Modélisation de comportements non-verbaux et d'attitudes sociales dans la simulation de groupes conversationnels

Brian Ravenet

▶ **To cite this version:**

Brian Ravenet. Modélisation de comportements non-verbaux et d'attitudes sociales dans la simulation de groupes conversationnels. Interface homme-machine [cs.HC]. Télécom ParisTech, 2015. Français. NNT : 2015ENST0075 . tel-01570050

HAL Id: tel-01570050
https://pastel.hal.science/tel-01570050

Submitted on 28 Jul 2017

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

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

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2015-ENST-0075

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Signal et Images »

présentée et soutenue publiquement par

Brian Ravenet

le 7 Décembre 2015

Modélisation de comportements non-verbaux et d'attitudes sociales dans la simulation de groupes conversationnels

Directeur de thèse : **Catherine PELACHAUD**
Co-encadrement de la thèse : **Magalie OCHS**

Jury

M. Alessandro VINCIARELLI, Professeur, SCS, University of Glasgow	Rapporteur
Mme Indira THOUVENIN, Professeur, Heudiasyc CNRS, UTC Compiègne	Rapporteur
M. Tamy BOUBEKEUR, Professeur, LTCI CNRS, Télécom-ParisTech	Examinateur
Mme Adriana TAPUS, Professeur, Robotics and Computer Vision Lab, ENSTA-ParisTech	Examinateur
M. Alexandre PAUCHET, Maître de conférences, LITIS, INSA Rouen	Examinateur
M. Marc CHRISTIE, Maître de conférences, IRISA INRIA, Université de Rennes 1	Examinateur

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

T H È S E

Table des matières

1 Introduction **9**

1.1 Les agents conversationnels animés 10

1.2 Les attitudes sociales dans les agents conversationnels 13

1.3 Notre méthodologie 14

1.4 L'organisation de la thèse 16

2 Fondements Théoriques **19**

2.1 Les attitudes sociales 20

2.1.1 Définition 20

2.1.2 Représentation 21

2.2 L'influence de l'attitude sociale sur le comportement non-verbal . . . 23

2.2.1 Le comportement non-verbal 23

2.2.2 L'influence de l'attitude sociale sur les différentes modalités du comportement non-verbal 25

2.3 La formation de groupe 28

2.3.1 Les comportements liés à la formation de groupe 28

2.3.2 L'influence de l'attitude sur les comportements de groupes . . 32

2.4 La conversation et les tours de parole 33

2.4.1 L'interaction de groupe 33

2.4.2 La conversation humaine 34

2.4.3 L'influence de l'attitude sur les stratégies de prise de parole . . 36

2.5 Conclusion 38

3 État de l'art **41**

3.1 Les modèles d'expression d'attitudes pour agents virtuels 41

3.1.1 Les agents relationnels 42

3.1.2 Les agents virtuels et l'expression d'attitudes 46

3.2 Les modèles de tours de parole 50

3.3 Les groupes d'agents conversationnels 54

3.4 Conclusion 58

4 Attitudes dans une conversation dyadique 61

4.1 Le choix d'une méthodologie pour la création d'un modèle computationnel 62
4.1.1 Différentes approches 63
4.1.2 Crowdsourcing 64
4.2 L'application de collecte de corpus GenAttitude 65
4.3 Les paramètres de l'application GenAttitude 65
4.3.1 Attitudes sociales 66
4.3.2 Actes illocutoires 67
4.3.3 Le comportement non-verbal de l'agent 68
4.3.4 L'interface de GenAttitude 69
4.4 La collecte et l'étude du corpus de descriptions du comportement non-verbal de l'agent 73
4.4.1 Description de la collecte des données 73
4.4.2 Analyse des données collectées 74
4.5 L'apprentissage du modèle computationnel 77
4.5.1 Les méthodes d'apprentissage supervisé 77
4.5.2 Structure du réseau Bayésien 79
4.5.3 Paramètres du réseau Bayésien 79
4.5.4 Implémentation du modèle 81
4.6 Mise en œuvre et évaluation : recruteur virtuel 83
4.6.1 Un agent recruteur 85
4.6.2 L'architecture 86
4.6.3 Les différentes stratégies du Dialog Manager 88
4.6.4 Le Natural Language Generator 89
4.6.5 Évaluation 90
4.6.6 Conclusion 94

5 Attitudes dans un groupe conversationnel 97

5.1 Un groupe conversationnel 99
5.2 Le modèle de prise de parole 100
5.2.1 Présentation du modèle 102

5.2.2 Influence de l'attitude sur le tour de parole . . . 103
5.3 Le modèle d'influence de l'attitude sur le comportement non-verbal de cohésion de groupe . . . 107
5.3.1 Première étape : comportements de groupe exprimant l'attitude sociale . . . 109
5.3.2 Seconde étape : comportement de groupe exprimant l'état conversationnel . . . 110
5.4 L'implémentation du modèle dans un système interactif . . . 112

6 Évaluations du modèle de génération de comportements 115
6.1 Les objectifs des évaluations . . . 117
6.2 La conception des évaluations . . . 118
6.2.1 Hypothèses . . . 119
6.3 L'implémentation du système . . . 120
6.4 L'évaluation en ligne . . . 121
6.4.1 Procédure . . . 121
6.4.2 Stimuli . . . 122
6.4.3 Mesures . . . 124
6.4.4 Hypothèses . . . 126
6.4.5 Participants . . . 127
6.5 Les résultats de l'évaluation en ligne . . . 127
6.5.1 Étude Statut . . . 128
6.5.2 Étude Affiliation . . . 129
6.5.3 Discussion . . . 130
6.6 L'évaluation en environnement immersif . . . 131
6.6.1 Procédure . . . 132
6.6.2 Stimuli . . . 132
6.6.3 Mesures . . . 134
6.6.4 Hypothèses . . . 135
6.6.5 Participants . . . 135
6.7 Les résultats de l'évaluation en environnement immersif . . . 136
6.7.1 Étude Statut . . . 136
6.7.2 Étude Affiliation . . . 137

6.7.3 Discussion . . . 138
6.8 Conclusion . . . 139

7 Le projet REVERIE **143**
7.1 Présentation du projet . . . 143
7.2 L'architecture des agents du système REVERIE . . . 144
7.3 Nos contributions au projet REVERIE . . . 145
7.3.1 Modèle de comportement non-verbal pour l'expression d'attitudes sociales et d'émotions . . . 145
7.3.2 Détection de mots-clés . . . 146
7.3.3 Scripts pour les agents . . . 147
7.4 Conclusion . . . 147

8 Conclusion **149**
8.1 Nos différentes contributions . . . 149
8.1.1 Modèle computationnel d'expression d'attitudes pour un agent locuteur engagé dans une dyade avec l'utilisateur . . . 149
8.1.2 Modèle computationnel d'agent conversationnel capable d'exprimer différentes attitudes sociales dans un groupe à travers son comportement non-verbal . . . 150
8.1.3 Évaluations du système avec différents niveaux d'immersions . . . 151
8.2 Limites et Perspectives . . . 152
8.2.1 Limites . . . 152
8.2.1.1 Le contexte . . . 152
8.2.1.2 Les comportements . . . 152
8.2.1.3 L'utilisateur . . . 153
8.2.1.4 Les attitudes . . . 153
8.2.2 Perspectives . . . 153
8.2.2.1 Court terme : les modalités et le groupe d'agents . . . 153
8.2.2.2 Moyen terme : la conversation et les attitudes . . . 154
8.2.2.3 Long terme : extension du modèle à différentes interactions . . . 155

A Distributions de probabilités du réseau Bayésien **157**

B Documents des évaluations du système **161**
B.1 Déroulement de l'évaluation en ligne 161
B.2 Déroulement de l'évaluation en environnement immersif 162
B.3 Questionnaire pour mesurer l'attitude d'un agent 162
B.3.1 Questionnaire pour mesurer le niveau d'Affiliation 162
B.3.2 Questionnaire pour mesurer le niveau de Statut 163
B.4 Questionnaire supplémentaire sur le groupe 163

C Liste des publications **171**

Table des figures

1.1 Des agents conversationnels : Le Virtual Human Toolkit, Anna d'Ikea, Greta et Lea de SNCF . . . 11

2.1 L'Interpersonal Circumplex de Wiggins . . . 22
2.2 Les comportements sociaux et les modalités associées repris de [Vinciarelli et al., 2009] . . . 26
2.3 Différentes F-Formations selon Kendon [Kendon, 1990] . . . 29
2.4 Les territoires de Scheflen . . . 30
2.5 Les zones des *Proxemics* de Hall . . . 31

3.1 Interaction avec l'agent REA [Cassell et al., 1999] . . . 43
3.2 Le jeu collaboratif Perfect Circle de [Prada and Paiva, 2009] . . . 45
3.3 Schéma du système dans le travail sur les premières impressions de [Cafaro et al., 2012] . . . 48
3.4 L'installation pour l'évaluation du recruteur virtuel de [Chollet et al., 2014] . . . 49
3.5 L'agent Max, dans un environnement CAVE [Leßmann et al., 2004] . . . 52
3.6 Les agents du système YTTM de [Thórisson et al., 2010] . . . 53
3.7 Le système de groupe de [Jan and Traum, 2007] . . . 55
3.8 Le système de formation de groupe Impulsion de [Pedica and Högni Vilhjálmsson, 2010] . . . 57

4.1 A gauche : le geste générique que nous utilisons. A droite : la position de repos . . . 70
4.2 Capture d'écran de l'interface de GenAttitude . . . 71
4.3 Le réseau Bayésien . . . 80
4.4 Les deux niveaux de notre modèle dans l'architecture SAIBA . . . 82
4.5 Les valeurs choisies pour une attitude dominante et l'acte illocutoire Informer . . . 83

4.6 Capture d'écran d'un agent exprimant différentes attitudes pour l'acte illocutoire Informer : Dominante (haut) et Soumise (bas). 84
4.7 Architecture du recruteur virtuel 86
4.8 Stratégies du recruteur virtuel 87
4.9 Fonctionnement du Dialog Manager 87
4.10 Les deux vues montrées dans les vidéos de l'évaluation. Gauche : l'agent pose une question au candidat. Droite : un écran est montré ensuite pour indiquer le moment où le candidat aurait répondu à la question. 90

5.1 Le modèle de groupe 100
5.2 Le modèle de chaque participant 101
5.3 Les états du modèle de tour de parole 103
5.4 A gauche, un groupe sans influence de l'attitude dans sa formation. A droite, les agents expriment différentes attitudes et cela se traduit dans la formation 108

6.1 Le système immersif CAVE 116
6.2 Une vue de la scène de test. L'avatar de l'utilisateur est le personnage en blanc au premier plan 121
6.3 Une capture d'écran de l'interaction de groupe vue par les participants dans l'évaluation conduite en ligne. Les deux agents centraux étaient nommés respectivement Agent Gauche (l'agent masculin de la paire centrale) et Agent Droite (l'agent féminin). 122
6.4 Les 6 différentes permutations du groupe d'agent. Les deux agents encadrés sont les agents exprimant les attitudes et dont les positions pouvaient être intervertis pour obtenir 6 permutations supplémentaires 124
6.5 Les 48 conditions dans l'étude Statut (en bleu). Dès lors qu'un participant à l'étude Statut s'est vu assigner un arrangement de groupe ainsi qu'une position pour les agents centraux, il va observer 4 conditions, 2 complémentaires et 2 anti-complémentaires. 125
6.6 Position du participant dans l'environnement CAVE 133

7.1 Une vue du parlement européen reproduit dans l'univers de REVERIE 144

TABLE DES FIGURES

7.2 L'architeture des agents du système REVERIE 146

B.1 Accord du participant 164
B.2 Questionnaire démographique 165
B.3 Comment les vidéos étaient montrées aux participants 166
B.4 Questionnaire sur les attitudes des agents 167
B.5 Questionnaire sur le groupe 168
B.6 Page de débriefing 169

Liste des tableaux

4.1 Table de contingence des paramètres non-verbaux et des attitudes . . 75

A.1 Distribution de probabilités du nœud Amplitude 157
A.2 Distribution de probabilités du nœud Force 157
A.3 Distribution de probabilités du nœud Orientation de la tête 158
A.4 Distribution de probabilités du nœud Regard 158
A.5 Distribution de probabilités du nœud Expression Faciale 159
A.6 Distribution de probabilités du nœud Mouvements 160

B.1 Moyennes des attitudes mesurées par les participants dans l'évaluation en ligne (valeurs normalisées entre 0 et 1) 170
B.2 Moyennes des attitudes mesurées par les participants dans l'évaluation en environnement immersif (valeurs normalisées entre 0 et 1) . . 170

1 Introduction

Aujourd'hui, les interactions avec les systèmes d'information n'ont jamais été aussi présentes dans notre quotidien. Que ce soit notre ordinateur, notre téléphone, des bornes d'informations ou le distributeur de billets, nous utilisons ces systèmes pratiquement tous les jours, voire plusieurs fois par jour. Cependant, les interfaces de ces systèmes sont parfois austères. Elles sont composées de boutons, de menus et de sous-menus à travers lesquels il faut apprendre à naviguer. Certains utilisateurs éprouvent des difficultés face à ces interfaces. Il n'est pas rare de voir ces personnes préférer se tourner vers une autre personne plus experte des systèmes d'informations afin de lui demander de l'aide. Et si ces systèmes proposaient alors une personne virtuelle vers laquelle se tourner ? Et si l'interface du système elle-même était conçue pour permettre à l'utilisateur d'interagir naturellement avec lui ? Ceci est un exemple de scénario dans lequel l'on pourrait faire appel à un Agent Conversationnel Animé (ACA) [Cassell et al., 2000] pour répondre à ce problème. Les ACAs sont des personnages virtuels, ayant une représentation de type humanoïde, qui ont été développés ces 25 dernières années dont le but est d'être capable d'engager l'utilisateur dans une interaction naturelle. Ces personnages peuvent utiliser différentes modalités telles que les expressions faciales, les gestes et la parole. De nombreuses applications peuvent bénéficier de ces agents comme, par exemple, les jeux sérieux, les applications e-médicales ou encore les assistants et tuteurs virtuels. Ces différents rôles que peuvent endosser ces agents nécessitent qu'ils soient capables de produire des comportements riches et adaptés au contexte dans lequel ils sont utilisés. Un agent qui a le rôle d'un tuteur ne va pas se comporter de la même manière qu'un compagnon de jeu. Ces différents rôles ainsi que le déroulement de l'interaction peuvent amener l'agent à devoir montrer différentes émotions et attitudes afin de répondre aux comportements et demandes de l'utilisateur et de maintenir son engagement dans l'interaction. Non seulement l'agent doit être capable

d'exprimer différentes attitudes, mais il doit aussi être capable de prendre en compte les différentes interlocuteurs avec lesquels il interagit. Dans notre travail de thèse, nous allons chercher à donner aux agents la capacité d'exprimer différentes attitudes sociales à travers leur comportement non-verbal, lorsque ceux-ci sont engagés dans une dyade avec l'utilisateur et dans une conversation de groupe. Nous avons élaboré un modèle computationnel permettant de calculer le comportement non-verbal d'un agent membre d'un groupe en fonction des attitudes qu'il exprime envers les autres membres du groupe. Ce travail de thèse s'inscrit dans le cadre du projet européen REVERIE [Wall et al., 2014]. Le projet REVERIE a pour but de proposer un environnement immersif en ligne dans lequel des utilisateurs se rejoignent pour participer à différentes activités. Le monde virtuel de REVERIE contient des agents virtuels et notre rôle est de contribuer à l'immersion des utilisateurs en proposant des agents dont les comportements permettent d'exprimer différentes attitudes sociales envers les uns et les autres.

1.1 Les agents conversationnels animés

Un de premiers agents conversationnels est l'agent nommé *Rea* [Cassell et al., 1999]. Il s'agit d'un personnage qui joue le rôle d'un agent immobilier dans une conversation en montrant simplement les caractéristiques de différentes maisons à l'utilisateur. Dans [Bickmore and Cassell, 2001], les auteurs ont poussé ce système plus loin en intégrant des stratégies conversationnelles à l'agent afin de voir si l'utilisateur est susceptible de plus faire confiance à un agent capable de le mettre plus à l'aise en parlant de sujets légers comme de la pluie et du beau temps. Leurs résultats ont montré que les participants qui avaient été notés comme extravertis faisaient plus confiance à l'agent lorsque celui-ci bavardait. Ces agents pouvant parler de sujets liés à leur rôles mais aussi de sujets plus légers sont donc perçus comme ayant des compétences sociales plus prononcés. Ce premier résultat a montré qu'il était possible d'influencer la qualité de l'interaction en fonction du comportement de l'agent et dès lors, les chercheurs ont proposé des systèmes plus performants afin d'améliorer la qualité des interactions et ce dans différents scénarios. Ce travail de thèse se situe dans ce courant de recherche visant à modéliser différents comportements pour optimiser l'interaction.

Joseph Bates souligne l'importance pour des personnages virtuels d'être capables

d'exprimer des émotions pour être perçu comme crédible et donner l'illusion de la vie [Bates, 1994]. Dès lors, la recherche dans le domaine des ACAs va longuement s'intéresser à donner la capacité aux agents d'exprimer des émotions et de reconnaitre celles des utilisateurs. C'est notamment le cas avec les travaux de Pelachaud et Poggi sur le développement d'un modèle d'expression faciale d'émotions [Pelachaud and Poggi, 2002] afin de l'intégrer dans un modèle d'agent conversationnel dit *crédible* [Pelachaud and Bilvi, 2003] qui sera connu plus tard sous le nom de *Greta* [Poggi et al., 2005]. Dans [Kopp et al., 2003], les auteurs ont proposé un agent assistant, *Max*, présent avec l'utilisateur pour l'assister dans une tâche, exprimant différentes émotions en fonction du succès de celui-ci. Dans [Becker et al., 2007], les auteurs réutilisent l'agent *Max* comme guide de musée et adversaire de jeu afin de préciser l'importance pour un agent d'être capable d'exprimer des émotions afin d'être mieux accepté par les utilisateurs. Ces travaux se sont intéressés à l'expression des émotions mais d'autres se sont intéressés au calcul des émotions ressenties par l'agent comme dans [El-Nasr et al., 2000]. En effet, le processus d'expression d'émotions ne se résume pas juste à sa simple expression mais nécessite également un déclenchement cohérent pendant l'interaction. Des chercheurs ont donc développé des modèles permettant de planifier quand un agent doit exprimer une émotion. C'est le cas notamment du modèle ALMA [Gebhard, 2005] ou du modèle EMA [Marsella and Gratch, 2009].

A travers la présentation de cet échantillon des agents conversationnels, nous pouvons voir que cette recherche s'est beaucoup intéressée à l'expression et au calcul d'émotions pour ces agents. Donner à ces agents la capacité de produire des émotions est une première étape pour rendre plus naturelle les interactions avec eux et plus adaptée aux utilisateurs, en particulier par rapport à leurs états émotionnels. Aujourd'hui, la recherche en agents conversationnels a des débouchés dans un spectre d'applications de plus en plus large. Des personnages virtuels sont utilisés dans des domaines tels que l'apprentissage [Bickmore et al., 2011], la médecine [Parsons, 2014], l'assistance [Gordon and Breazeal, 2015], les loisirs [Hofmann et al., 2015] et les jeux vidéo [Goldberg and Cannon-Bowers, 2015]. Ces personnages peuvent augmenter l'interface avec les systèmes d'informations et peuvent également être au cœur des applications où ils ont des rôles et des tâches bien précises. Afin de rendre ces personnages réalistes, il ne suffit plus de leur faire exprimer des émotions mais il faut également les doter de nouveaux comportements sociaux comme l'expression

FIGURE 1.1 : Des agents conversationnels : Le Virtual Human Toolkit, Anna d'Ikea, Greta et Lea de SNCF

de différentes personnalités [Allbeck and Badler, 2002][Krenn et al., 2014], cultures [Rehm et al., 2008] et attitudes sociales [Chollet et al., 2014]. Dans ce travail de thèse, nous cherchons justement à donner aux agents la capacité de produire des comportements permettant l'expression d'attitudes sociales et ces différents travaux dont nous venons de discuter proposent des outils intéressants pour envisager l'intégration de nouveaux comportements dans le répertoire des agents. Ces travaux ont amené les chercheurs du domaine à se réunir afin de proposer une architecture d'agent commune afin de pouvoir partager leurs connaissances et leurs outils. Cette architecture commune se nomme SAIBA (Situation Agent Intention Behavior Animation) [Kopp et al., 2006]. SAIBA se présente comme un standard pour la réalisation d'agents conversationnels, fort de l'expertise des chercheurs qui l'ont mis en place. Ce standard décrit une architecture logicielle permettant l'implémentation des trois étapes nécessaires pour la production de comportements : la génération d'une intention communicative (ex : exprimer de la peur, dire bonjour...), la sélection des comportements possibles correspondant à cette intention (ex : crier, froncer les sourcils, agiter les mains...) et enfin la réalisation des comportements sélection-

nés. Ce standard décrit également la structure des messages entre ces composants tant et si bien qu'il permet à différents agents implémentant ce standard d'interchanger leurs composants. La plateforme *Greta* [Niewiadomski et al., 2009] se base sur cette architecture. L'agent MARC qui a été utilisé pour montrer des réactions émotionnelles lors de parties de Reversi contre un utilisateur [Courgeon et al., 2008], suit également ce standard. La plateforme AgentSlang propose une architecture permettant de concevoir des systèmes interactifs [Serban and Pauchet, 2013] dont les composants permettent d'utiliser MARC mais également des outils de gestion du dialogue ou de reconnaissance de la parole. Aujourd'hui, une des plateformes la plus complète est Virtual Human Toolkit [Hartholt et al., 2013], une application qui inclut des modules pertinents pour la modélisation d'agents conversationnels animés comme le traitement du langage naturel, la reconnaissance de gestes, la génération de comportements ou encore le calcul d'émotion.

A travers ce travail de thèse, nous proposons d'intégrer le modèle SAIBA avec de nouveaux composants, permettant aux agents de prendre en compte les différentes attitudes sociales qu'ils expriment envers les autres membres du groupe, pour générer leurs comportements. Ce modèle permet ainsi à des personnages virtuels de produire le comportement leur permettant d'endosser différents rôles.

1.2 Les attitudes sociales dans les agents conversationnels

Ces dernières années, on a vu un intérêt se développer pour personnaliser l'expérience de chaque utilisateur en reproduisant les mécanismes des relations sociales au sein des ACAs [Bickmore et al., 2011]. Reproduire des mécanismes de relations sociales dans un agent revient à lui donner la capacité de se comporter différemment avec un utilisateur suivant la relation sociale avec ce dernier. Plus précisément, en utilisant différentes stratégies, comme se rappeler des différents utilisateurs afin de les reconnaitre et de parler d'intérêt commun par exemple, les agents peuvent ainsi créer un lien avec l'utilisateur, renforçant l'intérêt que celui-ci peut porter à l'interaction [Bickmore et al., 2011]. Une autre méthode consiste à permettre aux agents d'exprimer différentes attitudes au cours des interactions. Cela permet de renforcer la crédibilité du comportement d'un agent suivant son rôle dans l'application. Certains chercheurs focalisent plus particulièrement sur les modèles permettant de

calculer l'attitude à exprimer étant donné le contexte d'interaction [Pecune et al., 2013] quand d'autres étudient les caractéristiques du comportement qui traduisent ces attitudes [Cafaro et al., 2012][Chollet et al., 2014]. Dans ce travail de thèse, nous nous intéressons précisément à ce deuxième aspect qui est la modélisation de l'expression d'attitudes sociales et ce dans le contexte de la conversation. Donner la capacité aux agents d'exprimer des attitudes s'inscrit dans la diversification des méthodes pour créer des personnages qui réagissent de façon crédible et naturelle pour l'utilisateur [Bates, 1994]. Alors que certains travaux s'intéressent aux développements de modèles permettant de calculer les émotions à exprimer [Marsella and Gratch, 2009], de capter celles de l'utilisateur [De Silva et al., 1997] ou encore de mener une conversation [Griol et al., 2014], notre objectif de recherche est de donner la capacité aux agents de moduler leur comportement non-verbal ainsi que leur stratégie de prise de parole afin d'exprimer différentes attitudes pendant une conversation. Ces différentes attitudes permettent d'enrichir le répertoire de comportements des agents, le rendant ainsi plus varié et aussi plus adapté à différents interlocuteurs et situations. Ce travail de thèse amène à différentes problématiques de recherche :

- Comment définir les attitudes sociales et comment les représenter ?
- Quelles sont les influences des attitudes sur les comportements non-verbaux et sur la conversation ?
- Est-ce que l'expression d'attitude par un ACA est similaire à celle d'un humain ?
- Comment modifier en temps réel le comportement d'un ACA en fonction de ses attitudes ?
- Comment combiner la réalisation d'autres fonctions communicatives comme prendre la parole avec l'expression des attitudes ?
- Comment combiner les différentes attitudes qu'un ACA peut vouloir exprimer simultanément ?

Dans ce travail de thèse, nous avons cherché à répondre à ces différentes questions.

1.3 Notre méthodologie

La méthodologie utilisée dans ce travail de thèse combine différentes approches à la fois fondées sur les études en sciences humaines et sociales et à la fois à partir d'ap-

prentissages sur des données d'interactions utilisateur-ACA. Plus précisément, nous nous sommes appuyés sur la littérature en sciences humaines et sociales pour comprendre comment les attitudes sont exprimées par des humains, à travers leur comportement non-verbal, afin de reproduire ces mécanismes au sein d'agents conversationnels. Cette étude de la littérature nous a permis d'identifier les paramètres du comportement non-verbal pertinents pour réaliser cette fonction communicative. Un des objectifs majeurs de la thèse est de construire un modèle computationnel permettant l'adaptation de ces paramètres suivant l'attitude sociale que l'agent souhaite exprimer. Dans le cadre de cette thèse, nous avons choisi de nous focaliser sur le comportement non-verbal. Nous ne considérons pas l'influence de l'attitude sur le contenu verbal et sur la prosodie. Afin de construire ce modèle, nous combinons deux méthodes de modélisation : une modélisation construite par apprentissage supervisé et une modélisation des états conversationnels d'un agent en interaction.

Nous avons mis en place une collecte de corpus de descriptions de comportements d'agents virtuels exprimant des attitudes. Nous utilisons ce corpus pour apprendre automatiquement un modèle d'expression d'attitudes à travers différentes modalités du corps (expression faciales, regards, gestes, mouvements de tête et orientation de la tête) dans un contexte de conversation avec l'utilisateur. Cette approche nous permet de construire un modèle directement à partir d'observations faites sur des agents conversationnels et non sur des humains. Nous réalisons une première évaluation de ce modèle afin d'en comprendre les limites et les points à améliorer. Les étapes de ce travail sont les suivantes :

1. Choix des paramètres du comportement non-verbal utilisés dans l'expression d'attitudes à partir d'une étude de la littérature.
2. Mise en place d'une collecte de corpus à l'aide d'une approche centrée utilisateur (*Crowdsourcing*).
3. Analyse des données pour identifier les corrélations entre l'attitude et le comportement non-verbal.
4. Apprentissage du modèle à partir des corrélations et des données du corpus (réseau Bayésien).
5. Evaluation du modèle.

Dans l'objectif de pouvoir simuler un agent dans un groupe, nous choisissons

une approche à base d'états et de transitions afin de pouvoir décrire un modèle s'adaptant à différentes configurations de groupes et aux attitudes des membres envers chacun. Cette seconde modélisation se focalise alors sur les comportements plus particulièrement liés à la formation de groupe (regards, orientation du corps et distances avec les autres) et à la prise de parole (alternance entre les rôles de locuteur et d'auditeur, interruptions et comportements non-verbaux associés). Les étapes de ce travail sont les suivantes :

1. Choix d'une représentation du groupe et de ses paramètres.
2. Sélection des paramètres liés à la formation de groupes et modélisation de l'influence de l'attitude sur eux.
3. Modélisation des états conversationnels d'un agent, des comportements associés, des transitions entre les états et de l'influence de l'attitudes sur ces transitions.

Nous avons proposé une architecture d'agent social incluant nos modèles d'attitudes sociales, de formation de groupe et de tour de parole. Cette architecture repose sur Unity3D. Le système résultant prend la forme d'un jeu sérieux dans lequel un utilisateur peut déplacer un avatar dans un environnement virtuel afin d'observer et de rejoindre des groupes d'agents animés par nos modèles. Une évaluation du système est conduite dans laquelle il est demandé à des participants de reconnaitre les attitudes exprimées par certains agents au sein d'un groupe. Cette évaluation est réalisée dans des environnements impliquant différents niveaux d'immersion (navigateur Internet et système de réalité virtuelle CAVE). Les étapes de ce travail sont les suivantes :

1. Définition de l'architecture de l'agent social qui intègre les modèles.
2. Définition d'un protocole d'évaluation de ce système afin de tester sa capacité à produire les comportements correspondant à différentes attitudes.
3. Réalisation d'une évaluation à travers un navigateur Internet.
4. Réalisation d'une évaluation dans un système immersif de type CAVE.

Ce modèle est conçu afin d'être utilisé dans différents scénarios. Il vise à être utilisé afin de peupler un environnement virtuel de groupes de personnages qui simulent une conversation tout en maintenant des attitudes entre eux, ceci afin de renforcer la crédibilité du monde virtuel dans lequel l'utilisateur évolue.

1.4 L'ORGANISATION DE LA THÈSE

Ce rapport de thèse est organisé de la façon suivante. Tout d'abord, dans le chapitre 2, nous présentons les fondements théoriques en sciences humaines et sociales sur lesquels nous nous appuyons, ceci afin d'apporter les définitions aux termes que nous utilisons dans le reste de la thèse. Cela inclut plus précisément la littérature sur le comportement non verbal, la conversation et les groupes ainsi que l'expression d'attitudes interpersonnelles. Ensuite, dans le chapitre 3, nous présentons les travaux existants sur les ACAs qui se sont intéressés aux problématiques suivantes : génération de tours de parole, expressivité des agents, groupes d'agents et expressions d'attitudes interpersonnelles. Nous expliquons alors comment notre travail vient s'inscrire dans la continuité de ces travaux et comment le modèle que nous proposons se distingue des autres travaux. Dans le chapitre 4, nous décrivons notre première expérimentation, grâce à laquelle nous avons construit un modèle computationnel d'expression d'attitudes à travers le comportement non verbal d'un agent dans des interactions dyadiques avec l'utilisateur. Le chapitre 5 s'intéresse à l'extension du modèle précédent pour des situations de groupes en considérant des comportements liés au maintien de la formation de groupe. Notre approche se base sur l'intégration d'un système d'influence construit à partir de la littérature. Nous y décrivons ensuite le modèle de tour de parole influencé par les attitudes qui s'appuie sur une représentation des états convesationnels d'un agent. Dans le chapitre 6, nous décrivons les deux évaluations que nous avons menées pour vérifier la validité du modèle. Le projet européen dans lequel s'est inscrit ce travail de thèse ainsi que notre contribution au sein de ce projet sont présentés dans le chapitre 7. Enfin dans le dernier chapitre, nous discutons des résultats et des perspectives de ce travail.

2 Fondements Théoriques

L'objectif du travail de cette thèse est de proposer un modèle computationnel permettant à un agent, participant à une conversation de groupe, d'exprimer différentes attitudes sociales à travers son comportement non-verbal. Pour développer notre modèle, nous avons basé nos travaux sur la littérature en sciences humaines et sociales, mettant en évidence la définition des attitudes sociales et l'influence de celles-ci sur les comportements non-verbaux produits dans des groupes conversationnels d'humains. Les comportements d'un agent conversationnel pouvant se rapprocher de ceux d'un humain, l'objectif est de construire un modèle qui s'appuie sur cette littérature pour reproduire ces mécanismes dans un agent conversationnel. C'est pourquoi nous proposons à travers ce chapitre une étude de la littérature sur l'expression d'attitudes sociales à travers 3 dimensions : le comportement non-verbal, la formation de groupe et les stratégies de prise de parole. Plus précisément, cette étude de la littérature nous permet d'identifier quels sont les paramètres de ces comportements importants pour la réalisation de l'expression d'attitudes sociales, quelles sont leurs différentes valeurs et quelles sont leurs fonctions. Ces paramètres du comportement non-verbal incluent notamment les différents signaux pouvant être réalisés comme les gestes et les expressions faciales ainsi que leur intensité et leur durée. L'objectif est alors d'identifier les différentes valeurs de ces paramètres permettant d'exprimer différentes attitudes.

Ce chapitre est organisé comme suit. Dans une première section, nous cherchons à définir ce qu'est une attitude sociale et nous présentons la formalisation de ce concept que nous utilisons dans ce travail. Puis, dans la section suivante, nous étudions la communication non-verbale. Nous présentons quels sont les paramètres de cette communication, les différents signaux, et comment ceux-ci sont utilisés pour exprimer une attitude. Dans la troisième section, nous nous intéressons à la conversation et aux tours de paroles. Nous expliquons comment différentes stratégies de

prises de parole dans une conversation peuvent exprimer différentes attitudes. Dans la quatrième section, nous décrivons les mécanismes de la formation de groupe et comment différentes attitudes mènent à différentes formations.

2.1 Les attitudes sociales

2.1.1 Définition

La première étape dans ce travail de thèse est de définir et d'expliquer ce que sont les attitudes sociales. Il est rapporté par Hogg [Hogg, 2000] que les chercheurs qui se sont intéressés à ce sujet convergeaient vers une définition commune qui était qu'une attitude est un ensemble de croyances envers un objet, un concept ou une personne pouvant être positives ou négatives. Dans [Chindamo et al., 2012], les auteurs précisent différents termes qui ont même été utilisé par les chercheurs pour représenter ce même concept (par exemple attitude, évaluation ou positionnement). Dans notre travail, l'objectif est de faire reproduire par des agents des attitudes telles que celles exprimées envers d'autres personnes, qui peuvent être qualifiées dans ce cas d'interpersonnelles.

Scherer [Scherer, 2005] propose d'ailleurs une définition de ces attitudes sociales sous l'appellation *interpersonal stance*. Scherer explique qu'une attitude est "un style affectif utilisé de manière stratégique ou naturelle dans une interaction avec une personne ou un groupe de personnes". On retrouve dans sa définition l'implication qu'une attitude exprimée par une personne l'est envers autre chose, ici une personne ou un groupe de personnes. De plus, Chindamo explique que ces attitudes (ou *stances*) peuvent s'exprimer à l'aide du comportement verbal (ex. le contenu des phrases) et du comportement non-verbal (ex. les gestes et expressions faciales) [Chindamo et al., 2012]. En d'autres termes, lors d'une interaction, une personne produit des comportements, verbaux ou non-verbaux, spécifiques en fonction de ce qu'elle souhaite exprimer et de ses différents interlocuteurs. Par exemple elle peut être amicale avec un ami ou agir de façon autoritaire avec son employé. Cela signifie qu'une attitude ne remplace pas l'intention communicative, mais vient la compléter. Par exemple, l'intention de saluer une autre personne peut se traduire par des comportements très différents en fonction de l'interlocuteur : une personne peut dire "Salut" en faisant un hochement de tête ou alors elle peut aussi dire "Bonjour" en

agitant bien haut la main. Ces différents comportements sont donc bien choisis par la personne, d'une façon consciente ou inconsciente [Scherer, 2005], suivant le contexte (ex. interlocuteur et lieu). C'est cette capacité de variation dans le comportement non-verbal que nous voulons donner à un agent, afin qu'il puisse adapter naturellement son comportement en fonction de ses interlocuteurs (agents et utilisateurs) et en particulier en fonction de l'attitude qu'il adopte face à ces derniers.

2.1.2 Représentation

Afin de donner aux agents virtuels la capacité d'exprimer de telles attitudes, il est nécessaire de commencer par choisir une représentation à manipuler. Encore une fois, différentes représentations ont été proposées par les chercheurs qui se sont intéressées à la dimension interpersonnelle des interactions.

Schutz [Schutz, 1958] propose une représentation, qu'il appelle *FIRO* (pour "Fundamental Interpersonal Relations Orientation"). Dans cette représentation, les attitudes d'une personne sont représentées sous la forme de trois besoins à satisfaire qui sont l'*inclusion*, le *contrôle* et l'*appréciation*. L'inclusion correspond au besoin de participer aux interactions du groupe, Le contrôle correspond au besoin de contrôler les interactions et l'appréciation correspond au besoin d'être proches et familier avec les autres. Par exemple, une personne peut avoir un besoin élevé en inclusion et appréciation et faible voire inexistant en contrôle.

Burgoon et Hale [Burgoon et al., 1984] proposèrent 12 critères pour décrire une attitude envers une autre personne. Ces critères sont la *dominance* et la *soumission*, l'*émotivité*, le *calme*, la *similarité* avec l'autre, le *caractère formel* de l'interaction avec l'autre, la *réactivité*, la *confiance*, la *superficialité* de la relation, l'*appréciation* et l'*hostilité*, l'*engagement* et l'*inclusion*.

Mais la représentation qui reste la plus utilisée aussi bien dans la communauté des Sciences Humaines et Sociales que dans la communauté des ACA est la représentation à deux dimensions d'Argyle [Argyle, 1988]. Dans cette représentation, la première dimension est la dimension de *Statut* ou de *Dominance* qui représente la différence de statut social. Plus la valeur est élevée, plus l'attitude exprimée correspond à une attitude dominante et plus la valeur est faible, plus l'attitude exprimée correspond à une attitude soumise. La seconde dimension est la dimension d'*Affiliation* ou d'*Appréciation* qui correspond à la proximité avec l'autre. Plus la valeur est éle-

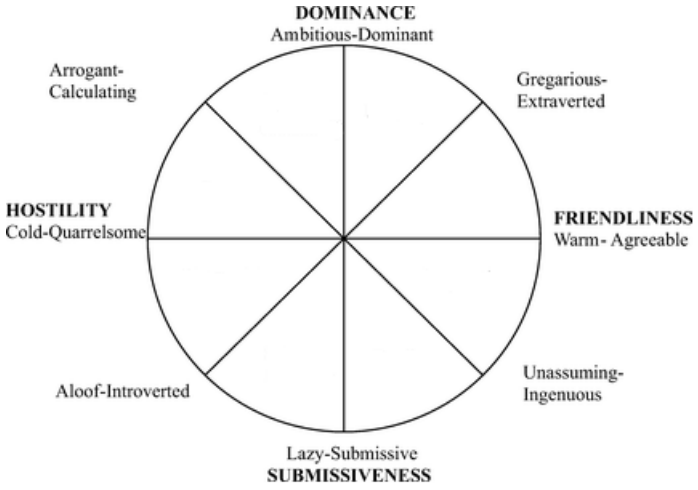


FIGURE 2.1 : L'Interpersonal Circumplex de Wiggins

vée, plus l'attitude exprimée correspond à une attitude amicale et plus la valeur est faible, plus l'attitude exprimée correspond à une attitude hostile. Ces deux dimensions correspondent à celles utilisées dans le modèle *Interpersonal Circumplex* décrit par Leary [Leary, 1958] et Wiggins [Wiggins, 2003]. Dans cette représentation, les attitudes sont représentées sur un continuum circulaire construit par ces deux dimensions (voir figure 2.1).

