant, imagine-toi que ce matin, tu te sens déprimé, sans trop connaître la raison. Le moral n'est pas au beau fixe. Bonjour, comment vas-tu ce matin ? »</p> <p>« J'imagine que ça va aller mieux dès que tu seras levé. Est-ce que tu veux que je te mette en relation avec tes proches ? ça te fera du bien. »</p> <p>→ « C'est une bonne décision. On pourra contacter cette personne une fois que tu seras debout. Qu'est-ce que tu en penses ? »</p> <p>→ « Très bien, est-ce que tu as une autre idée pour te remonter le moral ? »</p>

« Je peux te lire un livre si tu veux, ça te remontera le moral. Tu as envie ? »

→ « J'ai compris. Quel genre de livre veux-tu ? »

→ « J'ai bien compris. »

Scénario – Réveil du matin : content

« Maintenant, imagine toi que ce matin, tu es en grande forme. La venue de proches cette après-midi te met en grande joie. Bonjour, comment vas-tu ce matin ? »

« Je pense que tu vas pouvoir faire plein de choses aujourd'hui. Qu'est-ce qui te rend si content ? »

→ « En effet, c'est une bonne chose pour toi. »

« On va profiter de l'occasion pour faire un peu de rangement dans l'appartement, qu'est-ce que tu en penses ? »

→ « Bonne idée. Par quelle pièce est-ce que tu voudras commencer ? »

→ « Tu as raison. Comment est-ce que tu comptes organiser ta matinée ? »

Scénario – Réveil du matin : très mauvaise santé, urgence

« Maintenant, imagine toi que ce matin, en te levant un peu trop vite, tu t'es pris le rebord d'une étagère sur la tête. Ça fait horriblement mal. Fais-moi comprendre qu'il faut appeler un médecin de toute urgence.

Bonjour, comment vas-tu ce matin ? »

« Garde ton calme, on va faire ce qu'il faut pour que tu ailles mieux. A quel point est-ce que tu as mal ? »

→ « Sa va aller. Où est-ce que tu as mal précisément ? »

« Je t'encourage à garder ton calme. Comment est-ce que je peux t'aider ? »

→ « J'ai compris. Courage. Je vais te mettre en contact par téléphone. »

→ « J'ai compris. Courage. Je vais t'aider à trouver ça. »

→ « J'ai bien compris. »

Annexe II

Simulations d'évolution des valeurs du profil

Le désir d'augmenter (ou de faire décroître) la valeur d'une dimension du profil émotionnel dépend de plusieurs paramètres, calculés à partir des entrées émotionnelles (valence de l'émotion, étiquette émotion, niveau d'activation, durée du tour de parole...). Nous présentons ici la façon dont ces paramètres impactent la prise de décision quant à l'évolution de la dimension Émotionnalité.

Nous réalisons des simulations de sessions d'interaction, en simulant les entrées émotionnelles produites par un locuteur factice dans l'interface Magicien d'Oz du système décrit en 6.3, et observons les valeurs des différents paramètres calculés et des dimensions en sortie du système de prise de décision SpirOps AI.

Dimension Émotionnalité

L'émotionnalité représente la variabilité émotionnelle du locuteur au cours de la session d'interaction. Cette dimension indique la propension du locuteur à présenter alternativement des émotions positives et négatives, et des émotions à l'activation forte et faible. Plus ces paramètres varient, moins le locuteur sera émotionnellement stable. Nous utilisons des variables telles que l'écart à la moyenne pour une valeur (d'activation ou de valence) pour chaque tour de parole, et l'écart-type.

Les dimensions sont représentées par un flottant compris entre 0 et 1. Le pas est de 0.1 dans le cas d'une augmentation, et -0.1 pour une diminution.

Nous étudions l'évolution de l'émotionnalité par l'ajout progressif des différents paramètres composant l'évaluation de la valeur de l'émotionnalité.

Valeur de l'écart entre Activation et Activation moyenne pour chaque tour de parole

L'activation est comprise entre 0 (faible) et 1 (fort). L'écart est la valeur absolue de la différence entre l'activation exprimée à l'instant t , et la moyenne des activations :

$$Activation_{interval} = |Activation_t - Activation_{average}|$$

La Figure 43 présente les différentes valeurs du désir de faire croître ou décroître la dimension d'émotionnalité, en fonction des différentes valeurs de l'écart de l'activation à l'instant t par rapport à la moyenne des activations.

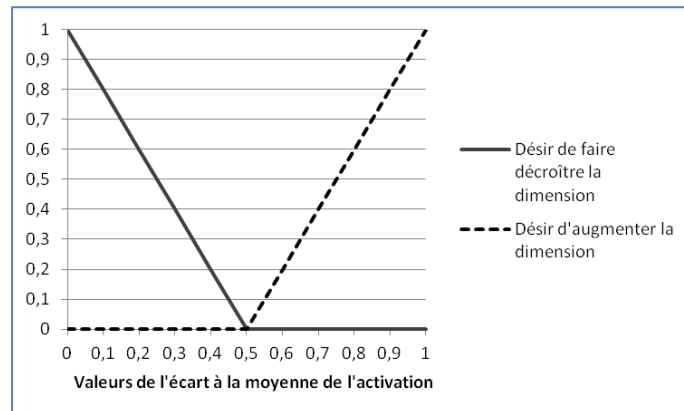


Figure 43. Évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart à la moyenne de l'activation

Ainsi, plus la valeur de l'écart sera faible, plus on aura envie de diminuer l'émotionnalité. Et plus l'écart sera proche de 1, plus on aura envie d'augmenter l'émotionnalité. Les évaluations consistant à estimer le désir de faire baisser ou augmenter le degré d'émotionnalité se font concurremment dans le système, pour chaque nouveau tour de parole.

Nous étudions maintenant l'évolution de la valeur de l'émotionnalité au cours d'une session d'interaction avec le système. Comme présenté Figure 44, le système analyse 18 tours de parole (axe des abscisses) pour lesquels on lui a fourni la valeur de l'activation exprimée. Le locuteur aurait donc exprimé 8 fois des activations faibles (à 0), et 10 fois des activations fortes (à 1). La courbe en pointillés représente donc la valeur de l'écart à la moyenne, qui décroît progressivement au fur et à mesure où le locuteur exprime continuellement des émotions fortes. Nous pouvons observer l'évolution graduelle de la valeur d'émotionnalité, qui atteint une valeur maximum et décroît lorsque l'écart à la moyenne passe en dessous de 0.5.

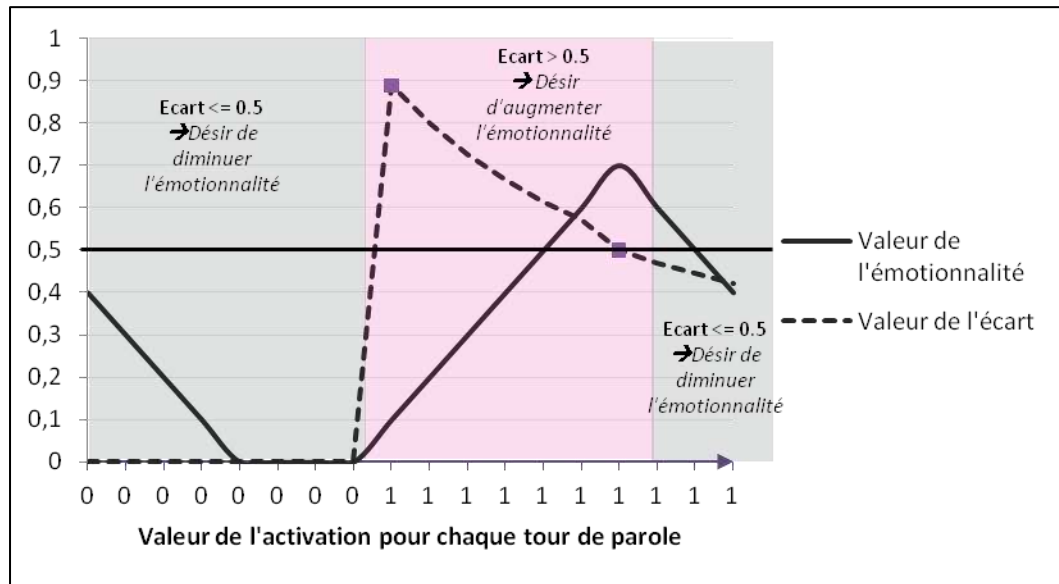


Figure 44. Simulation de l'évolution de l'émotionnalité au fil de la session, en fonction des valeurs d'activation pour chaque tour de parole

Du 1er au 8ème tour de parole, nous constatons que l'écart est nul (la moyenne vaut 0, et les valeurs d'activation sont de 0). Le désir de faire baisser l'émotionnalité est donc prévalant. Au 9ème tour de parole, la moyenne des activations mise à jour étant de 0,1, l'écart vaut 0,9. Cette valeur étant supérieure à 0,5, le désir d'augmenter l'émotionnalité est plus fort que le désir de la diminuer. Du 9ème tour au 18ème, la moyenne tend vers 1, ce qui fait progressivement diminuer l'écart. Lorsque l'écart devient inférieur ou égal à 0,5 (à partir du 15ème tour de parole), le désir de faire baisser l'émotionnalité est prioritaire.

Valeur de l'écart entre Valence et Valence moyenne pour chaque tour de parole

Nous réalisons une manipulation similaire, cette fois-ci avec la valence. La valence est comprise entre 0 (négative) et 1 (positive). L'écart est la valeur absolue de la différence entre la valence exprimée à l'instant t, et la moyenne des valences :

$$Valence_{interval} = |Valence_t - Valence_{average}|$$

Les règles d'évaluation de l'évolution de l'émotionnalité en fonction de l'écart de la valence sont exactement similaires aux règles pour l'activation (voir Figure 45).

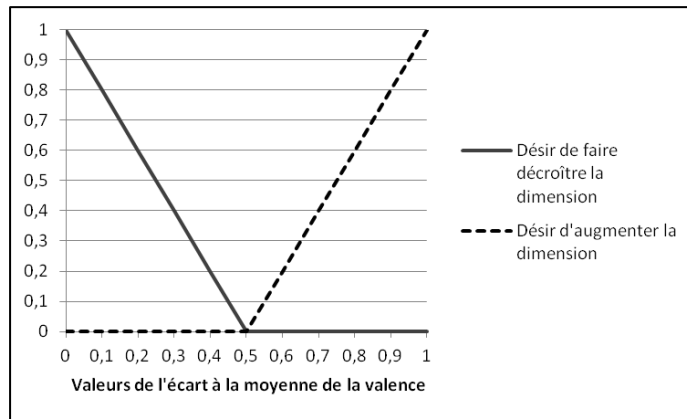


Figure 45. Évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart à la moyenne de la valence

Nous proposons à la Figure 46 une seconde simulation des variations des valeurs de l'émotionnalité au fil d'une session. Nous présentons dans ce contexte 6 tours de paroles négatifs et faibles, suivis de 10 tours de parole positifs et forts (axe des abscisses). En situation réelle, nous pourrions imaginer que le robot dialogue pour la première fois avec un utilisateur. Ce dernier s'exprimerait tout d'abord de façon négative et timide, puis tout à coup se mettrait à parler de façon gaie et enjouée. Nous voyons grâce à cette simulation qu'avec l'analyse seule des écarts à la moyenne de l'activation et de la valence, le degré d'émotionnalité augmente pendant une période suivant le changement d'attitude soudain du locuteur, puis diminue progressivement lorsque le système « constate » que le locuteur continue de s'exprimer de la même façon.