Utiliser la représentation à deux dimensions d'Argyle offre l'avantage d'être plus simple à manipuler que les autres tout en gardant une grande efficacité dans la description d'attitude. Les deux dimensions de Statut et d'Affiliation sont aisément compréhensibles et elles peuvent être de ce fait facilement utilisées pour représenter différentes attitudes. C'est un point important dans notre travail car dans la construction et la validation de notre modèle, nous faisons appel à des participants qui doivent pouvoir comprendre les concepts que nous leur faisons manipuler sans nécessairement être des experts du sujet. Par ailleurs, Gurtman [Gurtman, 2009] souligne l'efficacité de cette représentation pour décrire la personnalité dans le contexte des interactions. C'est pourquoi nous avons choisi d'utiliser également cette représentation. De plus, de nombreux travaux en sciences humaines et sociales sur l'étude de comportement [Burgoon et al., 1984, Burgoon and Le Poire, 1999, Mehrabian, 1969, Carney et al., 2005] ainsi que dans le domaine des ACAs [Gillies and Ballin, 2003][Chollet et al., 2014] se sont basés sur ces dimensions pour leurs études. Ces travaux nous seront utiles pour la construction de notre modèle.

Dans la section suivante, nous détaillons comment l'expression d'attitudes sociales influence le comportement non-verbal d'une personne.

2.2 L'influence de l'attitude sociale sur le comportement non-verbal

Les humains ont à leur disposition différentes modalités lorsqu'ils ont besoin de communiquer avec d'autres, comme la parole, les gestes ou l'écrit par exemple. La parole, grâce au langage, permet potentiellement d'encoder et d'exprimer toutes les informations que nous souhaitons communiquer. Cependant, dans nos interactions quotidiennes, lorsque nous sommes face à face avec nos interlocuteurs, nous encodons naturellement une partie de ce que nous souhaitons exprimer dans notre comportement non-verbal, ce comportement non-verbal servant ainsi de complément à la parole, permettant d'accentuer, d'atténuer ou de réguler ce qui est dit [Knapp et al., 2013]. Argyle [Argyle, 1988] présente cinq fonctions du comportement non-verbal : l'expression d'émotions, l'expression d'attitudes, la régulation des tours de parole, la présentation de soi et la réalisation de certaines coutumes. La fonction qui nous intéresse dans ce travail de thèse est l'expression d'attitudes et c'est la capacité de réaliser celle-ci, à travers le comportement non-verbal donc, que nous souhaitons donner aux agents virtuels.

2.2.1 Le comportement non-verbal

Les différentes modalités du comportement non-verbal (par exemple les expressions faciales, les gestes ou la posture) ont été étudiées afin de comprendre leurs différentes fonctions communicatives et quand et comment elles sont utilisées dans la communication.

Ekman et Friesen [Ekman and Friesen, 1976] ont proposé une nomenclature pour décrire les mouvements du visage appelée FACS. Dans cette nomenclature, les mouvements possibles sont représentées par des actions (comme lever les sourcils ou étirer les lèvres par exemple) au nombre de 28. Chaque action est également associée aux muscles du visage impliqués dans le mouvement. Plus tard, en suivant cette nomenclature, Ekman [Ekman, 1979] a étudié comment les sourcils sont utilisés pour communiquer des émotions et des signaux de régulations de la conversation. Il

explique que les mouvements de lever et baisser les sourcils sont utilisés pour accentuer un mot ou un point de conversation et en fin de phrase par exemple. Il explique également que le lever de sourcils peut être utilisé pour exprimer du doute ou de la surprise. A travers ce travail, l'on découvre qu'une même modalité du comportement (ici les sourcils) peur servir à encoder différentes fonctions communicatives.

Les gestes ainsi que leurs formes, leurs nombres et leurs rôles ont aussi été étudiés afin d'identifier leurs fonctions communicatives. Ekman et Friesen proposèrent une classification des gestes [Ekman and Friesen, 1969] en fonction de leur rôle dans l'accompagnement de la parole. Cette classification est la suivante :

- Les emblèmes : des gestes qui n'ont pas besoin de la parole pour communiquer un sens car ils représentent un sens à eux seuls, comme le pouce levé vers le haut ou le V de la victoire par exemple.
- Les illustrateurs : des gestes qui accompagnent la parole et imitent le sens de ce qui est dit, comme faire un geste vers le haut tout en parlant d'une croissance par exemple.
- Les gestes de régulation de la conversation : ils permettent de faire comprendre quand on prend la parole ou quand on souhaite la donner.
- Les gestes d'expression d'émotions : Un mouvement brusque des bras lors de l'expression de la peur par exemple.
- Les adaptateurs : des gestes de contrôle que l'on fait sur son corps ou sur un objet, comme jouer avec un stylo ou se gratter la tête par exemple.

McNeill proposa plus tard une autre classification [McNeill, 1992]. Selon lui, il existerait seulement quatre catégories de gestes qui sont les suivantes :

- Les gestes iconiques, imitant le contenu de ce qui est dit afin de l'accentuer.
- Les gestes métaphoriques, représentant les concepts de ce qui est dit.
- Les gestes déictiques, comme pointer du doigt ou montrer une direction.
- Les gestes *beats*, des petits gestes rythmiques de la main ou des doigts accompagnant le rythme de la parole.

Les raisons qui poussent les chercheurs à chercher des classifications et à étudier ces différentes modalités sont motivées par l'étude de ces comportements afin d'en comprendre les significations. Dans certains cas, nous utilisons consciemment nos gestes, nos regards et nos expressions faciales comme support pour la compréhension de ce que l'on souhaite dire, mais dans la plupart des cas, ces modalités sont

utilisés sans même que nous ne nous en rendions compte et il est très difficile de les contrôler, rendant ainsi possible pour notre interlocuteur l'identification de certains traits comme notre personnalité, notre état émotionnel ou notre attitude sociale [DePaulo, 1992].

2.2.2 L'influence de l'attitude sociale sur les différentes modalités du comportement non-verbal

Plusieurs chercheurs dont Darwin [Darwin, 2002] et Ekman [Ekman, 1999] ont théorisé comment sont exprimés les émotions dans nos expressions faciales et nos gestes. Cependant, les émotions ne sont pas les seules à pouvoir influencer le comportement non-verbal. Dans [Vinciarelli et al., 2009], les auteurs, à travers une étude de la littérature dans le domaine des comportements sociaux, listent quelles modalités sont associées à quels comportements sociaux (voir figure 2.2). D'ailleurs, certains chercheurs se sont intéressés plus précisément à l'expression et la perception des attitudes sociales à travers le comportement non-verbal [Argyle, 1988][Mehrabian, 1969]. Leurs résultats ont montré qu'il existe des variations de comportement lorsqu'une personne exprime différentes attitudes. Les travaux de Mehrabian [Mehrabian, 1969], Carney [Carney et al., 2005], Burgoon [Burgoon et al., 1984][Burgoon and Le Poire, 1999] et leurs collègues contiennent énormément d'informations sur ces variations grâce aux résultats de différentes études qu'ils ont conduits.

En ce qui concerne les gestes, ils ont notamment trouvé que produire plus de gestes et des gestes plus larges correspond à une attitude dominante ou amicale [Burgoon and Le Poire, 1999, Carney et al., 2005]. Lorsque l'on touche quelqu'un pendant une conversation (sur l'avant-bras ou l'épaule par exemple), en plus de la fonction communicative qui peut y être associée (comme vouloir dire à quelqu'un de reculer ou de s'approcher par exemple), cela peut correspondre à une attitude dominante ou amicale [Burgoon et al., 1984][Carney et al., 2005][Isbister, 2006],[Yabar and Hess, 2007]. A l'inverse cependant, l'utilisation de gestes de type adaptateur [Ekman and Friesen, 1969] qui consiste à se toucher soi-même (se tenir le bras, se gratter la tête, se frotter machinalement les mains) ou à manipuler des objets (jouer avec un stylo, une feuille en papier) se retrouve principalement dans l'expression d'une attitude soumise [Briton and Hall, 1995][Carney et al., 2005] mais peut également être associé à une attitude hostile [Burgoon and Le Poire, 1999] dans certains cas.

	Example Social Behaviours							Tech.		
Social Cues	emotion	personality	status	dominance	persuasion	regulation	rapport	speech anlysis	computer vision	biometry
Physical appearance										
height			√	√					√	√
attractiveness		√	√	√	√		√		√	√
body shape		√		√					√	√
Gesture and posture										
hand gestures	√	√			√	√	√		√	√
posture	√	√	√	√	√	√	√		√	√
walking		√	√	√					√	√
Face and eyes behaviour										
facial expressions	√	√	√	√	√	√	√		√	√
gaze behaviour	√	√	√	√	√	√	√		√	
focus of attention	√	√	√	√	√	√	√		√	
Vocal behaviour										
prosody	√	√		√	√		√	√		
turn taking	√	√	√	√		√	√	√		
vocal outbursts	√	√		√	√	√	√	√		
silence	√		√				√	√		
Space and Environment										
distance	√	√	√		√		√		√	
seating arrangement				√	√		√		√	

FIGURE 2.2 : Les comportements sociaux et les modalités associées repris de [Vinciarelli et al., 2009]

Ce dernier point souligne l'importance de considérer les modalités ensemble dans l'expression d'attitudes. Il peut être difficile de discriminer les attitudes en étudiant le comportement non-verbal modalité par modalité.

L'orientation de la tête peut également varier selon les attitudes exprimées. Orienter sa tête vers l'autre, pencher légèrement la tête vers le haut et secouer légèrement la tête sont des comportements associés avec une attitude dominante alors que les comportements inverses comme pencher sa tête vers le base et orienter sa tête ailleurs que vers l'autre peuvent donc traduire une attitude soumise voire hostile [Briton and Hall, 1995][Carney et al., 2005].

Par rapport à la posture, se pencher en avant vers l'autre [Burgoon et al., 1984][Burgoon and Le Poire, 1999][Carney et al., 2005][Isbister, 2006][Yabar and Hess, 2007], se tenir droit [Carney et al., 2005], s'orienter directement vers l'autre [Mehrabian, 1969][Burgoon et al., 1984][Burgoon and Le Poire, 1999][Carney et al., 2005][Isbister, 2006][Yabar and Hess, 2007] sont des comportements pouvant être associés avec de la dominance ou de l'amicalité alors que les comportements inverses comme se reculer, s'orienter sur le côté et se tenir moins droit sont plutôt associés à des attitudes soumises ou hostiles.

Le regard offre des variations subtiles en termes d'expressions d'attitudes. S'engager dans des regards mutuels est un signe de dominance ou d'amicalité alors que simplement faire attention aux autres traduit plutôt une attitude soumise (voire amicale) [Burgoon et al., 1984][Carney et al., 2005][Yabar and Hess, 2007][Mehrabian, 1969]. De plus, il est noté qu'un regard trop insistant peut traduire une attitude plutôt hostile [Carney et al., 2005][Burgoon et al., 1984][Mehrabian, 1969]. Les évitements de regards quant à eux traduisent une attitude soumise voire hostile [Burgoon and Le Poire, 1999][Carney et al., 2005][Mehrabian, 1969].

Enfin et bien évidemment, les expressions faciales permettent également d'exprimer différentes attitudes. La capacité de produire plus d'expressions faciales variées est associé à des attitudes dominantes et amicales [Carney et al., 2005][Burgoon and Le Poire, 1999][Hess et al., 2005]. Cependant toutes les expressions ne sont pas associées à la dominance et à l'amicalité. Les expressions d'émotions de valence négative sont associés à des attitudes de valence négative et inversement. Exprimer du dégout est plutôt associé à une attitude hostile [Carney et al., 2005][Hess and Thibault, 2009] de même qu'exprimer une expression de tristesse traduit une attitude soumise

[Hess and Thibault, 2009]. Les sourires restent principalement associé à une attitude amicale [Burgoon et al., 1984][Briton and Hall, 1995][Yabar and Hess, 2007].

Toutes ces études et ces différences traduisent bien l'importance de l'attitude sociale dans la production et la perception du comportement non-verbal d'une personne. Il s'agit là de comportements non-verbaux servant de support à la conversation. Il est important de rappeler que l'interprétation de ces comportements non-verbaux n'a pas été étudiée de manière isolée. Cela veut dire que chaque modalité n'est pas considérée indépendamment des autres mais en interaction les unes avec les autres. D'ailleurs, dans [Burgoon and Le Poire, 1999], les auteurs rappellent "qu'une modalité du comportement non-verbal est continuellement accompagné par les autres et c'est l'ensemble qui permet de délimiter et de clarifier le sens" ("No nonverbal cue is an island. It is continously surrounded by a host of nonverbal behaviors which together may delimit and clarify meaning.").

Dans la prochaine section, nous nous intéressons aux comportements liés à la formation de groupes et comment ceux-ci sont également influencés par l'expression d'attitudes sociales.

2.3 La formation de groupe

Quand on parle de groupe conversationnel, on parle de groupe de personnes réunies dans le but de communiquer les unes avec les autres. Dans ce travail, nous nous intéressons à des groupes de personnes de tailles modérées qui se réunissent en face-à-face. Dans cette section, nous définissons ce qu'est un groupe conversationnel en abordant les deux aspects d'un groupe, le rassemblement de personnes d'une part et leurs interactions d'autre part.

Tout d'abord, il convient de définir ce que l'on entend par groupe de personnes : un ensemble d'au moins deux personnes qui interagissent entre elles peut être appelé un groupe [McGrath, 1984]. Cela voudrait dire qu'une foule ou une audience peuvent également être appelé "groupe" mais ce ne sont pas des formations que nous avons considérées dans ce travail. Dans ces formations, tous les membres ne sont pas en interactions directes avec les autres.

Nous nous intéressons uniquement à des groupes de tailles réduites. Certains chercheurs les appellent "petit groupe" et leur donnent une limite de taille d'une vingtaine de personnes [Beebe and Masterson, 2009]. Cependant, plus la taille d'un

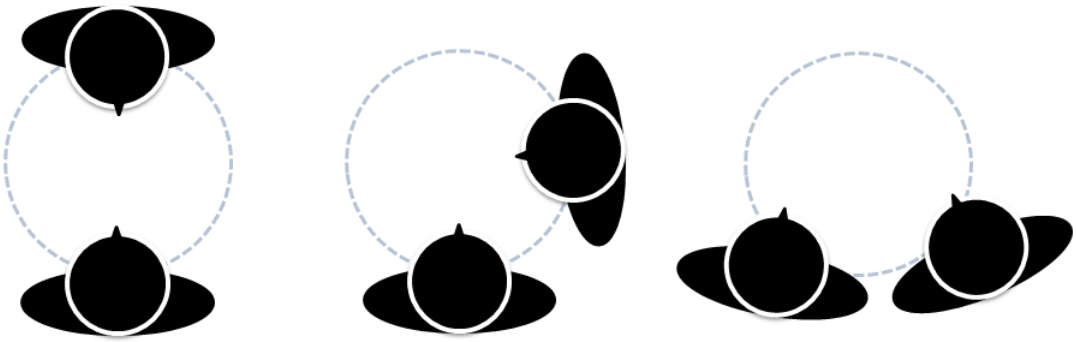


FIGURE 2.3 : Différentes F-Formations selon Kendon [Kendon, 1990]

groupe augmente, moins chaque membre peut interagir avec tous les autres. Nous prenons alors pour définition de groupe conversationnel la définition de Homans [Homans, 2013] où il explique qu'un groupe est un ensemble de personnes suffisamment peu nombreuses pour qu'elles puissent toutes directement communiquer les unes avec les autres, en face-à-face, sans intermédiaire.

2.3.1 Les comportements liés à la formation de groupe

Lors de rassemblements, les personnes s'organisent dans l'espace dans des formations structurées et des chercheurs ont remarqué et identifié les caractéristiques physiques de ces formations. Kendon [Kendon, 1990] a proposé une représentation de celles-ci, qu'il a nommé F-Formation, où les gens sont décrits comme essayant de partager un espace commun, au centre du groupe, et de se répartir autour de cet espace. Comme les gens essayent de se donner une quantité égale de l'espace commun, cela donne à la formation de groupe cet aspect circulaire (voir figure 2.3). Deux personnes qui interagissent peuvent ne pas se trouver face à face par exemple mais côte à côte, orientant toutes deux leurs corps vers un point commun.

Scheflen et Ashcraft [Scheflen and Ashcraft, 1976] proposent une autre description de ces formations en décomposant les différents territoires au sein de celles-ci. D'après eux, un rassemblement humain se compose d'abord de deux zones concentriques que sont le *nucleus*, l'espace dans lequel se trouve les participants du groupe et la *region*, autour du *nucleus*, l'espace dans lequel se trouvent les autres personnes

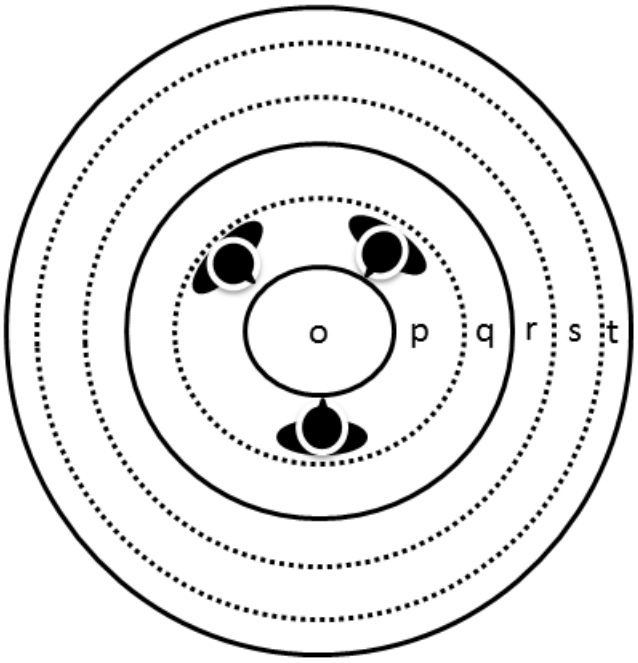


FIGURE 2.4 : Les territoires de Scheflen

qui sont spectateurs ou qui sont en train de rejoindre la formation. Ces deux zones sont elles-mêmes décomposées en 3 sous-zones. De la plus centrale à la plus éloignée, le *nucleus* se compose des zones *o*,*p* et *q* et la *region* se compose des zones *r*, *s* et *t* telles que décrites dans la figure 2.4. La zone *o* correspond à l'espace partagé par les membres du groupes, autour duquel ils se réunissent, et la zone *p* correspond à la zone dans laquelle se trouvent les participants. La zone *q* sert de zone tampon pour les personnes essayant de rejoindre les participants. En ce qui concerne la *region*, les zones *r* et *s* correspondent à deux niveaux de spectateurs du groupe et la zone *t* est utilisé pour les personnes rejoignant ou évitant la formation. Les auteurs précisent que pour des petits groupes conversationnels, Les zones de la *region* peuvent ne pas être nécessaires et être confondues avec la zone la plus excentrée du *nucleus*, la zone *q*. Le *nucleus* se compose alors du centre *o*, qui peut se confondre avec le centre de la F-formation de Kendon, de la zone *p* dans laquelle se trouvent les participants et de la zone *q* dans laquelle se trouvent les personnes qui ne sont pas des participants, qui observent le groupe ou juste passent à côté pour l'éviter.

Une autre théorie qui permettrait d'expliquer comment la formation d'un groupe

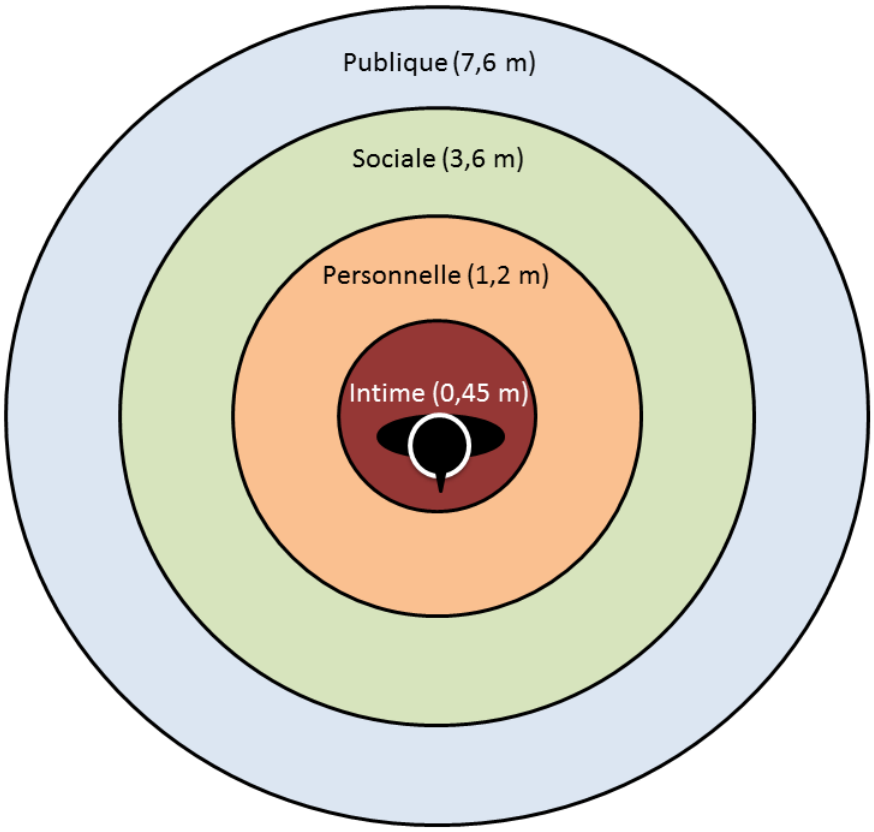


FIGURE 2.5 : Les zones des *Proxemics* de Hall

émerge à partir du comportement de ses membres est la théorie des *Proxemics* de Hall [Hall, 1969]. Hall s'intéresse non pas à l'espace du groupe mais à l'espace autour des personnes. Il décrit cet espace comme étant constitué de quatre zones concentriques qui sont utilisées pour communiquer différemment suivant les interlocuteurs. De la zone la plus proche à la plus lointaine, Hall a nommé ces zones : intime, personnelle, sociale et publique (voir figure 2.5). Suivant cette théorie, la plupart des interactions sociales avec des connaissances ont lieu dans la zone sociale, alors que les zones personnelles et intimes sont réservées pour les contacts avec la famille et les amis très proches.

Ces différentes théories s'intéressent aux distances qui peuvent exister entre les membres d'une interaction. On remarque d'ailleurs une certaine complémentarité entre celles-ci. Les théories de Kendon [Kendon, 1990] et Scheflen [Scheflen and Ashcraft, 1976] permettent de représenter l'espace de la formation alors que la théorie

de Hall [Hall, 1969] permet de représenter l'espace occupé par une personne. Ces différentes théories avancent l'idée qu'une formation de groupe existe grâce aux comportements produits par les participants afin de maintenir celle-ci : les membres réajustent leur distance entre eux ainsi que leur orientation afin de respecter les formations et les *Proxemics*.

Pour résumer, les groupes sont caractérisés par leurs membres et par leur organisation autour d'un espace commun. Chaque membre contribue à la formation géométrique en ajustant ses distances et ses orientations avec les autres membres.

2.3.2 L'influence de l'attitude sur les comportements de groupes

Nous avons vu comment les personnes utilisent leurs distances et orientations entre elles au sein d'un groupe pour maintenir la formation de celui-ci. Mais ces comportements peuvent également révéler les attitudes sociales exprimées par une personne envers les autres.

Mehrabian [Mehrabian, 1969] décrit les regards, la posture et la distance comme des comportements importants qui impactent l'évaluation de l'attitude lors d'une interaction. Dès lors, ces composantes du comportement ne sont plus utilisées uniquement pour maintenir une formation de groupe mais peuvent également encoder simultanément des attitudes sociales. Mehrabian a présenté différentes études qui se sont intéressées à la relation entre attitudes interpersonnelles et une variété de comportements non-verbaux, incluant les *Proxemics* des travaux de Hall. Par exemple, il explique que lorsqu'une personne à une relation peu amicale avec une autre, cette personne évite de regarder l'autre. Au contraire, plus la personne apprécie l'autre, plus elle va regarder l'autre. Cependant il est précisé que cette corrélation entre le niveau d'appréciation et la quantité de regard est vraie jusqu'à un certain point. Trop de regards reçus peuvent amener la personne regardée à ressentir une attitude un peu moins amicale que lorsque la quantité de regards est plus modérée. Un phénomène similaire décrit la quantité de regard en fonction de l'expression d'une attitude soumise. Mehrabian rapporte que dans une configuration à trois personnes, la personne la plus dominante est regardée plus que les autres.

L'orientation du corps peut également permettre d'exprimer une attitude dans une conversation. Mehrabian [Mehrabian, 1969] explique que les personnes orientent leur torse plus directement vers les personnes envers qui est exprimé le moins de

dominance.

En ce qui concerne le maintien des distances avec les autres, une personne se tiendra plus proche d'une autre en fonction de l'attitude amicale qu'elle lui exprime. Cependant, une personne aura tendance à se tenir plus éloignée d'une autre si elle lui exprime une différence de statut (de la dominance ou de la soumission).

Dans une étude plus récente [Cristani et al., 2011], les auteurs ont trouvé que les attitudes sociales peuvent être inférées à partir des distances interpersonnelles entre les membres d'un groupe. En utilisant les *Proxemics* de Hall et la F-Formation de Kendon pour modéliser les formations de groupes, les auteurs ont rapporté également que les gens avec une plus grande différence de statut maintiennent une plus grande distance entre eux.

Dans [Sanchez-Cortes et al., 2013], les auteurs identifient un leader dominant émergeant dans un petit groupe en utilisant des indices de la voix et des regards à travers l'analyse d'un corpus annoté d'interactions de petits groupes. Ils ont trouvé notamment qu'une personne est perçue comme plus dominante si elle est regardée plus que les autres.

Nous avons vu dans cette section comment les attitudes sociales sont également exprimées à travers les comportements de formation de groupe. Dans la section suivante, nous verrons comment les attitudes influencent également la dynamique de la conversation à travers les stratégies liées aux tours de parole.

2.4 La conversation et les tours de parole

2.4.1 L'interaction de groupe

Un groupe ne se résume pas seulement à sa formation dans l'espace. Un groupe se définit également par les activités qui ont lieu entre les membres de celui-ci. McGrath [McGrath, 1984] propose un cadre descriptif des composants d'un tel groupe. Il commence par expliquer qu'un groupe, qui participe à un processus d'interactions entre ses membres, est d'abord défini par ceux-ci. Chaque membre a un ensemble de propriétés (descriptive comme le genre, la personnalité ou l'âge). Ensuite les relations et les rôles entre les membres définissent la structure du groupe. Ceci constitue la première composante du groupe. La seconde composante contient la description du contexte du groupe et de sa tâche. Un groupe implique que ses membres sont

en train de réaliser une tâche dans un environnement. Ces deux composantes vont définir le cadre des interactions entre les membres du groupe. Ces interactions sont définies elles-mêmes en trois couches. La première couche est la couche de communication. Chaque comportement d'un membre peut être vu comme une communication de celui-ci vers tous les autres membres. Dans la seconde couche, cette communication est porteuse d'une action (l'interaction, dépendante de la tâche du groupe) et d'informations interpersonnelles. Enfin la dernière couche concerne l'impact de ce comportement sur la tâche du groupe et sur les autres membres (par exemple changement de leurs buts ou de leurs relations sociales).

Alors que McGrath donne un cadre de haut niveau pour décrire le contexte dans lequel les interactions de groupes ont lieu, Bales [Bales, 1950] présente un ensemble de douze catégories pour décrire les différentes interactions sociales qui peuvent avoir lieu dans un groupe : les interactions socio-émotionnelles positives (montrer de la solidarité, montrer du relâchement et acquiescer), les interactions socio-émotionnelles négatives (montrer de l'animosité, montrer de la tension et désapprouver), des questions orientées tâches (demander une direction, demander une opinion et demander une suggestion) et les réponses orientées tâches (donner une direction, donner une opinion et donner une suggestion). Il affirme que toutes interactions humaines peut se trouver classer dans une et une seule de ces catégories. Ces catégories ne sont pas liées à un contexte et sont utiles pour une description générale des interactions.

Ces travaux montrent l'importance au sein des interactions de groupe de la dimension sociale. Dans le modèle de McGrath, les relations entre les membres du groupe sont au cœur de la structure de celui-ci et des actions entre les membres. Dans la catégorisation de Bales, la dimension sociale est indispensable à la description des interactions entre les membres d'un groupe. Nous allons voir comment cette dimension sociale s'applique dans le cas d'un groupe conversationnel.

2.4.2 La conversation humaine

La littérature en sciences humaines et sociales présente plusieurs représentations et définitions de ce qu'est une conversation. Les travaux de Sacks, Schegloff et Jefferson [Sacks et al., 1974] figurent parmi les travaux fondateurs dans le domaine. Les auteurs présentent un modèle de description de la conversation où celle-ci y est décrite comme une séquence de tours de parole avec comme stratégie, suivie par

l'ensemble des participants, la minimisation des silences entre les tours et la fluidité des transitions entre eux. D'après les auteurs, le tour de parole d'un participant est composé d'unités de construction du tour (*turn-constructional-unit* TCU). Ces unités peuvent être des phrases entières, des séquences de mots ou juste des mots. A la fin de chacune de ces unités de construction se trouvent des points de transition (*transition-relevance places* TRP) à partir desquels un autre participant peut prendre le tour de parole. Les participants sont capables d'anticiper l'apparition des points de transition, minimisant ainsi les silences, et ils cherchent à appliquer un ensemble de règles visant à minimiser le nombre de locuteurs concurrents. Par exemple, à la fin d'un TCU, si le locuteur courant fait référence à un autre participant, ce participant et seulement lui est invité à prendre le tour suivant.

Ces travaux ont été fondateurs pour tenter d'offrir des outils d'analyse et de donner une structure à la conversation. Cependant, même si ces travaux sont toujours de référence de par leur côté initiateur [Sidnell, 2011], certaines limites ont été relevées par les chercheurs qui ont suivis.

Dans [O'Connell et al., 1990], les auteurs soulignent le fait que le modèle de Sacks, Schegloff et Jefferson ne considère pas les normes culturelles et les indices multimodaux propres aux participants pour la reconnaissance des TCU et des TRP.

Dans [Power and Dal Martello, 1986], Les auteurs ont pointé du doigt le manque de considération pour le comportement non-verbal dans les mécanismes de transmission du tour de parole et l'utilisation exclusive de descriptions syntaxiques pour localiser les TRP. Ce dernier point implique que les participants ne s'intéressent qu'à la forme du texte pour y localiser les points de transitions. Les auteurs précisent que cela pourrait donner lieu à des tours de paroles incomplets. Par exemple, dans la phrase "Je veux [] manger []", un TRP existe après "je veux" et "manger". Sans chercher à interrompre l'autre, il ne devrait pas être possible de prendre la parole après "je veux", le tour de parole ne serait pas complet. Le texte seul peut ne pas être suffisant pour identifier les points de transitions entre les participants.

Duncan [Duncan, 1972] a décrit des caractéristiques du comportement non-verbal qui peuvent être utilisées par les participants pour indiquer leur désir de prendre ou de donner le tour de parole. Par exemple une personne peut, afin d'indiquer à quelqu'un qu'elle souhaite donner le tour de parole, changer l'intonation de sa voix et relâcher le mouvement de ses gestes.

Ces travaux amènent l'idée qu'un tour de parole n'est pas un concept pré-identifié qui construit l'interaction, mais plutôt se construit au fur et à mesure de l'interaction, à partir des comportements de ses participants. La théorie de Clark des actions conjointes, appliquée aux tours de parole telle que décrite dans [Clark, 1996], présente le tour de parole comme émergeant de l'action conjointe de parler et d'écouter et propose de ce fait cette autre représentation de la conversation. Les tours de parole ne sont pas alors les buts à atteindre, mais ils émergent naturellement lorsque les actions conjointes de parler et d'écouter réussissent. Cette représentation permet également de pouvoir considérer que dans une conversation, plusieurs personnes peuvent parler en même temps.

De ce fait, dans cette thèse, une représentation de la conversation proche de celle de Clark sera utilisée. Cette représentation permet de considérer que plusieurs participants puissent parler en même temps, ce qui est nécessaire pour l'expression de certaines attitudes. Le modèle présenté donnera aux participants alternativement les rôles de locuteur et d'auditeur. Le passage d'un rôle à l'autre se faisant en fonction des attitudes et des activités de l'ensemble des participants.

2.4.3 L'influence de l'attitude sur les stratégies de prise de parole

Au sein d'une conversation, il est naturel d'observer que tous les participants ne prennent pas la parole autant et de la même façon. Cela peut dépendre du type de conversation que les participants ont mais également des relations entre eux et des attitudes qu'ils souhaitent exprimer. Dans cette section, nous présentons des travaux de la littérature en sciences humaines et sociales qui se sont intéressés à ces différences.

Dans [O'Connell et al., 1990], les auteurs rapportent plusieurs travaux sur ce sujet qui expliquent par exemple qu'une conversation est construite dans un contexte social. Ce contexte définit les normes que les participants s'engagent à respecter explicitement ou implicitement dans le cadre de la conversation comme les temps de parole de chacun et les rôles par exemple. Suivant le contexte (ex : repas de famille, réunion de travail), les participants vont peut-être accepter les interruptions et les superpositions des tours de parole.

Cappella et ses collègues [Cappella et al., 1985] rapportent de nombreux travaux qui s'intéressent aux différences de comportements et de stratégies de prise

de parole et comment ces différences sont perçues en termes de dominance et d'affiliation. L'intérêt est particulièrement mis sur les moyens mis en œuvre par les participants pour effectuer les transitions entre les rôles de locuteur et d'auditeur. Les auteurs expliquent qu'une personne qui tient le tour de parole plus souvent peut être perçue comme plus dominante et plus amicale. De plus, ils rapportent que cette relation entre la perception de la dominance et la quantité de parole reste vraie même lorsque le contenu du discours est moins pertinent. En d'autres termes, la quantité d'interventions est plus importante pour la perception de l'attitude que la qualité de celles-ci. Ils notent que la perception de la dominance augmente également si les participants sont actifs lorsqu'ils ont le rôle d'auditeur, c'est-à-dire qu'ils produisent des *backchannels*. En revanche, ils notent toutefois qu'une monopolisation excessive du tour de parole peut faire baisser le niveau de perception d'amicalité.

Kennedy et Camden [Kennedy and Camden, 1983] ont regardé les différences entre les hommes et les femmes en termes d'interruptions lors de conversation. Ils ont trouvé que les femmes se faisaient interrompre plus souvent lorsqu'elles souriaient et lorsqu'elles évitaient de regarder dans les yeux la personne prenant le tour de parole. Ces différents comportements semblent surtout refléter une différence d'attitude. Dans [Smith-Lovin and Brody, 1989], les auteurs ont remarqué aussi des différences entre les hommes et les femmes dans les comportements d'interruption mais surtout ils rapportent que les tentatives d'interruptions sont liées à l'expression de dominance ou d'hostilité. En effet, ils expliquent que les participants masculins projetaient une différence de statut sur le genre de leur interlocuteur, ce que ne faisaient pas les participants féminins.

Goldberg [Goldberg, 1990] a présenté les différentes perceptions d'attitude liées aux tentatives d'interruptions. Lorsqu'un individu tente de prendre la parole en même temps qu'un autre, suivant son but et son comportement, son attitude sera perçue différemment. Un individu faisant beaucoup d'hésitations, pour exprimer ses opinions ou demander celles des autres sera perçu comme soumis. Un individu faisant beaucoup d'hésitations pour demander les opinions des autres sur le sujet courant sera perçu comme amical. A l'inverse, un individu faisant peu d'hésitations et qui souhaite exprimer ses opinions ou commenter celles des autres sera perçu comme plus hostile. Enfin, un individu qui n'hésite pas, faisant des phrases complètes afin d'orienter la conversation et d'exprimer ses opinions sera perçu comme dominant.

Dans [Dunbar and Burgoon, 2005], les auteurs ont conduit une étude sur des personnes en couple (mariés et non mariés) et rapportent que leur perception de la dominance de leur partenaire est corrélée au nombre d'interruptions qu'ils subissent pendant une conversation.

Dans [Gifford and Hine, 1994], les auteurs rapportent les résultats d'une expérience dans laquelle les temps de parole sont utilisés afin de mesurer la perception d'attitudes. Ils ont trouvé notamment que les gens perçoivent une attitude dominante et extraverti lorsque leur interlocuteur parle beaucoup. Ce paramètre peut également être un signe d'une attitude hostile. A l'inverse, une attitude soumise ou introvertie est perçue lorsque l'interlocuteur parle peu. Dans [Mast, 2002], l'auteur a identifié une corrélation entre le temps de parole et la perception de la dominance. Ainsi, de nombreux travaux ont montré qu'une personne qui exprime une attitude dominante parle plus [Gatica-Perez, 2009].

2.5 Conclusion

Dans ce chapitre nous avons tenté de mettre en évidence les phénomènes qui ont lieu dans une conversation de groupe et en particulier nous nous sommes intéressés à l'expression d'attitudes sociales. Nous avons débord défini ce qu'est une attitude et nous avons vu que l'expression d'attitudes s'effectue à plusieurs niveaux du comportement non-verbal. Tout d'abord, l'attitude s'exprime à travers le comportement non-verbal qui accompagne la conversation. Gestes, postures, expressions faciales, regards sont autant de paramètres qui peuvent être porteur de l'attitude de la personne qui parle ou qui écoute. Dans une formation de groupe, les personnes maintiennent celles-ci par la production individuelle de comportements permettant de réguler les distances et les orientations entre elles. Ces comportements non-verbaux peuvent également se voir influencés en fonction des attitudes exprimées. Enfin, nous avons également vu comment les stratégies de prises de parole (temps de parole, contrôle de la conversation, interruptions...) peuvent également révéler différentes attitudes. C'est pourquoi nous allons considérer ces paramètres dans notre travail. En particulier, notre modèle permettra de faire changer la quantité de prise de parole d'un agent dans une conversation en fonction de ses attitudes exprimées. Il sera également possible pour un agent d'interrompre un autre agent ou de se faire interrompre.

Dans le chapitre suivant, nous étudions plus précisément les travaux qui se sont

intéressés à la modélisation de ces mécanismes au sein d'agents conversationnels.

Résumé du chapitre

- L'attitude sociale représente un style conversationnel adopté naturellement ou stratégiquement par une personne envers ses interlocuteurs.
- Bien que différentes représentations existent, nous avons choisi une représentation communément utilisée qui décrit une attitude par deux dimensions, une dimension de Statut et une dimension d'Affiliation.
- L'attitude est exprimée à différents niveaux du comportement non-verbal et nous avons vu en particulier comment celle-ci s'exprime à travers les comportements accompagnant la conversation, la formation de groupe et les stratégies de prise de parole.

3 État de l'art

L'étude et la reproduction des comportements humains par des agents virtuels ou physiques humanoïdes est l'un des sujets de recherche de la communauté des agents conversationnels depuis plusieurs décennies [Cassell et al., 2000]. Cependant, l'étude de l'expression d'attitudes est un sujet qui a été peu étudiée dans le contexte de groupes d'agents conversationnels. Les problématiques de cette étude peuvent être adressées sur différents aspects : l'expression d'attitudes, les mécanismes de tours de parole et à la formation de groupe. Par conséquent, nous présentons dans ce chapitre des travaux qui se sont intéressés à ces composantes. Nos travaux de recherches sont d'ailleurs fondés sur certains de ces travaux.

Dans la première section, nous présentons des agents qui ont été dotés de la capacité d'exprimer différents comportements en fonction de l'état de leur relation ou de leur attitude envers leur interlocuteur. Dans la deuxième section, nous discutons des modèles de tours de paroles qui ont été proposés dans la communauté. Enfin dans la troisième section, nous présentons des travaux qui se sont intéressés à la modélisation de groupes de personnages virtuels.

3.1 Les modèles d'expression d'attitudes pour agents virtuels

Dans cette section, nous nous intéressons aux travaux qui visent à modéliser les mécanismes des attitudes sociales afin de les implémenter au sein des agents virtuels. Comme défini par Scherer, une attitude est une intention qui peut être spontanée ou stratégique [Scherer, 2005]. Elle correspond à l'expression d'un alignement avec l'autre, autrement dit elle peut refléter l'état d'une relation sociale. Nous commençons par présenter des travaux précurseurs pour notre travail qui se sont intéressés à donner à des agents conversationnels la capacité d'entretenir des relations sociales

avec l'utilisateur.

3.1.1 Les agents relationnels

Dans [Bickmore and Cassell, 2001], les auteurs introduisent la notion d'agents relationnels, des agents capables de créer une relation avec l'utilisateur qui dépasse le cadre de la réalisation d'une tâche. Les agents dits relationnels sont capables d'entretenir des relations sociales avec les utilisateurs. Les auteurs cherchent à donner des capacités relationnelles à un agent afin d'améliorer la confiance que l'utilisateur a envers celui-ci, ceci afin d'améliorer la qualité de l'interaction [Bickmore and Cassell, 2001]. Les auteurs proposent donc un modèle de dialogue, étendant celui de l'agent immobilier virtuel REA [Cassell et al., 1999], permettant à un agent d'engager l'utilisateur dans un dialogue social. Le dialogue social ne se limite pas juste en un échange orienté tâche, mais contient également des stratégies relationnelles et des éléments de conversation de la vie de tous les jours comme parler de la météo ou de l'actualité par exemple. Leurs résultats montrent qu'un tel agent faisant preuve de stratégies conversationnelles permet d'améliorer le niveau de confiance des utilisateurs ayant un profil extraverti. Ce premier travail est donc intéressant car il a permet de montrer qu'il est possible de changer la perception qu'ont certains utilisateurs d'un agent en changeant son comportement tout en maintenant l'objectif de la conversation. Cependant ce modèle ne s'intéresse pas à représenter différentes attitudes ou relations et il ne fonctionne que pour le scénario d'agent immobilier de REA.

La suite de ces travaux sur les agents relationnels aboutirent au développement de Laura [Bickmore and Picard, 2005]. Laura est un coach de fitness qui a été conçu pour être utilisé sur une durée de 30 jours, et qui fournit un support pour la reprise d'une activité sportive. Le modèle relationnel implémenté dans cet agent permet à celui-ci de se montrer de plus en plus amical avec l'utilisateur au fur et à mesure des jours. Laura utilise ainsi une représentation en étape de sa relation avec l'utilisateur. Plus le nombre de jours où l'utilisateur a utilisé l'agent est grand, plus l'agent montre un comportement amical. Cela se traduit par l'agent qui se tient plus proche de l'écran pour montrer sa proximité avec l'utilisateur, incite plus de regard mutuel et fait plus de gestes des bras et de la tête. Ces variations de comportements ont été implémentées sous forme de règles, basées sur la littérature en comportement

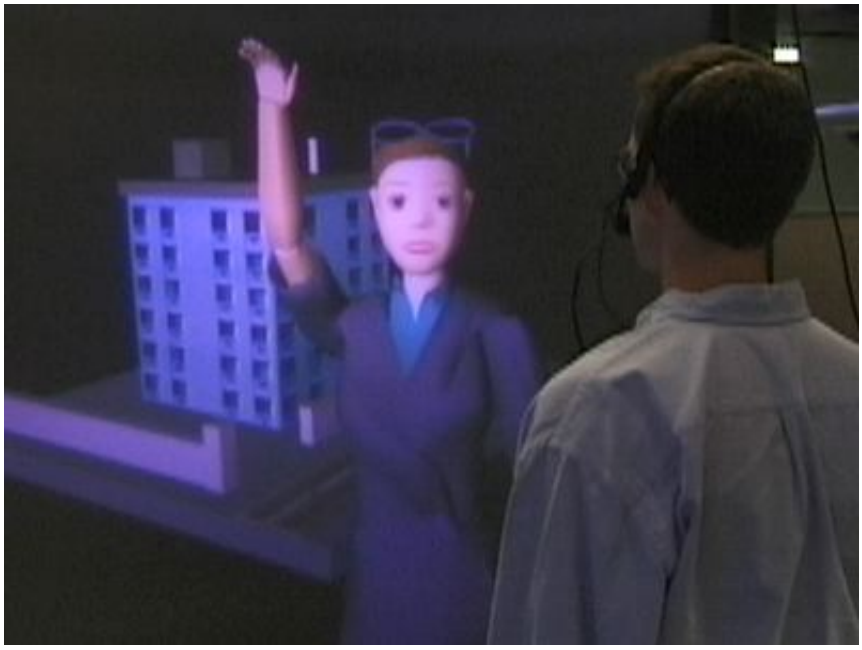


FIGURE 3.1 : Interaction avec l'agent REA [Cassell et al., 1999]

humain. Les auteurs décrivent ensuite une étude qu'ils ont menée où un groupe de participants interagissait avec l'agent muni du modèle relationnel et un autre groupe interagissait avec l'agent sans le modèle relationnel. Leurs résultats montrent des valeurs significativement supérieures pour leurs mesures d'appréciation, de confiance et de respect de l'agent par les utilisateurs lorsque l'agent est muni du modèle relationnel. Dans ce travail différents niveaux de relations sont possibles, mais celles-ci sont représentées par une seule dimension et il n'est pas possible de représenter des relations négatives.

Dans Avatar Arena [Rist et al., 2003], les auteurs ont proposé un modèle pour simuler des conversations entre 3 agents, affichés sur 3 moniteurs différents disposés en cercle. Ces agents ont pour tâche de négocier entre eux un rendez-vous. Chaque agent a une attitude envers chaque autre agent, représentée par trois valeurs (positive, négative et neutre) sur une dimension d'appréciation. Au fur et à mesure de la discussion, en apprenant les préférences des uns et des autres sur leurs disponibilités, les attitudes se mettent à jour, en cas de dissonance, suivant le modèle de *Congruity* proposé dans [Osgood and Tannenbaum, 1955]. Ces différentes attitudes changent ensuite le contenu de la conversation pour refléter celles-ci. Dans ce système, les auteurs ont cherché à modéliser la dynamique des attitudes et comment

celles-ci peuvent évoluer en suivant un modèle théorique. Cependant, leur modèle ne fonctionne que dans ce contexte de négociation de rendez-vous.

D'autres travaux par la suite ont cherché à développer d'autres représentations et mécanismes de calcul des relations. Dans [Kasap et al., 2009], la relation avec l'utilisateur est représentée par les deux dimensions de Statut et d'Affiliation et évolue en fonction de l'humeur et des émotions produites au sein de l'agent EVA pendant les interactions avec l'utilisateur. Ce modèle est conçu à partir de la littérature en sciences humaines et sociales. Dans ce système, l'agent se souvient des interactions passées et de la valence des émotions produites pendant celles-ci afin de calculer la valeur de sa relation avec l'utilisateur. Par exemple, si l'agent ressent de la gratitude envers l'utilisateur, cela va augmenter son appréciation et faire baisser sa dominance. Ces règles s'inspirent du modèle de calcul des émotions appelé OCC [Ortony et al., 1988] proposé par Ortony, Clore et Collins. La relation est ensuite utilisée dans le calcul des émotions futures et suivant l'émotion courante de l'agent, différentes expressions faciales et gestes peuvent être utilisés par l'agent. Bien que les émotions permettent de générer de la variabilité dans le comportement de l'agent, les relations n'ont pas d'effet direct sur ceux-ci. De plus, le modèle d'influence suit un scénario prédéfini (dans [Kasap et al., 2009], l'agent est un professeur de géographie).

Dans [Ochs et al., 2009], les auteurs proposent une approche similaire mais permettant de s'abstraire du scénario. Ils proposent un modèle de calcul de l'état socio-émotionnel d'un agent afin de créer des personnages de jeux vidéo plus crédibles. Leur travail s'appuie sur la littérature en sciences humaines et sociales pour construire leur modèle. Dans celui-ci, l'état socio-émotionnel d'un agent se décompose en trois parties : la personnalité, les émotions ressenties et les relations sociales avec les autres. La personnalité de l'agent a une influence sur les émotions qu'il ressent et celles-ci changent l'état de ses relations sociales. Les auteurs utilisent une représentation des relations composée de quatre dimensions : l'Affiliation, le Statut, la Familiarité et la Solidarité. Lorsque qu'un agent ressent une émotion causée par un autre agent, cette émotion produit une augmentation ou une diminution de la valeur de chacune de ces dimensions (excepté la Familiarité). Par exemple, les émotions *joie*, *espoir*, *admiration* et *fierté* font augmenter l'Affilation alors que les émotions *détresse*, *colère*, *peur* et *honte* la font baisser. Ce modèle permet de représenter des relations positives et négatives et comment celles-ci évoluent en fonction des émotions de l'agent. Cepen-

dant, ce modèle ne décrit pas comment ces relations influencent les comportements de l'agent puisqu'il ne permet que de mettre à jour son état interne.

Dans ces deux travaux précédents, les relations ont été calculées à partir des émotions produites au sein de l'agent. Une autre approche pour le calcul des relations est proposée dans [Prada and Paiva, 2009] où celles-ci sont calculées à partir des actions des autres. Dans ces travaux, les auteurs proposent une application multi-agent dans laquelle un utilisateur peut participer à un jeu avec un groupe d'agents appelé *Perfect Circle* (voir figure 3.2). Ces agents peuvent réaliser différentes actions afin de faciliter ou non la complétion des tâches demandées au sein du jeu. Par exemple, les agents peuvent encourager ou décourager l'utilisateur ou un autre agent en train de réaliser une tâche suivant leur configuration initiale. Ils peuvent également aider ou gêner à la complétion des tâches. Ces différentes actions ont des impacts sur les relations entre chaque agent, représentées encore une fois par le modèle à deux dimensions de Statut et d'Affiliation, grâce à un système de règles construit en suivant la "Balance Theory" de Heider [Heider, 1958].

Dans cette section, nous avons vu des travaux qui ont cherchés à modéliser des relations sociales et leur dynamique au sein des agents. Les auteurs se sont inspirés de [Osgood and Tannenbaum, 1955] et de [Heider, 1958] afin de proposer des modèles permettant de reproduire les mécanismes humains dans le calcul des relations. La plupart de ces travaux se limite à simuler la dynamique des relations que dans un contexte ou un scénario prédéfini comme REA [Bickmore and Cassell, 2001], Laura [Bickmore and Picard, 2005], EVA et Avatar Arena [Rist et al., 2003]. De plus, bien que certains de ces modèles permettent d'influencer la prise de décision en fonction des relations, comme Avatar Arena [Rist et al., 2003] et Perfect Circle [Prada and Paiva, 2009], peu de ces modèles se sont intéressés à l'influence des relations sur le comportement non-verbal des agents (excepté Laura [Bickmore and Picard, 2005]), préférant se focaliser sur le calcul de l'état interne de l'agent, comme le modèle proposé dans [Ochs et al., 2009].

Dans cette thèse, nous voulons compléter ces travaux en proposant un modèle permettant de calculer le comportement non-verbal d'un agent exprimant une attitude sociale. Ces travaux calculent des relations sociales, or l'expression d'une relation sociale implique l'expression d'un alignement avec l'autre, faisant écho à la définition d'attitudes de Scherer [Scherer, 2005].

FIGURE 3.2 : Le jeu collaboratif Perfect Circle de [Prada and Paiva, 2009]

Dans la section suivante, nous verrons des travaux qui se sont intéressés à cette expression d'attitudes dans le comportement non-verbal d'agents conversationnels.

3.1.2 Les agents virtuels et l'expression d'attitudes

Le projet Demeanour est un des premiers travaux qui a cherché à modéliser l'expression d'attitudes entre les personnages d'une application, à travers leurs comportements non-verbaux [Gillies et al., 2004]. Dans cette application, il est possible de configurer les relations (utilisant la représentation à deux dimensions de Statut et d'Affiliation d'Argyle [Argyle, 1988]) entre un petit groupe d'agents et d'observer le changement dans leur comportement non-verbal (regard, gestes et postures). Cet outil a notamment été utilisé pour de l'improvisation théâtrale [Ballin et al., 2004] où les utilisateurs réunis dans une même scène configuraient non pas directement le comportement de leur avatar mais les relations de celui-ci avec les autres. Le système se base sur ce que les auteurs appellent un réseau de comportements. Il s'agit en réalité d'un ensemble de règles qui permettent de déterminer une attitude à exprimer en fonction de la personnalité de l'agent et de ses relations voulues avec les autres. Dès lors qu'une attitude est déterminée, le système choisit au hasard, parmi une sélection correspondante à l'attitude voulue, le comportement non-verbal. La limite principale de ce modèle est que les liens entre relations et comportements sont définis à partir d'observations informelles. De plus, le système intègre un nombre restreint de modalités.

Les travaux présentés dans [Bee et al., 2009] et [Bee et al., 2010] se sont intéressées à la perception de la dominance chez un agent exprimant différents types de comportements. Dans [Bee et al., 2009], leur agent était capable d'avoir différentes positions de tête, différentes directions de regard et d'exprimer plusieurs émotions. Ils ont réalisé une étude afin d'identifier les valeurs de dominance pour tous ces comportements. Ils ont trouvé, par exemple, qu'une tête penchée vers le bas exprime moins de dominance et que les expressions faciales des émotions de joie, dégout colère et l'expression faciale neutre augmentaient la perception de la dominance. Les mêmes auteurs dans [Bee et al., 2010] ont réalisé une seconde étude similaire mais en incorporant du comportement verbal généré par un moteur de dialogue capable de générer différentes phrases en fonction des valeurs des paramètres "extraverti" et "agréable". Ils ont trouvé que les différentes attitudes, la dominance exprimée par

le comportement non-verbal et les attitudes extraverti et agréable par le comportement verbal, ont été bien reconnues. Cependant, les auteurs ont utilisé un agent avec une apparence unique (un majordome). Il serait nécessaire d'étendre leur travail à différentes apparences d'agents.

Dans [Lee and Marsella, 2011], les auteurs ont cherché à modéliser le comportement non-verbal non plus juste du locuteur mais également des personnages auditeurs, simples spectateurs d'une conversation, en fonction des relations les uns avec les autres. Les relations étaient encore une fois représentées par les deux dimensions de Statut et d'Affiliation. Pour cela, ils ont demandé à des acteurs de jouer leur scénario afin de construire une table de correspondance entre les actes du langage et les rôles conversationnels de chaque personnage et le comportement non-verbal joué par les acteurs. Cette table a ensuite été implémentée sous la forme de règles. Par exemple, un personnage dominant avec un autre va lui poser une question en se tenant droit et en regardant l'autre droit dans les yeux. Ce travail est intéressant car il intègre les autres rôles conversationnels que celui du locuteur. Les auteurs ont cherché à modéliser également le comportement correspondant aux rôles d'auditeurs et de spectateurs d'une conversation. Cependant, leur modèle est basé sur un unique scénario et sa méthode de conception ne permet pas de l'étendre.

Dans [Cafaro et al., 2012], les auteurs ont étudié la perception qu'ont des utilisateurs d'un agent capable de produire différents comportements de salutations. Dans leur étude, les participants représenté par un avatar avançaient (à la première ou à la troisième personne) vers l'agent et s'arrêtaient en arrivant dans sa zone personnelle (selon les Proxemics de Hall [Hall, 1969]). Cette phase d'approche durait 12,5 secondes. Les auteurs manipulaient la capacité de l'agent de produire des regards, des sourires et de faire un pas vers le participant pendant cette phase d'approche, afin d'identifier quand l'agent est perçu comme extraverti et agréable. Les différents comportements ont été conçus à partir des travaux de Kendon [Kendon, 1990]. Leurs résultats ont montré qu'un agent produisant plus de sourire et de regard est perçu comme plus amical et qu'un agent qui se rapproche est perçu comme plus extraverti.

Dans [Chollet et al., 2014], les auteurs ont réalisé l'annotation d'un corpus de vidéos de simulation d'entretiens d'embauches entre un recruteur et un jeune afin de construire un modèle computationnel de génération de comportements non-verbaux exprimant des attitudes représentées par le modèle à deux dimensions de Statut

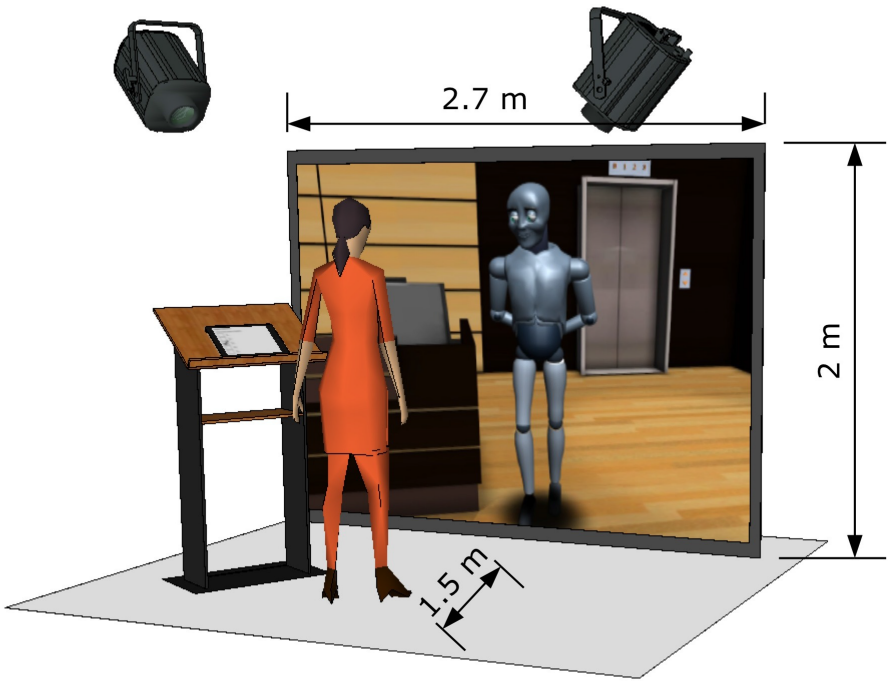


FIGURE 3.3 : Schéma du système dans le travail sur les premières impressions de [Cafaro et al., 2012]

et d'Affiliation. Le modèle proposé génère des séquences de comportements non-verbaux exprimant les intentions communicatives de l'agent, puis, grâce à un réseau Bayésien dynamique construit à partir des annotations du corpus, le modèle insère des signaux supplémentaires afin d'exprimer les attitudes. Une étude a été menée qui a montré la capacité du système à exprimer l'amicalité, l'hostilité et la dominance mais pas la soumission (voir figure 3.4).

Certains de ces travaux ont cherché à étudier l'expression d'attitudes dans le comportement non-verbal et d'autres ont cherché à proposer un modèle pour générer le comportement non-verbal en fonction de ces attitudes. Nous identifions certaines limites que nous souhaitons adresser dans notre travail. Dans les travaux de Bee et ses collègues [Bee et al., 2009][Bee et al., 2010], les auteurs ont étudié la perception de l'attitude dominante seulement et se sont limités, pour le comportement non-verbal, au buste d'un agent avec une apparence spécifique. Dans [Cafaro et al., 2012], les auteurs ont étudié la perception d'attitudes agréables et extraverties dans les premières secondes d'une interaction seulement. Dans [Lee and Marsella, 2011], le modèle de comportement n'a été conçu que pour un scénario unique et avec un

FIGURE 3.4 : L'installation pour l'évaluation du recruteur virtuel de [Chollet et al., 2014]

groupe d'acteurs qui ont dû livrer leur interprétation subjective des attitudes. Dans [Chollet et al., 2014], le modèle a été appris à partir d'un corpus annoté d'interactions humaines dans un contexte d'entretiens d'embauche.

Dans notre modèle de comportement des attitudes sociales, nous proposons d'adresser ces limites. Le modèle que nous proposons est en partie construit sur un corpus de descriptions de comportements d'agents virtuels. Nous prenons en compte les deux dimensions Statut et Affiliation. Nous considérons un ensemble de modalités incluant tout le haut du corps. Enfin, le modèle est conçu pour être utilisable dans différents contextes.

Dans la section suivante, nous présentons des travaux sur la modélisation des tours de parole.

3.2 LES MODÈLES DE TOURS DE PAROLE

Les attitudes d'un agent s'expriment envers un interlocuteur. Elles sont donc exprimées lorsque l'agent est en interaction. Or, les attitudes peuvent avoir une influence sur les comportements de prises de parole comme vu dans le chapitre 2.

Dans cette section, nous allons nous intéresser aux travaux dans le domaine des agents virtuels qui ont cherché à modéliser des mécanismes de tours de parole et nous verrons également comment différentes stratégies de prises de parole ont été implémentées.

Dans [Ter Maat et al., 2010], les auteurs ont fait varier les stratégies de tours de parole d'un agent en utilisant une installation en Magicien d'Oz et ont mené une étude afin de mesurer la perception qu'ont des utilisateurs des attitudes de l'agent. Les différentes stratégies correspondaient à différentes façons de prendre le tour de parole à la fin de celui de l'utilisateur. Soit les auteurs faisaient prendre à l'agent le tour de parole un peu avant la fin de celui de l'utilisateur, soit au moment exact où l'utilisateur avait fini, soit après une pause. Leurs résultats ont montré que l'agent qui prend le tour en avance, pouvant interrompre l'utilisateur, est perçu comme moins agréable et plus assuré alors que l'agent qui laisse une pause est perçu comme plus agréable et moins dominant. Ce travail, bien que ne proposant pas de modèle de tours de parole, permet toutefois de voir que varier faire varier ces comportements par un agent lui permet d'exprimer différentes attitudes.

Dans les travaux de Raux et al. [Raux and Eskenazi, 2009], les auteurs ont proposé un modèle de tour de parole entre un utilisateur et un système de dialogue qui fonctionne suivant une machine à états. Les différents états représentent qui a la parole (3 états : soit l'agent, soit l'utilisateur, soit les deux) et les transitions représentent des actions pour passer d'un état à l'autre (ex. l'utilisateur prend la parole, l'agent rend la parole etc.). Chacune de ces actions se voit attribuer un coût en suivant le principe émit par Sacks dans [Sacks et al., 1974] : "chaque participant cherche à minimiser les blancs entre les tours de parole et les superpositions". Par exemple, les actions qui amènent le système à parler en même temps que l'utilisateur (comme prendre la parole alors que l'utilisateur est en train de parler) ont un coût plus élevé que les actions permettant un seul interlocuteur à la fois (comme s'interrompre si l'utilisateur se met à parler). Ce système ne permet cependant d'obtenir qu'un profil de comportement et ne permet de représenter que des conversations dyadiques avec l'utilisateur.

Dans [Traum and Rickel, 2002], Traum et Rickel ont proposé d'étendre le système de dialogue présent dans le projet *Mission Rehearsal Exercise* (MRE) [Hill et al., 2001]. Pour cela, ils ont proposé une représentation théorique de la conversation très

riche, se décomposant en plusieurs niveaux (prendre contact, demander l'attention et converser par exemple), identifiant ainsi les différentes actions qui interviennent dans une conversation. Ces actions sont par exemple *demander le tour de parole* ou *prendre le tour de parole*. Les auteurs ont également défini les signaux non-verbaux associés à ces actions de gestion des tours de parole suivant les travaux de Duncan [Duncan, 1972] et des travaux précédents dans la communauté des agents virtuels. Leur représentation théorique permet de considérer les conversations multi-agents. Chaque agent doit maintenir l'état de la conversation en mémoire ainsi que la liste des participants et l'identité de celui qui a le tour de parole. Cependant, ce modèle n'explicite pas comment les différents actions se déclenchent. D'autre part, il est très ciblé pour la réalisation du projet MRE.

Leßmann et al. [Leßmann et al., 2004] ont proposé d'ajouter à l'agent MAX un système de tours de parole basé également sur des états et des transitions (figure 3.5). Ce système permet à l'agent Max d'évoluer entre 4 états différents, Max parle, L'utilisateur parle, Max et l'agent parle en même temps et aucun des deux ne parlent. Un ensemble de règles permet à l'agent de passer d'un état à l'autre en fonction de l'action conversationnelle de l'utilisateur et des objectifs de l'agent. Ce modèle diffère des autres car il choisit de s'éloigner du formalisme de Sacks [Sacks et al., 1974] en permettant à l'agent et à l'utilisateur de parler en même temps ou de laisser des blancs entre les tours. Cependant il ne prend en compte qu'une conversation dyadique avec l'utilisateur.

Un des travaux les plus complets dans le domaine des agents virtuels et des tours de parole est le modèle YTTM proposé par Thórisson et ses collègues [Thórisson et al., 2010]. Bien que ce travail ait commencé par un modèle pour conversation dyadique seulement [Thórisson, 2002], il a été étendu pour prendre en compte les conversations multi-participants (voir figure 3.6). Ce modèle est également basé sur le modèle de tours de parole de Sacks [Sacks et al., 1974] et sur les travaux de Duncan sur les comportements de gestion du tour de parole [Duncan, 1972]. Un des points forts de ce modèle est qu'il s'agit d'un modèle qui supervise un seul agent. C'est-à-dire que ce sont les comportements générés individuellement par chaque agent du groupe qui vont produire l'émergence des tours de parole. Dans ce modèle, un agent possède plusieurs contextes qui s'activent en fonction de ce qu'il perçoit (un autre agent parle par exemple) et de ses buts (prendre la parole par exemple). A

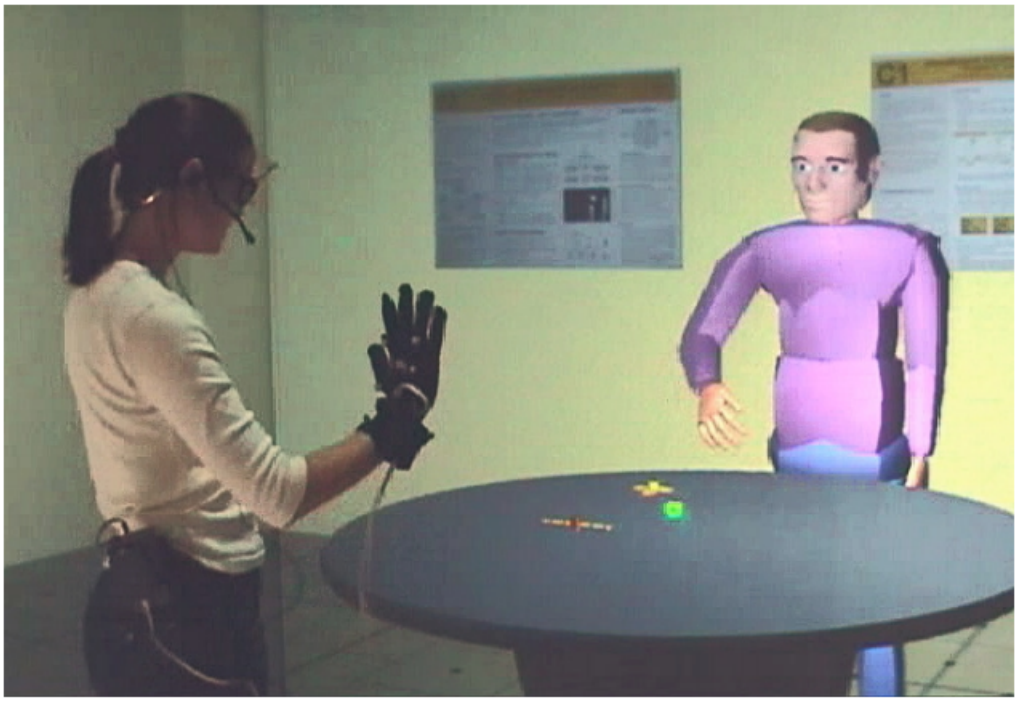


FIGURE 3.5 : L'agent Max, dans un environnement CAVE [Leßmann et al., 2004]

chaque contexte correspond un ensemble de comportements permettant de simuler la conversation. Par exemple, il existe un contexte appelé *je-souhaite-prendre-la-parole* qui déclenche un mouvement de main pour signaler le désir de l'agent de prendre la parole lorsqu'il est activé. Dans ce modèle, il est également possible de faire varier la rapidité avec laquelle l'agent souhaite reprendre la parole, permettant de définir différents profils d'agent. Cependant, ce système cherche à minimiser la superposition de tours de parole or ceci est nécessaire pour l'expression d'attitudes.

Nous avons vu dans cette section que les systèmes gérant les tours de parole proposés dans la communauté favorisent une représentation des mécanismes de tours de parole à l'aide d'états et de transitions. La plupart de ces systèmes suivent d'ailleurs les recommandations de Sacks [Sacks et al., 1974] pour gérer les transitions entre les états afin de s'assurer qu'il n'y ait pas plusieurs locuteurs en même temps comme [Raux and Eskenazi, 2009] et [Leßmann et al., 2004]. Par conséquent, ils ne permettent pas d'exprimer différentes attitudes à travers les stratégies des tours de parole puisque l'expression de certaines attitudes peuvent amener plusieurs personnes à parler en même temps (voir Chapitre 2). De plus, peu de ces systèmes ne

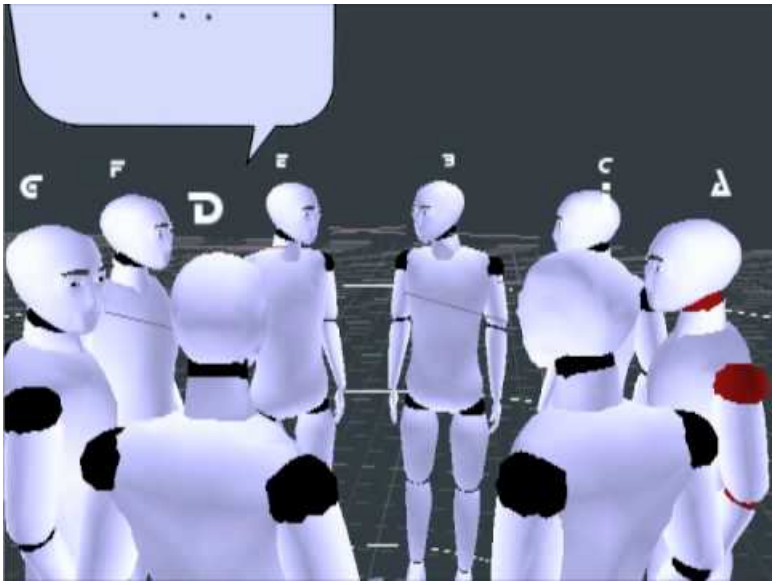


FIGURE 3.6 : Les agents du système YTTM de [Thórisson et al., 2010]

permettent de considérer une interaction de groupe, excepté [Thórisson et al., 2010].

Nous proposons d'adresser ces deux limites dans notre modèle. Plutôt que de considérer l'approche de Sacks, nous allons baser notre représentation des tours de parole sur la description des actions conjointes de Clark [Clark, 1996]. Notre modèle se concentrera donc sur la réalisation des actions de parler et d'écouter sans chercher à minimiser les blancs entre les tours ou les superpositions de locuteurs. De plus, nous prendrons en compte les attitudes sociales pour faire varier la durée des tours de parole et permettre aux agents de s'interrompre et de prendre le tour de parole avant que le locuteur ait fini de parler. Nous allons également prendre en compte des interactions de groupe dans notre modèle.

Dans la section suivante, nous présentons les travaux qui se sont intéressés aux groupes d'agents.

3.3 LES GROUPES D'AGENTS CONVERSATIONNELS

Notre travail a pour objectif de proposer un modèle d'expression d'attitudes pour agents conversationnels membre d'un groupe. Notre objectif principal est l'expression d'attitude et celle-ci peut s'inscrire dans les comportements produits par les agents pour maintenir leur formation de groupe. C'est pourquoi dans cette section

nous nous intéressons aux travaux dans le domaine de la modélisation de groupes afin d'y identifier les caractéristiques importantes, qu'ils incluent l'expression d'attitudes ou non. Certains chercheurs s'intéressent également à la modélisation de foules [Thalmann, 2007] mais nous ne considérons pas ces travaux ici. Nous allons présenter des travaux qui ont cherché à représenter des petits groupes tels que définis dans le chapitre 2.

Dans [Jan and Traum, 2005], les auteurs ont proposé un modèle pour la simulation de groupes conversationnels permettant de peupler le décor de l'environnement du système *Mission Rehearsal Exercise* (MRE) [Hill et al., 2001]. Ce modèle permet aux agents de produire le comportement non-verbal imitant des échanges de tours de parole entre les membres. Chaque agent est paramétrable pour déterminer s'il souhaite prendre la parole souvent ou s'il est susceptible d'être interrompu par exemple. Ces mêmes auteurs ont dans [Jan and Traum, 2007] proposé un modèle pour la formation de groupes se basant sur les travaux de Kendon [Kendon, 1990] et de Hall [Hall, 1969]. Ils ont conçu un ensemble de règles mécaniques qui calculent les forces d'attraction entre les participants d'une même conversation en fonction de leurs rôles conversationnels pour les positionner dans le groupe. Par exemple, lorsqu'un agent se met à écouter un autre agent, il va chercher à se rapprocher comme pour mieux l'écouter. Leurs travaux, bien que permettant de configurer différents profils d'agent à travers les distances possibles entre les agents, n'ont pas cherché pas à modéliser l'influence de l'attitude sociale sur les comportements produits.

Dans [Ballin et al., 2004], les auteurs expliquent comment ils ont construit le modèle permettant de faire varier les postures et l'orientation des personnages du système Demeanour (voir section 3.1.2) en fonction des attitudes. Tout d'abord une attitude est calculée à partir de la personnalité de l'agent et du comportement des autres agents. Ensuite, à chaque attitude correspond une posture particulière et la valeur de l'attitude est utilisée pour déterminer l'intensité de la posture. Dans ce travail, il n'y a pas de gestion de la distance entre les personnages. Les personnages ne s'éloignent pas ou ne se rapprochent pas les uns des autres. Cependant, dès lors qu'une attitude est déterminée pour un personnage envers un autre, les comportements possibles pour cette attitude peuvent inclure des postures telles que se pencher en avant (comme pour montrer son intérêt) ou au contraire se pencher en arrière (comme pour montrer son aversion) en plus de s'orienter ou non directement vers

FIGURE 3.7 : Le système de groupe de [Jan and Traum, 2007]

l'autre. L'attitude est utilisée pour régler l'intensité de ces comportements. Cependant, l'attitude n'influence que la posture et la taille du groupe gérée par Demeanour est restreinte.

Dans les travaux de Prada et Paiva [Prada and Paiva, 2005], les auteurs ont conçu un système multi-agents dans lequel un utilisateur prend part avec un groupe de personnages virtuels à un jeu de résolution de problèmes (voir section 3.1.1). Les auteurs ne s'intéressent pas au comportement non-verbal, uniquement à la prise de décisions, cependant il est intéressant de regarder ce travail car ils proposent une modélisation du groupe basée sur la formalisation de McGrath [McGrath, 1984] qui leur permet de prendre en compte les relations sociales entre les membres au cœur de leur formalisation. Chaque agent possède une relation avec chaque autre agent. Il se voit également attribué d'une position sociale dans le groupe, une importance, en fonction des relations qu'il a avec les autres et des actions qu'il a accomplies. Le système de prise de décisions proposé par les auteurs se présente comme un ensemble de règles logiques qui prennent en compte cette dimension sociale dans le déclenchement des actions. Par exemple, pour qu'un agent puisse initier une interaction avec le groupe, celui-ci doit montrer un assez haut niveau d'extraversion ou

alors une bonne position sociale dans le groupe. Ce travail offre une représentation intéressante du groupe et de la dynamique sociale entre les membres. Cependant, les auteurs n'ont pas cherché à modéliser le comportement non-verbal des agents. Les travaux de Rehm et Endrass [Rehm and Endrass, 2009] s'inspirent des travaux de [Prada and Paiva, 2005] pour proposer un système permettant de simuler également les dynamiques des relations sociales au sein d'un groupe d'agents. Pour cela, ils ont proposé une correspondance entre les interactions sociales de Bales [Bales, 1950] (voir chapitre 2) et les relations sociales ainsi que la personnalité nécessaires pour pouvoir réaliser ces interactions. Par exemple, une personne agréable qui a une relation amicale avec une autre a plus de chance de montrer son soutien ou d'être d'accord avec l'autre. Ils ont ensuite implémenté un ensemble de règles s'inspirant des travaux sur la *Congruity* [Osgood and Tannenbaum, 1955] pour mettre à jour les valeurs des relations sociales d'un agent en fonction des actions des autres. Un des points intéressants de leur approche est que les auteurs ont pris en compte la distance entre les agents, se basant sur les travaux de [Kendon, 1990] et de Hall [Hall, 1969] pour calculer la force de ces influences. Bien que ces travaux soient intéressants du point de vue de la modélisation et de la représentation de la dynamique sociale, ils ne se sont pas intéressés à l'influence exercée sur les comportements non-verbaux.

Pedica et Vilhjálmsson [Pedica and Högni Vilhjálmsson, 2010] ont implémenté un modèle qui arrange les agents virtuels dans des formations en se basant sur les travaux de Kendon [Kendon, 1990], de Scheflen [Scheflen and Ashcraft, 1976] et de Hall [Hall, 1969]. Les agents maintiennent continuellement leur formation, réajustant leur distance et orientation en cas d'arrivée ou de départ de membres dans le groupe. Leurs comportements incluent également des regards lancés aux nouveaux arrivants dans le groupe. Leur approche mélange un système de règles et un système de forces mécaniques. Le système de forces permet de maintenir la formation de groupe en cherchant à respecter un certain nombre de contraintes (comme rester à une certaine distance d d'un autre participant) de façon continue. Le système de règles permet de faire varier les contraintes à intervalles réguliers (par exemple, augmenter la distance d avec un autre agent s'il s'approche trop). Leur modèle a été évalué avec des conversations simulées et leurs résultats ont montré que l'utilisation de ces comportements de cohésion de groupe améliore la crédibilité du groupe [Pedica et al., 2010]. Cependant, les auteurs ne sont pas intéressés à la

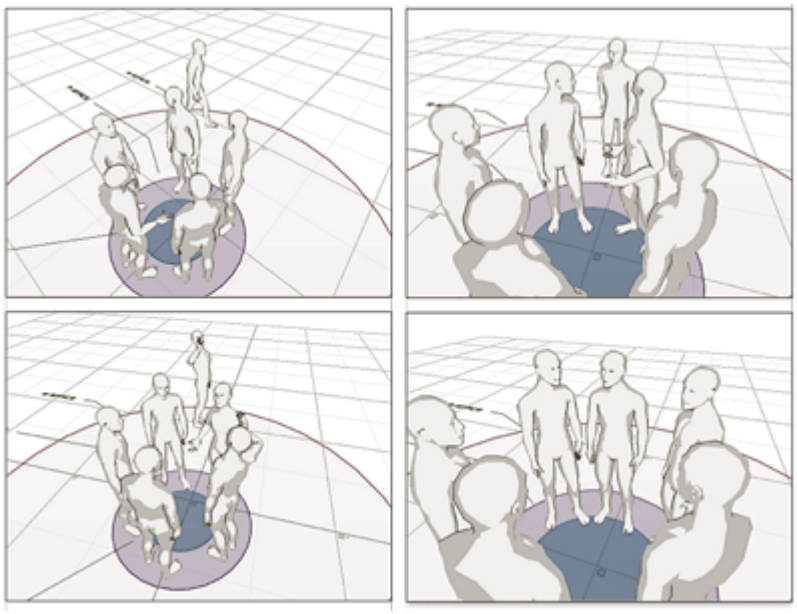


FIGURE 3.8 : Le système de formation de groupe Impulsion de [Pedica and Högni Vilhjálmsson, 2010]

définition de différents profils pour chaque personnage et il n'y a aucune influence de l'attitude sur le comportement des agents.