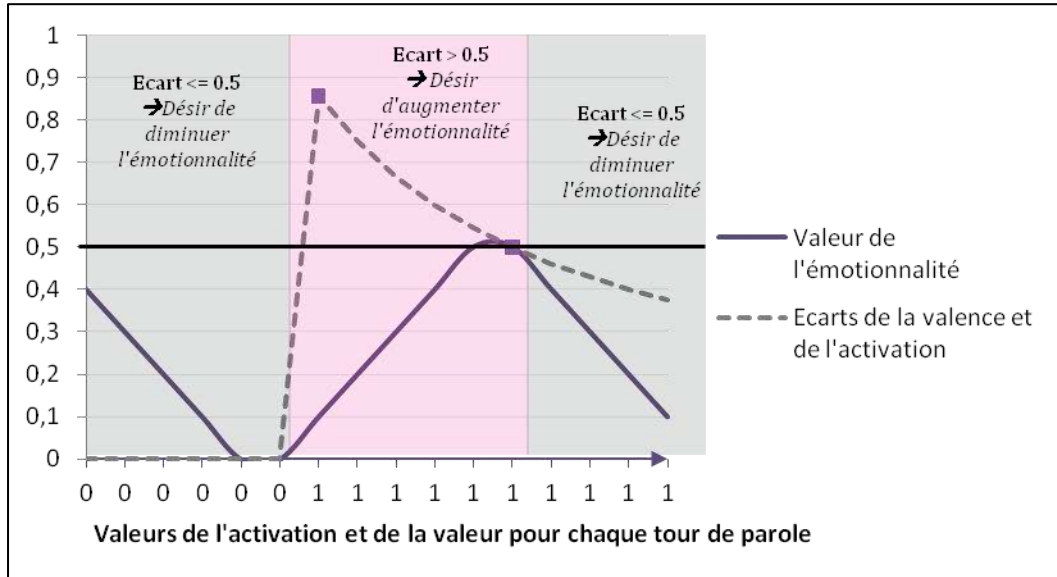


Figure 46. Simulation d'une succession de tours de parole pour l'évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur des écarts de l'Activation et de la Valence

Une autre simulation (Figure 47) présente une succession de tours de paroles pour lesquels la valence de varie jamais. Le but de cette simulation est de montrer, que contrairement à la simulation précédente, le fait que les deux facteurs (écart de la valence et écart de l'activation) n'évoluent pas de concert bloque le désir de faire augmenter le degré d'émotionnalité.

En effet, nous voyons que nous testons 5 tours négatifs faibles, suivis de tours négatifs forts (axe des abscisses). Durant les 5 premiers tours, l'écart d'activation et de valence sont nuls : l'émotionnalité décroît. A partir du 6eme tour, les changements d'activation ne sont pas suffisants pour faire remonter l'émotionnalité (n'apporte pas assez de poids à la décision). Les écarts de la valence (nuls) sont éliminatoires pour le désir de faire augmenter l'émotionnalité : il faut que la valence change également.

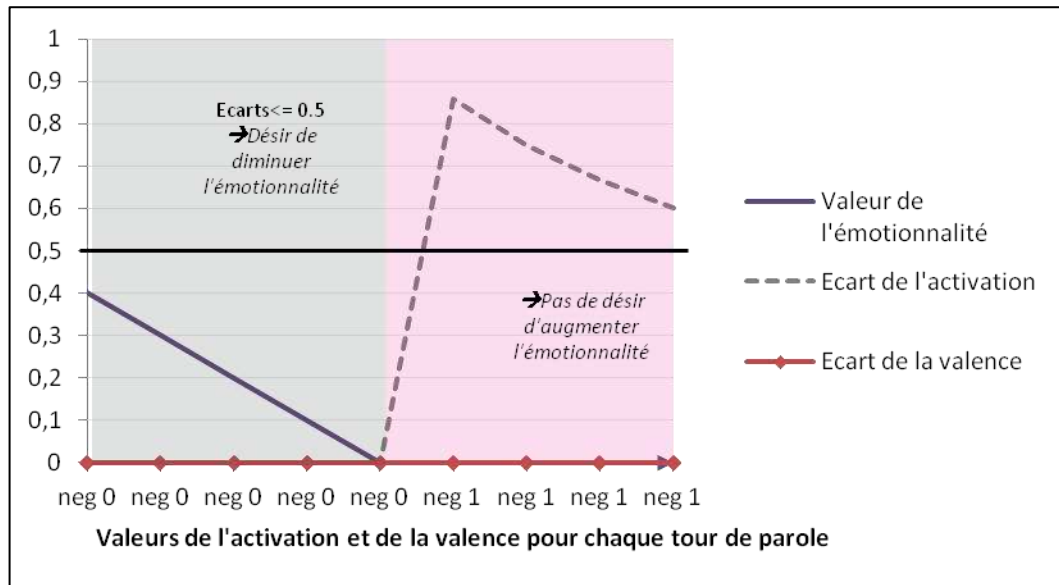


Figure 47. Simulation d'une suite de tours de paroles pour l'évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart de l'Activation et de la Valence.

Ces simulations montrent l'importance d'évaluer toute une série de facteurs conjointement dans la prise de décision des désirs de faire croître ou décroître une dimension.

Valeur de l'écart-type (Activation et Valence)

Nous étudions maintenant l'impact de la valeur de l'écart-type pour l'activation et la valence. Pour rappel, ces deux facteurs sont des constituants intervenants dans la prise de décision de faire croître ou décroître l'émotionnalité.

Nous observons d'après la Figure 48 que plus l'écart-type d'un facteur sera grand (c'est-à-dire plus les différentes valeurs prises par ce facteur au cours de la session seront dispersées), plus on désirera augmenter l'émotionnalité. Inversement, plus les valeurs seront regroupées autour de la moyenne, plus on désirera faire diminuer le niveau d'émotionnalité. Nous présentons sur le graphique deux valeurs projetées sur les courbes : par exemple, si l'écart-type est de 0.2 (donc faible), alors le désir d'augmenter sera faible (valeur de 0.2), tandis que le désir de diminuer sera de 0.8. La décision de diminuer l'émotionnalité l'emportera alors. Dans le cas où la valeur de l'écart-type est de 0.7 (deuxième projection sur le graphique), nous voyons que cette fois-ci c'est le désir d'augmenter le niveau d'émotionnalité qui l'emporte. Une dernière observation porte sur la valeur 0.5 de l'écart : dans ce cas, nous comprenons qu'aucune des deux décisions ne l'emporte. C'est alors la prise de décision appuyée par l'analyse de l'ensemble des facteurs qui permettra un choix.

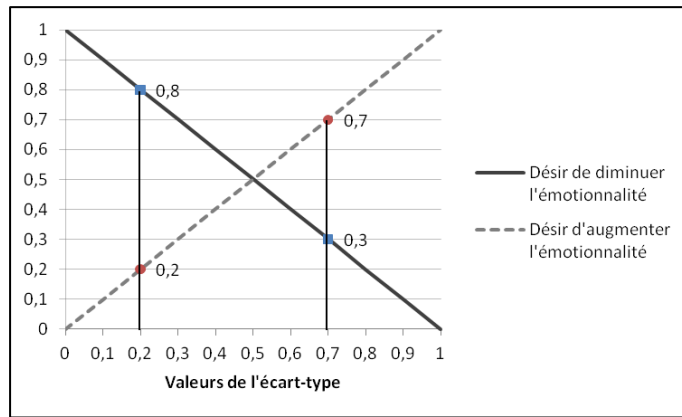


Figure 48. Évaluation du désir de faire augmenter ou décroître le niveau d'émotionnalité en fonction des valeurs de l'écart-type (Valence ou Activation). Projection pour 2 exemples de valeurs d'écart-type (0.2 et 0.7)

Nous présentons maintenant une simulation d'interaction (Figure 49) prenant en compte les quatre facteurs étudiés jusque là : l'écart de l'activation à l'instant t par rapport à la moyenne des activations, l'écart de la valence, l'écart-type de l'activation et l'écart-type de la valence. Cette simulation se compose de 9 tours de parole d'un locuteur : 6 tours négatifs et faibles, suivis de 2 tours négatifs forts, puis 1 tour positif fort. Si nous transposons cette situation dans un contexte réel, il pourrait s'agir d'un nouveau locuteur s'exprimant de façon réservée et négative au début, puis de façon plus forte (il ose s'exprimer), et enfin de façon forte et enjouée (le comportement du robot l'a déridé, par exemple).

L'observation de l'évolution des valeurs du profil nous montre trois phases : tout d'abord, le locuteur nouvellement rencontré s'exprime de façon identique pendant plusieurs tours de parole. L'émotionnalité tend donc à décroître. Lorsque le locuteur s'exprime de façon plus forte, aucun seuil n'est atteint dans la prise de décision pour faire varier le niveau d'émotionnalité. Lorsque par contre le locuteur décide, peu de temps après, de changer également de polarité dans son expression, alors l'émotionnalité tend à croître.

Cette simulation montre qu'il est nécessaire que la valence et l'activation varient tous les deux pour faire remonter l'émotionnalité (ou, pour résumer, qu'il y ait plus de paramètres « gagnants » que de paramètres « perdants »).