Lee et Marsella dans [Lee and Marsella, 2011] ont cherché à concevoir le comportement non-verbal de différents personnages d'une application scénarisée immersive (voir section 3.1.2). Bien que leur approche ait permis d'étudier les différences de comportement suivant à la fois les relations sociales et les rôles dans la conversation (locuteur, interlocuteur et observateur), leur étude ne s'est pas intéressée à la dynamique du groupe et à la formation de celui-ci.

Enfin, Damian et ses collègues [Damian et al., 2011] ont développé un système permettant de configurer le comportement d'agents au sein d'un groupe. Ces comportements incluent différentes distances, différentes orientations, différentes durées de regard et différentes valeurs pour les paramètres d'expressivité des gestes tels que l'amplitude, la fluidité et la vitesse. Contrairement aux travaux de Pedica et Vilhjálmsson [Pedica and Högni Vilhjálmsson, 2010], ce système ne calcule pas de façon continue la position et l'orientation de chaque membre mais uniquement à la réception d'évènements, comme un changement dans le nombre de membres du groupe par exemple. Les auteurs ont ensuite mené une étude afin de vérifier la capacité de

leur système à produire des comportements correspondant à de l'extraversion et au genre de l'agent en se basant sur la littérature. Leurs résultats ont montré que le niveau d'extraversion des agents était bien reconnu quel que soit le genre de l'agent. Ce modèle permet cependant de configurer uniquement un seul profil de comportement à travers la valeur d'extraversion et ne permet pas de montrer des comportements différents suivant l'interlocuteur de l'agent.

Peu de ces travaux se sont intéressés à l'expression d'attitudes dans le comportement non-verbal relatif à la formation de groupes. Les travaux de [Prada and Paiva, 2005] et [Rehm and Endrass, 2009] sont intéressants pour construire une modélisation qui intègre la dimension sociale au cœur de celle-ci mais ils ne s'intéressent pas au comportement non-verbal. Les travaux de [Pedica and Högni Vilhjálmsson, 2010] et de [Jan and Traum, 2007] se sont intéressés à la dynamique des comportements de formation de groupe mais ne se sont pas intéressés à la modélisation de différents profils ou de l'expression d'attitudes. Les travaux de [Ballin et al., 2004] et [Lee and Marsella, 2011] se sont intéressés à l'expression d'attitude mais les paramètres du comportement considérés sont restreints et dans le cas de [Lee and Marsella, 2011], leur résultat ne concerne qu'un scénario. Le travail de [Damian et al., 2011] est le plus intéressant cependant il n'a considéré qu'une dimension d'extraversion et ne permet pas de configurer différentes attitudes envers chaque autre agent, il s'agit plus de l'expression d'une personnalité globale. Dans notre travail, nous proposons d'adresser ces limites avec un modèle permettant de prendre en compte différentes attitudes pour chaque autre agent, représentées sur deux dimensions, afin d'influencer différentes modalités du comportement non-verbal de formation du groupe.

3.4 Conclusion

Dans ce chapitre, nous avons vu différents travaux qui se sont intéressés à la modélisation de l'expression d'attitudes, de la conversation et de la formation de groupes. L'expression d'attitudes a été peu étudié dans le contexte des groupes et de la conversation. Nous proposons un modèle pour adresser ces limites à travers les travaux de cette thèse. L'objectif de notre modèle est de permettre à des agents d'exprimer des attitudes sociales à travers leur comportement dans un contexte de conversation et ce quel que soit le contenu de la conversation. C'est pourquoi nous ne considérons pas le contenu verbal dans ce travail. Pour cela, nous allons utiliser un

plus grand nombre de modalités que les travaux existants, permettant de combiner l'expression d'attitudes dans les comportements accompagnant la conversation avec l'expression d'attitudes dans les comportements de formation de groupe. De plus, nous proposons également de réguler la dynamique de la conversation des agents en prenant en compte les attitudes dans leurs stratégies de prises de parole. Pour cela nous adoptons un formalisme dans la représentation des tours de parole qui s'éloigne des travaux existants afin de prendre en compte la superposition des locuteurs, indispensable dans l'expression d'attitudes.

Dans le chapitre suivant, nous présentons la première partie de notre travail qui a consisté à concevoir un modèle d'expression d'attitudes pour un agent en interaction avec un utilisateur.

Résumé du chapitre

- Nous avons vu les différents travaux qui ont cherché à faire exprimer des attitudes à des agents. Une des limites majeures est le nombre de modalités considérées qui reste restreint.
- Les travaux existants liés à la conversation et à la prise de parole se basent sur un modèle théorique qui ne favorise pas la superposition de tours de parole, ne permettant pas l'expression d'attitudes sociales pour générer différentes stratégies.
- Les travaux liés à la formation de groupe prennent rarement en compte l'expression d'attitudes ou ne permettent pas de prendre en compte différentes attitudes pour chaque interlocuteur.

4

Attitudes dans une conversation dyadique

Plusieurs travaux de recherches [Bates, 1994][Reilly, 1996] mettent en évidence le fait qu'un agent est perçu plus crédible dès lors qu'il peut exprimer des comportements sociaux. Ces comportements peuvent être l'expression d'émotions, de différentes intentions communicatives ou encore d'attitudes. Dans notre travail, nous nous intéressons plus particulièrement à l'expression d'attitudes. Contrairement aux émotions [Dias and Paiva, 2005][Gebhard, 2005][Adam et al., 2009][Laird, 2012] et à la personnalité [Rousseau et al., 1996][Li et al., 2007][Romano and Wong, 2004][Read et al., 2007], les modèles computationnels d'attitudes sociales ont été peu explorés (voir Chapitre 3).

Or une personne peut exprimer des attitudes à travers ce qu'elle dit mais également à travers ce qu'elle fait [Argyle, 1988]. Plus précisément, lorsqu'une personne parle, elle peut transmettre le même message de différentes façons, en adoptant différentes attitudes, en fonction du contexte de la conversation. Par exemple, une personne peut utiliser une attitude amicale avec une personne proche et saluer chaleureusement un ami en lui souriant, lui serrant fermement la main et en lui parlant fort alors que cette même personne peut utiliser une attitude plus dominante avec un employé en ayant une expression faciale plus neutre et en gardant plus de distance. L'intention est la même, saluer, et pourtant le comportement non-verbal est très différent. Ce sont ces variations que nous souhaitons comprendre et implémenter au sein d'un agent conversationnel à travers le travail décrit dans ce chapitre.

Afin d'enrichir la gamme de comportement des agents et de leur donner plus de crédibilité, l'objectif de nos travaux de recherches est de développer un modèle computationnel qui permettra de calculer le comportement non-verbal d'un agent en fonction de l'attitude qu'on veut lui faire exprimer. Dans une conversation, une personne exprime une attitude lorsqu'elle parle mais aussi lorsqu'elle écoute. Ce chapitre décrit la première partie de notre travail, dans laquelle nous nous sommes

concentrés sur l'expression d'attitude du locuteur dans une conversation dyadique. Le modèle que nous proposons doit faire face à différents défis. Ce modèle doit pouvoir fonctionner quel que soit le genre et l'apparence de l'agent, il doit pouvoir combiner l'expression d'attitudes avec la réalisation de l'intention communicative de l'agent, il doit pouvoir gérer la dimension multimodale de l'expression d'attitudes et les attitudes exprimées par les agents équipées du modèle doivent être reconnues par les utilisateurs.

Dans une première section, nous expliquons la méthodologie que nous avons utilisée dans le but de construire ce modèle computationnel ainsi que les paramètres du comportement non-verbal de l'agent que nous manipulons. Dans la section suivante, nous présentons l'étude que nous avons réalisée afin de collecter des données nécessaires à l'apprentissage de notre modèle et nous discutons les résultats. Nous présentons ensuite le modèle computationnel que nous avons construit à partir des données de l'étude. Enfin, une implémentation de ce modèle, une évaluation perceptive expérimentale de celui-ci ainsi que les résultats sont discutés dans une dernière section.

4.1 Le choix d'une méthodologie pour la création d'un modèle computationnel

Bien qu'il existe des scénarios où les comportements des personnages virtuels soient contrôlés entièrement par un humain (i.e. des avatars), un des enjeux dans le domaine des agents virtuels intelligents est de leur donner la capacité de réagir et de produire des comportements de façon *autonome*. L'autonomie implique de concevoir des mécanismes de déclenchement de ces comportements. Cela est rendu possible par le développement de modèles computationnels qui vont être utilisés pour calculer les comportements d'un agent en fonction d'un ensemble de paramètres. Différents types de modèles existent. Certains produisent de la parole en fonction de ce qui a été dit précédemment [McTear, 2004], d'autres calculent l'état émotionnel d'un agent virtuel [Dias and Paiva, 2005] et d'autres encore sont utilisés pour déterminer la prochaine action à réaliser [Nouri and Traum, 2014]. Dans notre travail, l'objectif est de développer un modèle permettant de calculer les signaux non-verbaux à utiliser afin d'exprimer une attitude. Pour construire ce type de modèle, différentes approches ont été proposées.

4.1.1 Différentes approches

La première approche consiste à créer un ensemble de règles comportementales à partir de la littérature sur le comportement humain. Ces règles, de la forme SI ALORS, décrivent l'impact de paramètres (tels que l'état interne de l'agent) sur le déclenchement de réactions (par exemple : SI amical ALORS sourire). L'agent Laura [Bickmore and Picard, 2005] utilise ce type d'approche. Le système se souvient du nombre de fois qu'un utilisateur a interagi avec Laura. Plus ce nombre est élevé, plus Laura montre des signes amicaux avec l'utilisateur. Elle se tient plus proche de l'écran, fait plus de sourires ou encore plus de gestes (voir chapitre 3). Cette approche à base de règles a l'avantage de permettre un développement flexible du modèle. Il est possible d'introduire de nouvelles règles sans remanier tout le corpus de règles. On peut également corriger des règles existantes. Cependant cette approche peut conduire à un modèle très déterministe qui manque de variabilité. Si une règle peut être appliquée, elle le sera. Par conséquent, si une même séquence devait être rejouée elle le serait de la même façon. De plus, lorsque le modèle contient peu de règles, il est aisément compréhensible, mais dès lors que le nombre de règles devient important, il perd en lisibilité, ce qui peut amener à l'introduction de règles contradictoires.

Une autre approche consiste à utiliser des méthodes d'apprentissage sur un corpus de données afin d'en extraire le modèle ; par exemple, à partir d'un corpus de vidéos d'interactions. Les vidéos sont annotées suivant un schéma d'annotation et les données de l'annotation sont ensuite utilisées pour apprendre le modèle. Cette méthode est utilisée dans [Buisine et al., 2006] notamment où les auteurs comparent la perception qu'ont des utilisateurs entre des émotions exprimées dans des vidéos et des émotions exprimées par un agent reproduisant ces mêmes émotions grâce à un modèle construit sur des annotations provenant de ces mêmes vidéos. Les auteurs se sont basés sur un corpus existant mais il est également possible de créer ses propres vidéos. Comme dans [Chollet et al., 2014] où les auteurs ont participé à l'enregistrement de simulations d'entretiens d'embauche qu'ils ont annotées par la suite. Les annotations ont ensuite été fournies à un algorithme d'apprentissage qui a produit un réseau Bayésien décrivant les probabilités d'observer certaines séquences de comportements en fonction de l'attitude d'un agent. Ces approches permettent d'obtenir des modèles qui proposent de la variabilité dans les comportements gé-

nérés. En utilisant ces approches probabilistes, on ajoute une part d'aléatoire dans la sélection des comportements. Elles offrent également un grand contrôle sur les paramètres puisqu'il est possible de choisir le schéma d'annotation que l'on souhaite. Cependant, le travail de sélection du schéma adéquat peut être difficile afin de s'assurer que celui retenu convienne au problème que l'on cherche à représenter. Ces approches sont aussi couteuses en temps et en ressources. Afin d'obtenir des résultats pertinents et un modèle performant, il est nécessaire de collecter un assez grand nombre de données à fournir à l'algorithme d'apprentissage. Dans le cas d'une collecte de données par annotation de vidéos, le travail de codage minutieux rend la tâche fastidieuse pour une seule personne et il faut alors pouvoir obtenir l'aide d'autres personnes. Une autre solution pour la collecte de données consiste à déléguer la création et l'annotation du corpus aux utilisateurs d'une plateforme en ligne. Cette technique, de plus en plus populaire pour collecter des données, est connue sous le nom de *Crowdsourcing* [Yuen et al., 2011].

4.1.2 Crowdsourcing

La méthode de *Crowdsourcing* permet le contrôle des paramètres de l'annotation, mais elle permet également de collecter aisément un grand nombre de données. Cette technique permet de distribuer, grâce à Internet, une tâche à un grand nombre d'utilisateurs grâce à des plateformes accessibles en ligne. Une des plateformes la plus connue est Amazon Mechanical Turk qui propose un système de rémunération (très peu élevée) aux participants d'une tâche [Mason and Suri, 2012]. Il est également possible d'utiliser des plateformes gratuites ou de concevoir sa propre plateforme. C'est notamment l'approche qui a été utilisée dans [Ochs et al., 2012] pour construire un modèle de génération de différentes types de sourires pour un agent virtuel. La tâche demandée aux utilisateurs était de configurer les sourires d'un agent virtuel en manipulant des paramètres de l'expression de sourires. Nous avons choisi d'utiliser cette approche pour notre collecte de données. Elle permet de contrôler les paramètres que nous souhaitons utiliser et permet également de collecter un grand nombre de données. De plus, nous allons pouvoir demander aux utilisateurs de configurer directement le comportement non-verbal de l'agent virtuel au lieu d'essayer de faire correspondre des comportements humains.

Dans les sections suivantes, nous présenterons la plateforme en ligne que nous

avons conçu afin de réaliser cette collecte de type *Crowdsourcing*.

4.2 L'APPLICATION DE COLLECTE DE CORPUS GENATTITUDE

Dans une perspective méthodologique centrée sur l'utilisateur, un corpus de données créé par des utilisateurs sur des agents exprimant des attitudes sociales a été collecté grâce à l'approche dite de *Crowdsourcing*. Pour ce faire, nous avons développé l'application *GenAttitude*. *GenAttitude* est une application accessible en ligne depuis un navigateur Internet. Son interface permet aux utilisateurs de configurer les animations d'un agent virtuel en manipulant les valeurs de certains paramètres de son comportement non-verbal. La tâche demandée aux utilisateurs était de configurer le comportement de l'agent afin que celui-ci exprime une attitude sociale donnée. Un des avantages de cette approche est de permettre la collecte d'un corpus d'animations d'agents et non de comportements humains. Dans les systèmes qui se basent sur l'annotation de vidéos d'interactions, le modèle de comportement pour un agent est appris sur des données correspondant à des humains. Puis, nous apprenons le modèle de génération de comportements de l'agent à partir d'un corpus de comportements d'agents. Les animations générées par le modèle seront ainsi similaires aux animations qui auront été configurées par les utilisateurs. Un travail d'annotation manuelle peut se révéler laborieux pour une personne ou couteux en annotateurs. En déployant *GenAttitude* sur Internet, nous permettant de déléguer la création du corpus, nous pouvons recruter en peu de temps un grand nombre de participants et donc un large corpus d'animations.

Dans la section suivante, nous présentons les paramètres que nous manipulons à travers l'application GenAttitude.

4.3 LES PARAMÈTRES DE L'APPLICATION GENATTITUDE

Pour développer notre propre plate-forme de collecte de données nous avons dû choisir les paramètres à considérer. Tout d'abord, nous souhaitons obtenir un modèle permettant de générer le comportement non-verbal d'un agent qu'il soit masculin ou féminin. La littérature a montré qu'il peut y avoir des différences de comportements entre les hommes et les femmes dans l'expression d'attitudes [Briton and Hall, 1995][Kennedy and Camden, 1983]. Nous considérons également que l'expres-

sion d'attitudes par un locuteur est réalisée en combinaison avec un acte illocutoire, un acte du langage, puisque certains travaux ont souligné l'importance du lien qui peut exister entre les deux [Siebel, 2003]. Dans ce cas, comment savoir si une variation de comportement est due à l'attitude ou à une intention communicative ? Par exemple, si je pose une question à quelqu'un en le pointant du doigt, est-ce lié à ce que je suis en train de dire (par exemple j'indique que je fais référence à l'autre, à ce qu'il représente) ou essaie-je de montrer de la dominance ? Les situations que nous présentons alors aux utilisateurs à travers *GenAttitude* ne sont pas juste paramétrées par l'attitude demandée, mais également par un acte illocutoire ainsi que par le genre de l'agent (masculin ou féminin). Dans cette section, nous présentons l'ensemble de nos paramètres ainsi que leurs valeurs possibles. Nous commençons par les paramètres des situations, décrivant le contexte dans lequel un agent doit exprimer une attitude (son genre, attitude sociale et acte illocutoire), puis nous décrirons les paramètres du comportement non-verbal, ceux qui permettent l'expression de ces attitudes et qui seront manipulables par les utilisateurs.

4.3.1 Attitudes sociales

Comme indiqué dans le chapitre 2, la représentation des attitudes que nous adoptons est fondée sur celle d'Argyle [Argyle, 1988]. Dans cette représentation les attitudes sont constituées de deux dimensions. Une dimension de Statut et une dimension d'Affiliation. Dans la dimension de Statut, les valeurs vont de *Soumis* à *Dominant* et dans la dimension d'Affiliation les valeurs vont de *Hostile* à *Amical*. Comme attitude possible pour le système *GenAttitude*, les quatre valeurs extrêmes des dimensions ont été considérées. Nous nous sommes limités à ces quatre attitudes pour plusieurs raisons. Premièrement, il est difficile d'estimer comment ces dimensions fonctionnent en interaction (par exemple, une attitude dominante et amicale). Il est également difficile d'estimer les différentes intensités des attitudes (par exemple comment faire la différence entre une attitude légèrement dominante et une attitude très dominante). Enfin, Leary [Leary, 1958], dans sa description de l'Interpersonal Circumplex, explique qu'une attitude est représentée par un point sur la périphérie du cercle. Cela implique par exemple que lorsque l'attitude est complètement dominante, elle ne peut contenir ni de l'hostilité, ni de l'amicalité. C'est pourquoi dans *GenAttitude* l'agent pourra exprimer soit une attitude dominante, soit une attitude

soumise, soit une attitude amicale, soit une attitude hostile.

4.3.2 Actes illocutoires

Nous considérons dans notre travail qu'une attitude est exprimée en accompagnement d'un acte illocutoire. Comme expliqué précédemment, des chercheurs se sont penchés sur l'importance de l'expression d'attitude dans la production d'actes illocutoires [Siebel, 2003]. Les actes illocutoires sont une catégorie d'actes du langage. Un acte du langage peut être défini comme une action d'un individu utilisant ses mots pour provoquer un changement dans son environnement. C'est une notion qui a d'abord été introduite par John Austin [Austin, 1975]. Austin a proposé 3 grandes classes d'actes du langage que nous rappelons brièvement ici :

1. **Les actes locutoires**, qui correspondent au simple fait de prononcer une phrase grammaticalement correcte (par exemple, la phrase "Pouvez-vous m'ouvrir la porte s'il vous plait ?").
2. **Les actes illocutoires**, qui nous intéressent ici et qui correspondent au sens de ce qui est dit (par exemple dans la phrase "Pouvez-vous m'ouvrir la porte s'il vous plait ?", je demande à ce qu'on m'ouvre la porte).
3. **Les actes perlocutoires**, qui correspondent aux effets que l'acte illocutoire a sur les croyances et les buts de celui à qui s'adressent la phrase (par exemple dans la phrase "Pouvez-vous m'ouvrir la porte s'il vous plait ?", le receveur peut voir de ma part un besoin d'assistance ou encore il peut se décider à m'ouvrir la porte).

Austin a ensuite voulu proposer une classification des actes illocutoires en 5 grandes classes. Searle repris son travail et proposa plus tard une version plus robuste à discriminer les différents actes [Searle, 1976]. Cette classification des actes illocutoires sur laquelle nous nous appuyons pour cadrer l'expression d'attitudes est la suivante :

1. **Assertifs**, actes qui exposent les croyances et les connaissances de celui qui les prononce (par exemple "J'ai vu un film hier soir.").
2. **Directifs**, actes qui engagent l'autre à réaliser une action (par exemple "Tu dois me rendre mon DVD demain !").
3. **Promissifs**, actes qui engagent celui qui les prononce à réaliser une action (par exemple "J'irai au cinéma ce soir.").

4. **Expressifs**, actes qui exposent les émotions et les opinions de celui qui les prononce (par exemple "Ce film est très bon.").
5. **Déclaratifs**, actes qui ont un pouvoir de changer l'état du monde (par exemple "Je vous déclare Mari et Femme !").

Nous avons choisi un ensemble d'actes du langage pour illustrer nos situations qui couvrent ces 5 catégories. *Informer* comme assertif, *Demander* comme directif, *Accepter* comme promissif, *Approuver* et *Désapprouver* comme expressifs et *Refuser* comme déclaratif. Dans notre système, notre agent se limite à produire que des actes issus de cette sélection. Ce choix a été motivé principalement dans le but de couvrir toutes les catégories et d'avoir des actes assez génériques et de différentes valences.

4.3.3 Le comportement non-verbal de l'agent

Comme vu précédemment dans le chapitre 2, les attitudes sont exprimées à travers le comportement non-verbal d'un individu. Lorsque une personne parle, elle effectue, de façon consciente ou inconsciente, de subtiles et moins subtiles altérations de son comportement qui peuvent trahirent, entre autre chose, l'expression d'une attitude interpersonnelle [Argyle, 1988]. Par exemple : une personne dominante fait de plus grands gestes, une personne soumise penche plus la tête vers le bas, une personne amicale sourit plus souvent et une personne hostile fronce plus les sourcils [Burgoon et al., 1984, Burgoon and Le Poire, 1999, Carney et al., 2005]. L'objectif de notre travail est de proposer un modèle capable de donner à un agent cette capacité de varier son comportement en fonction de l'attitude qu'il exprime. Bien que la qualité de la voix semble pouvoir être affectée par l'attitude qu'une personne exprime [Gobl et al., 2003], dans ce travail de thèse, nous nous concentrons sur les paramètres du comportement non-verbal. Nous ne considérons pas non plus le contenu verbal dans ce travail et par conséquent notre agent ne parle pas. La sélection des paramètres est fondée sur une analyse des études théoriques et empiriques décrites dans le chapitre 2. A partir de cette analyse, nous avons sélectionné les paramètres suivants :

1. **Le type d'expressions faciales :** les trois expressions faciales possibles pour notre agent sont : un sourire de joie (positif), un froncement de sourcils (négatif) ou une expression neutre [Carney et al., 2005][Hess et al., 2005].

2. **Les gestes :** les quatre valeurs pour ce paramètre sont : la présence d'un mouvements de tête seul (un hochement de tête ou un secouement de tête), la présence d'un geste des bras seul (ouverture des 2 bras, paume horizontale orientée vers le haut, voir figure 4.1), les deux à la fois ou aucun des deux [Carney et al., 2005].
3. **Amplitude des gestes des bras :** l'amplitude des gestes correspond à l'espace qu'occupent les bras lors de la réalisation des gestes. L'amplitude peut être faible, moyenne ou forte [Carney et al., 2005].
4. **Force des gestes des bras :** la force des gestes correspond à l'accélération du geste pendant le mouvement. Cette accélération peut être faible, moyenne ou forte [Carney et al., 2005].
5. **Orientation de la tête :** La tête peut être orientée devant, vers le haut, vers le bas ou sur le côté [Burgoon and Le Poire, 1999].
6. **Regard :** le regard peut être droit vers la caméra ou vers le côté (évitement) [Briton and Hall, 1995][Burgoon et al., 1984].

L'objectif est donc de déterminer les valeurs pour ces paramètres suivant l'attitude que doit exprimer l'agent. Puisque les mouvements de bras et de têtes peuvent aussi exprimer l'acte illocutoire, nous les avons prédéfinis pour être cohérent avec les actes du langage que nous considérons. Lorsque l'acte est soit *Refuser* ou *Désapprouver*, le mouvement de tête possible pour l'agent est un secouement de la tête (comme pour dire non). Autrement il s'agit d'un simple hochement de tête. Pour ce qui est du geste des bras, nous avons choisi un geste neutre qui peut être produit pour tous nos actes illocutoires. Il s'agit d'un geste avec les paumes des mains ouvertes vers le haut, les mains placées devant la personne. Lorsqu'il n'y a aucun geste des bras, nous considérons la position de repos suivante : les deux mains sont jointes devant le ventre (voir Figure 4.1).

4.3.4 L'interface de GenAttitude

GenAttitude se présente comme une application dans laquelle les utilisateurs ont la tâche de configurer le comportement non-verbal d'un agent conversationnel en changeant les valeurs des paramètres du comportement non-verbal que nous considérons. Pour ce type d'application et de tâche, les animations de l'agent doivent

FIGURE 4.1 : A gauche : le geste générique que nous utilisons. A droite : la position de repos

pouvoir se mettre à jour en temps réel dès que les valeurs des paramètres sont modifiées. Plusieurs techniques sont possibles. Aujourd'hui, les technologies utilisées sur Internet permettent d'envisager d'avoir un moteur 3D qui tourne dans le navigateur Internet des participants. Des moteurs comme Unity3D proposent des extensions pour navigateurs permettant de faire fonctionner des applications 3D en temps réel. Cependant cela nécessite que le plugin soit installé sur l'ordinateur du participant et ce type de plugin n'est pas très répandu. Une autre solution est l'utilisation de WebGL. WebGL est une interface de programmation pour HTML5 basé sur OpenGL qui permet de faire des applications 3D dans les navigateurs. Cette technologie était toutefois trop récente pour notre développement (jusqu'au 17 octobre 2013, Internet Explorer ne le supportait pas). Nous avons choisi alors de nous inspirer de l'approche présentée dans [Ochs et al., 2012]. Les auteurs ont développé une application Flash qui au lieu de jouer en temps réel les animations, jouait des vidéos qui avaient été préparées pour chacune des configurations des paramètres. Cette approche permettait à l'application d'être compatible avec la plupart des navigateurs, Flash étant une technologie très répandue à ce moment, et permettait de proposer une application suffisamment riche pour être ludique du point de vue de l'utilisateur. Nous

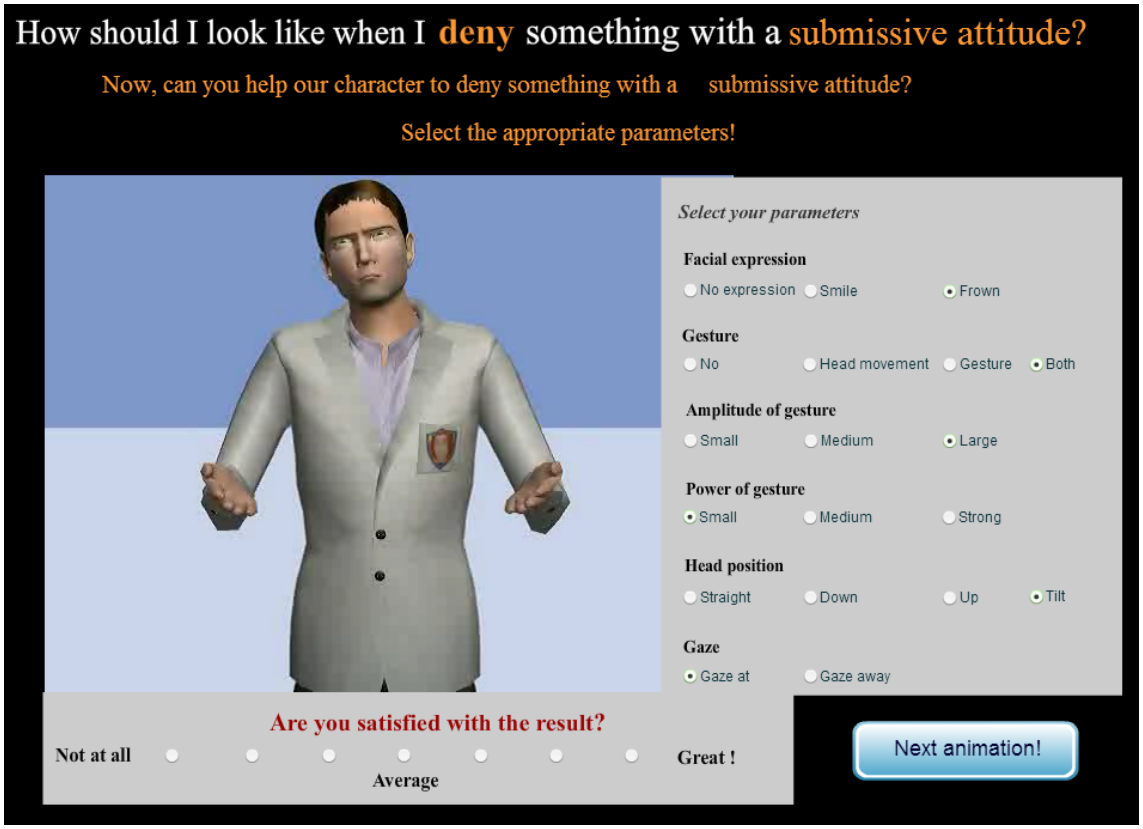


FIGURE 4.2 : Capture d'écran de l'interface de GenAttitude

avons donc décidé de proposer une application similaire, permettant de jouer une vidéo pour chacune des configurations des paramètres du comportement choisie par l'utilisateur.

GenAttitude a donc été réalisé en utilisant la technologie Adobe Flash et la plateforme d'agent virtuel VIB/Greta [Pecune et al., 2014]. L'architecture modulaire de VIB/Greta nous a permis de développer des composants additionnels pour générer les animations du comportement non-verbal de l'agent et ainsi produire toutes les vidéos nécessaires à notre application.

L'interface de *GenAttitude*, illustrée dans la Fig. 4.2, est inspirée par celle proposée dans [Ochs et al., 2012]. En haut de l'écran se situe la description de la tâche courante. Il y est expliqué pour quelle attitude et pour quel acte illocutoire l'utilisateur doit configurer le comportement de l'agent. Sur le côté, on trouve les différents paramètres du comportement non-verbal qu'il est possible de configurer. L'utilisateur peut changer les valeurs de ces paramètres en utilisant des boutons radios. La liste des paramètres ainsi que leurs valeurs possibles sont ceux décrits dans la section précédente. Il n'est pas possible de sélectionner une valeur pour l'amplitude et la force des gestes des bras si la valeur du paramètre des gestes est égale à "Mouvement de la tête seul" ou "Aucun mouvement de tête et geste des bras". Afin d'éviter un biais d'une configuration par défaut, les paramètres sont présélectionnés au hasard. Sur le côté gauche se trouve une vidéo qui joue en boucle le résultat de la configuration des paramètres. Chaque fois qu'une valeur est changée, une nouvelle vidéo est rechargée afin de montrer le résultat à l'utilisateur. Enfin, en bas de l'écran, une échelle de Likert à 7 points est présente afin de permettre à l'utilisateur de noter sa satisfaction vis-à-vis de l'animation qu'il a ainsi créée.

Étant donné le nombre de paramètres et de valeurs possibles, nous avons généré 1440 vidéos pour GenAttitude. Nos actes ont trois valences possibles : positif (*Accepter* et *Approuver*), négatif (*Refuser* et *Désapprouver*) et neutre (*Informer* et *Demander*) résultant en trois animations distinctes de mouvements de têtes. Ce qui donne lorsqu'il y a les gestes des bras : 3 mouvements de têtes $*$ 3 expressions faciales $*$ 4 orientations de la tête $*$ 2 regards $*$ 3 amplitudes $*$ 3 forces $=$ 648. Dans le cas où il n'y a pas de gestes des bras, nous avons : 3 mouvements de têtes $*$ 3 expressions faciales $*$ 4 orientations de la tête $*$ 2 regards $=$ 72). Ce qui nous donne un total de $648 + 72 = 720$ animations possibles, 1440 pour les deux agents

(masculin et féminin). Pour ce faire, nous avons développé un module additionnel au sein de la plateforme VIB/Greta. Ce module génère toutes les combinaisons de signaux (expressions faciales, gestes, amplitudes, forces, orientations de la tête, regards et mouvements de tête pour toutes les valeurs des paramètres) pour chacun de nos actes illocutoires puis capture une vidéo de l'animation pour chacune des combinaisons.

Avant de mettre en place la collecte de données avec *GenAttitude*, 6 pré-tests ont été réalisés afin de s'assurer que l'interface était facile à prendre en main et que les tâches étaient bien comprises. Les retours furent très bons et très encourageants, les participants ayant apprécié l'aspect ludique de notre interface et ils nous ont permis de corriger certaines formulations dans la description de la tâche donnée aux participants qui pouvaient être mal comprises.

Dans la section suivante, nous présentons la procédure de collecte des données.

4.4 LA COLLECTE ET L'ÉTUDE DU CORPUS DE DESCRIPTIONS DU COMPORTEMENT NON-VERBAL DE L'AGENT

Chaque participant a utilisé *GenAttitude* pour configurer le comportement non-verbal d'un agent dans six situations différentes. Chaque situation correspondait à une attitude sociale, un acte illocutoire et un genre de l'agent (masculin ou féminin). Par exemple, un participant peut se voir demander de configurer le comportement d'un agent féminin, devant exprimer de la dominance alors qu'elle pose une question. Afin d'éviter un effet de l'ordre de présentation sur les résultats, tout a été contrebalancé. Le participant va donc pouvoir configurer les paramètres du comportement non-verbal de l'agent, et donc son animation, en choisissant les valeurs qui lui semblent correspondre le plus à chaque situation.

4.4.1 Description de la collecte des données

Cent soixante-dix utilisateurs ont utilisé GenAttitude sur une période d'une semaine. Nous avons collecté 925 descriptions de comportements (dont 311 réalisées par des participants de genre féminin). Chaque description correspond à :

- un acte illocutoire ;
- une attitude interpersonnelle ;

– un genre de l'agent ;
– une configuration des paramètres du comportement non-verbal.

Nous avons eu des participants qui venaient de différents pays (Maroc, États-Unis d'Amérique, Allemagne, Pays-Bas, Belgique, Irlande, Taïwan, Roumanie, Tunisie, Italie, France, Inde, Chine, Canada, Lituanie, Argentine, Brésil, Royaume-Uni, Espagne, Grèce, Turquie et Afrique du Sud) mais la plupart étaient de France (70%), âgés de 17 à 62 ans ($M = 28.8$, $ET = 2.8$). Chaque participant devait noter sa satisfaction vis-à-vis de l'animation générée sur une échelle de Likert à 7 points. Les résultats ont montré une satisfaction globalement positive ($M = 5.35$, $ET = 1.16$, $Med = 6$).

4.4.2 Analyse des données collectées

Afin d'étudier le corpus collecté et la dépendance entre les attitudes, les actes, le genre et les paramètres non-verbaux, nous avons utilisé le test statistique du χ^2. Le test statistique du χ^2 est utilisé pour tester la dépendance de deux variables qualitatives (où les valeurs représentent des catégories et non pas des quantités). Ce test compare la distribution des effectifs d'un tableau de contingence des deux variables avec une distribution théorique dans laquelle les variables seraient indépendantes. Le résultat du χ^2 indique si les deux variables sont dépendantes, et nous avons utilisé la mesure du V de Cramer (qui est la racine carrée du χ^2 divisé par le χ^2 maximum) pour nous renseigner sur la force de cette dépendance (entre 0 et 1). Pour définir nos variables, nous les avons séparées en deux groupes, indépendantes et dépendantes. Les variables indépendantes sont les suivantes :

– **ACTE** (Informer, Demande, Approuver, Désapprouver, Accepter, Refuser) ;
– **ATTITUDE** (Dominant, Soumis, Amical, Hostile) ;
– **GENRE** (Masculin, Féminin).

Les variables dépendantes sont les suivantes :

– **EXPRESSION_FACIALE** (positive, négative, neutre) ;
– **GESTES** (bras seulement, tête seulement, bras et tête, aucun) ;
– **FORCE** (faible, normale, forte) ;
– **AMPLITUDE** (faible, normale, forte) ;
– **ORIENTATION_TETE** (droite, bas, côté, haut) ;
– **REGARD** (droit, côté).

TABLE 4.1 : Table de contingence des paramètres non-verbaux et des attitudes

Variable	Valeur	Dominant	Amical	Hostile	Soumis
EXPRESSION FACIALE	négative	**44,64%**	9,40%	**79,20%**	16,81%
	neutre	29,18%	8,55%	12,39%	**46,98%**
	positive	26,18%	**82,05%**	8,41%	36,21%
GESTES	bras seulement	27,47%	**30,77%**	24,34%	20,69%
	tête seulement	29,18%	28,21%	26,99%	**34,91%**
	bras et tête	**33,05%**	29,49%	**31,86%**	23,28%
	aucun	10,30%	11,54%	16,81%	21,12%
FORCE	moyenne	24,82%	39,01%	23,62%	26,47%
	faible	26,95%	**41,84%**	19,69%	**53,92%**
	forte	**48,23%**	19,15%	**56,69%**	19,61%
AMPLITUDE	moyenne	29,08%	**35,46%**	22,05%	17,65%
	faible	31,21%	29,08%	29,13%	**56,86%**
	forte	**39,72%**	**35,46%**	**48,82%**	25,49%
ORIENTATION TETE	bas	22,32%	19,23%	**30,53%**	**58,62%**
	devant	30,47%	27,78%	23,45%	13,36%
	côté	12,45%	**39,32%**	24,78%	18,10%
	haut	**34,76%**	13,68%	21,24%	9,91%
REGARD	côté	15,88%	17,52%	28,32%	49,14%
	droit	**84,12%**	**82,48%**	**71,68%**	**50,86%**

Les résultats montrent des dépendances significatives entre les variables ATTITUDE et EXPRESSION_FACIALE ($\chi^2(6) = 451{,}65$; $p < 0{,}005$; Cramer's V= 0,49), ATTITUDE et GESTES ($\chi^2(9) = 24{,}0$; $p < 0{,}005$; Cramer's V= 0,09), ATTITUDE et FORCE ($\chi^2(6) = 69{,}9$; $p < 0{,}005$; Cramer's V= 0,26), ATTITUDE et AMPLITUDE ($\chi^2(6) = 34{,}0$; $p < 0{,}005$; Cramer's V= 0,182), ATTITUDE et ORIENTATION_TETE ($\chi^2(9) = 167{,}3$; $p < 0{,}005$; Cramer's V= 0,25) et ATTITUDE et REGARD ($\chi^2(3) = 81{,}69$; $p < 0{,}005$; Cramer's V= 0,29). En d'autres termes, l'attitude à exprimer a eu une influence sur la configuration de tous les paramètres du comportement non-verbal choisis par les utilisateurs.

Nous avons aussi voulu vérifier si nous trouvions des dépendances entre les actes illocutoires et les paramètres du comportement non-verbal. Seulement EXPRESSION_FACIALE ($\chi^2(10) = 24{,}9$; $p < 0{,}05$; Cramer's V= 0,11) et GESTES ($\chi^2(15) = 89{,}39$; $p < 0{,}005$; Cramer's V= 0,18) ont montré une dépendance significative avec ACTES.