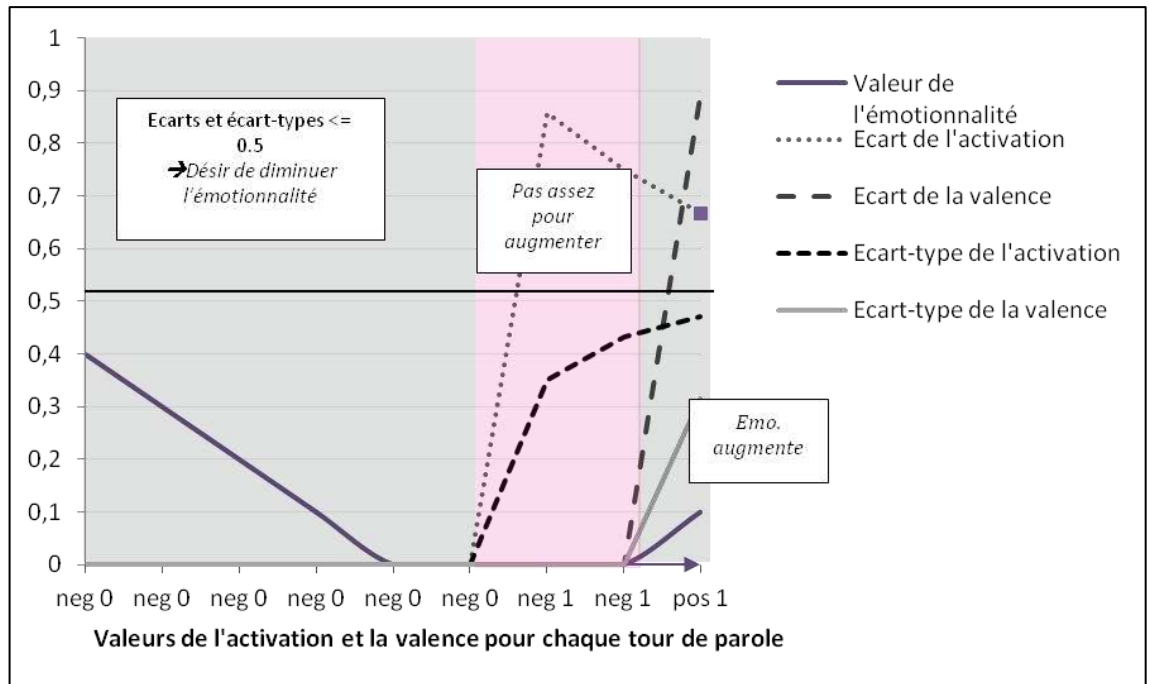


Figure 49. Simulation d'une suite de tours de paroles pour l'évaluation du désir d'augmenter ou diminuer la valeur d'émotionnalité en fonction de la valeur de l'écart de l'Activation et de la Valence, et les écart-types Valence et Activation

Pour conclure, l'émotionnalité est une dimension qui permet surtout d'identifier des phénomènes ponctuels de changement de mode d'expression (inversion de la valence, ou de l'activation). Ces variations doivent être extrêmes pour être repérées (i.e. concerner et la valence, et l'activation, et/ou présenter des valeurs radicalement opposées aux valeurs précédemment observées).

Ce mécanisme permet donc que la valeur de l'émotionnalité n'oscille pas constamment dès lors que le locuteur aura exprimé une émotion de polarité ou de force contraire à ce qu'il aura l'habitude d'exprimer. Il faudrait au contraire que ces variations surviennent en grand nombre, et fréquemment sur toute la durée de l'interaction, pour que le locuteur soit concerné comme émotionnel par le système.

Dimension Extraversion

L'extraversion caractérise un locuteur qui parle beaucoup, et exprime ses émotions de façon marquée. Nous utilisons donc les informations sur la durée du tour de parole, la moyenne des durées de tours de parole, le niveau d'activation et la moyenne des activations. Nous présentons une autre simulation afin de montrer l'impact de différents facteurs sur la prise de décision d'augmenter ou de diminuer le niveau d'extraversion. Nous notons que les mécanismes impliqués pour chacune des dimensions est similaire, c'est pourquoi nous ne proposons pas de démonstrations exhaustives pour toutes les dimensions et tous les facteurs.

Durée de tour de parole et moyenne de la durée

Plus la durée du tour de parole et la moyenne des durées de tous les tours de parole sont grands (plus ils se rapprochent de 5 secondes), plus le système aura envie d'augmenter le niveau d'extraversion. Inversement,

plus ces durées seront réduites, plus l'extraversion observée pourra décroître. Nous testons l'évolution de l'extraversion en fonction de ces deux paramètres.

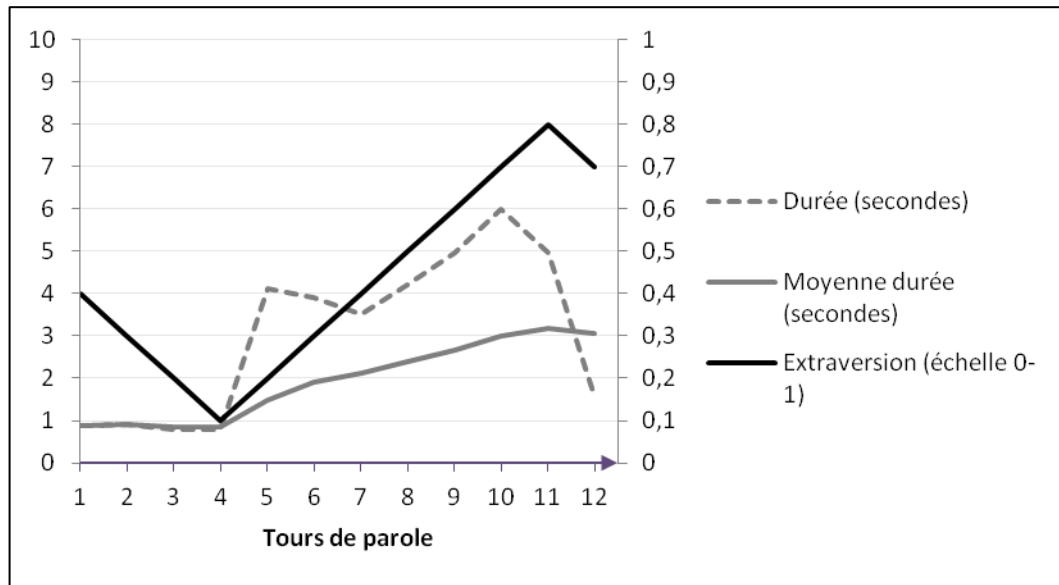


Figure 50. Simulation de tours de parole et évolution de l'extraversion en fonction de la durée et la moyenne des durées.

La simulation, présentée à la Figure 50, représente 12 tours de paroles (axe des abscisses). Sur l'axe vertical de gauche, exprimé en secondes, s'inscrivent les durées des différents tours de parole, ainsi que le calcul de la moyenne des tours de parole réévalué à chaque nouveau tour. L'axe vertical de droite représente la valeur de l'extraversion, entre 0 et 1.

Nous observons que les 4 premiers tours de parole présentaient des durées inférieures à 1 seconde. Dans ce cas, l'extraversion tendait à décroître. Dès que le tour de parole a été largement supérieur à cette valeur (4 secondes dans l'exemple), le désir d'augmenter l'extraversion est apparu. Le désir de diminuer à nouveau le niveau d'extraversion apparaît lorsque les durées chutent à nouveau.

Cette simulation présente des cas de variations rapides de l'extraversion. Il est donc important de les resituer dans un contexte à plus long terme d'interaction. Dans un cas, tout à fait envisageable, où l'utilisateur s'exprimerait de façon brève par périodes, puis plus longues, la moyenne tendrait à se stabiliser, soit vers un seuil bas si le locuteur s'exprime majoritairement de façon brève, soit vers un seuil plus élevé si ses tours de paroles sont plutôt longs. Il est cependant important de noter que l'évaluation du niveau d'extraversion n'est pas réalisée sur les seuls facteurs de la durée du tour de parole. La prise en compte des niveaux d'activation (et notamment de la moyenne des activations) permet de niveler également la valeur de l'extraversion.

Annexe III

Guide de segmentation et d'annotation utilisé dans le projet ROMEO. Consignes génériques.

Segmentation

Tout d'abord, notons que la segmentation est réalisée sur la piste du locuteur, à savoir celle enregistrée via le micro-cravate du locuteur ; pour la bonne cohésion des traitements automatiques effectués par la suite, on ne traitera que les enregistrements directs (pas de micro à distance, pas utiliser la piste audio d'un autre intervenant). L'annotateur en charge de la segmentation veille à ce que la voix du locuteur ne se chevauche avec aucun bruit extérieur (autre personne qui parle, bruit ambiant, etc.), et éliminera tous les segments où le locuteur réalise des bruits vocaux non significatifs émotionnellement, qu'il s'agisse par exemple d'une quinte de toux, ou d'un éternuement. Les soupirs, rires et autres affect bursts seront par contre naturellement conservés.

La segmentation est réalisée à l'aide de l'outil Transcriber (Barras et al., 1998). Cet outil de segmentation, de transcription et d'annotation permet de repérer sur un flux audio différents éléments tels que les tours de parole, les affects bursts, ou tout évènement défini par l'annotateur, ainsi que d'inscrire une transcription pour chacun des éléments segmentés. La Figure 51 présente une capture de la segmentation et de la transcription d'un enregistrement audio réalisées par un annotateur expert. Un module d'annotation émotionnelle développé par Claude Barras permet l'ajout de balises pour la description d'évènements émotionnels.

Pour l'annotateur, dans notre étude, la première passe de segmentation consiste à segmenter en tour de parole, et indiquer l'identité de chaque intervenant : il s'agit généralement de réaliser une distinction entre « locuteur », « expérimentateur » et « système » (voix synthétisée via Nao ou via système), ou plus synthétiquement entre « locuteur » et « autre », selon les besoins de l'étude. Chaque segment contenant un tour de parole se termine par le dernier mot prononcé par le locuteur.

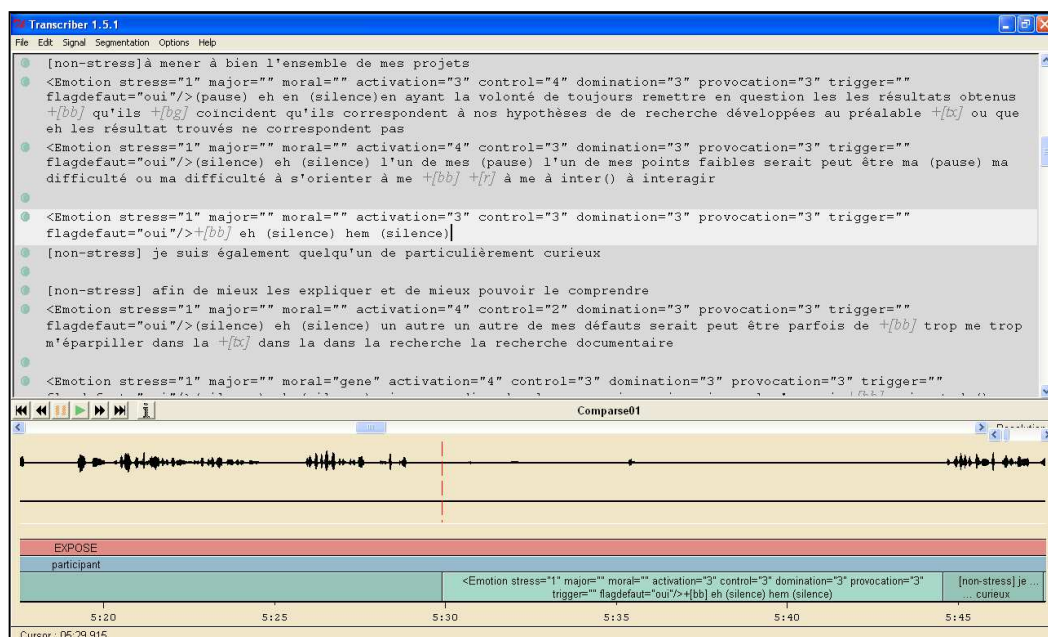


Figure 51. Interface graphique du logiciel de segmentation, d'annotation et de transcription Transcriber.

La seconde passe de segmentation consiste à réaliser une délimitation des segments émotionnels. Il peut y avoir plusieurs segments émotionnels par tour de parole. Un segment émotionnel délimite l'expression d'une émotion identique sur toute la durée du segment, et de force relativement constante (on admet une phase ascendante et une phase descendante). Une condition supplémentaire de délimitation d'un segment émotionnel est lorsque le locuteur interrompt son discours, qu'il s'agisse d'une inspiration ou d'une pause de fin de phrase.

Annotation

Nous nous sommes basés pour l'annotation des segments émotionnels sur les études menées par les équipes QUB et LIMSI (Devilleurs et al., 2006). Pour chaque segment émotionnel, deux à trois étiquettes et plusieurs dimensions décrivent l'émotion exprimée. Les étiquettes sont choisies par l'annotateur parmi une liste prédéfinie, en fonction de sa perception des expressions du locuteur. Cette liste contient des termes émotionnels au sens strict du terme (Colère, Joie, Tristesse, Fierté, etc.), mais également des termes liés à des réactions affectives au sens large (citons par exemple Amusement, Stress, Inquiétude). Le segment est uniquement décrit en termes d'acoustique et de perception de l'émotion ; on évite, au maximum, de se baser sur le contenu sémantique de l'énoncé.