Par ailleurs, aucune dépendance n'a été trouvée entre le genre de l'agent et les paramètres du comportement. Une analyse supplémentaire n'a révélé aucune dépendance entre le genre du participant et les paramètres non plus. En d'autres termes, les utilisateurs n'ont pas défini de comportements différents suivant leur genre ou celui de l'agent.

L'analyse du corpus a révélé principalement des relations entre les attitudes exprimées et les valeurs choisies pour les paramètres. Nous détaillons les valeurs les plus choisies par attitude dans les paragraphes suivants.

Dominant :

L'attitude dominante est décrite principalement par une expression faciale négative, la présence des mouvements de têtes et de bras, pas de regard sur le côté, et une orientation haute de la tête. Les gestes ont une large amplitude et sont exprimés avec force.

Soumis :

L'attitude soumise est décrite principalement par une expression faciale neutre, des mouvements de têtes seuls, et une orientation de la tête vers le bas. Il y a une légère préférence pour le regard droit mais celle-ci est très faible. Si les gestes des bras sont activés, les valeurs de force et d'amplitude sont faibles.

Amical :

Cette attitude est décrite par les gestes des bras seuls, une expression faciale positive, la tête sur le côté et le regard droit. Pour l'amplitude des gestes, les valeurs normale et forte sont les plus choisies et pour la force, la valeur faible est la plus choisie.

Hostile :

Cette attitude est décrite par une expression faciale négative, des mouvements des bras et de la tête et pas de regard sur le côté. Les gestes sont larges et exprimés avec force, comme dans l'attitude dominante. Cependant, contrairement à l'attitude dominante, l'orientation de la tête est vers le bas.

Ces résultats, décrits dans le Tableau 4.1, montrent des similarités avec les résultats de la littérature sur la perception des attitudes chez les humains. En effet, les individus attendent d'une personne dominante plus de gestes larges et une tête orientée vers le haut [Carney et al., 2005], d'une personne soumise moins d'expressions faciales [Carney et al., 2005] et de la part d'une personne amicale plus de sourires [Burgoon et al., 1984]. Cependant, nos résultats montrent aussi quelques différences avec la littérature. Aucune influence du genre n'a été significative. Aussi, plus de regard sur le côté était attendu de la part d'un agent hostile ou soumis. Ce dernier point pourrait s'expliquer par le fait que l'agent était montré faisant face à la caméra dans un environnement générique et qu'il pouvait être difficile d'interpréter vers quoi le regard de l'agent se tournait.

A partir de ces résultats, nous avons proposé un modèle computationnel qui utilise un réseau Bayésien dont les paramètres ont été appris sur nos données collectées. Ce modèle est décrit dans la section suivante.

4.5 L'APPRENTISSAGE DU MODÈLE COMPUTATIONNEL

L'étude du corpus décrite dans la section précédente a révélé des dépendances entre l'attitude et les paramètres du comportement non-verbal de l'agent. Le travail décrit dans cette section consiste à utiliser ces dépendances afin de construire un modèle computationnel permettant de déterminer le comportement non-verbal d'un agent virtuel exprimant une attitude sociale. En d'autres termes, ce modèle sera construit à partir des données du corpus grâce à l'utilisation d'une méthode d'apprentissage supervisé.

4.5.1 Les méthodes d'apprentissage supervisé

Les méthodes d'apprentissage supervisé sont utilisées afin de construire un outil de prédiction à partir d'observations. En d'autres termes, les observations passées sont utilisées pour construire un algorithme de prédiction pour des observations futures [Kotsiantis et al., 2007]. Ces outils sont notamment utiles pour résoudre des problèmes de classifications ou de prises de décisions. Différentes méthodes existent comme les modèles de Markov cachés, les réseaux de neurones, les machines à vecteurs de support, les arbres de décision ou les réseaux Bayésiens par exemple. Dans

[Ochs et al., 2012], les auteurs ont cherché à construire un modèle permettant de générer différents types de sourires : amusé, poli et gêné. Pour cela, ils ont utilisé les données qu'ils ont collectées sur les caractéristiques morphologiques des différents sourires afin de construire un arbre de décision. Dans un arbre de décision, chaque nœud (même la racine) représente une variable prédictive du système (les paramètres morphologiques du sourire dans notre exemple) et les feuilles représente les classes à prédire (le type du sourire dans notre exemple). Pour déterminer la classe d'un nouvel individu, on parcourt l'arbre depuis la racine en suivant les distributions de la variable à prédire à chaque nœud. Lorsque l'on arrive dans une feuille, l'on obtient une prédiction pour la classe de l'individu. Cet arbre est ensuite utilisé pour trouver différentes configurations pour chaque type de sourire. Dans [Busso et al., 2007], les auteurs ont utilisé un système à base de modèles de Markov cachés (des automates à états) pour générer des mouvements de tête à partir d'informations sur la prosodie (pitch de la voix, sa dérivée et sa dérivée seconde) pour différentes états émotionnels. Un modèle de Markov caché a été appris pour chaque état émotionnel. Ces modèles sont particulièrement bien adaptés pour la prédiction de séquences.

Ces méthodes proposent différentes techniques afin de construire la structure du modèle et d'apprendre ses paramètres. Elles n'offrent pas les mêmes avantages et ne sont pas toujours indiquées suivant l'objectif que l'on souhaite modéliser. Par exemple, les arbres de décisions permettent d'obtenir un modèle très lisible mais le nombre de branches peut rapidement devenir trop élevé, générant un temps de parcours de l'arbre plus long. Les réseaux de neurones offrent l'avantage d'être capables d'apprendre avec peu de données un modèle fiable, cependant la méthode nécessite de définir, pour la fonction d'apprentissage, les valeurs initiales des paramètres. De plus, les poids appris par le système ne sont pas compréhensibles par l'utilisateur. De même pour les modèles de Markov cachés, il est nécessaire de fixer la typologie de l'automate pour son apprentissage ce qui peut ne pas être évident.

Les réseaux Bayésiens sont des graphes orientés acycliques où les nœuds représentent les variables et les arêtes entre deux nœuds représentent une dépendance probabiliste entre deux variables ($P(A|B)$) [Koller and Friedman, 2009]. Nous avons retenu le réseau Bayésien pour notre modèle car c'est une représentation qui fonctionne bien avec des variables catégorielles. La structure d'un réseau Bayésien est basée sur les dépendances conditionnelles entre les variables, or nous avons déjà

identifié dans la section 4.4 des dépendances entre l'attitude et le comportement non-verbal. De plus, le réseau Bayésien offre les avantages de fournir un modèle lisible, capable de générer de la variabilité, aisément combinable avec d'autres modèles et qui peut être évolutif.

Nous présentons d'autres travaux qui ont utilisés cette représentation. Dans [Bergmann, 2012], le but était de calculer la forme des gestes à produire en fonction du contexte de l'interaction et des gestes précédents. Plus précisément, comme nœuds d'entrée pour représenter ce contexte, ont été choisis le geste précédent, le contenu du discours et des paramètres du sujet de conversation comme la position et la forme des objets. Comme nœuds de sortie, ont été choisis des paramètres des gestes (par exemple l'orientation de la main, le mouvement, et la forme). Dans [Pelachaud and Poggi, 2002], le réseau est utilisé pour animer les expressions faciales d'un agent. En entrée se trouvent les intentions communicatives (performatives et émotions) et en sortie se trouvent des modalités du visage comme les sourcils, la forme de la bouche, l'orientation de la tête, les mouvements de têtes et le regard. Un modèle de génération de comportements automatiques pour avatars a été proposé dans [Kendira, 2010] fonctionnant grâce à un modèle Bayésien dynamique. Ce modèle a été conçu pour permettre la génération de comportements non-verbaux (des gestes uniquement) en fonction de l'activité de l'utilisateur (audio et manipulation de la souris) dans le contexte d'une narration collaborative.

Nous présentons en détail la structure de notre réseau dans la section suivante.

4.5.2 Structure du réseau Bayésien

L'objectif est de modéliser l'influence des attitudes et des actes illocutoires sur les paramètres du comportement non-verbal. Par conséquent, le réseau Bayésien proposé à la structure suivante : les nœuds d'entrée sont l'attitude interpersonnelle et l'acte illocutoire, les nœuds de sortie sont les paramètres du comportement non-verbal étudiés avec GenAttitude (Section 4.2) et enfin un arc orienté existe entre un nœud d'entrée et un nœud de sortie si une dépendance statistique entre les deux variables a été trouvée. Ces dépendances correspondent à celles identifiées dans la section 4.4. Il existe donc un arc entre le nœud d'attitude et chaque nœud des paramètres du comportement non-verbal. Il existe également un arc entre le nœud de l'acte illocutoire et le nœud de l'expression faciale et entre le nœud de l'acte

illocutoire et le nœud de l'activation des gestes. La structure est montrée figure 4.3.

4.5.3 Paramètres du réseau Bayésien

Apprendre les paramètres du réseau Bayésien consiste à apprendre la distribution des probabilités conditionnelles de chaque nœud. Pour estimer ces paramètres, nous avons utilisé les données collectées du corpus (Section 4.4) afin de faire une estimation directe des probabilités conditionnelles du modèle (grâce au logiciel Weka [Hall et al., 2009]). En d'autres termes, pour chaque nœud de sortie du modèle, sa distribution de probabilités est obtenue en calculant la fréquence de chaque valeur pour les différentes valeurs des nœuds d'entrée.

Cependant, afin de prendre en compte le niveau de satisfaction des utilisateurs dans l'estimation des paramètres, un sur-échantillonnage du corpus a d'abord été réalisé. Ce sur-échantillonnage des données permet de donner un poids plus important aux descriptions qui ont reçu une satisfaction plus élevée par les participants qui ont utilisé *GenAttitude* : chaque description est présente dans le corpus n fois, où n est le niveau de satisfaction reporté avec cette description. En conséquence, une description avec un niveau de satisfaction de 7 sera présente 7 fois dans le nouveau corpus sur-échantillonné. Ce nouveau corpus contient 4947 entrées.

Le détail des distributions de probabilités de chaque nœuds se trouvent en annexe A.

Avantages de ce modèle :

Le réseau Bayésien permet d'utiliser les données collectées de deux façons. Ce réseau permet de choisir une attitude et un acte illocutoire afin d'obtenir les probabilités pour chaque valeur des paramètres du comportement non-verbal. Ensuite, ces probabilités peuvent être utilisées afin de configurer le comportement non-verbal correspondant aux attitudes et actes désirés. Baser la configuration du comportement sur ces probabilités permet d'introduire de la variabilité en utilisant celles-ci pour introduire de l'aléatoire dans la sélection des valeurs pour les paramètres du comportement.

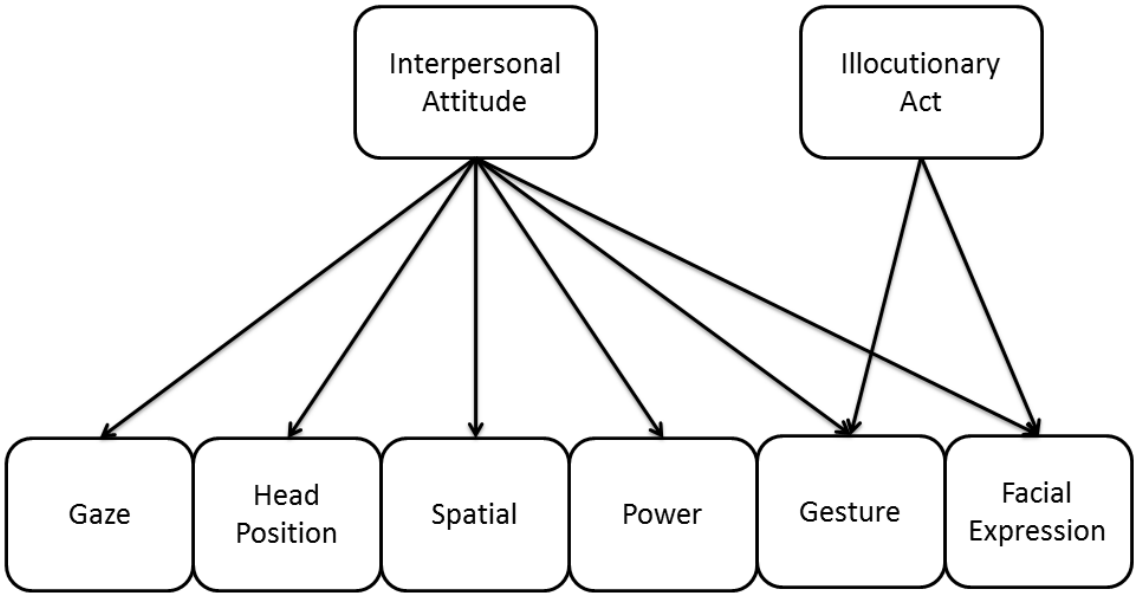


FIGURE 4.3 : Le réseau Bayésien

4.5.4 Implémentation du modèle

Le modèle a été implémenté au sein de la plateforme VIB/Greta. Pour rappel, l'architecture de VIB/Greta est basée sur SAIBA et dans cette architecture, une intention communicative est transformée en liste de signaux multimodaux grâce au *Behavior Planner* [Kopp et al., 2006]. Ensuite ces signaux sont transformés en une animation par le *Behavior Realizer*. Dans le cas de la plateforme VIB/Greta, les animations générées suivent le standard MPEG-4. Le *Behavior Realizer* génère des paramètres d'animation pour le corps (appelés Body Animation Parameters) et pour le visage (appelés Facial Animation Parameters). De plus, les agents de la plate-forme VIB/Greta disposent de paramètres d'expressivité leur permettant d'avoir de la variabilité dans leur comportement pour une même intention [Mancini and Pelachaud, 2007]. Ces paramètres sont les suivants :

- Overall activation (OAC) : L'utilisation ou non de chacune des modalités.
- Spatial extent (SPC) : l'amplitude des mouvements.
- Temporal (TMP) : la vitesse des mouvements.
- Fluidity (FLD) : La fluidité des mouvements.
- Power (PWR) : l'accélération des mouvements.

– Repetitivity (REP) : la génération de mouvements répétés et rythmiques.

Le modèle que nous avons implémenté influence le processus SAIBA à deux niveaux, avant la génération de signaux par le *Behavior Planner* et juste après celle-ci. Son fonctionnement est le suivant.

Tout d'abord, le modèle ne s'active que si une attitude sociale est définie pour l'agent et si l'intention communicative traitée correspond à l'un des six actes illocutoires que nous considérons. Au premier niveau, le modèle commence par récupérer l'ensemble des valeurs pour ses paramètres du comportement correspondant à l'attitude qu'il doit exprimer. Pour cela, il utilise les probabilités conditionnelles du réseau et l'attitude sociale désirée. Ensuite, en fonction de la valeur sélectionnée pour le paramètre des mouvements de la tête et des gestes des bras, le modèle modifie le paramètre d'expressivité d'*Overall Activation* des gestes et de la tête. Par exemple, si la valeur choisie correspond à des mouvements de la tête seule (sans les bras), le modèle met la valeur 0 au paramètre OAC des bras et 1 au paramètre OAC de la tête.

Puis le *Behavior Planner* du système Greta sélectionne les signaux pour réaliser l'intention communicative. Le paramètre d'*Overall Activation* est utilisé lors de cette sélection afin de choisir ou non si les modalités de tête et de gestes sont activées.

Au second niveau, le modèle vient modifier cette liste avant qu'elle ne soit envoyée au *Behavior Realizer* en fonction des valeurs sélectionnées par le réseau Bayésien. Par exemple, si des gestes sont activés, leurs valeurs de *Spatial Extent* et de *Power* sont changées pour correspondre aux valeurs obtenues par le réseau. Ou encore, il est possible d'insérer une expression faciale de joie et un basculement de tête si les valeurs obtenues par le réseau Bayésien l'indiquent. Ce processus est illustré figure 4.4.

Nous proposons d'illustrer notre implémentation avec un exemple concret. L'agent doit effectuer un acte illocutoire de type *Informer* avec une attitude dominante. En suivant notre réseau Bayésien, La combinaison avec la probabilité la plus haute est un visage neutre, des gestes des bras, larges et forts, la tête orientée vers le haut et le regard droit (voir figure 4.5). Par conséquent, le premier niveau de notre modèle met la valeur 0 au paramètre OAC de la tête et la valeur 1 au paramètre OAC des bras. Le système VIB/Greta sélectionne donc un ensemble de signaux incluant le mouvement des bras pour le *Behavior Realizer*. Le second niveau de notre modèle

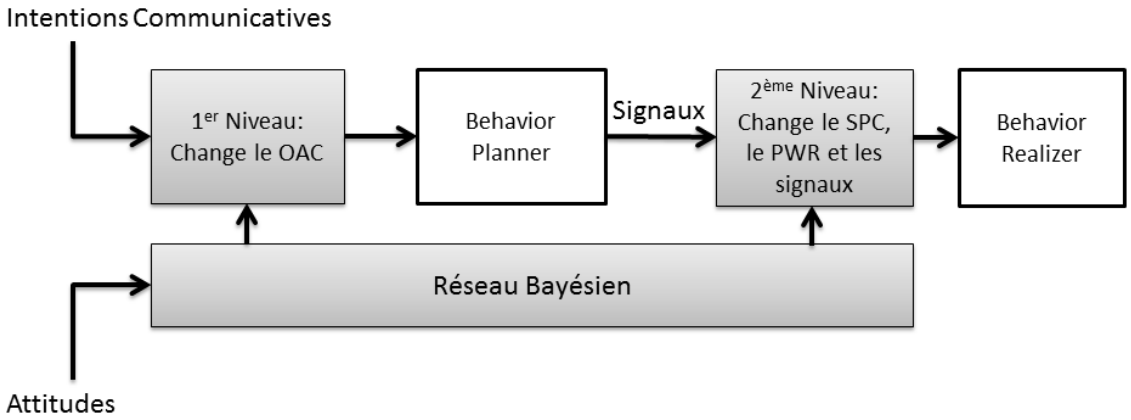


FIGURE 4.4 : Les deux niveaux de notre modèle dans l'architecture SAIBA

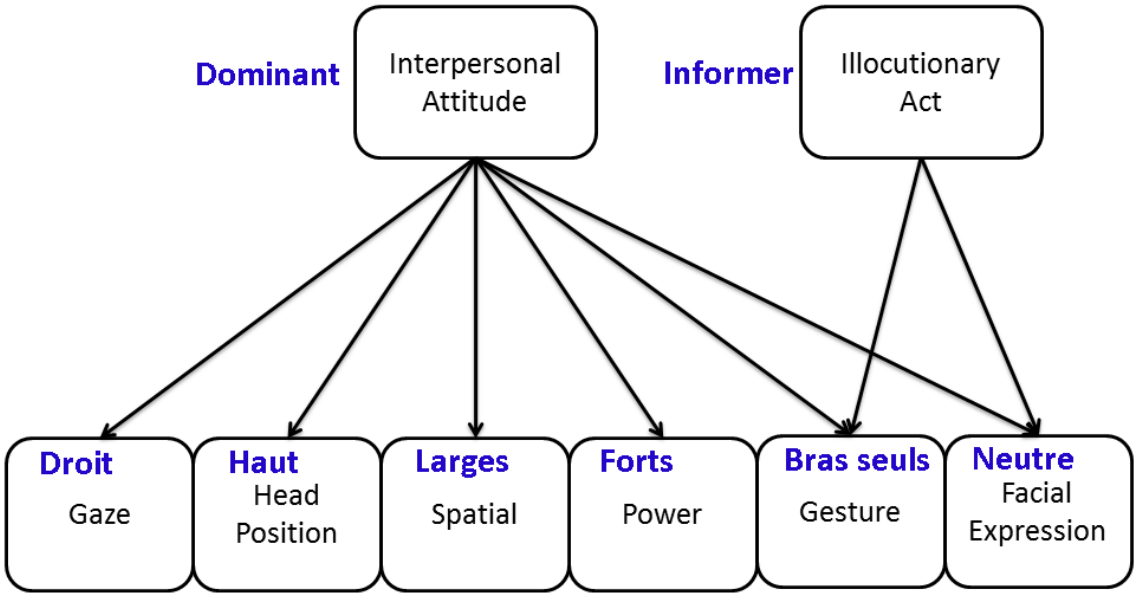


FIGURE 4.5 : Les valeurs choisies pour une attitude dominante et l'acte illocutoire Informer

met la valeur 1 au paramètre SPC et au paramètre PWR du geste. Il vérifie également qu'aucune expression faciale n'est présente dans la sélection de signaux et insère une orientation de la tête vers le haut. Pour le même acte du langage mais avec une attitude soumise, la combinaison avec la probabilité la plus haute est une expression faciale positive, pas de gestes des bras et une tête orientée vers le bas. Le modèle intervient de la même façon que précédemment. Ces exemples sont montrés Fig. 4.6.

Dans la section suivante, nous décrivons l'application de recruteur virtuel que nous avons développée grâce à notre modèle ainsi qu'une évaluation de ce système.

4.6 Mise en œuvre et évaluation : recruteur virtuel

Afin de réaliser une première évaluation de notre travail, nous avons intégré notre modèle d'attitude dans une application concrète. En collaboration avec Zoraida Callejas, nous avons proposé une architecture de recruteur virtuel capable de faire varier ses attitudes, à travers son comportement non-verbal et verbal, en fonction de la difficulté voulue pour l'entretien et du niveau d'anxiété de l'utilisateur [Callejas et al., 2014]. Ce recruteur virtuel a pour but de fournir une plateforme d'entrainement aux entretiens d'embauche à destination de jeunes en insertion professionnelle. Nous avons ensuite évalué la capacité de ce recruteur à produire des comportements qui expriment des attitudes en réalisant une étude en ligne conduite en anglais.

4.6.1 Un agent recruteur

Les entretiens d'embauche sont un excellent cadre de démonstration pour l'expression d'attitude. Tout d'abord, les entretiens d'embauche continuent d'être la méthode de sélection de candidats la plus utilisée [Macan, 2009], et la préparation de ces entretiens joue un rôle important dans la qualité de la performance du postulant. Différentes études ont montrées les bénéfices de l'entrainement pour ce genre d'entretiens. Par exemple, dans [Huffcutt et al., 2011], les auteurs ont montré que l'entrainement aux entretiens et que l'expérience d'entretiens précédents avaient un effet positif sur l'assurance du postulant et réduisait son anxiété. Dans [Tross and Maurer, 2008], les auteurs ont trouvé que l'accompagnement permettait aux postulants de produire des réponses plus structurées, de développer des compétences

FIGURE 4.6 : Capture d'écran d'un agent exprimant différentes attitudes pour l'acte illocutoire Informer : Dominante (haut) et Soumise (bas).

d'entretien fondamentales comme les comportements de regard et la capacité de répondre aux questions usuelles. De précédents travaux ont montré que l'entrainement des jeunes adolescents aux techniques d'entretiens était bénéfique pour leur carrière [Creed et al., 2008].

Le système que nous proposons permet aux jeunes de pouvoir s'entrainer avec des entretiens de différentes difficultés, dans lesquels un agent conversationnel joue le rôle du recruteur. Dans [Huffcutt et al., 2011], les auteurs expliquent que la personnalité du recruteur a un impact significatif sur les performances des postulants. Ceux-ci ont une expérience plus difficile avec un recruteur plus fermé et/ou moins agréable. Il devient alors important, lorsque l'on conçoit un recruteur virtuel de lui donner la capacité de mettre le postulant plus ou moins à l'aise afin de le faire travailler sur différents niveaux de difficulté. Dans cette section, nous décrivons le recruteur virtuel que nous avons développé [Callejas et al., 2014] en utilisant le modèle computationnel d'attitude décrit dans la section précédente. Ce recruteur est capable de reconnaitre le niveau d'anxiété de l'utilisateur afin d'adapter la difficulté de l'entretien en modifiant son attitude sociale envers le postulant. Ce changement d'attitude se traduit par des actes illocutoires et un comportement verbal et non-verbal différent. Dans notre modélisation de l'expression d'attitudes, nous nous sommes intéressés uniquement au comportement non-verbal, cependant, dans le but de pouvoir évaluer le modèle en situation, nous avons collaboré avec Zoraida Callejas au développement d'un modèle de dialogue capable d'adapter également le contenu verbal en fonction des attitudes.

4.6.2 L'architecture

L'architecture de cet agent, basée sur la plateforme VIB/Greta, suit l'architecture d'agent SAIBA [Kopp et al., 2006] et est montrée Figure 4.7. En lieu et place de l'*Intent Planner*, nous avons un *Dialog Manager*, qui se charge de calculer le prochain acte illocutoire et l'attitude à adopter par l'agent. Le *Behavior Planner* de cet agent SAIBA est celui que nous avons augmenté avec le réseau Bayésien et pour finir nous utilisons le *Behavior Realizer* de la plateforme VIB/Greta. Des composants additionnels viennent compléter cette architecture. Un détecteur du niveau d'anxiété de l'utilisateur est connecté au *Dialog Manager*, un *Natural Language Generator* permet de générer les phrases de l'agent et un module de *text-to-speech* permet de

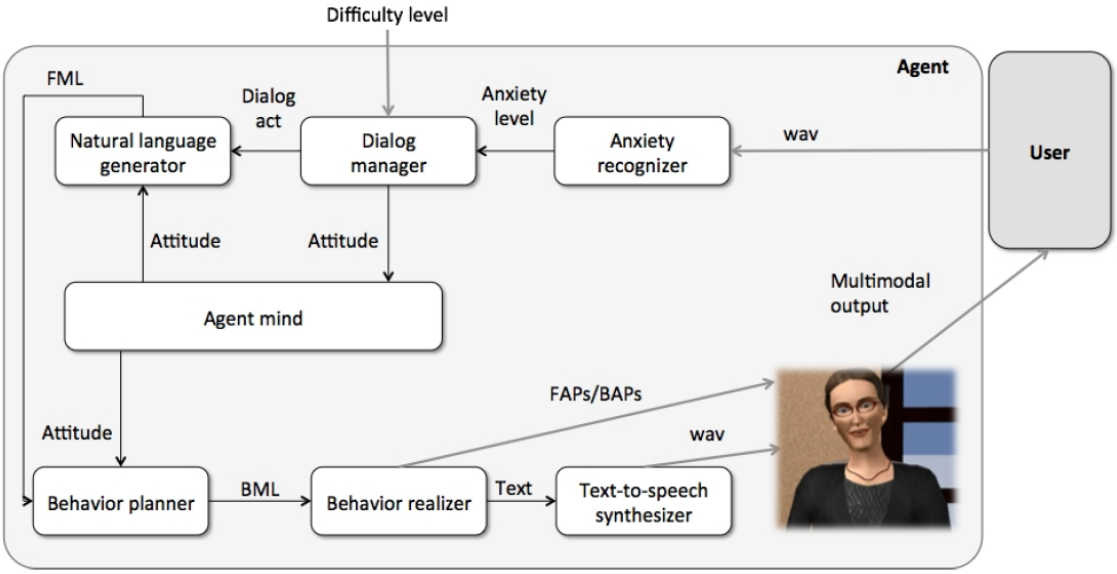


FIGURE 4.7 : Architecture du recruteur virtuel

Anxiety level in the previous turn	Difficulty level	System objective for the current turn
Low	1 or 2	COMFORT
Low	3 or 4	If *tendency* = decreasing then COMFORT 90% - CHALLENGE 10% else then COMFORT 10% - CHALLENGE 90%
Low	5 or 6	If *tendency* = neutral then COMFORT else then CHALLENGE
Medium	1 or 2	COMFORT
Medium	3 or 4	If *tendency* = increasing then COMFORT 10% - CHALLENGE 90% Else then COMFORT 90% - CHALLENGE 10%
Medium	5 or 6	CHALLENGE
High	1 or 2	COMFORT
High	3 or 4	COMFORT 50% - CHALLENGE 50%
High	5 or 6	CHALLENGE

FIGURE 4.8 : Stratégies du recruteur virtuel

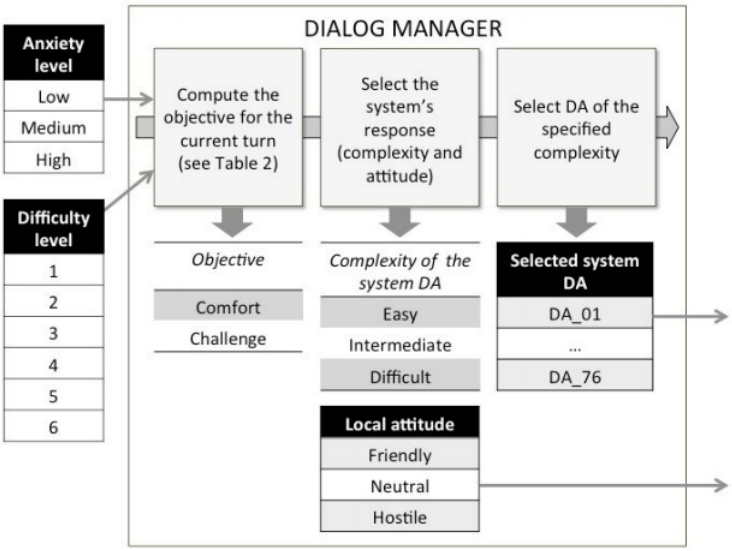


FIGURE 4.9 : Fonctionnement du Dialog Manager

produire la voix de celui-ci. Le *Dialog Manager* choisit à chaque tour de parole quelle stratégie (réconforter ou mettre en difficulté) adopter face à l'utilisateur en fonction du niveau d'anxiété de celui-ci et de la difficulté voulue pour l'entretien (de 1 à 6). Ce processus est dépeint dans la figure 4.8. Nous n'avons pas cherché à intégrer une solution particulière de détection d'anxiété, nous partons du principe que cette valeur nous est fournie entre 0 et 1 afin que nous puissions déterminer trois catégories discrètes, basse (en-dessous de 0,25), moyenne (entre 0,25 et 0,75) et haute (au-dessus de 0,75). Par exemple, il est possible de calculer ce niveau en analysant les signaux audiovisuels ou physiologiques de l'utilisateur [Baur et al., 2013]. La plate-forme SSI fournit d'ailleurs un ensemble d'outils permettant de mesurer les émotions et les attitudes d'un utilisateur à travers sa voix et son comportement non-verbal [Wagner et al., 2013]. Le *Dialog Manager*, à travers sa stratégie, choisit le prochain tour de parole et l'attitude mais c'est le *Natural Language Generator* qui choisit le contenu de la phrase en fonction de l'attitude. Cette phrase est ensuite envoyée, sous la forme d'un fichier FML [Heylen et al., 2008], au *Behavior Planner* qui utilise notre réseau Bayésien afin de choisir le comportement non-verbal de l'agent pour réaliser ce tour de parole. Pour terminer, ces signaux sont envoyés au Text-To-Speech et au Behavior Realiser qui vont produire l'animation finale de l'agent. Dans les sections suivantes, nous détaillons comment fonctionnent le *Dialog Manager* et le *Natural*

Language Generator.

4.6.3 Les différentes stratégies du Dialog Manager

Le *Dialog Manager* que nous avons utilisé pour cette implémentation est celui développé en collaboration avec Zoraida Callejas [Callejas et al., 2014]. Il prend en entrée un niveau de difficulté et le niveau d'anxiété de l'utilisateur qu'il reçoit à chaque tour de parole de l'utilisateur. Plus la difficulté est élevée, plus le système va essayer de mettre l'utilisateur en difficulté (par exemple en essayant de le déstabiliser avec les questions posées). A l'inverse, plus la difficulté est basse, plus le système va essayer de mettre l'utilisateur à l'aise. De plus, la stratégie de dialogue est aussi influencée par la variation de l'anxiété de l'utilisateur. Elle est mesurée par la pente de la régression linéaire de toutes les mesures d'anxiété depuis le début de l'interaction. Ces deux valeurs sont utilisées pour choisir la difficulté du prochain acte de dialogue (la question que le recruteur virtuel va poser) et l'attitude sociale avec laquelle le produire. La difficulté de l'acte de dialogue se traduit par le focus sur les points négatifs et l'ouverture laissée à la réponse. Par exemple, une question va donc être plus difficile à y répondre si elle se concentre sur les faiblesses du candidat et si la réponse nécessite une longue explication plutôt qu'une réponse concise. Ce processus est basé sur les deux phases de la gestion de l'anxiété chez l'humain : la détection d'une menace et comment les individus évaluent la disponibilité et l'efficacité de leurs moyens pour y faire face [Beck and Clark, 1997]. Dans notre cas, cela correspond à la présence de contenu négatifs dans la phrase et à la probabilité élevée qu'une question ouverte rende les candidats incertains de leur capacité à élaborer une réponse satisfaisante. Par exemple, l'acte de dialogue "Why have you been out of work so long ?" (Pourquoi avez-vous été si longtemps sans travail ?) est une question difficile, car elle se concentre sur l'aspect négatif et est ouverte. A l'inverse, la question "When did you finish your previous job ?" (Quand avez-vous fini votre précédent emploi ?) est plus facile à répondre car il s'agit d'une question fermée sans jugement de la part du recruteur. Ce processus est décrit dans la Figure 4.9. Le système continue de poser des questions jusqu'à ce que l'utilisateur reste dans un état d'anxiété moyen ou bas sur plusieurs tours de parole. Le nombre de tours est plus élevé pour une difficulté plus haute. Une fois l'acte de dialogue et l'attitude choisis, le *Natural Language Generator* choisit une phrase qui satisfait ces deux valeurs,

et notre *Behavior Planner* augmenté de notre modèle computationnel d'attitudes calcule le comportement non-verbal de l'agent.

4.6.4 Le Natural Language Generator

Le *Natural Language Generator* est le second composant que nous avons développé avec Zoraida Callejas [Callejas et al., 2014]. Il s'agit d'un sélectionneur de phrases qui se base sur l'attitude calculée par le *Dialog Manager* pour choisir la phrase qui l'exprime le mieux. Afin de refléter cette attitude, nous avons suivi la méthode du projet PERSONAGE [Mairesse and Walker, 2011] qui décrit comment faire varier la production du langage en fonction de la personnalité représentée à l'aide du modèle OCEAN [McCrae and Costa, 1987][McCrae and John, 1992]. Bien que le modèle de PERSONAGE ait été conçu à partir du modèle de personnalité OCEAN, des travaux ont montré que certaines caractéristiques de l'extraversion et de l'introversion pouvaient s'appliquer pour générer des comportements respectivement amicaux et hostiles [Isbister, 2006]. Les paramètres de la production du langage que nous faisons varier en fonction de l'attitude sont donc le nombre de références à soi-même, l'utilisation de pronoms, la diversité du vocabulaire, la préférence pour les noms plutôt que les verbes, le formalisme des expressions, la longueur des phrases et la préférence pour le contenu négatif plutôt que positif. En ce qui concerne les références à soi-même, l'agent peut parler directement de lui ("I will interview you today" / Je vais vous interviewer aujourd'hui) ou il peut parler de lui comme faisant partie de la société ("What do you know about us ?" / Que savez-vous de nous ?). De cette façon, une attitude hostile est rendue en parlant moins de soi-même et en utilisant le nom de la compagnie. Pour une attitude amicale, plus de pronoms, moins de synonymes et un langage moins formel sont utilisés. De cette façon, les phrases donnent l'impression d'être moins réfléchies et plus détendues. L'attitude hostile est rendue en utilisant un langage plus formel avec un vocabulaire plus riche et moins de pronoms, ce qui donne des phrases plus longues. Il y a plus de négations et une préférence pour du contenu négatif, à l'inverse de l'attitude amicale. Dans cette dernière, on retrouve également plus de verbes que de noms dans les phrases, traduisant une préférence pour l'action.

Nous avons créé une base de données de 228 phrases qui contient au moins une version amicale, hostile et neutre de chaque acte de dialogue.

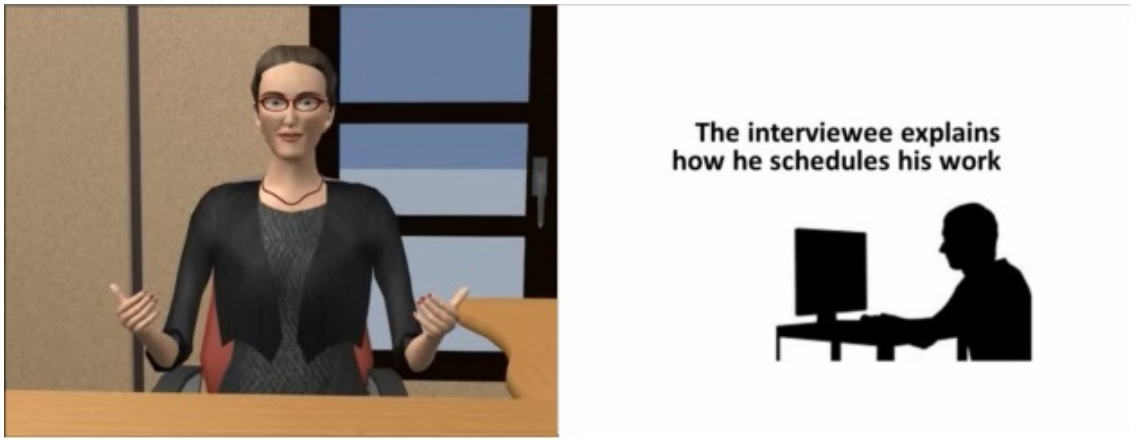


FIGURE 4.10 : Les deux vues montrées dans les vidéos de l'évaluation. Gauche : l'agent pose une question au candidat. Droite : un écran est montré ensuite pour indiquer le moment où le candidat aurait répondu à la question.

4.6.5 Évaluation

Procédure :

Afin d'évaluer le modèle verbal et non-verbal d'attitudes sociales, nous avons conduit une étude perceptive. L'objectif de cette étude est de mesurer la capacité d'un recruteur virtuel équipé de notre modèle à transmettre différentes attitudes sociales à travers son comportement verbal et non-verbal. Nous n'avons pas cherché à évaluer la capacité du système à changer sa difficulté en fonction du niveau d'anxiété de l'utilisateur mais nous nous sommes focalisés sur l'expression des attitudes. Nous avons formulé les trois hypothèses suivantes :

- **H1 :** Le comportement non-verbal amical (respectivement hostile) de l'agent donne à l'utilisateur l'impression que l'agent exprime une attitude amicale (respectivement hostile).
- **H2 :** Le comportement verbal amical (respectivement hostile) de l'agent donne à l'utilisateur l'impression que l'agent exprime une attitude amicale (respectivement hostile).
- **H3 :** Le comportement multimodal (verbal et non-verbal) amical (à l'inverse, hostile) renforce la perception de l'attitude amicale (respectivement hostile).

Plus précisément, l'évaluation a pour but de montrer que le modèle d'agent proposé transmet correctement les attitudes sur une modalité (H1 et H2) et sur plusieurs

(H3). Nous ne prenons pas en compte dans cette évaluation le niveau d'anxiété de l'utilisateur et la difficulté de l'entretien qui relèvent de la stratégie du *Dialog Manager*.

Stimuli : Nous avons réalisé des vidéos montrant ce que produit notre modèle dans des entretiens simulés pour 4 conditions :

- Verbale : L'attitude de l'agent est exprimée uniquement à travers son comportement verbal.
- Non-verbale : L'attitude de l'agent est exprimée uniquement à travers son comportement non-verbal.
- Multimodale : L'attitude de l'agent est exprimée à travers son comportement verbal et non-verbal.
- Contrôle : L'agent n'exprime pas d'attitude particulière (le système VIB/Greta génère les signaux sans influence de notre modèle).