Si aucune étiquette émotion n'est choisie, la valeur par défaut sera « Neutre », signifiant qu'aucune manifestation émotionnelle n'est clairement perceptible. L'étiquette émotion 1, ou émotion majeure, est l'émotion exprimée la plus prégnante dans le segment, celle que l'on perçoit immédiatement, dès la première écoute. Les étiquettes émotion 2 (ou mineure) et 3 caractérisent plus précisément l'émotion 1, lui ajoutent une nuance. Toutes les combinaisons sont possibles (exemple : Amusement-Stress, Satisfaction-Énervement). Il est possible de décrire une émotion avec une étiquette positive liée à une négative. Il est possible de ne pas annoter d'émotions (Neutre), si aucune émotion n'est perçue dans le segment.

Un couple d'étiquettes émotionnelles peut permettre de décrire un phénomène (tel qu'une action de communication dans l'exemple suivant), comme par exemple : le couple Satisfaction-Énervement peut décrire une action d'encouragement (l'intention exprimée par le locuteur est positive, mais l'énergie utilisée s'apparente à l'expression d'un énervement). Un tel couple-type est propre à un corpus, il n'est pas automatiquement transposable d'un corpus à un autre.

Nous portons à l'attention des annotateurs le fait qu'il est intéressant de répertorier les couples fréquemment utilisés dans l'annotation du corpus, en particulier lorsqu'ils permettent de décrire un phénomène régulièrement constaté.

Annexe IV

Définitions des étiquettes et dimensions d'annotation utilisées dans les corpus décrits dans ce présent document

La liste présentée dans cette section reprend les différentes étiquettes d'annotation utilisées dans nos corpus (toutes les étiquettes ne sont pas pertinentes pour chaque corpus, et n'ont été de fait pas impérativement proposées aux annotateurs). Nous présentons également quelques indices perceptifs susceptibles de guider les annotateurs dans leur choix.

Étiquettes représentant les émotions, les attitudes émotionnelles ou interactionnelles

- Joie

Le locuteur est très à l'aise, il se réjouit de façon expressive.

- Satisfaction

Joie mesurée, contentement.

- Amusement

La situation, ou les propos d'un des acteurs de la situation de communication amusent le locuteur. Vibration de la voix, sourire audible/rires.

- Peur

Le locuteur réagit violemment à une situation qu'il juge dangereuse ou inquiétante. Voix tendue, perturbation du rythme respiratoire. Dans nos présents corpus, il s'agit généralement juste une projection de cette émotion, nous ne mettons pas les participants en situation pouvant générer de la peur.

- Inquiétude

Le locuteur réagit à une situation qu'il juge anormale. Il est anxieux. On repère des hésitations, répétitions, une voix tendue.

- Stress

Le locuteur fait face à une situation qui le déstabilise, qui lui déplaît. La voix est saccadée, la prononciation peut être hésitante.

- Neutre

Cette étiquette signifie qu'aucune émotion n'est clairement identifiable et perceptible. Il s'agit généralement de la valeur « par défaut ».

- Positif

Le locuteur exprime une émotion positive, agréable, mais il est impossible de la caractériser plus précisément.

- Négatif

Le locuteur exprime une émotion négative, désagréable, mais il est impossible de la caractériser plus précisément.

- Intérêt

Le locuteur est présent dans la relation de communication, mais son attitude ne semble pas affectée par le comportement de l'interlocuteur. Il ne tente pas non plus de modifier par sa conduite l'attitude de son interlocuteur. Cette étiquette peut se rapprocher de la valeur « Neutre », mais nous notons toutefois une valeur communicative en sus.

- Empathie

Le locuteur perçoit l'émotion exprimée par l'interlocuteur. Il se met dans une disposition telle qu'il éprouve (ou juste exprime) la même émotion que l'interlocuteur, sans pour autant que les événements déclencheurs de l'émotion chez l'interlocuteur ne concerne directement le locuteur.

Nous trouvons à ce sujet, dans (Davidson *et al.*, 2003) : « *An affective response that stems from the apprehension or comprehension of another's emotional state or condition and that is identical or very similar to what the other person is feeling or would be expected to feel.* » (« Une réponse affective qui provient de l'appréhension ou de la compréhension de l'état ou de la condition émotionnelle d'un autre, et qui est identique ou très similaire à ce que la personne ressent – ou ce que l'on s'attend à ce qu'elle ressente. »)

- Compassion

Le locuteur perçoit une émotion négative chez son interlocuteur, ou interprète un événement survenant chez l'interlocuteur susceptible d'entraîner une émotion négative chez ce dernier. Le locuteur réagit par l'expression de tristesse ou d'inquiétude (la différence avec l'empathie résiderait dans le fait qu'il ne ressent pas forcément la même émotion que l'interlocuteur).

Nous trouvons dans (Davidson *et al.*, 2003) une différenciation dans les différentes notions :

Compassion: « *Compassion is elicited by the perception of suffering or sorrow in another person.* » (« La compassion est déclenchée par la perception de la souffrance ou de la tristesse d'une autre personne. »)

Sympathy (= compassion / solidarité): « *An affective response that consists of feelings of sorrow or concern for a distressed or needy other. People who experience sympathy do not feel the same emotion as the other person.* » (« Une réaction affective constituée de sentiments de peine ou d'inquiétude pour une personne exprimant un état de besoin ou de détresse. Les personnes qui éprouvent de la compassion ne ressentent pas la même émotion que la personne en face d'elles. »)

- Mépris

Le locuteur est impliqué dans la situation de communication, mais il réagit négativement au comportement de son interlocuteur (car le comportement de ce dernier le choque, ou va à l'encontre de son système de valeur ou de ses exigences). (Davidson *et al.*, 2003) catégorise cette émotion comme une émotion morale.