Nous avons créé un total de 7 vidéos : deux vidéos (amicale et hostile) pour les conditions verbale, non-verbale et multimodale et une vidéo pour la condition de contrôle. Dans chaque vidéo, le scénario du dialogue est le même. Il consiste en 10 tours de parole pour l'agent et 9 tours de parole pour le candidat (représentés par le même écran de texte disant simplement "Le candidat répond au recruteur"). De cette façon, chaque vidéo montre différents actes du langage du recruteur qui pourraient intervenir dans un entretien, et nous ne focalisons pas sur ce que répond le candidat. L'ordre des tours de parole et des phrases est le même pour chaque vidéo, la différence se trouvant dans la forme des phrases et du comportement non-verbal (mouvements de tête et de regard, gestes et expressions faciales) en fonction de l'attitude.

Mesures :

Notre étude a été conduite en ligne en Anglais. Nous avons créé des vidéos qui alternent une vue de l'agent prononçant des phrases et un écran texte indiquant que le candidat a répondu (voir figure 4.10). Nous avons conçu un questionnaire qui accompagnait chaque vidéo afin que les participants puissent indiquer leur perception de l'attitude. Sur des échelles de Likert à 5 points, nous avons demandé aux participants d'évaluer leur accord (de "Totally disagree / Pas d'accord du tout" à "Totally agree / Complètement d'accord") avec les phrases suivantes : (1) "The virtual recruiter behavior is believable / Le comportement du recruteur virtuel est

crédible", (2) "The virtual recruiter gives the impression to want to hire the interviewee / Le recruteur donne l'impression de vouloir engager le candidat", (3) "The virtual recruiter gives the impression to want to fail the interviewee / Le recruteur donne l'impression de vouloir faire échouer le candidat", (4) "The virtual recruiter tries to make the interviewee at ease / Le recruteur cherche à mettre le candidat à l'aise", (5) "The virtual recruiter wants to destabilize the interviewee / Le recruteur veut déstabiliser le candidat", (6) "The virtual recruiter expresses an hostile attitude / Le recruteur exprime une attitude hostile", (7) "The virtual recruiter expresses a friendly attitude / Le recruteur exprime une attitude amicale" et (8) "The virtual recruiter expresses a dominant attitude / Le recruteur exprime une attitude dominante". Les phrases (2), (4) et (7) correspondent à de l'attitude amicale et les phrases (3), (5) et (6) à de l'attitude hostile.

Participants : 110 personnes (dont 48 femmes) ont participé à cette évaluation en ligne avec un âge moyen de 34 (ET=12). Les participants venaient majoritairement de France (N=60) mais parlaient anglais. Chaque participant a vu et noté 4 vidéos (une vidéo pour chaque condition) et l'ordre de présentation de celles-ci a été contrebalancé.

Résultats : Nous avons pu collecter 440 évaluations de vidéo. Nous avons réalisé des tests ANOVA pour chercher les effets entre l'attitude de l'agent et les réponses des participants dans chacune des conditions et nous avons ensuite réalisé des tests post-hoc HSD-Tukey afin de déterminer le sens de ces effets. Nous avons trouvé des effets différents suivant les conditions.

Dans la condition non-verbale, les tests ANOVA ont révélés des effets de l'attitude sur les réponses aux questions (2) $F(2{,}220) = 6{,}7$ $p < 0{,}001$, (3) $F(2{,}220) = 12{,}8$ $p < 0{,}001$, (4) $F(2{,}220) = 22{,}3$ $p < 0{,}001$, (5) $F(2{,}220) = 9{,}9$ $p < 0{,}001$, (6) $F(2{,}220) = 15{,}7$ $p < 0{,}001$, (7) $F(2{,}220) = 28$ $p < 0{,}001$ et (8) $F(2{,}220) = 12{,}7$ $p < 0{,}001$. Un test post-hoc HSD-Tukey a indiqué que l'attitude de l'agent était bien perçue comme plus amicale aux questions (2), (4) et (7) quand l'agent exprime une attitude amicale que quand il exprime une attitude hostile ou que dans la condition de contrôle ($p < 0{,}05$). L'attitude de l'agent a été perçue comme plus hostile aux questions (3), (5) et (6) quand l'agent exprimait une attitude hostile que quand il exprimait une attitude amicale ($p < 0{,}001$). La différence n'est pas significative face à la condition de contrôle. L'attitude amicale a donc bien été reconnue mais

l'attitude hostile a été partiellement reconnue. H1 est donc partiellement validée.

Dans la condition verbale, un effet de l'attitude a été identifié sur les réponses aux questions (2) $F(2,219) = 3,9$ $p < 0,05$, (4) $F(2,219) = 5,9$ $p < 0,005$ et (8) $F(2,219) = 3,1$ $p < 0,05$. Un test post-hoc HSD-Tukey a révélé que l'attitude était perçue comme plus amicale aux questions (2), quand l'attitude de l'agent est amicale plutôt que hostile ($p < 0,05$), et (4), quand l'attitude de l'agent est amicale plutôt que hostile ($p < 0,05$) et que dans la condition de contrôle ($p < 0,01$). L'attitude n'a pas été perçue comme plus hostile dans les réponses aux questions (3), (5) et (6). Seulement l'attitude amicale a été partiellement reconnue. H2 n'est donc pas validée.

Dans la condition multimodale, les tests ANOVA ont révélés des effets de l'attitude sur les réponses aux questions (2) $F(2,221) = 14,7$ $p < 0,001$, (3) $F(2,221) = 17,1$ $p < 0,001$, (4) $F(2,221) = 28,9$ $p < 0,001$, (5) $F(2,221) = 25,1$ $p < 0,001$, (6) $F(2,221) = 39,1$ $p < 0,001$, (7) $F(2,221) = 45,7$ $p < 0,001$ et (8) $F(2,221) = 21,5$ $p < 0,001$. Un test post-hoc HSD-Tukey a révélé que l'attitude de l'agent était bien perçue comme plus amicale aux questions (2), (4) et (7) quand l'agent exprimait une attitude amicale que quand il exprimait une attitude hostile ou que dans la condition de contrôle ($p < 0,01$). Ils ont également révélé que l'attitude de l'agent était bien perçue comme plus hostile aux questions (3), (5) et (6) quand l'attitude de l'agent était hostile que quand celle-ci était amicale ou que dans la condition de contrôle ($p < 0,05$). Par conséquent, les attitudes ont été bien reconnues dans cette condition. H3 est donc validée.

En ce qui concerne la perception de la crédibilité, aucune différence n'a été décelée. Le recruteur virtuel est perçu en moyenne comme crédible (M=3,1 sur une échelle de Likert à 5 points) et dominant (M=2,99 sur une échelle de Likert à 5 points). Ceci pourrait être expliqué par le fait que le rôle d'un recruteur renvoie une image dominante. Nous avons également trouvé des différences significatives entre les genres des participants. Les femmes ont trouvé l'agent moins hostile ($M_h = 2,20$) $F(1,439) = 5,6$ $p < 0,05$ et plus crédible ($M_c = 3,27$) $F(1,439) = 7,9$ $p < 0,005$ que les hommes ($M_h = 2,46$; $M_c = 2,98$). Ceci pourrait être expliqué par le fait que l'agent était féminin mais cette explication reste à vérifier.

L'attitude amicale a été bien reconnue mais l'attitude hostile a été partiellement reconnue dans la condition non-verbale. Par conséquent, notre hypothèse H1 est

partiellement validée. Le comportement non-verbal seul a permis d'exprimer une attitude amicale mais pas une attitude hostile. Cela pourrait signifier que ces deux attitudes ne s'expriment pas avec les mêmes modalités. Dans la construction des exemples, nous avons varié les phrases (c'est-à-dire le texte) et le comportement non-verbal. Par contre nous n'avons pas modifié l'intonation de la voix Concernant le comportement verbal, notre hypothèse H2 n'a pas été validée, l'attitude amicale seule ayant été partiellement reconnue. Encore une fois, le comportement verbal seul non plus n'a pas permettre d'exprimer toutes les attitudes. Notre hypothèse H3 a été validée quant à elle car les attitudes amicale et hostile ont bien été perçues dans la condition multimodale. Ces résultats montrent l'importance de considérer plusieurs modalités simultanément pour transmettre une attitude.

4.6.6 Conclusion

Dans ce chapitre, nous avons donné aux agents la capacité d'exprimer des attitudes. Nous avons créé une plate-forme en ligne afin de faire construire un corpus de descriptions de comportement non-verbal d'un agent par des participants. Ce corpus, construit par les utilisateurs de notre application, contient des descriptions de comportement d'agent correspondant à l'expression de différentes attitudes lors de la production de différents actes du langage. En analysant les données de ce corpus, nous avons pu constater certaines similarités avec l'expression d'attitudes chez les humains. Ensuite à partir des données du corpus, nous avons construit un réseau Bayésien que nous utilisons dans un modèle computationnel de génération de comportement non-verbal. Ce modèle permet à un agent de produire des comportements non-verbaux différents suivant l'attitude qu'il souhaite exprimer.

Enfin, nous avons intégré ce modèle avec un modèle de production d'actes du langage dans un recruteur virtuel. Une évaluation de ce système a montré que les attitudes étaient mieux reconnues lorsque la parole était aussi influencée par l'attitude et pas que par le comportement non-verbal. Il est possible que l'ensemble des modalités du comportement que nous avons considéré ne soit pas suffisant. Afin d'exprimer correctement des attitudes, l'ensemble des modalités de la communication doivent, peut-être, être utilisé comme l'intonation de la voix par exemple. Dans ce premier travail, nous avons considéré un agent interagissant seul face à un utilisateur : les résultats sont très encourageants. Dans la suite, nous considérons l'agent au cœur

d'un groupe conversationnel et nous allons pouvoir considérer plus de modalités dans l'expression d'attitudes. Nous avons alors décidé d'intégrer les comportements de regard, l'orientation du corps ainsi que la distance interpersonnelle dans la gestion de l'espace conversationnel. Cette gestion de l'espace permettra également aux agents d'exprimer leurs attitudes dans des situations de groupes. Nous souhaitons nous abstraire du contenu des conversations afin de pouvoir proposer un modèle pouvant s'adapter au maximum de situations. Cependant les résultats décrits dans ce chapitre ont montré l'importance de l'influence des attitudes dans la parole afin de correctement retranscrire celles-ci. Nous avons alors décidé de proposer un modèle de tour de parole qui permet à un agent de changer sa stratégie de prise de parole en fonction de l'attitude qu'il souhaite exprimer. Dans les chapitres suivants, nous présentons nos travaux afin de construire ce modèle ainsi que les évaluations de celui-ci que nous avons conduites.

Résumé du chapitre

- Une collecte de corpus, décrivant le comportement non-verbal d'un agent exprimant différentes attitudes, basée sur une méthodologie de *Crowdsourcing* a été réalisée.
- Un modèle computationnel, pour le calcul du comportement non-verbal d'un agent exprimant différentes attitudes, appris sur les données du corpus a été proposé.
- Une évaluation du modèle dans le contexte d'un agent recruteur a souligné l'importance de considérer plusieurs modalités dans l'expression d'attitudes.

5 Attitudes dans un groupe conversationnel

Dans le chapitre précédent, nous avons présenté le travail réalisé afin de proposer un modèle de comportement non-verbal exprimant une attitude pour un agent locuteur. Le contexte dans lequel nous avons étudié cet agent était une interaction avec l'utilisateur seulement ; jusqu'à présent nous n'avons pas abordé les situations de conversations avec plusieurs interlocuteurs. De plus, comme nous avons limité notre agent au rôle du locuteur de la conversation, nous avons ainsi étudié uniquement la production de comportements lorsque celui-ci a la parole. L'objectif du travail décrit dans ce chapitre continue le travail précédent en s'intéressant à des conversations de groupe et aux différents comportements que peuvent produire des agents afin d'exprimer leurs attitudes face aux divers membres du groupe. Un de nos objectifs est de permettre aux agents d'exprimer une attitude différente envers chacun de leurs interlocuteurs. Dans ce cas, plusieurs questions se posent que nous nous proposons d'adresser : Quelles sont les stratégies comportementales à mettre en œuvre pour permettre ces différentes expressions simultanément ? Est-ce que les comportements sont utilisés tour à tour pour exprimer les différentes attitudes où alors expriment-ils toutes les différentes attitudes qu'un agent peut exprimer en même temps ?

Étendre ainsi le travail à des situations de groupe implique également de considérer de nouveaux facteurs dans l'expression d'attitude. Jusqu'à présent, nous nous sommes surtout intéressés au comportement non-verbal de l'agent. L'évaluation précédente a révélé des faiblesses dans la capacité à exprimer des attitudes quand seul le comportement non-verbal était utilisé pour exprimer des attitudes. Dans l'optique d'étendre l'expression d'attitudes à des situations de groupes, nous considérons également des modalités supplémentaires. Afin de proposer un modèle capable d'exprimer des attitudes quelle que soit la conversation, nous nous abstrayons encore du contenu verbal. Cependant, nous avons vu que les attitudes étaient mieux reconnues lorsque le comportement verbal et le comportement non-verbal étaient influencés conjoin-

tement. Afin d'adresser cette limitation, nous proposons de considérer les stratégies de prise de parole dans les comportements influençables par les attitudes (vouloir la parole souvent, attendre son tour, interrompre les autres...), quel que soit le contenu verbal.

Un groupe conversationnel implique également que celui-ci respecte une certaine géométrie, c'est-à-dire la position et l'orientation de ses membres. Nous proposons de jouer sur cette géométrie en fonction des attitudes exprimées par les membres du groupe. En particulier, nous voulons changer les distances entre les membres, les orientations du corps et les directions du regard qui permettent d'assurer au groupe de maintenir sa cohésion (à laquelle chaque membre participe par son comportement), afin de permettre d'exprimer les attitudes. Il nous faut être vigilant et ne pas permettre des distances entre certains membres qui viendraient briser cette cohésion.

Ces évolutions et ces ajouts impliquent certaines difficultés ; leurs intégrations demandent de définir un algorithme assurant que chaque information soit communiquée. Ainsi, comment un agent peut exprimer différentes attitudes simultanément (par exemple comment un agent A peut exprimer une attitude dominante envers un agent B et une attitude soumise envers un agent C) ? Comment les différentes modalités que nous considérons se synchronisent-elles pour permettre l'expression d'attitudes (par exemple couper la parole et faire de grands gestes pour exprimer une attitude dominante) ? Comment résoudre des situations de conflits qui peuvent émerger quand différentes modalités cherchent à exprimer différentes fonctions communicatives (par exemple se tourner vers son interlocuteur pour lui parler et se tourner vers un agent plus dominant) ?

Dans ce chapitre, nous proposons un modèle qui cherche à répondre à ces différentes questions pour permettre l'expression d'attitudes sociales par des agents membres d'un groupe conversationnel. Dans la première section de ce chapitre, nous définissons et proposons une représentation d'un groupe conversationnel. Chaque agent, membre du groupe donc, sera équipé du modèle computationnel de production de comportement non-verbal développé au chapitre précédent augmenté de deux composants additionnels qui viendront ajouter de nouveaux comportements à l'agent relatifs à la situation de groupe. Le premier de ces deux composants, présenté dans la section 5.2 et que nous appellerons modèle de prise de parole, permettra aux agents de prendre tour à tour la parole en fonction des activités (*parler* et *écouter*) et des

attitudes des uns et des autres. Le second, présenté dans la section 5.3 et que nous appellerons modèle de cohésion de groupe, permettra d'influencer de nouvelles modalités du comportement non-verbal afin de pouvoir exprimer une attitude à travers la position et la posture de l'agent face au groupe auquel il appartient. L'ensemble des trois modèles est conçu pour fonctionner simultanément en collaboration afin de produire le comportement individuel d'un agent. Enfin dans la section 5.4, nous détaillons l'implémentation et l'intégration de ces modèles.

5.1 Un groupe conversationnel

Nous nous intéressons à modéliser des groupes de taille réduite, dont le nombre de participants ne dépasse pas la vingtaine [Beebe and Masterson, 2009]. Ces groupes sont définis comme étant un rassemblement de personnes où chaque participant peut interagir directement avec les autres [Homans, 2013]. Les travaux de McGrath [McGrath, 1984] ont cherché à identifier les caractéristiques de ces groupes et à proposer une représentation des interactions qui y ont lieu. Nous nous inspirons de ce travail de représentation pour construire notre architecture de groupe conversationnel. Les concepts que nous manipulons s'inscrivent très bien dans cette formalisation.

McGrath explique qu'un groupe est avant tout chose défini par ses membres. Cela signifie que l'existence même du groupe dépend du comportement de ses membres et n'existe pas sans eux. C'est parce que des personnes se rapprochent les unes des autres qu'apparait un groupe et non l'inverse. Il explique alors que la propriété fondamentale d'un groupe sont ses membres qui le composent ainsi que la structure bâtie par eux. Cette structure représente les rôles de chaque membre dans le groupe ainsi que les relations les uns avec les autres. Dans notre modèle, nous proposons d'utiliser cette représentation et de considérer que la structure du groupe est donnée par les attitudes exprimées par les membres. Chaque agent peut exprimer une attitude différente envers chaque autre membre du groupe et celles-ci ne sont pas nécessairement réciproques.

La deuxième propriété d'un groupe selon McGrath est la tâche que ses membres cherchent à accomplir en se réunissant ainsi, que ce soit une simple conversation ou un match de sport. Cette tâche intervient dans un contexte comme une réunion de travail ou une sortie entre amis. Dans notre représentation, nous limitons la tâche à la conversation mais nous choisissons de nous abstraire du contexte.

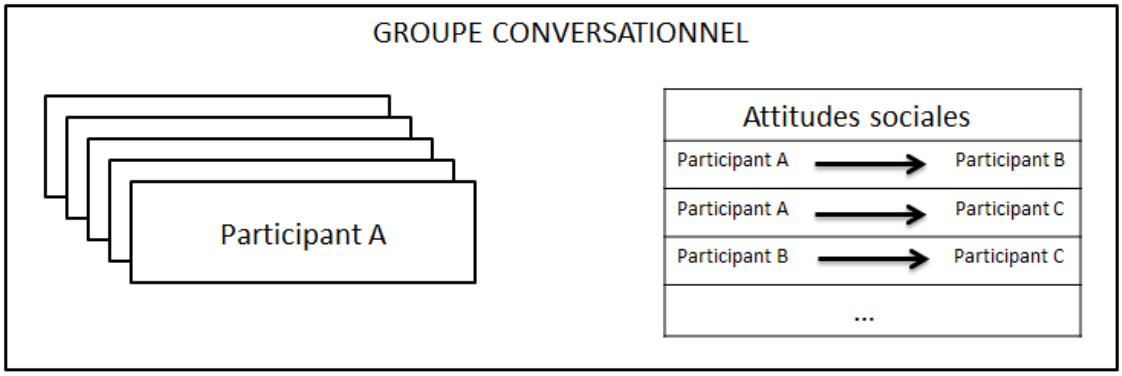


FIGURE 5.1 : Le modèle de groupe

Enfin, la troisième propriété décrit les interactions possibles entre les membres du groupe. Dans le cas d'un travail collectif, il peut s'agir d'aider un autre membre du groupe et dans le cas d'une conversation, il peut s'agir de parler à un autre membre ou de l'écouter par exemple. Nous choisissons de nous limiter exactement à ces deux actions que sont *parler* et *écouter*. Notre modèle de prise de parole se chargera de faire alterner l'action des agents entre celles-ci.

Notre architecture de groupe est montrée dans la figure 5.1. Afin de se rapprocher le plus possible de la représentation de McGrath, l'entité groupe conversationnel de notre modèle a un rôle très limité. Ce composant permet uniquement d'identifier les membres et la structure du groupe. Ce sont les comportements individuels de chaque membre qui vont faire émerger la dynamique du groupe. Chaque agent va réaliser des comportements différents en fonction des attitudes qu'il exprime envers les autres et des comportements des autres. Chaque participant alterne entre les rôles de locuteur et d'auditeur, grâce à un modèle de prise de parole ; il produit différents comportements non-verbaux dont la génération est réalisée par le *Behavior Planner* (gestes et expressions faciales) et déforme la formation du groupe sans briser sa cohésion en fonction des attitudes qu'il souhaite exprimer. Ce fonctionnement est montré dans la figure 5.2.

5.2 LE MODÈLE DE PRISE DE PAROLE

Un tour de parole correspond à l'intervalle de temps entre le moment où un participant à une conversation prend la parole et celui où il la rend [Sacks et al.,

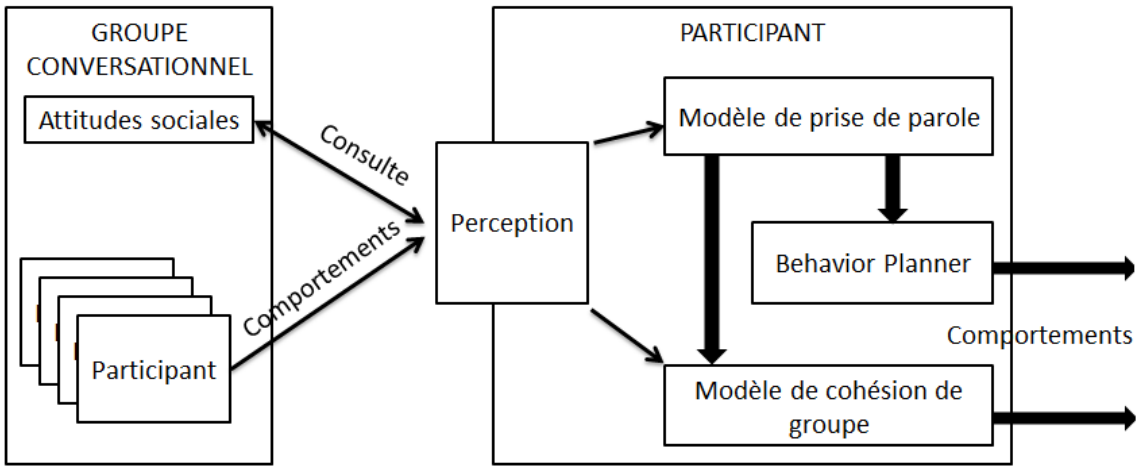


FIGURE 5.2 : Le modèle de chaque participant

1974]. Nous avons vu précédemment comment Sacks, Schegloff et Jefferson [Sacks et al., 1974] représentaient la conversation comme construite par les tours de parole de ses participants dans le chapitre 2. D'après ces auteurs, la stratégie inhérente à toute conversation est la minimisation des silences entre les tours et la fluidité des transitions, chaque participant cherchant à ce qu'il n'y ait qu'un locuteur à la fois. Dans une vision différente, nous avons également vu comment Clark [Clark, 1996] propose une représentation de la conversation où les tours de parole émergent de celle-ci et ne sont pas ses éléments constituants. Les participants cherchent avant tout à réaliser les actions de parler et d'écouter et c'est la réussite de ces actions conjointes qui fait émerger un tour de parole.

Nous avons vu que l'expression d'attitude peut amener les participants à prendre la parole en même temps (voir chapitre 3). Pour cette raison, la formalisation de Sacks et de ses collègues ne nous permet pas de donner aux agents la capacité d'exprimer des attitudes à travers leur stratégie de prise de parole. Sacks et ses collègues avancent l'idée que tous les participants d'une conversation cherchent continuellement à ce qu'il n'y ait qu'un locuteur à la fois. Or certains travaux ont montré qu'une personne qui exprime une attitude dominante peut parler en même temps qu'une autre, pour l'interrompre ou pour prendre le contrôle de la conversation [Goldberg, 1990]. Nous avons donc choisi de concevoir le modèle de prise de parole en s'inspirant de la représentation de Clark. De ce fait, nos agents vont alterner entre les rôles de locuteur et d'auditeur en fonction de leurs attitudes et du comportement des autres.

L'idée directrice de notre modèle n'est pas de réaliser systématiquement des tours de parole mais de faire alterner les agents entre ces rôles sans nécessairement s'assurer de la complétion de leurs tours de paroles. Grâce à cela, nous adressons certaines limites du modèle de Sacks, Schegloff et Jefferson mais aussi de Clark. Nous considérons une dimension sociale à travers l'attitude dans le mécanisme d'échange de la parole, ce que la formalisation de Sacks et celle de Clark ne proposent pas, comme souligné dans [O'Connell et al., 1990], et nous autorisons plusieurs agents à parler en même temps et donc des conversations pouvant avoir plusieurs locuteurs à la fois ; situations possibles quand les attitudes sont exprimées [Goldberg, 1990]. Ainsi notre modèle de tour de parole fait émerger les différents états (prise de parole, interruption, donner le tour de parole, etc.) en fonction de ce que souhaitent communiquer les agents mais aussi en fonction de leurs attitudes envers chacun. Différentes stratégies de gestion des tours de parole apparaissent suivant les attitudes des membres du groupe.

5.2.1 Présentation du modèle

Nous avons vu dans le chapitre 2 que l'attitude peut s'exprimer à travers les stratégies de prise de parole. Dans la réalisation de ce modèle, nous avons considéré que tous les membres du groupe ont l'intention de communiquer avec les autres et nous n'avons pas considéré le contenu des phrases mais seulement un désir de produire du son (une séquence de mots inintelligibles). Un agent réussit à prendre la parole en fonction des attitudes interpersonnelles qu'il souhaite exprimer envers les autres. Nous avons modélisé les tours de paroles comme une machine à états, inspiré par les travaux de Thórisson [Thórisson et al., 2010]. La définition des états se base sur les rôles conversationnels de Goffman [Goffman, 1981] et sont les suivants :

1. *own-the-speech* : l'agent est dans cet état (avoir le tour de parole) lorsqu'il est le seul à parler.
2. *compete-for-the-speech* : l'agent est dans cet état (se disputer le tour de parole) lorsqu'il parle en même temps que quelqu'un d'autre.
3. *end-of-speech* : l'agent est dans cet état (fin du tour de parole) lorsqu'il finit de parler, afin de donner le tour de parole.

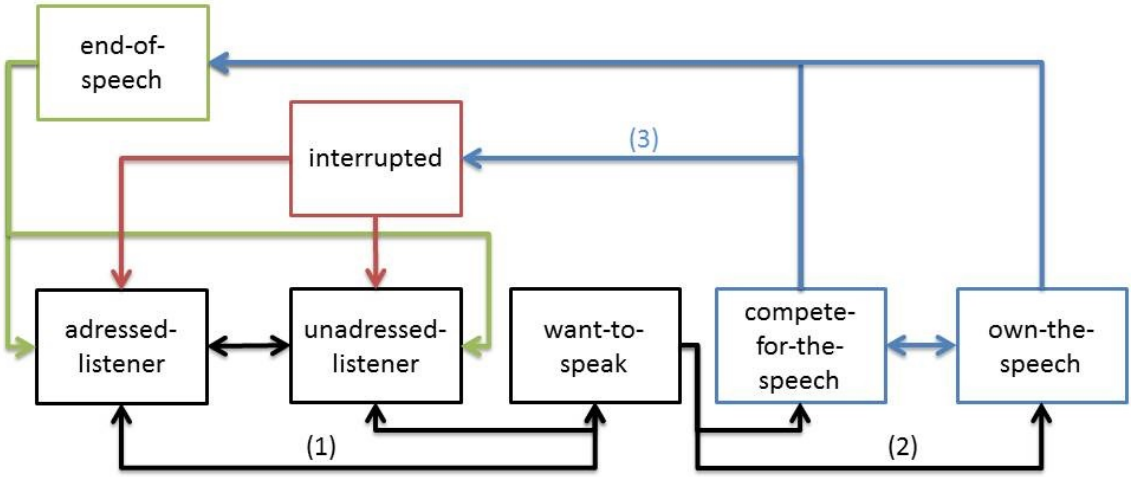


FIGURE 5.3 : Les états du modèle de tour de parole

4. *interrupted* : l'agent est dans cet état (interrompu) lorsqu'il s'est fait interrompre.
5. *unaddressed-listener* : l'agent est dans cet état (auditeur indirect) lorsqu'il écoute la conversation et que personne ne lui parle directement.
6. *addressed-listener* : l'agent est dans cet état (auditeur direct) lorsqu'il écoute la conversation et qu'une autre personne lui parle directement.
7. *want-to-speak* : l'agent est dans cet état (désir de prendre le tour de parole) lorsqu'il écoute et qu'il souhaite prendre la parole.

Les états et les transitions entre eux sont montrés dans la Figure 5.3.

Un agent démarre dans l'état *unaddressed-listener*. Toutes les 250 ms, le système vérifie si une transition depuis l'état courant est activable. Ce temps est le temps minimum chez un humain pour réagir à un stimuli [Kosinski, 2008]. Le fonctionnement des transitions entre les états est détaillé dans la section suivante.

5.2.2 Influence de l'attitude sur le tour de parole

Le modèle permet à un agent de changer d'état, afin de passer du rôle de locuteur à celui d'auditeur et inversement, en fonction des attitudes que cet agent exprime, du temps écoulé depuis la dernière fois qu'il a parlé et des autres agents actuellement en train de parler. Afin de représenter les attitudes d'un agent, nous

utilisons toujours la représentation des attitudes d'Argyle [Argyle, 1988] sur les deux dimensions Statut et Affiliation où chaque dimension va de -1 à 1. La valeur minimale de l'axe Statut représente une attitude soumise (respectivement une attitude hostile sur l'axe Affiliation) et la valeur maximale représente une attitude dominante (respectivement une attitude amicale).

Nous détaillons maintenant les différentes transitions. Entre les états *unaddressed-listener* (qui correspond à l'état où l'agent écoute sans être adressé par un autre) et *addressed-listener* (qui correspond à l'état où l'agent écoute en étant adressé par un autre) existent deux transitions permettant de passer d'un de ces états à l'autre. Ces transitions s'activent si l'agent est directement adressé par un autre ou non. Cela permet à l'agent de suivre si un autre agent lui parle afin de le regarder le cas échéant.

Deux transitions existent à partir des états *unaddressed-listener* et *addressed-listener* (où l'agent est un auditeur) vers l'état *want-to-speak* (où l'agent souhaite prendre la parole) qui s'activent à l'aide d'une fonction de transition. Cette fonction se base sur la dernière fois que l'agent a parlé (nous appelons cette valeur $Last$) et des attitudes qu'il exprime envers les autres ($Stat$ et Aff comme respectivement la valeur moyenne de Statut et la valeur moyenne d'Affiliation exprimées envers les autres). D'après [Cappella et al., 1985] et [Goldberg, 1990], plus la valeur de Statut et d'Affiliation est grande, plus une personne souhaite prendre la parole fréquemment. A l'inverse, une attitude soumise entraine un désir de parler plus faible. Une attitude hostile n'a pas d'effet particulier sur le désir de prendre la parole. Dans notre système, nous considérons que tous les agents ont l'intention de parler. Cela signifie que même s'ils expriment une attitude soumise, ils ont tout de même un désir de parler, bien que minimal. Nous avons donc défini de façon empirique un délai minimal ($Delay$) avant d'avoir le désir de parler de nouveau. Par conséquent, la fonction de transition $f_{li,wa}$ (depuis un état d'auditeur vers l'état *want-to-speak*), où Now est la date courante, est la suivante :

$$f_{li,wa}(Last,Stat,Aff) = \begin{cases} 1 & if(Now \geq Last + \\ & Delay * (1 - \frac{1}{2}(Stat + norm(Aff))) \\ 0 & else \end{cases} \quad (5.1)$$

La fonction $norm()$ est la normalisation de $[-1,1]$ à $[0,1]$. Cela permet à une attitude

de type hostile de ne pas faire augmenter le temps avant le prochain désir de prendre la parole. Cette fonction active donc la transition (renvoie 1) lorsque suffisamment de temps (plus ou moins long en fonction des attitudes, une attitude plus dominante et amicale entrainant un temps plus court) s'est écoulé depuis la dernière fois que l'agent a parlé.

Depuis l'état *want-to-speak* (l'agent a le désir de parler) vers l'état *own-the-speech* (l'agent a le tour de parole), la transition s'active immédiatement si aucun autre agent n'est en train de parler. Cela signifie que si l'agent souhaite prendre la parole et que personne d'autre n'est déjà en train de parler, il peut commencer librement à parler. Si un ou plusieurs agents étaient déjà en train de parler, alors cette transition ne peut s'activer.

C'est pourquoi il existe deux autre transitions depuis l'état *want-to-speak* vers l'état *compete-for-the-speech* (l'agent parle en même temps qu'un autre). La première transition s'active si un des agents actuellement en train de parler est en train de donner la parole à l'agent (cet autre agent est dans l'état *end-of-speech* et il s'adresse à l'agent). Si cette transition ne peut s'activer, l'agent peut essayer de prendre part à la conversation si ses attitudes lui permettent soit d'essayer d'interrompre les autres agents (en exprimant de la dominance par exemple) soit de parler en même temps qu'eux (en exprimant de l'amicalité par exemple). La fonction qui active cette transition se base sur la liste des autres agents actuellement en train de parler et sur les attitudes que l'agent leur exprime. Soit $Stat_{sp}$ et Aff_{sp}, les valeurs moyennes de Statut et d'Affiliation exprimées envers les locuteurs courants. Une personne qui exprime de la dominance interrompt les autres plus facilement [Goldberg, 1990]. Exprimer de l'amicalité ou de l'hostilité entraine la possibilité de parler en même temps que les autres [Goldberg, 1990, O'Connell et al., 1990]. Par conséquent, la fonction $f_{wa,sp}$ (de l'état *want-to-speak* vers l'état *compete-for-the-speech*) est la suivante :

$$f_{wa,sp}(Stat_{sp},Aff_{sp}) = \begin{cases} 1 & if(Stat_{sp} + |Aff_{sp}| > 0) \\ 0 & else \end{cases} \tag{5.2}$$

Pour que cette fonction active la transition (renvoie 1), il faut donc que l'agent exprime suffisamment de dominance ou d'amicalité ou d'hostilité envers les autres agents en train de parler pour prendre la parole en même temps qu'eux.

Entre les états *own-the-speech* et *compete-for-the-speech*, deux transitions permettent de passer de l'un à l'autre si (au moins) un autre agent ou non parle en même temps que l'agent. Cela permet à l'agent, quand il entre dans l'état *compete-for-the-speech*, de regarder l'autre agent qui lui dispute le tour si nécessaire. En entrant la première fois dans un de ces états, le modèle choisi aléatoirement à qui s'adresse l'agent en fonction de l'amicalité exprimée envers les autres membres du groupe. Ceci est une simplification qui ne tient pas en compte de l'objectif du dialogue puisque nous ne nous intéressons pas au contenu verbal.

Depuis l'état *compete-for-the-speech* vers l'état *interrupted* (lorsque l'agent s'est fait interrompre), le processus est similaire à celui de la fonction $f_{wa,sp}$ mais cette fois ci, le test est inversé, car l'agent n'essaye plus d'interrompre mais essaye de ne pas être interrompu. Par conséquent, notre fonction $f_{sp,in}$ est la suivante :

$$f_{sp,in}(Stat_{sp},Aff_{sp}) = \begin{cases} 1 & if(Stat_{sp} + |Aff_{sp}| \leq 0) \\ 0 & else \end{cases} \tag{5.3}$$

Une fois dans l'état *interrupted*, l'agent interrompt son tour de parole et regarde le ou les agents qui l'ont interrompu.

Depuis les états *compete-for-the-speech* et *own-the-speech* vers *end-of-speech*, les transitions s'activent à la fin du tour de parole de l'agent, lorsqu'il n'a plus rien à dire. Cela veut dire qu'il ne se sera pas fait interrompre.

Enfin, depuis les états *interrupted* et *end-of-speech* vers *unaddressed-hearer* et *addressed-hearer*, ces transitions s'activent après avoir produit les comportements associés à ces états. Dans l'état *interrupted*, l'agent interrompt son tour de parole et regarde le ou les autres agents qui l'ont interrompu. Dans l'état *end-of-speech*, l'agent se tourne vers l'autre agent à qui il donne le tour de parole.

Dans cette section, nous avons présenté notre modèle de prise de parole influencé par l'expression d'attitudes sociales. Nous nous sommes inspirés des travaux existants en modélisation des mécanismes de tours de parole [Thórisson et al., 2010] et de la littérature en sciences humaines et sociales sur le sujet des tours de parole [Sacks et al., 1974][Clark, 1996] et de l'expression d'attitude [Goldberg, 1990][Cappella et al., 1985] afin de proposer un modèle permettant à un agent participant à une conversation de prendre ou de laisser la parole en fonction des attitudes qu'il exprime. Ce modèle s'inscrit dans une architecture plus globale. Dans la section

suivante, nous décrivons le composant en charge de modifier les comportements non-verbaux de l'agent relatifs à la formation de groupe, aux états conversationnels et aux attitudes sociales.

5.3 Le modèle d'influence de l'attitude sur le comportement non-verbal de cohésion de groupe

Nous présentons maintenant un modèle permettant d'influencer le comportement non-verbal utilisé dans le maintien de la cohésion de groupe. La cohésion de groupe est l'ensemble des moyens mis en œuvre par ses membres pour maintenir l'existence de celui-ci. Afin de concevoir notre modèle, nous avons d'abord cherché à identifier les comportements utilisés par les humains pour maintenir cette cohésion. Comme vu dans le chapitre 2, Kendon [Kendon, 1990], Scheflen [Scheflen and Ashcraft, 1976] et Hall ont décrit différents modèles de gestion de la distance entre les personnes. Kendon décrit comment les gens se réunissent autour d'un espace commun. Il explique que les personnes dans un groupe se trouvent arrangés autour d'un cercle et que leur corps est orienté vers le centre de ce cercle. Il souligne également l'importance des comportements de regards qui ont lieu durant un échange. Scheflen propose de voir plus loin que ce cercle et propose de découper cet espace commun (l'intérieur du cercle) aux membres en différentes zones. Ces zones sont concentriques et permettent d'identifier plus précisément quel espace est utilisé pour les membres du groupe et quel espace est utilisé pour les non-membres. Hall, quant à lui, s'est intéressé aux espaces autour de la personne. Il explique par exemple qu'une personne laisse entrer une autre personne dans sa zone proche si celle-ci est un amie ou de la famille. Il apparait alors clairement que les distances que l'on met entre soi et les autres sont un indice fort quant à la relation exprimée envers eux. Par conséquent, les propriétés du comportement de cohésion de groupe que nous considérons dans l'expression d'attitudes sont *la distance entre les membres du groupe, l'orientation du corps et les direction du regard*. Mehrabian [Mehrabian, 1969] dans ses travaux a d'ailleurs cherché les différences de comportements en fonction des attitudes sur ces propriétés. Il explique par exemple qu'une personne dominante reçoit plus d'attention, de regards alors que l'on tourne son corps plus directement vers une personne soumise. Duncan [Duncan, 1972] explique comment les comportements de regard sont utilisés dans une conversation afin de souligner les débuts et les fins de phrases ou de tours

FIGURE 5.4 : A gauche, un groupe sans influence de l'attitude dans sa formation. A droite, les agents expriment différentes attitudes et cela se traduit dans la formation de parole.

Le modèle de cohésion de groupe a donc pour objectif de déterminer les distances que l'agent souhaite entre lui et chaque autre membre du groupe, vers qui il souhaite s'orienter et qui il souhaite regarder (voir figure 5.4). Cependant, celui-ci doit également prendre en compte le fait que l'orientation du corps et la direction du regard peuvent être utilisées par le modèle de prise de parole. Si l'agent souhaite s'adresser à quelqu'un ou lui donner le tour de parole, il doit être en mesure de pouvoir le regarder et se tourner vers lui [Duncan, 1972]. Nous devons donc nous assurer que notre modèle puisse gérer des situations de conflits sur ces comportements. Pour cela nous avons décidé de donner la priorité à la réalisation de la fonction communicative associée à l'état conversationnel de l'agent, c'est-à-dire au modèle de prise de parole.

Notre modèle réalise le calcul des valeurs pour les propriétés du comportement de cohésion de groupe en deux étapes successives. Dans la première étape, le modèle calcule les valeurs pour nos trois propriétés (distance, orientation et regard) afin d'exprimer les attitudes et ce sans prendre en compte l'état conversationnel de l'agent. Il s'agit alors du comportement de base de l'agent, réalisant uniquement l'expression d'attitudes sociales. Dans la seconde étape, immédiatement après la première, le modèle laisse le module de prise de parole décider de vers qui l'agent souhaite s'orienter ou qui il souhaite regarder ou les deux afin de réaliser les comportements liés à la prise de parole, en particulier lorsque l'agent prend la parole, souhaite la prendre, la donne ou est interrompu.

5.3.1 Première étape : comportements de groupe exprimant l'attitude sociale

Contrairement au modèle décrit dans le chapitre 4, nous ne nous sommes pas basés sur une méthode d'apprentissage pour construire ce modèle mais avons choisi une méthode basée sur des règles. Nous avons vu précédemment qu'il existe une grande variété de configurations de groupe possibles (nombres d'individus, attitudes entre eux, placement dans le groupe...) et une collecte de données capable de prendre en compte toutes ces variables serait très couteuse. C'est pourquoi le modèle que nous proposons se base sur des règles de haut niveau, conçues à partir des résultats décrits dans la littérature en Sciences Humaines et Sociales. Les règles de ce premier composant ainsi que leur fonctionnement sont décrites ci-dessous.

D'après Hall [Hall, 1969], Les zones personnelles et sociales d'un individu, qui représentent les zones où ses interactions sociales ont lieu, s'étendent entre 0,46m et 2,1m depuis celui-ci. Le paramètre de distance interpersonnelle de notre modèle prend des valeurs entre cet intervalle. Un facteur d'interpolation α est calculé pour cet intervalle en se basant sur la dominance et l'appréciation exprimées. Une différence de statut (une expression de dominance ou de soumission) entraine une plus grande distance entre deux personnes alors que de l'amicalité entraine une plus petite distance [Mehrabian, 1969][Cristani et al., 2011]. L'hostilité n'est pas reportée comme étant un facteur dans la modification des distances. Soit Ag_i un agent et soit $Stat_{ij}$ et Aff_{ij} le statut et l'affiliation exprimés par cet agent envers un autre agent Ag_j, la fonction $d_i(Ag_j)$ pour calculer la distance interpersonnelle qu'un agent Ag_i souhaite avec un agent Ag_j est :

$$d_i(Ag_j) = 0.46 + \alpha(2.1 - 0.46) \text{ avec } \alpha = \frac{(|Stat_{ij}| + (\frac{1-Aff_{ij}}{2}))}{2} \tag{5.4}$$

Pour la direction du regard, nous calculons un potentiel pour un agent Ag_i de regarder tous les autres agents Ag_j. Ensuite, ces potentiels sont utilisées pour choisir aléatoirement un seul agent à regarder. Nous effectuons un processus similaire pour déterminer vers quel autre agent Ag_j un agent Ag_i oriente son corps. Dans [Mehrabian, 1969] et [Cristani et al., 2011], il est montré qu'un individu qui regarde plus un autre est perçu comme exprimant de la dominance ou de l'amicalité. Cependant il est aussi expliqué que trop de regards tendent à baisser cette perception [Mehrabian, 1969]. Nous avons voulu reproduire cet effet en utilisons une interpolation

trigonométrique, en utilisant la fonction Sinus, qui fait décroitre le potentiel calculé lorsque les attitudes tendent vers leur maximum (dominante et amicale). La fonction $p_{gi}(Ag_j)$ pour calculer le potentiel d'un agent Ag_i de regarder un autre agent Ag_j est la suivante :

$$p_{gi}(Ag_j) = norm(1 + \frac{sin(\frac{1}{2} + 2 \times Aff_ij) + sin(\frac{1}{2} + 2 \times (-Stat_ij))}{2}) \quad (5.5)$$

La fonction $norm()$ est la normalisation de [0,2] vers [0,1].

Pour l'orientation du corps, nous considérons uniquement un effet de la dimension de Statut. Mehrabian avance l'idée qu'une personne se tourne vers la personne envers qui elle exprime le plus de soumission [Mehrabian, 1969]. Par conséquent, nous faisons une simple conversion du niveau de Statut exprimé pour déterminer le potentiel pour un agent de s'orienter vers un autre. La fonction $p_{bi}(Ag_j)$ pour calculer le potentiel d'un agent Ag_i de s'orienter vers un agent Ag_j est la suivante :

$$p_{bi}(Ag_j) = \frac{1 - Stat_ij}{2} \quad (5.6)$$

Cette fonction renvoie des valeurs dans l'intervalle [0,1].

Nous avons montré comment nous calculons des valeurs pour les paramètres de distances interpersonnelles, d'orientation du corps et la direction du regard d'un agent conversationnel membre d'un groupe et exprimant différentes attitudes envers les autres membres. Ce modèle est basé sur des travaux de la littérature [Mehrabian, 1969] et sur des travaux en agents conversationnels animés [Cristani et al., 2011][Pedica et al., 2010][Damian et al., 2011].

Dans la section suivante, nous décrivons la deuxième étape du processus de calcul qui permet de récupérer les modalités du regard et de l'orientation du corps pour réaliser les fonctions communicatives de prise de parole.

5.3.2 Seconde étape : comportement de groupe exprimant l'état conversationnel

Dans la section 5.2.2, nous avons décrit les différents états conversationnels et comment un agent passe de l'un à l'autre. Ici, nous présentons les différents comportements liés à ces états et comment ils viennent compléter les comportements reflétant les attitudes sociales. Ces comportements peuvent entrer en conflit avec

les comportements exprimant les attitudes attitudes ; conflit entendu comme intervenant au niveau des modalités impliquées dans ces comportements de différentes fonctions communicatives. Nous proposons de donner la priorité aux comportements permettant la réalisation des fonctions communicatives de prise de parole. Après la première étape qui a produit des valeurs pour les paramètres du comportement exprimant différentes attitudes, la seconde étape que nous décrivons ici manipule l'orientation du corps et la direction du regard afin de permettre les comportements associés aux états conversationnels (se tourner et regarder son interlocuteur par exemple). Cette seconde étape peut également permettre de déclencher des expressions faciales et des gestes correspond à l'état conversationnel dans lequel se trouve l'agent en utilisant le premier modèle de comportement décrit dans le chapitre 4.

Pour commencer, dans l'état *unaddressed-listener*, la seconde étape ne vient pas modifier les valeurs. Seul la première étape de génération de comportements de groupe influencés par les attitudes a produit des valeurs pour les paramètres du comportement (c'est-à-dire les distances interpersonnelles, l'orientation du corps et la direction du regard fonctionnent comme décrits dans la section 5.3.1).

Les différents comportements pour chaque autre état sont décrits dans la suite. En entrant dans l'état *want-to-speak*, le modèle choisi pour l'agent de regarder le locuteur actuel pour lui indiquer le désir de prendre la parole [Duncan, 1972].

Dans les états *own-the-speech* ou *compete-for-the-speech*, qui correspondent aux états dans lesquels l'agent est en train de parler, le composant *Behavior Planner* reçoit du modèle de prise de parole une phrase à réaliser (un auditeur et une suite de mots à prononcer). En fonction de cette phrase et des attitudes exprimées entre les membres du groupe, il va générer le comportement non-verbal correspondant aux attitudes grâce au modèle décrit dans le chapitre 4 et va choisir comme cible pour l'orientation du regard et du corps, l'auditeur [Mehrabian, 1969].

Dans l'état *compete-for-the-speech*, la cible pour la direction du regard est l'agent locuteur vers lequel est exprimé l'attitude la plus soumise [Mehrabian, 1969].

Dans les états *own-the-speech* et *end-of-speech*, la cible pour la direction du regard est l'auditeur [Mehrabian, 1969] [Duncan, 1972]. Dans l'état *interrupted*, le comportement de l'agent est interrompu et la cible du regard et d'orientation du corps deviennent le locuteur actuel vers lequel est exprimé le plus de soumission [Goldberg, 1990].

Dans cette section, nous avons présenté comment fonctionne le composant permettant d'utiliser les modalités d'orientation du corps et de la direction du regard dans la réalisation des fonctions communicatives de la prise de parole. Ce modèle s'appuie sur la littérature en Sciences Humaines et Sociales [Mehrabian, 1969][Duncan, 1972][Goldberg, 1990]. Dans la section suivante, nous décrivons l'implémentation du modèle complet, à partir du modèle décrit dans le chapitre 4 et du modèle présenté dans ce chapitre.

5.4 L'IMPLÉMENTATION DU MODÈLE DANS UN SYSTÈME INTERACTIF

Dans les sections précédentes nous avons vu l'architecture globale de notre agent, composée principalement de trois modèles de génération de comportements. Deux modèles ont été présentées dans ce chapitre, le modèle de prise de parole ainsi que le modèle de comportement de cohésion de groupe. Le troisième modèle d'attitudes sociales a été présenté dans le chapitre précédent. L'ensemble des comportements produits par ces trois modèles simultanément permet à des agents équipés de cette architecture et membres d'un groupe de participer à une conversation dans laquelle ils sont capables d'exprimer différentes attitudes envers leurs interlocuteurs.

L'implémentation du système complet a été réalisée grâce au moteur de jeux vidéo Unity3D et de deux autres systèmes, la plateforme d'agent virtuel VIB/Greta [Pecune et al., 2014] et la bibliothèque de comportement de groupe Impulsion AI, développée par Pedica et ses collègues. [Pedica et al., 2010].

Impulsion est une bibliothèque de scripts et d'outils pour le moteur de jeux vidéo Unity3D permettant la simulation dans une scène 3D d'un groupe de personnages virtuels. Le modèle d'Impulsion se base sur les travaux de Kendon [Kendon, 1990], Scheflen [Scheflen and Ashcraft, 1976] et Hall [Hall, 1969] et permet à des personnages virtuels de se réunir en de petites formations comme celles qui nous intéressent dans ce travail de thèse. En utilisant Impulsion, on peut créer un groupe d'agents qui réagissent dynamiquement au changement du nombre de participants dans le groupe afin de se repositionner et conserver la formation. Dans la version originale d'Impulsion, il n'y a aucune considération pour l'expression d'attitude et tous les agents se trouvent à la même distance du centre du groupe, donnant à leur rassemblement une forme totalement circulaire. En travaillant de concert avec les auteurs

de cette bibliothèque (Hannes Vilhjálmmson et Claudio Pedica), nous avons étendu Impulsion afin de prendre en compte l'expression d'attitude dans cette formation. Désormais, en plus de la liste des entités membres du groupe, la bibliothèque de scripts Impulsion peut aussi prendre en compte les attitudes entre chaque membre du groupe. Cette nouvelle version du groupe prend en compte ces attitudes pour calculer les distances interpersonnelles désirées entre chaque membre suivant notre modèle et Impulsion se charge de placer les personnages du groupe afin de respecter au mieux ces distances.

Le moteur de comportement d'Impulsion distribue les directions de regards et d'orientations d'un agent vers tous les autres suivant un système de règles paramétrables. Dans la configuration par défaut d'Impulsion, les règles sont paramétrées de telle sorte que les regards et les orientations sont distribués de façon équilibrée. Cependant, il est possible de les paramétrer de telle sorte que les agents se mettent à regarder un agent ou un avatar (contrôlé par un utilisateur) qui viendrait rejoindre le groupe par exemple [Pedica and Vilhjálmsson, 2012]. Cela permet à des agents animés par Impulsion de simuler des comportements de cohésion de groupe. Nous avons étendu alors cette fonctionnalité pour pouvoir spécifier une cible (un autre agent ou avatar) pour calculer la direction du regard. Cependant, dans Impulsion, il n'y avait pas de distinction entre la direction du regard et l'orientation du corps. Un agent qui en regarde un autre, tourne la tête et le torse en même temps. Nous avons donc séparé les modalités du torse et de la tête afin de pouvoir les orienter différemment et nous avons permis à notre système de spécifier une cible pour le torse et une autre pour le regard à Impulsion.

Impulsion fournit son propre jeu de personnages 3D à utiliser dans le système. Nous avons alors intégré les personnages de la plateforme d'agent VIB/Greta dans le système Impulsion de telle manière qu'ils puissent être animés par Impulsion et par le *Behavior Planner* que nous avons conçu à partir du réseau Bayésien présenté dans le chapitre 4, simultanément.

Enfin, nous avons réalisé l'implémentation du modèle de prise de parole dans Unity3D. Ce modèle de prise de parole, en fonction de l'état conversationnel de l'agent et de ses attitudes, envoie des phrases à produire à VIB/Greta et peut spécifier l'interlocuteur à Impulsion afin de déterminer vers qui le locuteur doit s'adresser et ainsi obtenir les orientations du corps des agents. Dans cette implémentation,

nous avons construit un répertoire de phrases fictives, afin que les agents puissent produire du son sans se soucier du contenu verbal de leur discours. Ces phrases fictives sont des séquences de syllabes qui ne forment pas de contenu intelligible que nous utilisons afin de pouvoir simuler le déroulement d'une conversation.

Le système implémenté permet à un utilisateur de naviguer dans la scène en contrôlant un avatar à la première ou à la troisième personne. Bien que le modèle permette en théorie d'intégrer l'utilisateur à la conversation par une simple détection de son activité de parole, nous ne l'avons pas pris en compte pour l'instant puisque nous ne nous sommes pas intéressés au contenu verbal. Par conséquent l'utilisateur peut rejoindre un groupe et les agents lui feront de la place pour qu'il fasse parti de la formation mais il ne peut être qu'un simple observateur de la conversation.

Dans ce chapitre, nous avons proposé une architecture pour un agent conversationnel, incluant différents modèles de comportements, lui permettant de produire des comportements non-verbaux, de prise de parole et de cohésion de groupe exprimant différentes attitudes sociales. Ce modèle permet de définir des attitudes différentes pour chaque interlocuteur de l'agent et nous avons proposé une approche pour éviter les conflits, qui peuvent émerger lorsqu'une même modalité est utilisée pour différentes fonctions communicatives, en donnant la priorité au modèle de prise de parole.

Dans le chapitre suivant, nous présenterons les deux évaluations que nous avons conduites afin d'évaluer notre modèle.

Résumé du chapitre

- Une architecture pour un agent membre d'un groupe conversationnel, lui permettant de prendre en compte les attitudes qu'il exprime à travers sa production de comportement, a été présentée.
- Cette architecture inclut un modèle de prise de parole, un modèle de production de comportements pour la formation de groupe et le modèle de comportement présenté dans le chapitre précédent.
- L'intégration et l'implémentation de cette architecture ont été réalisées au sein de la plateforme Unity3D, à l'aide des outils Impulsion et VIB/Greta.

6 Évaluations du modèle de génération de comportements

Lorsque l'on propose un modèle computationnel afin de reproduire chez des agents virtuels certaines fonctions communicatives du comportement humain, une évaluation de celui-ci doit être réalisée afin de vérifier que ces fonctions sont correctement exprimées par les agents et bien reconnues par les utilisateurs. Dans ce chapitre, nous présentons les deux évaluations que nous avons menées afin de vérifier la validité du modèle. Les évaluations proposées reposent chacune sur une étude perceptive : chaque participant a dû observer une série de scènes générées par notre modèle. La première évaluation a été conduite en ligne, les participants regardant des vidéos montrant un groupe animé par notre modèle sur leur terminal personnel. C'est généralement à travers leur écran d'ordinateur, de téléphone ou de tablette que sont présentées les applications aux utilisateurs et nous cherchons à vérifier que les utilisateurs puissent reconnaitre les attitudes dans ces conditions. Cependant, on voit apparaitre aujourd'hui de nombreux systèmes de réalité virtuelle qui propose un degré d'immersion dans les univers virtuels plus élevés. C'est pourquoi nous avons également conduit une deuxième évaluation dans laquelle nous avons mis les participants au cœur d'un système immersif de réalité virtuelle dans lequel ils étaient membres d'un groupe animé par notre modèle afin de vérifier que les attitudes étaient toujours bien reconnues dans cette configuration. Pour cela nous avons utilisé comme système de réalité virtuelle un système CAVE : il s'agit d'un environnement physique constitué de 4 parois (une au sol, une au fond et deux sur les côtés) sur lesquelles sont affichées des projections stéréoscopiques de la scène 3D de telle façon qu'un utilisateur portant des lunettes stéréoscopiques et se trouvant sur la paroi du sol à l'illusion de se trouver au milieu de la scène 3D (voir figure 6.1). La position des lunettes de l'utilisateur correspond au point de vue de la scène et est suivie continuellement afin de mettre à jour en temps réel les points de vue à projeter sur chaque paroi pour maintenir l'illusion de l'immersion.

FIGURE 6.1 : Le système immersif CAVE

Dans ce chapitre, nous décrivons les deux évaluations et comment elles ont implémentées et conduites. Dans une première section, nous expliquons plus en détail les objectifs de ce travail d'évaluation. Dans une deuxième section, nous présentons la méthodologie sur laquelle nous nous reposons pour concevoir nos évaluations. Dans une troisième section, nous décrivons l'implémentation du système interactif à partir duquel nous avons généré les scènes à évaluer. Dans une quatrième section nous décrivons d'abord l'évaluation que nous avons conduite en ligne et discutons des résultats. Dans la section suivante nous nous intéressons à l'évaluation dans l'environnement immersif. Enfin, dans une sixième section, nous discutons les résultats des deux évaluations et les comparons.

6.1 LES OBJECTIFS DES ÉVALUATIONS

Le travail présenté dans cette thèse a mené à la conception d'un modèle permettant de générer le comportement non-verbal d'un agent membre d'un groupe afin que celui-ci puisse exprimer différentes attitudes au sein d'une conversation simulée. Or il convient de vérifier si le modèle proposé est valide, c'est-à-dire est-ce qu'il permet bien de représenter ces attitudes ? Premièrement, notre travail s'est appuyé sur la littérature en sciences humaines et sociales sur les comportements humains. Dès lors, il est important de vérifier si ces mêmes comportements reproduits par des personnages virtuels sont toujours bien reconnus comme exprimant les attitudes voulues.

Deuxièmement, les interfaces avec les environnements 3D ne cessent d'évoluer. Aujourd'hui plus que jamais, les écrans font partie de notre quotidien et il est devenu beaucoup plus naturel pour n'importe qui, et pas seulement un petit groupe d'initiés, d'interagir avec un environnement virtuel. De plus, l'on voit apparaitre de plus en plus d'applications en réalité virtuelle, notamment grâce à l'arrivée sur le marché de casque comme l'Oculus Rift, ou des écrans auto-stéréoscopiques qui permettent d'avoir un degré d'immersion supérieur aux affichages traditionnels [Bowman et al., 2007].

C'est pourquoi notre évaluation a les objectifs suivants. En premier lieu nous cherchons bien évidemment à vérifier que les attitudes exprimées par un agent utilisant notre modèle sont bien reconnues. Cependant cette vérification se fera sur deux niveaux. Nous testerons d'abord le système sur un affichage traditionnel, de type

écran d'ordinateur, grâce à une évaluation que nous conduirons à travers le navigateur Internet des participants. Puis nous testerons le système d'un environnement immersif, plongeant ainsi les participants au cœur de la scène.

Dès lors, nous nous posons les questions suivantes :

- Est-ce que les attitudes, exprimées à travers le comportement non-verbal des agents, vont être correctement reconnues par les participants ?
- Est-ce qu'il y a des différences entre un affichage sur écran et un système immersif dans la capacité des participants à reconnaitre ces attitudes ?
- Est-ce que le genre des agents ou des participants influent sur la perception des attitudes ?

Au travers des évaluations que nous avons conduites et que nous présentons dans ce chapitre, nous répondrons à ces questions. Dans la section suivante, nous présentons la conception de nos évaluations.

6.2 La conception des évaluations

Une des problématiques dans l'évaluation d'un tel système est le nombre de configurations possibles. La taille du groupe, sa composition, les attitudes entre chaque membre sont autant de facteurs pouvant mener à de nombreuses configurations différentes. C'est pourquoi nous nous restreignons à un sous-ensemble des configurations possibles, qui reste riche et varié, en proposant une méthodologie qui se focalise sur l'étude des interactions entre les attitudes de deux participants simultanément.

Tester toutes les configurations de groupe possibles (différentes tailles de groupe, compositions du groupe et attitudes entre les membres) est une tâche impossible car le nombre de configurations est très large. De plus, cela ne présente qu'un intérêt limité. Tout d'abord parce que lorsque des individus se rassemblent en groupe, ils n'expriment pas constamment des attitudes différentes envers tous les autres membres du groupe. De ce fait, tester exactement toutes les configurations possibles se révèlent être un travail très couteux et peu pertinent. C'est pourquoi nous avons décidé de réduire le nombre de conditions possible en simplifiant la portée de notre étude. Nous avons choisi de représenter un rassemblement typique que l'on pourrait trouver dans les applications mentionnées dans le chapitre 1 et avons fixé le nombre de membres à quatre, deux agents masculins et deux agents féminins. Ceci nous permettant tout de même d'équilibrer entre les genres et l'apparence des agents.

Potentiellement, un groupe de 4 agents qui expriment chacun différentes attitudes, représentées sur deux dimensions (i.e. Statut et Affiliation), envers tous les autres membres impliquent également un nombre exponentiel de possibilités. C'est pourquoi nous simplifions ici aussi en utilisant une échelle discrète (nous ne considérons que les extrêmes de chaque axe) : on obtient ainsi 3^{24} configurations possible pour un groupe de 4 agents.

Ce nombre étant toujours très élevé, nous avons cherché comment nous pourrions réduire le nombre de conditions. Nous avons alors décidé de nous appuyer sur la théorie de la Complémentarité Interpersonnelle (CI) [Sadler and Woody, 2010]. Cette théorie stipule que pour qu'une interaction entre deux personnes puisse perdurer, celles-ci doivent exprimer des attitudes opposées sur l'axe de Statut (par exemple Dominant et Soumis), ou des attitudes similaires sur l'axe d'Affiliation (par exemple Amicale et Amicale). Il s'agit là d'attitudes dites *complémentaires*. Des attitudes opposées sur l'axe d'Affiliation et des attitudes similaires sur l'axe de Statut sont appelées des attitudes *anti-complémentaires*. Nous avons alors décidé d'étudier cet effet et de ne considérer que des conditions où tous les agents expriment des attitudes neutres envers les autres, sauf deux agents (côte à côte, que nous appelons Agent Gauche et Agent Droite) qui expriment également des attitudes spécifiques l'un envers l'autre, complémentaires et anti-complémentaires. Cette méthodologie nous permet de tester si la perception de l'attitude d'un agent en interaction avec un autre est influencée par l'attitude de son interlocuteur. Comme nous considérons 12 permutations du groupe possibles et 4 situations de complémentarités des attitudes sur chacun des 2 axes, nous obtenons 96 configurations possibles.

Nous avons alors décidé de conduire deux études en parallèle au sein de chaque évaluation, l'une s'intéressant uniquement à des attitudes exprimées sur l'axe de Statut et l'autre sur l'axe d'Affiliation, ceci afin de pouvoir tester les situations complémentaires ainsi que les situations anti-complémentaires. Dans chacune des évaluations, ces deux études sont nommées respectivement *Étude Statut* et *Étude Affiliation*.

6.2.1 Hypothèses

Au regard des questions de recherche que nous nous posons dans ce travail, nous avons formulé les hypothèses suivantes :

- H1.s : L'Agent Gauche est perçu(e) comme plus dominant(e) lorsqu'il(elle) exprime de la dominance que lorsqu'il(elle) exprime de la soumission.
- H1.a : L'Agent Gauche est perçu(e) comme plus amical(e) lorsqu'il(elle) exprime de l'amicalité que lorsqu'il(elle) exprime de l'hostilité.
- H2.s : L'Agent Droite est perçu(e) comme plus dominant(e) lorsqu'il(elle) exprime de la dominance que lorsqu'il(elle) exprime de la soumission.
- H2.a : L'Agent Droite est perçu(e) comme plus amical(e) lorsqu'il(elle) exprime de l'amicalité que lorsqu'il(elle) exprime de l'hostilité.
- H3 : Prenant en compte la théorie CI, les participants devraient reconnaitre les attitudes plus facilement dans les conditions complémentaires que dans les conditions anti-complémentaires.

Nous vérifierons que ces hypothèses se vérifient à travers les deux évaluations, celle conduite en ligne à travers le navigateur Internet des participants et celle conduite dans le système immersif. Dans la section suivante, nous détaillons l'implémentation du système pour réaliser ces évaluations.

6.3 L'IMPLÉMENTATION DU SYSTÈME

Comme décrit dans le chapitre 5, notre modèle complet a été implémenté dans le moteur Unity3D, utilisant la technologie de la plate-forme VIB/Greta et de la librairie Impulsion. C'est pourquoi nous avons développé au sein de ce même moteur notre système final. Ce système permet à un utilisateur de naviguer à la première ou à la troisième personne dans une scène 3D dans laquelle peuvent se trouver des groupes d'agents animés par notre modèle. Les agents prennent la parole et produisent des comportements non-verbaux en fonction de leurs attitudes. Pour éliminer le bais que le verbal peut avoir sur la perception d'attitudes, les agents ne sont capables de produire que des phrases fictives, des séquences de syllabes inintelligibles permettant de simuler le déroulement de la conversation. La scène se présente comme suit. Il s'agit d'une reconstitution d'un square au milieu d'un paysage urbain au centre duquel se trouve un kiosque. Les agents ainsi que l'avatar de l'utilisateur peuvent se trouver n'importe où dans le square et l'utilisateur est libre de déplacer son avatar (voir figure 6.2).

Pour les besoins des évaluations, nous avons créé une scène particulière dans laquelle se trouve un seul groupe d'agents, dont les comportements sont générés par

FIGURE 6.2 : Une vue de la scène de test. L'avatar de l'utilisateur est le personnage en blanc au premier plan

notre modèle, et nous avons également configuré l'avatar de l'utilisateur. Celui-ci désormais ne permet à l'utilisateur de voir la scène qu'à la première personne seulement et sa position et son orientation ont été figées face au groupe afin que l'utilisateur puisse observer le groupe du même point de vue dans toutes les conditions.

Dans la section suivante, nous présentons l'évaluation en ligne.

6.4 L'ÉVALUATION EN LIGNE

Dans cette section, nous décrivons comment nous avons mené l'évaluation en ligne de notre modèle et nous présentons les résultats de celle-ci.

6.4.1 Procédure

Dans chacune des deux études (Statut et Affiliation) de notre évaluation en ligne, nous avons demandé aux participants d'observer des vidéos mettant en scène 4 agents (2 masculins et 2 féminins, chacun avec une apparence bien distincte) dans une conversation de groupe. L'utilisateur est un observateur et les agents sont

FIGURE 6.3 : Une capture d'écran de l'interaction de groupe vue par les participants dans l'évaluation conduite en ligne. Les deux agents centraux étaient nommés respectivement Agent Gauche (l'agent masculin de la paire centrale) et Agent Droite (l'agent féminin).

positionnés d'une façon que deux agents apparaissent sur les côtés et que deux agents apparaissent au centre de l'image (comme montré Figure 6.3). Ce sont à ces deux agents centraux que nous ferons exprimer des attitudes envers l'un et l'autre et qui seront observés par les participants de notre évaluation.

6.4.2 Stimuli

Nous décrivons ici les conditions présentées aux participants à travers les vidéos. Nous avons utilisé notre système implémenté dans Unity3D pour générer les vidéos nécessaires à notre évaluation. Chaque vidéo durait une minute et présentait un groupe de 4 agents en train d'avoir une conversation simulée par notre modèle. Comme nous ne nous intéressons pas au contenu verbal, nous avons conçu un ensemble de phrases inintelligibles avec différentes durées (une attitude soumise ou amicale favorisent les interventions courtes alors qu'une attitude dominante favorise les interventions plus longues) que les agents pouvaient utiliser pour prendre la parole. Un possible biais de l'apparence et du positionnement des agents (par exemple vêtements et genre) peut intervenir. C'est pourquoi nous avons choisi un groupe

de personnages avec des apparences variées ainsi qu'appartenant aux deux genres (masculin et féminin) afin de prendre en compte ces facteurs pour nos conditions en faisant varier les deux agents observés parmi ce groupe. Nous appelons ces personnages Masculin1, Masculin2, Féminin1 et Féminin2. Les quatre personnages du groupe pouvaient être arrangés dans 6 différentes configurations tout en gardant la formation circulaire du groupe. Nous avons nommé ce facteur *GroupArrangement*. Ensuite, au sein d'une configuration, l'agent de gauche et l'agent de droite (ceux exprimant les attitudes, au centre de l'image) pouvaient également être intervertis afin de tester l'expression d'attitude des deux côtés. Nous avons nommé ce facteur *PairPosition*. Nous avions donc 12 différentes configurations de groupe possibles (voir figure 6.4). Même si nous ne nous intéressions qu'aux attitudes exprimées par les deux agents au centre du groupe, les deux agents sur les côtés continuaient de participer à la conversation exprimant et recevant des attitudes neutres avec les autres membres du groupe. Enfin, le dernier facteur pour nos conditions concerne les attitudes elles-mêmes. Dès lors qu'un participant s'est vu assigné un arrangement de groupe et une position pour les deux agents centraux, il observera 4 scènes de ce groupe (4 vidéos donc, plus une vidéo de tutoriel en début de session) dans lesquelles les agents centraux expriment différentes attitudes. Deux scènes ou les agents expriment des attitudes complémentaires et 2 scènes où les agents expriment des attitudes anti-complémentaires. De ces facteurs, nous obtenons 48 conditions pour chaque étude (Statut et Affiliation) : 6 arrangements du groupe $*$ 2 permutations $*$ 4 conditions de complémentarité. La figure 6.5 montre les conditions dans l'étude Statut (en bleu). Nous avons donc conçu 96 vidéos pour couvrir toutes nos conditions plus 12 vidéos (une pour chaque configuration de groupe) servant du tutoriel pour le participant.

Comme nous manipulons les attitudes des agents centraux (Agent Gauche et Agent Droite), nous avons défini les variables indépendantes suivantes :

- **ExpStatusL** correspondant au statut exprimé par l'Agent Gauche dans l'étude Statut (respectivement **ExpAffL** dans l'étude Affiliation).
- **ExpStatusR** correspondant au statut exprimé par l'Agent Droite dans l'étude Statut (respectivement **ExpAffR** dans l'étude Affiliation).

Chaque variable a deux niveaux, *Soumis* et *Dominant* (respectivement *Hostile* et *Amical*).

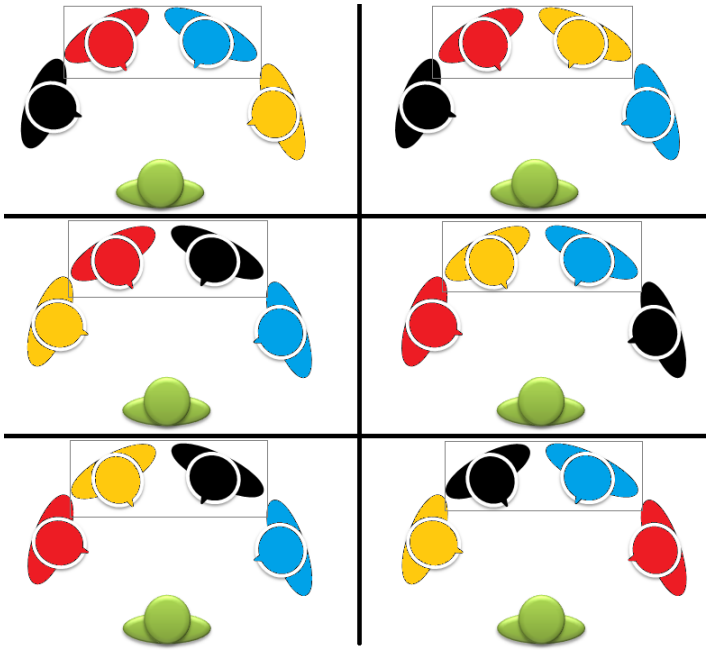


FIGURE 6.4 : Les 6 différentes permutations du groupe d'agent. Les deux agents encadrés sont les agents exprimant les attitudes et dont les positions pouvaient être intervertis pour obtenir 6 permutations supplémentaires

6.4.3 Mesures

Pour chaque vidéo, nous avons demandé aux participants de répondre à un questionnaire afin de mesurer leur perception des attitudes des deux agents. Afin de toucher le plus de participants possibles, nous avons conduit nos évaluations en anglais. Ce questionnaire a été conçu à partir d'adjectifs classés autour de l'*Interpersonal Circumplex* choisi dans [Wiggins et al., 1988] et a d'abord été utilisé dans [Funder et al., 2000]. Nous l'avons adapté afin de poser des questions relatives à nos agents. Pour chacun des deux agents (Agent Gauche et Agent Droite), nous posons 4 questions à propos de son attitude envers l'autre agent utilisant un de ces adjectifs (2 avec une valence positive et 2 avec une valence négative). Pour l'étude Statut, les adjectifs étaient les suivants : en contrôle (*controlling*), peu sûr (*insecure*), dominant (*dominating*) et cherche le réconfort (*looking for reassurance*). Pour la session *Affiliation*, les adjectifs étaient : chaleureux (*warm*), détaché (*detached*), aimable (*likeable*) et hostile (*hostile*). Une des questions était par exemple, "L'Agent Gauche est chaleureux avec l'Agent Droit" (*Left Agent expresses warmth towards the Right*

GroupArrangement	PairPosition	Complémentaire	Anti-Complémentaire
		Ag.Ga. - Ag.Dr.	Ag.Ga. - Ag.Dr.
Masculin1 - **Féminin1 - Féminin2** - Masculin2	Féminin1 - Féminin2	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
	Féminin2 - Féminin1	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
Masculin1 - **Féminin1 - Masculin2** - Féminin2	Féminin1 - Masculin2	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
	Masculin2 - Féminin1	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
Masculin2 - **Féminin1 - Masculin1** - Féminin2	Féminin1 - Masculin1	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
	Masculin1 - Féminin1	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
Féminin1 - **Masculin2 - Féminin2** - Masculin1	Masculin2 - Féminin2	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
	Féminin2 - Masculin2	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
Féminin1 - **Masculin2 - Masculin1** - Féminin2	Masculin2 - Masculin1	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
	Masculin1 - Masculin2	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
Masculin2 - **Masculin1 - Féminin2** - Féminin1	Masculin1 - Féminin2	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis
	Féminin2 - Masculin1	Dominant - Soumis	Dominant - Dominant
		Soumis - Dominant	Soumis - Soumis

FIGURE 6.5 : Les 48 conditions dans l'étude Statut (en bleu). Dès lors qu'un participant à l'étude Statut s'est vu assigner un arrangement de groupe ainsi qu'une position pour les agents centraux, il va observer 4 conditions, 2 complémentaires et 2 anti-complémentaires.

Agent). Nous avons également posé 3 questions à propos du groupe en général afin de mesurer les trois éléments suivants : la crédibilité du groupe (*believability of the group*), l'engagement de ses membres (*engagement of the members*) et sa richesse sociale (*social richness*). Toutes les réponses se faisaient sur des échelles de Likert à 5 points (de "Pas d'accord du tout"/*Completely disagree* à "Complètement d'accord"/*Completely agree*).

Par conséquent, à partir des mesures saisies par les participants, nous avons défini les variables dépendantes suivantes :

- **MeasureStatusL** correspondant à la perception par l'utilisateur du statut de l'Agent Gauche (respectivement **MeasureAffL** dans l'étude Affiliation).
- **MeasureStatusR** correspondant à la perception par l'utilisateur du statut de l'Agent Droite (respectivement **MeasureAffR** dans l'étude Affiliation).

Les variables dépendantes correspondant aux questions relatives au groupe sont les suivantes : **GroupBelievability**, **GroupEngagement** et **GroupSocialRichness**. Nous avons agrégé les réponses aux questions afin de produire une seule valeur normalisée pour chaque variable dépendante dans l'intervalle [0,1].

6.4.4 Hypothèses

Les hypothèses sont ici reformulées à l'aide des variables définies plus haut et nous proposons également une hypothèse supplémentaire relative à nos mesures sur le groupe :

- **H1.s (Agent Gauche)** : La valeur de MeasureStatusL est plus haute quand ExpStatusL est à la valeur Dominant contrairement à la valeur Soumis.
- **H1.a (Agent Gauche)** : La valeur de MeasureAffL est plus haute quand ExpAffL est à la valeur Amicale contrairement à la valeur Hostile.
- **H2.s (Agent Droite)** : La valeur de MeasureStatusR est plus haute quand ExpStatusR est à la valeur Dominant contrairement à la valeur Soumis.
- **H2.a (Agent Droite)** : La valeur de MeasureAffR est plus haute quand ExpAffR est à la valeur Amicale contrairement à la valeur Hostile.
- **H3 (Théorie CI)** : Prenant en compte la théorie CI, les participants devraient reconnaitre les attitudes plus facilement dans les conditions complémentaires que dans les conditions anti-complémentaires.

– **H4 (Groupe)** : Les valeurs pour GroupBelievability, GroupEngagement et GroupSocialRichness devraient être plus haute dans les situations complémentaires que dans les situations anti-complémentaires.

6.4.5 Participants

Dans chaque étude, chaque participant s'est vu attribuer 4 vidéos suivant les facteurs de notre évaluation. Un participant voyait d'abord une page de consentement et un questionnaire démographique (nationalité, âge et genre). Puis, nous lui montrons un tutoriel vidéo où tous les agents expriment une attitude neutre avec une question d'exemple. Enfin nous montrons les 4 vidéos dans un ordre contre balancé, chaque vidéo étant accompagnée par un questionnaire au bas de la page tel que décrit plus haut. A la fin, une page de débriefing est montrée expliquant au participant quels étaient les enjeux de cette évaluation. Nous avons recruté un total de 144 participants à travers des listes de diffusion, 72 dans chaque étude. Chaque vidéo a donc été vue 6 fois. Dans l'étude Statut, 58.34% des participants avaient entre 18 et 30 ans et 50% étaient des femmes, 48.61% étaient des hommes et 1.39% ne se prononcèrent pas. Dans l'étude Affiliation, 66.66% étaient âgées entre 18 et 30 ans, 56.94% étaient des femmes et 43.06% étaient des hommes. Nous avons eu des participants qui provenaient de différents profils culturels mais la majorité étaient de France (47.22% dans la session Statut et 52.68% dans la session Affiliation).

6.5 LES RÉSULTATS DE L'ÉVALUATION EN LIGNE

Dans cette section sont présentés les résultats de l'évaluation en ligne. Comme expliqué précédemment, deux études ont été conduites en parallèle au sein de chacune des évaluations, chacune s'intéressant à un axe de la représentation des attitudes. Afin de vérifier nos hypothèses, nous devons chercher les réponses suivantes :

– Y a-t-il un effet significatif des variables indépendantes sur les variables dépendantes, en d'autres termes, est-ce que changer l'expression d'attitude des deux agents a un effet sur la perception des attitudes des agents ? Et dans quel sens ?