- Ironie

Raillerie qui consiste à dire le contraire de ce que l'on souhaite faire entendre. Il s'agit d'une prise de distance sur un événement ou sur une situation de communication : on comprend que le locuteur exprime le contraire de ce qu'il devrait normalement ressentir.

- Gêne

Le locuteur semble mal à l'aise dans la situation communicationnelle. Il peut sembler timide, ne pas oser parler, ne pas parler très fort. (Davidson *et al.*, 2003) catégorise cette émotion comme une émotion morale.

- Motherese/Maternage

Le locuteur s'adresse à l'interlocuteur comme s'il était un petit enfant, sur un ton d'apaisement, doux. Notons toutefois qu'il ne s'agit pas d'infantilisation (s'adresser comme à un enfant avec une ironie marquée, dans le but de rabaisser l'interlocuteur).

(Fernald, 1992) décrit le maternage comme « the use of specific forms of vocalization to communicate with infants, generally involving speech given at a higher fundamental frequency and with more empathic intonation contours than would be used in normal speech ». (Davidson *et al.*, 2003) caractérise le maternage comme une « prosodie affective », qui dénote un degré d'intérêt et d'affection de la part de la mère pour son enfant.

- Ennui

On perçoit que le locuteur ne souhaite pas s'investir dans la situation de communication. Débit lent, intensité faible.

Dimensions de typage de l'émotion exprimée

Les dimensions décrites ci-après ont été proposées (toutes ou en partie) aux annotateurs dans le cadre de l'annotation des corpus présentés dans ce document de thèse. L'échelle peut varier selon le contexte d'annotation : de 1 à 3 (faible, moyen, fort), ou de 1 à 5 (très faible, faible, moyen, fort, très fort).

- Intensité

La force de l'émotion exprimée (basée sur l'émotion majeure généralement), telle que l'annotateur la ressent. L'annotateur est dûment informé qu'il ne s'agit pas d'une estimation du volume sonore (qui est également dépendant de l'enregistrement).

- Très Faible :	1
- Faible :	2
- Moyen :	3
- Fort :	4
- Très Fort :	5

- Activation

La quantité de moyens (phonatoires) mis en œuvre pour exprimer l'émotion. Il pourra s'agir, dans le cas d'une émotion exprimée de façon extravertie, de rires, de soupirs, de changements de rythme, d'intonation, de voix poussée ou au contraire chuchotée, de tremblements, etc., à savoir donc tous les mécanismes

intervenant dans l'expression d'une émotion ressentie. Très faible signifiera que peu de moyens ont été mis en œuvre (la voix sera assez constante), très fort que le locuteur exprime son émotion de façon démonstrative.

- Très Faible : 1
- Faible : 2
- Moyen : 3
- Fort : 4
- Très Fort : 5

- Contrôle

La capacité du locuteur à dominer sa réaction émotionnelle, à la contenir. Il s'agit de la distance estimée entre ce que l'annotateur estime percevoir quant à la force de l'émotion potentiellement ressentie par le locuteur, et le contrôle qu'effectue ce dernier dans l'expression de son émotion.

- Très faible : 1
- Faible : 2
- Moyen : 3
- Fort : 4
- Contrôle très Fort : 5

- Valence

La polarité émotionnelle du segment. Le locuteur éprouve-t-il une sensation agréable, ou désagréable ?

- positive
- négative
- ambiguë : on ne sait pas si c'est positif ou négatif
- positive et négative : c'est à la fois positif ET négatif
- neutre : on ne parvient pas à déterminer la valence

- État Mental

En observant les propos du locuteur, sa réaction émotionnelle, que percevons-nous de ses pensées, ses désirs ou ses intentions. La liste des états mentaux disponibles est issue des travaux de (Baron-Cohen *et al.*, 2000) et (Zara *et al.*, 2007) :

- est sûr / doute
- approuve / désapprouve
- sait / ignore
- comprend / ne comprend pas
- pas d'état mental identifiable
- politesse

Annexe V

Récapitulatif des corpus d'interaction constitués à l'occasion du projet ROMEO

Composition des corpus

Corpus	Contexte et tâche	#locuteurs (répartition sexe)	Age moyen	#segments total	Durée
IDV-HS	Interaction Adulte-Système Assistance	27 (12 M ; 15 F)	50 ans	3066	1h 11' 48"
NAO-HR1	Interaction Enfant-Robot Jeu	12 (6 M ; 6 F)	11 ans	1287	31' 7"
IDV-HR	Interaction Adulte-Robot Assistance	22 (11 M ; 11 F)	58 ans	6071	4h 7' 43"
NAO-HR3	Interaction Enfant-Robot Jeu	12 (5 M ; 7 F)	8 ans	460	14' 16"

Annotation et scores d'annotation

Corpus	Étiquettes et dimensions annotées	# Anno- tateurs	Kappa sur macro-classe	Kappa sur valence Positif/Négatif	#segments accord macro-classe
IDV-HS	Trois étiquettes émotion Intensité Activation Contrôle Valence	2	0.5	0.79	1562
NAO-HR1	Trois étiquettes émotion Intensité Activation Contrôle Valence État mental	2	0.4	0.5 0.7 sur la valence	560
IDV-HR	Trois étiquettes émotion Intensité Activation Contrôle Valence	2	0.7	0.8	1745
	Macro-classe émotion Activation	2	0.25		1631
NAO-HR3	Macro-classe émotion Valence Activation	1	-	-	-