– Y a-t-il une interaction entre les variables indépendantes ? Est-ce que la perception de l'attitude d'un agent influence la perception de l'attitude de l'autre ?

Au vu de ces questions, le test statistique le plus adapté est le test de l'analyse de la variance multivarié MANOVA. Ce test est particulièrement indiqué pour répondre à nos questions dans un design avec plusieurs variables dépendantes. Dans la section suivante, nous détaillons les résultats de notre analyse MANOVA dans l'étude Statut de l'évaluation en ligne.

6.5.1 Étude Statut

Afin de tester la validité des hypothèses H1.s, H2.s et H3.s, nous avons conduit une analyse MANOVA répétée (Double analyse de la variance multivariée). Nos variables dépendantes sont MeasureStatusL et MeasureStatusR et nos variables indépendantes sont ExpStatusL et ExpStatusR. Nous avons trouvé un effet principal de ExpStatusL ($WilksLambda = 0.50$, $F(2,70) = 35.7$, $p < 0.001$) et de ExpStatusR ($WilksLambda = 0.57$, $F(2,70) = 25.7$, $p < 0.001$). Cela signifie qu'exprimer différentes attitudes a bien un effet sur la perception de celles-ci. Aucune interaction significative entre ces deux variables n'a été trouvée. Ce qui veut dire que choisir des situations complémentaires ou anti-complémentaires n'a pas eu d'effet sur la perception des attitudes, rejetant ainsi l'hypothèse **H3**.

Comme l'hypothèse de la sphéricité n'est pas violée, nous avons conduit une analyse complémentaire à l'aide de mesures répétées uni-variées ANOVA pour chacune des deux variables dépendantes. Pour la variable MeasureStatusL, le test ANOVA a confirmé un effet significatif de la variable ExpStatusL ($F(1,71) = 63$, $p = .0$). Plus précisément, l'Agent Gauche a été perçu comme plus dominant lorsque la variable ExpStatusL était au niveau Dominant (M=.56, SE=.02) contrairement au niveau Soumis (M=.43, SE=.01). Pas d'autre interaction n'a été trouvée, par conséquent **H1.s** est **supportée**. Cela signifie que lorsque l'Agent Gauche exprime de la dominance, il est bien perçu comme plus dominant et que lorsqu'il exprime de la soumission, il est bien perçu comme plus soumis, quelque soit son genre et son apparence.

Pour la variable MeasureStatusR, le test ANOVA a confirmé un effet significatif de la variable ExpStatusR ($F(1,71) = 48$, $p = .0$). Plus précisément, l'Agent Droite a été perçu comme plus dominant lorsque la variable ExpStatusR était au niveau Dominant (M=.60, SE=.01) contrairement au niveau Soumis (M=.46, SE=.01). Pas d'autre interaction n'a été trouvée par conséquent **H2.s** est **supportée**. Cela signifie

également que lorsque l'Agent Droite exprime de la dominance, il est bien perçu comme plus dominant et que lorsqu'il exprime de la soumission, il est bien perçu comme plus soumis, quelque soit son genre.

En ce qui concerne les trois mesures relatives au groupe, nous avons conduit trois analyses répétées uni-variées ANOVA. A part pour un effet de la variable ExpStatusR sur la crédibilité du groupe ($p = 0.03$) de petite taille ($\eta_p^2 = .06$), pas d'autre effet significatif n'a été trouvé (tous les $p > .13$). D'ailleurs cet effet n'a pas été trouvé sur l'Agent Gauche. Nous avons conduit une autre analyse MANOVA avec comme facteurs inter-sujets, l'arrangement du groupe et le genre du participant. Nous n'avons pas trouvé d'interactions significatives (tous les $p > .38$).

Dans la section suivante, nous détaillons les résultats de notre analyse MANOVA dans l'étude Affiliation de l'évaluation en ligne.

6.5.2 Étude Affiliation

Nous avons conduit une analyse MANOVA similaire avec comme variables dépendantes MeasureAffL et MeasureAffR et comme variables indépendantes ExpAffL et ExpAffR pour tester la validité des hypothèses H1.a, H2.a et H3.a. Nous avons trouvé un effet principal de la variable ExpAffL ($WilksLambda = 0.42$, $F(2,70) = 48.4$, $p < 0.001$) et de la variable ExpAffR ($WilksLambda = 0.55$, $F(2,70) = 28$, $p < 0.001$). Cela signifie qu'exprimer différentes attitudes a bien un effet sur la perception de celles-ci. Pas d'interaction significative entre ces deux variables n'a été trouvée. Ce qui veut dire que choisir des situations complémentaires ou anti-complémentaires n'a pas eu d'effet sur la perception des attitudes, rejetant ainsi l'hypothèse **H3**.

Comme l'hypothèse de la sphéricité n'est pas violée, nous avons conduit une analyse complémentaire à l'aide de mesures répétées uni-variées ANOVA pour chacune des deux variables dépendantes. En ce qui concerne la variable MeasureAffL, le test ANOVA a confirmé un effet significatif de la variable ExpAffL ($F(1,71) = 93, p = .0$). Plus précisément, l'Agent Gauche a été perçu comme plus amical lorsque la variable ExpStatusL était au niveau Amical (M=.61, SE=.02) contrairement au niveau Hostile (M=.37, SE=.01). Pas d'autre interaction n'a été trouvée par conséquent **H1.a** est **supportée**. Cela signifie que lorsque l'Agent Gauche exprime de l'amicalité, il est bien perçu comme plus amical et que lorsqu'il exprime de l'hostilité, il est bien

perçu comme plus hostile, quel que soit son genre et son apparence.

Pour la variable MeasureAffR, le test ANOVA a confirmé un effet significatif de la variable ExpAffR ($F(1,71) = 54$, $p = .0$). Plus précisément, l'Agent Droite a été perçu comme plus amical lorsque la variable ExpAffR était au niveau Amical (M=.58, SE=.02) contrairement au niveau Hostile (M=.41, SE=.01). Pas d'autre interaction n'a été trouvée par conséquent **H2.a** est **supportée**. Cela signifie que lorsque l'Agent Droite exprime de l'amicalité, il est bien perçu comme plus amical et que lorsqu'il exprime de l'hostilité, il est bien perçu comme plus hostile, quel que soit son genre.

En ce qui concerne les trois mesures relatives au groupe, nous avons conduit trois analyses répétées uni-variées ANOVA. A part pour un effet de la variable ExpAffL sur la perception de l'engagement du groupe ($p = 0.01$) de petite taille ($\eta_p^2 = .09$), pas d'autre effet significatif n'a été trouvé (tous les $p > .12$). Ce petit effet pourrait s'expliquer par le fait que les agents exprimant les attitudes amicale et hostile tentent de prendre la parole plus souvent. Néanmoins cet effet n'a pas été reproduit sur l'Agent Droite. Nous avons également conduit une autre analyse ANOVA avec comme facteurs inter-sujets l'arrangement du groupe et le genre du participant. Nous n'avons pas trouvé d'interactions significatives (tous les $p > .17$).

6.5.3 Discussion

Dans cette évaluation que nous avons conduite en ligne, nous avons recueilli la perception qu'ont eue les utilisateurs des attitudes exprimées par deux agents (Amicale, Hostile, Soumise ou Dominante) grâce à notre modèle. L'analyse des données récoltées a révélé que lorsqu'un agent exprime une attitude, celle-ci est bien perçue comme telle. Nous validons ainsi nos deux hypothèses H1 et H2. Quel que soit le genre et l'apparence de l'agent, les comportements générés par notre modèle ont permis d'exprimer les attitudes voulues. De plus, nous n'avons pas mesuré d'interactions entre les variables, ce qui signifie que chaque attitude (celle de l'Agent Gauche ou celle de l'Agent Droite) est bien reconnue indépendamment de l'autre.

Au cœur de notre design, nous avons intégré la Complémentarité Interpersonnelle dans la configuration des conditions. Nos agents étant engagés dans une conversation avec d'autres agents, il nous a semblé intéressant de vérifier si les attitudes exprimées entre deux agents interagissaient entre elles puisqu'une attitude s'exprime comme un

alignement par rapport à l'autre (voir chapitre 2). Nos mesures n'ont montré aucune interaction (l'hypothèse H3 n'est pas validée). Nos mesures relatives aux groupes n'ont montré aucune différence non plus entre les situations complémentaires et anti-complémentaires (l'hypothèse H4 n'est pas validée). Il est possible que nous n'ayons pas considéré suffisamment de paramètres dans l'expression d'attitudes pour voir apparaitre des différences entre les situations, comme la prosodie ou les *backchannels* (signaux émis par les interlocuteurs pour marquer leur attention, leur accord ou désaccord etc. [Allwood et al., 2012]). De plus, les situations complémentaires sont décrites par la théorie comme favorisant la durée des interactions : nos agents étant engagés dans une interaction ayant une durée indéfinie, ils ne montraient pas de différences dans leur comportement permettant d'identifier leur désir de poursuivre ou d'interrompre les interactions.

Les hypothèses H1 et H2 ayant été validées, et les hypothèses H3 et H4 ayant été non validées, on pourrait conclure que les attitudes ont été bien reconnues lorsqu'elles étaient perçues indépendamment les unes des autres, comme en isolation du reste des agents et de l'interaction elle même, puisque nous n'avons pas mesuré de dépendances entre elles. Cependant, nous devons rester vigilants et préciser que certains comportements ne doivent pas être interprétés en isolation mais par rapport à l'autre, comme la distance, l'orientation ou la prise de parole par exemple. Nous pouvons donc dire que notre modèle a réussi à produire les comportements non-verbaux permettant à un agent de lui faire exprimer différentes attitudes sociales par rapport aux autres agents et que la complémentarité de la situation n'a pas d'influence sur la perception de ces attitudes.

Bien évidemment, ce travail présente certaines limites. L'utilisateur n'est pas pris en compte dans l'interaction de groupe, il n'y a pas de considération du contenu verbal et enfin le design de l'évaluation se limite à cette configuration de 4 agents dont 2 seulement qui expriment des attitudes sur une seule dimension. Notre deuxième évaluation nous permet de commencer à adresser la première limite. Nous introduisons l'utilisateur dans le groupe en le plongeant dans l'environnement avec les agents.

Dans la section suivante, nous conduisons la même évaluation mais cette fois-ci en plaçant les participants dans un environnement immersif de type CAVE.

6.6 L'ÉVALUATION EN ENVIRONNEMENT IMMERSIF

Dans cette section, nous présentons l'évaluation que nous avons conduite dans un environnement immersif de type CAVE. L'objectif de cette seconde évaluation est de vérifier que les attitudes des agents sont toujours bien reconnues alors que l'on passe de vidéos, au contenu prédéfini, à un système immersif fonctionnant en temps réel. Pour cela, nous sommes allés à Brest, au Centre Européen de Réalité Virtuelle (CERV). Nous avons porté notre système, conçu initialement pour fonctionner sur un écran d'ordinateur standard, dans l'environnement CAVE dont les chercheurs disposent au CERV. Grâce à des composants logiciels disponibles pour Unity3D et l'équipement du CERV, cela a été réalisé assez facilement en une semaine. Ce travail a consisté à remplacer la caméra virtuelle standard de la scène Unity3D par un système multi-caméras qui se déplace dans la scène en fonction de la position de l'utilisateur. Ce système multi-caméras consiste en un assemblage de 8 caméras virtuelles, permettant de projeter une image pour l'œil gauche et une image pour l'œil droit sur chacun des 4 écrans du système CAVE.

6.6.1 Procédure

Dans cette évaluation, nous avons conservé le même design que dans l'évaluation Web afin de pouvoir comparer les résultats avec deux niveaux d'immersion différents. Il y a toutefois deux différences. La première est la perspective de l'utilisateur face au groupe. Dans la première évaluation l'utilisateur regardait des vidéos du groupe alors que cette fois-ci, il est placé physiquement au sein du groupe, pouvant même regarder autour de lui et se pencher (dans la limite de ce que permet l'environnement CAVE, voir figure 6.6). Potentiellement, les utilisateurs pouvaient même se déplacer un peu autour du groupe mais nous leur avons demandé de ne pas se déplacer pour reproduire au maximum les conditions de l'évaluation en ligne. La seconde différence est que cette fois-ci les scènes tournent en temps réel et ne sont pas juste des vidéos. Nous avons choisi d'utiliser un design très proche afin de ne pas introduire trop de nouvelles variables et de pouvoir réaliser ainsi une comparaison pertinente des résultats.

FIGURE 6.6 : Position du participant dans l'environnement CAVE

6.6.2 Stimuli

Dans cette évaluation nous ne montrons plus des vidéos aux participants mais nous leur faisons assister à des scènes animées en temps réel par notre modèle d'une durée d'une minute chacune. Pour cela, nous avons équipé les participants d'une simple paire de lunettes stéréoscopiques dotées d'un marqueur pour les suivre dans l'espace, et nous les avons installé dans l'environnement CAVE de telle sorte qu'ils se retrouvent au cœur de la scène. Réaliser une telle évaluation requiert que les participants viennent au laboratoire. Il nous était donc difficile de recruter le même nombre de participants que dans l'étude en ligne (144). C'est pourquoi nous avons réduit nos besoins en ne considérant plus la variable *PairPosition* dans nos facteurs de conditions, celle correspondant à l'échange de position entre l'Agent Gauche et l'Agent Droite, mais uniquement *GroupArrangement* correspondant à la configuration du groupe. De cette façon, nous conservons un design dans lequel le genre des agents placés au centre (les agents étudiés donc) est équilibré. Cela donne tout de même 12 conditions différentes, 6 pour l'étude Statut et 6 pour l'étude Affiliation. Nous avons donc fixé le nombre de participants à 24 afin d'avoir exactement deux participants par conditions. Dans une condition, chaque participant assiste à 4 scènes correspondant à différentes configurations d'attitudes, 2 scènes avec des attitudes complémentaires et 2 scènes avec des attitudes anti-complémentaires de la même façon que dans l'évaluation en ligne. Chaque scène est donc vue 2 fois.

Par conséquent, nous avons défini les variables indépendantes suivantes, de la même façon que dans l'évaluation en ligne :

- **ExpStatusL** correspondant au statut exprimé par l'Agent Gauche dans l'étude Statut (respectivement **ExpAffL** dans l'étude Affiliation).
- **ExpStatusR** correspondant au statut exprimé par l'Agent Droite dans l'étude Statut (respectivement **ExpAffR** dans l'étude Affiliation).

Chaque variable a deux niveaux, *Soumis* et *Dominant* (respectivement *Hostile* et *Amical*).

6.6.3 Mesures

Après chaque scène, nous avons demandé aux participants de répondre au même questionnaire que dans l'évaluation en ligne afin de mesurer leur perception des

attitudes des deux agents. Par conséquent, à partir des mesures saisies par les participants, nous avons défini les variables dépendantes suivantes :

- **MeasureStatusL** correspondant à la perception par l'utilisateur du statut de l'Agent Gauche (respectivement **MeasureAffL** dans l'étude Affiliation).
- **MeasureStatusR** correspondant à la perception par l'utilisateur du statut de l'Agent Droite (respectivement **MeasureAffR** dans l'étude Affiliation).

Nous avons agrégé les réponses aux questions afin de produire une seule valeur normalisée pour chaque variable dépendante dans l'intervalle [0,1]. Dans cette évaluation, nous avons choisi de nous concentrer uniquement sur la perception des attitudes et nous n'avons pas posé de questions sur le groupe. Dans l'évaluation en ligne, seuls deux effets de petites tailles avaient été identifiés entre les attitudes des deux agents et les trois mesures relatives au groupe. Afin de simplifier la tâche des participants de l'expérience immersive, nous avons choisi de retirer ces mesures.

6.6.4 Hypothèses

Nous reprenons les hypothèses H1, H2 et H3 telles que définies dans l'évaluation en ligne. Seule l'hypothèse H4 (correspondant aux mesures sur le groupe) n'est pas reprise.

6.6.5 Participants

Dans chaque étude, chaque participant s'est vu attribué 4 scènes suivant les facteurs de notre évaluation. Un participant devait d'abord remplir un formulaire de consentement avec quelques questions démographiques (nationalité, âge et genre). Puis, nous lui montrons une scène d'exemple où tous les agents expriment une attitude neutre. Enfin nous montrons les 4 scènes dans un ordre contre balancé, chaque scène étant ensuite suivi d'un questionnaire sur les attitudes des agents à remplir sur un ordinateur à l'extérieur du CAVE. A la fin, un document de débriefing est donné au participant expliquant quels étaient les enjeux de cette évaluation. Nous avons recruté un total de 24 participants à travers le personnel et les étudiants de l'École National d'Ingénieurs de Brest, 12 pour chaque étude. Dans l'étude Statut, 58.3% des participants avaient entre 31 et 40 ans, 16.7% étaient des femmes et 83.3% étaient des hommes. Dans l'étude Affiliation, 66.66% étaient âgées entre 21 et 30

ans, 25% étaient des femmes et 75% étaient des hommes. Nous avons eu 2 participants provenant du Liban et de Madagascar, le reste venait de France (100% dans la session Statut et 83.3% dans la session Affiliation).

6.7 Les résultats de l'évaluation en environnement immersif

Dans cette section nous présentons les résultats de l'évaluation que nous avons conduite dans l'environnement immersif. De la même façon que dans l'évaluation en ligne, nous cherchons à savoir si les attitudes ont été bien perçues et s'il a une interaction entre les attitudes perçues de chaque agent.

6.7.1 Étude Statut

Nous avons conservé la même approche pour les mesures de cette seconde évaluation que dans le précédente. Afin de tester les hypothèses H1.s, H2.s et H3.s, nous avons conduit des mesures MANOVA 2x2 sur les variables dépendantes MeasureStatusL et MeasureStatusR avec comme variables indépendantes ExpStatusL et ExpStatusR. Nous avons trouvé un effet principal de la variable ExpStatusL ($WilksLambda = 0.36$, $F(2,10) = 8.9$, $p < 0.05$) et de la variable ExpStatusR ($WilksLambda = 0.29$, $F(2,10) = 12.5$, $p < 0.005$). Cela signifie que varier l'expression des attitudes des agents a bien eu un effet sur la perception de celles-ci. Aucune interaction significative entre ces deux variables n'a été trouvée. Ce qui veut dire que choisir des situations complémentaires ou anti-complémentaires n'a pas eu d'effet sur la perception des attitudes, rejetant ainsi l'hypothèse **H3**.

Comme l'hypothèse de la sphéricité n'est pas violée, nous avons conduit une analyse complémentaire à l'aide de mesures répétées uni-variées ANOVA pour chacune des deux variables dépendantes. Pour la variable MeasureStatusL, le test ANOVA a confirmé un effet significatif de la variable ExpStatusL ($F(1,11) = 18.24$, $p < .005$). Plus précisément, l'Agent Gauche a été perçu comme plus dominant lorsque la variable ExpStatusL était au niveau Dominant (M=.67, SE=.04) contrairement au niveau Soumis (M=.40, SE=.02). Aucune autre interaction n'a été trouvée, par conséquent **H1.s** est **supportée**. Cela signifie que lorsque l'Agent Gauche exprime de la dominance, il est bien perçu comme plus dominant et que lorsqu'il exprime de

la soumission, il est bien perçu comme plus soumis, quel que soit son genre et son apparence.

Pour la variable MeasureStatusR, le test ANOVA a confirmé un effet significatif de la variable ExpStatusR ($F(1,11) = 15.8$, $p < .005$). Plus précisément, L'agent Droit a été perçu comme plus dominant lorsque la variable était au niveau dominant (M=.66, SE=.04) contrairement au niveau soumis (M=.42, SE=.04). Pas d'autre interaction n'a été trouvée par conséquent **H2.s** est **supportée**. Cela signifie également que lorsque l'Agent Droite exprime de la dominance, il est bien perçu comme plus dominant et que lorsqu'il exprime de la soumission, il est bien perçu comme plus soumis, quel que soit son genre et son apparence.

Nous avons conduit une analyse MANOVA supplémentaire avec comme facteurs inter-sujets l'arrangement du groupe et le genre du participant. Aucune interaction significative n'a été trouvée (tous les $p > .07$).

Dans la section suivante, nous détaillons les résultats de notre analyse MANOVA dans l'étude Affiliation de l'évaluation immersive.

6.7.2 Étude Affiliation

Nous avons conduit une analyse MANOVA 2x2 similaire avec comme variables dépendantes MeasureAffL et MeasureAffR et comme variables indépendantes ExpAffL et ExpAffR. Nous avons trouvé un effet principal de la variable ExpAffL ($WilksLambda = 0.44$, $F(2,10) = 6.3$, $p < 0.05$) mais pas de la variable ExpAffR ($WilksLambda = 0.62$, $F(2,10) = 3$, $p > 0.05$). Aucune interaction significative entre les deux variables n'a été trouvée. Ce qui veut dire que choisir des situations complémentaires ou anti-complémentaires n'a pas eu d'effet sur la perception des attitudes, rejetant ainsi l'hypothèse **H3**.

Nous avons également conduit une analyse complémentaire qui s'est intéressé à des mesures répétées uni-variées ANOVA 2x2 pour chacune des variables dépendantes. Pour la variable MeasureAffL, le test ANOVA a confirmé un effet principal de la variable ExpAffL ($F(1,11) = 13.2$, $p < .005$). Plus précisément, l'Agent Gauche a été perçu comme plus amical lorsque la variable ExpStatusL était au niveau Amical (M=.6, SE=.05) contrairement au niveau hostile (M=.44, SE=.04). Pas d'autre interaction n'a été trouvée par conséquent **H1.a** est **supportée**. Cela signifie que lorsque l'Agent Gauche exprime de l'amicalité, il est bien perçu comme plus amical

et que lorsqu'il exprime de l'hostilité, il est bien perçu comme plus hostile, quel que soit son genre et son apparence .

En ce qui concerne la variable MeasureAffR, le test ANOVA a confirmé un effet principal de la variable ExpAffR ($F(1,11) = 6.2$, $p < .05$). Plus précisément, l'Agent Droit a été perçu comme plus amical lorsque la variable ExpAffR était au niveau Amical (M=.57, SE=.05) contrairement au niveau Hostile (M=.41, SE=.04). Pas d'autre interaction n'a été trouvée par conséquent **H2.a** est **supportée**. Cela signifie que lorsque l'Agent Droite exprime de l'amicalité, il est bien perçu comme plus amical et que lorsqu'il exprime de l'hostilité, il est bien perçu comme plus hostile, quel que soit son genre et son apparence.

Nous avons également conduit une analyse MANOVA supplémentaire avec comme facteurs inter-sujets l'arrangement du groupe et le genre du participant. Aucune interaction significative n'a été trouvée (tous les $p > .25$).

6.7.3 Discussion

De la même façon que dans l'évaluation en ligne, nous avons recueilli la perception qu'ont eue les utilisateurs des attitudes exprimées par deux agents (Amicale, Hostile, Soumise ou Dominante) grâce à notre modèle. L'analyse des données récoltées a révélé que lorsqu'un agent exprime une attitude, celle-ci est bien perçue par les participants. Nous validons donc également nos deux hypothèses H1 et H2 dans l'environnement immersif. Quel que soit le genre et l'apparence de l'agent, les comportements générés par notre modèle ont permis d'exprimer les attitudes voulues. De plus, nous n'avons pas mesuré d'interactions entre les variables, ce qui signifie que chaque attitude (celle de l'Agent Gauche ou celle de l'Agent Droite) est bien reconnue indépendamment de l'autre.

Nous avons conservé l'hypothèse selon laquelle les attitudes seraient mieux reconnues dans les situations complémentaires. Nous avons pensé que le niveau d'immersion pourrait avoir un effet sur la perception de l'interaction des attitudes. Cependant, nous n'avons pas mesuré un tel effet dans notre analyse des résultats. Par conséquent l'hypothèse H3 n'est toujours pas vérifiée, même dans un environnement plus immersif. Ceci pourrait être expliqué par le fait qu'en plus des paramètres manquants identifiés dans l'évaluation précédente (prosodie, *backchannels*, expression de la volonté de continuer l'interaction etc.), le fait d'introduire l'utilisateur dans l'es-

pace des agents ajoute de nouvelles contraintes. Bien que les agents fassent de la place pour l'humain dans le groupe (prise en compte spatiale), ils l'ignorent, ne lui adressent pas la parole et parlent un langage qui ne veut rien dire (pas de prise en compte au niveau de la conversation). Les participants ont pu ressentir fortement ce manque d'interaction des agents avec eux et ainsi être moins sensibles à la complémentarité des situations d'interaction.

Les hypothèses H1 et H2 ayant été validées, et l'hypothèse H3 ayant été non-validée, nous pouvons donc dire que notre modèle a réussi à produire les comportements non-verbaux permettant à un agent de lui faire exprimer différentes attitudes sociales par rapport aux autres agents et que la complémentarité de la situation n'a pas d'influence sur la perception de ces attitudes.

Alors que nous cherchions également à adresser la première limite de l'évaluation précédente, c'est-à-dire l'absence d'intégration de l'utilisateur dans l'environnement, ce travail présente toujours certaines limites. Tout d'abord, bien que nous ayons plongé l'utilisateur dans l'environnement, il ne peut toujours pas interagir avec le groupe. Puisque nous n'avons pas pris en compte le contenu verbal et que nos agents parlent dans un langage fictif, il est difficile de définir un scénario dans lequel l'utilisateur s'intègre à cette conversation. De plus, nous ne cherchions pas non plus à intégrer trop de nouvelles variables dans cette deuxième évaluation afin de ne pas trop s'éloigner du design de la première évaluation. D'ailleurs, le design de l'évaluation se limite toujours à cette configuration de 4 agents dont 2 seulement qui expriment des attitudes sur une seule dimension.

6.8 Conclusion

Dans ce chapitre, nous avons présenté deux évaluations de notre modèle ainsi que les résultats de celles-ci. Parce que les configurations possibles de groupes et d'attitudes étaient très nombreuses, nous avons proposé de vérifier la validité de l'expression des attitudes à l'aide d'un design se focalisant sur la complémentarité interpersonnelle de deux agents au sein d'un groupe conversationnel se composant de 4 agents (deux masculins et deux féminins). Nous avons également choisi de conduire ces évaluations à travers un affichage de vidéos dans le navigateur Internet et à travers un système immersif de réalité virtuelle de type CAVE, ceci afin d'étudier d'éventuelles différences dans la perception d'attitude selon que l'utilisateur soit

observateur de la scène ou immergé dans celle-ci.

A travers ces évaluations, nous avons cherché à répondre aux questions de recherches suivantes : Est-ce que les attitudes, exprimées à travers le comportement non-verbal des agents, vont être correctement reconnues par les participants ? Est-ce qu'il y a des différences entre un affichage sur écran et un système immersif dans la capacité des participants à reconnaitre ces attitudes ?

Les résultats ont montré que les attitudes des agents, suivant notre design ont été bien reconnues et ce dans les deux évaluations quel que soit le genre et l'apparence des agents. Alors que nous avons basé notre étude sur la théorie de la Complémentarité Interpersonnelle, nous n'avons pas mesuré d'interactions entre les variables, ce qui signifie que chaque attitude a été reconnue indépendamment de l'attitude exprimée par l'autre. Bien que le design soit légèrement différent entre les deux évaluations (nous n'avons pas considéré chaque paire d'agents dans les deux positions possibles), les conclusions dans les deux évaluations sont similaires. Par conséquent, nous n'avons pas observé de différences entre les deux évaluations. Nous pouvons donc dire que notre modèle permet de générer le comportement non-verbal d'agents exprimant des attitudes sur une même dimension (Affiliation ou Statut) les uns envers les autres. Notre modèle a été validé pour des applications sur écran ou des applications immersives. Il convient de préciser que même si aucune interaction n'a été mesurée entre les attitudes, certains comportements des agents se basent sur ceux des autres agents (comme la distance et l'orientation) et que par conséquent, l'expression d'attitudes d'un agent se fait en résonance avec ses interlocuteurs.

Cependant il existe plusieurs limites à notre approche que nous avons relevées tout au long de ce chapitre. Tout d'abord, nous n'avons considéré qu'un sous-ensemble des configurations possibles en choisissant de manipuler les attitudes sur un seul axe à la fois et en prenant uniquement les valeurs extrêmes des axes. Il serait intéressant d'étudier ce qu'il se passe lorsque l'on fait varier les attitudes de façon continue durant l'interaction. Par rapport à la Complémentarité Interpersonnelle, nous n'avons pas mesuré d'interactions entre les attitudes de chaque agent mais nous n'avons pas mesuré si les participants percevaient une différence dans la durée potentielle de l'interaction. Une telle différence pourrait nous renseigner sur la capacité du modèle a correctement exprimer les situations complémentaires et anti-complémentaires. Nous n'avons considéré qu'un groupe de 4 agents, avec au-

tant d'agents masculins que féminins or il serait intéressant de voir s'il n'y a pas de différences lorsqu'un groupe est composé uniquement d'agent masculins ou d'agents féminins par exemple, et ce avec différents nombres de participants. Les agents exprimant des attitudes étaient limités à deux et étaient toujours côte à côte. De plus, nous n'avons pas considéré le contenu verbal dans ces études. Comme les conversations n'étaient pas intelligibles, nous ne pouvions faire participer l'utilisateur à celle-ci, même dans le système immersif.

Néanmoins, le fait que les attitudes, suivant notre design, aient été bien reconnues nous indiquent que nous allons dans la bonne voie en ayant choisi d'étendre notre premier modèle en intégrant de nouvelles modalités. En effet, il est difficile de limiter l'expression d'attitudes à certaines modalités. L'attitude s'exprime à travers tous les comportements lors d'une interaction. Le modèle que nous avons proposé permet de simuler des groupes de personnes dans lesquels des agents peuvent exprimer des attitudes envers les uns et les autres. Plusieurs perspectives s'offrent à nous comme l'intégration du contenu verbal, de l'utilisateur dans la conversation ou encore de l'extension à de multiples agents exprimant des attitudes dans des configurations plus variées.

Nous présentons dans le prochain chapitre le projet européen dans lequel se travail de cette thèse s'est inscrit.

Résumé du chapitre

- Deux évaluations du système complet ont été réalisées afin de vérifier que les attitudes exprimées par les agents sont bien reconnues, une évaluation conduite en ligne dans le navigateur Internet des participants et une évaluation conduite dans un environnement immersif.
- Le même design a été suivi dans les deux évaluations, s'appuyant sur la théorie de la complémentarité interpersonnelle pour évaluer des conditions dans lesquelles deux agents s'expriment réciproquement des attitudes au sein d'un groupe de quatre.
- Les résultats ont montré que les attitudes des agents ont été bien reconnues dans toutes les conditions des deux évaluations.

7 Le projet REVERIE

Le travail de cette thèse a été supporté par le projet européen FP7 REVERIE (*REal and Virtual Engagement in Realistic Immersive Environments*) [Wall et al., 2014]. Dans ce chapitre, nous présentons les objectifs de ce projet et comment notre travail s'inscrit dans l'architecture du système REVERIE.

7.1 Présentation du projet

Le projet REVERIE a démarré fin 2011 et s'est conclu mi-2015. L'objectif de ce projet est de bâtir un environnement virtuel en ligne dans lequel les utilisateurs se rejoignent afin de participer à des activités sociales et éducatives (voir figure 7.1). Une des idées directrices du projet REVERIE est de créer un univers immersif pour les utilisateurs. Pour cela, le projet réunit différent partenaires académiques et industrielles, chacun spécialisé dans un domaine pertinent pour le projet, afin de proposer des solutions à la pointe de la technologie (le son, la 3D, les agents virtuels, la reconnaissance d'images et de mouvements, les réseaux etc.) pour plonger les utilisateurs dans l'univers de REVERIE. Les utilisateurs interagissent naturellement avec le système REVERIE en utilisant leurs voix et gestes qui sont captés par des micros et des caméras Kinect de Microsoft ; l'univers qui leur est proposé contient des personnages virtuels semi-autonomes capables de réagir de façon naturelle aux activités des utilisateurs. Ces agents sont dits semi-autonomes car ils peuvent réagir d'eux-mêmes à certains évènements mais restent contrôlés par un humain la plupart du temps. Dans le cadre de REVERIE, les agents sont capables de réagir aux signaux des utilisateurs de façon autonome, comme comprendre certains mots-clés et reconnaitre certains gestes mais ils doivent également suivre un scénario préétabli.

Deux cas d'utilisations ont été implémentés pour illustrer les capacités du système REVERIE, une visite du parlement européen et un jeu de *Jacques-a-dit*, tous

FIGURE 7.1 : Une vue du parlement européen reproduit dans l'univers de REVERIE

deux menés par un agent semi-autonome. Dans la section suivante, nous présentons l'architecture des agents semi-autonomes de REVERIE et comment notre travail s'inscrit dans cette architecture.

7.2 L'ARCHITECTURE DES AGENTS DU SYSTÈME REVERIE

Au cœur de la conception de REVERIE, se trouve le besoin de peupler son univers de personnages qui interagissent naturellement avec l'utilisateur. Pour cela, nous, ainsi que les différents partenaires du projet responsables de la conception de ces personnages (Centrum Wiskunde Informatica, Queen Mary College, Informatics and Telematics Institute, DCU Clarity, Blitz Game Studio, Alcatel-Lucent), avons proposé une architecture d'agents présentée dans [Kuijk et al., 2013]. Cette architecture consiste en un ensemble de composants, chacun ayant la charge de réaliser une tâche bien précise (voir figure 7.2) comme reconnaitre la parole de l'utilisateur ou animer l'agent virtuel. Ces composants sont répartis en 4 catégories :

- *Analyzers* : ces composants reçoivent les signaux de l'environnement et de l'utilisateur et les transforment en un format manipulable pour les autres composants.
- *Interpreters* : ces composants traitent les données provenant des *Analyzers* afin

de mettre à jour l'état interne de l'agent.

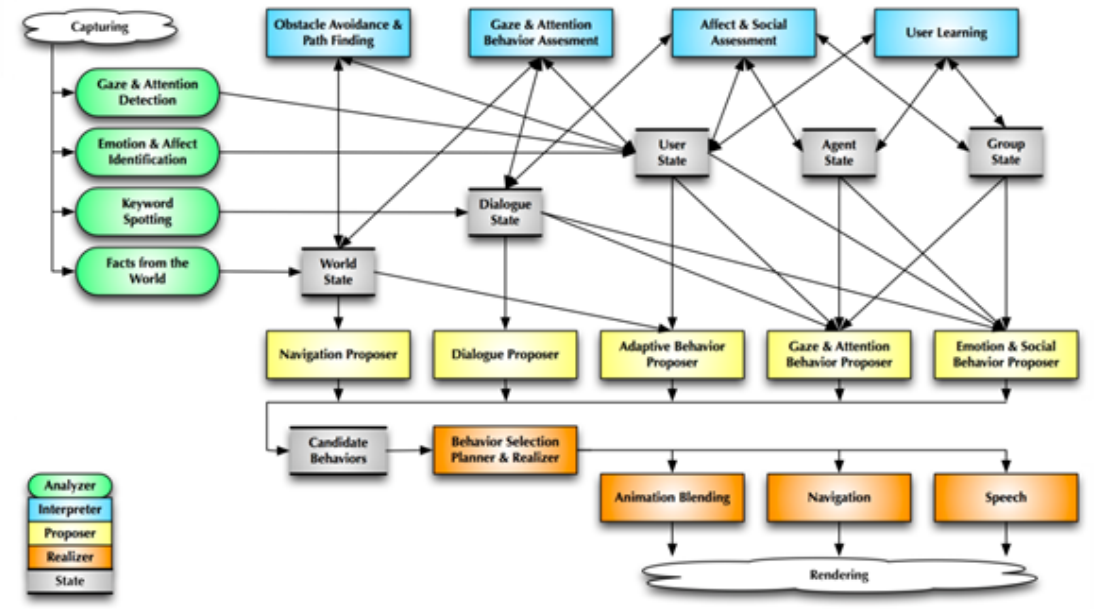
- *States* : ces composants conservent les informations connues de l'agent et son état interne.
- *Proposers* : ces composants proposent des comportements à partir de l'état interne de l'agent.
- *Realizers* : ces composants réalisent les comportements de l'agent, sa voix ainsi que son animation.

Cette architecture reprend les fondamentaux des architectures d'agent proposées dans la communauté d'agents virtuels [Kopp et al., 2006] et spécialise certains composants pertinents pour le projet REVERIE comme la détection de mots-clés. Notre contribution principale au sein du projet REVERIE a été de proposer le modèle que nous avons développé dans cette thèse pour réaliser la fonction du composant *Emotion & Social Behavior Proposer*. Nos autres contributions incluent le développement du composant *Keyword Spotting* qui reconnait certains mots-clés prononcés par l'utilisateur, l'intégration d'un outil de synthèse vocale (OpenMary) et la participation à la conception du script de l'agent lui permettant de jouer les scénarios de REVERIE, une visite virtuelle du parlement européen et une partie du jeu *Jacques-a-dit*. Dans la section suivante, nous détaillons ces contributions.

7.3 Nos contributions au projet REVERIE

7.3.1 Modèle de comportement non-verbal pour l'expression d'attitudes sociales et d'émotions

En tant que partenaire du projet REVERIE, notre engagement principal était de fournir un modèle capable de produire le comportement non-verbal d'un agent conversationnel dans un groupe, capable d'exprimer différentes attitudes sociales et émotions. Le résultat est le travail présenté dans cette thèse ; la conception du modèle ainsi que son évaluation sont détaillées dans les chapitres précédents. L'intégration du modèle que nous avons développé au cœur de la plateforme VIB/Greta nous a permis de proposer un modèle permettant l'expression d'attitudes sociales et d'émotions. Notre modèle a été utilisé pour construire les animations des agents suivant les attitudes que ceux-ci devaient avoir dans les différents scénarios de REVERIE. Au sein de l'équipe en charge du développement des agents du système



7

FIGURE 7.2 : L'architeture des agents du système REVERIE

REVERIE, nous avons également contribué à l'implémentation du système de communication entre les composants de l'architecture d'agents. Pour cela, nous avons conçu un protocole de communication entre les composants se basant d'abord sur RabbitMQ [Videla and Williams, 2012], un gestionnaire de messages et puis sur le protocole de communication développé au sein du projet REVERIE.

7.3.2 Détection de mots-clés

Notre seconde contribution est un système de détection de mots-clés prononcés par l'utilisateur. Nous avons d'abord développé une solution fonctionnant à l'aide de Google Speech API, une librairie en ligne de Google permettant la reconnaissance de parole. Cependant, cette solution présentait certaines limitations comme un temps de réponse trop lent et une possibilité d'indisponibilité en cas de coupure d'accès au service en ligne. C'est pourquoi nous avons par la suite proposé une seconde version de ce modèle qui s'appuie sur la librairie de reconnaissance de la parole CMU Sphinx [Lamere et al., 2003], une solution utilisable sans connexion Internet que nous avons pu implémenter au cœur du système REVERIE. Nous avons obtenu un bien meilleur temps de réponse.

7.3.3 Scripts pour les agents

Afin que les agents de REVERIE puissent remplir leur rôle dans les différents cas d'utilisation du système, nous et d'autres partenaires du projet (CTVC, Queen Mary College, Centrum Wiskunde Informatica) avons proposé de faire suivre à l'agent des scénarios prédéfinis. Pour cela, nous avons proposé une approche qui se base sur le modèle de dialogue orienté tâche Disco4Game [Rich and Sidner, 2012]. Le modèle Disco4Game permet de modéliser les buts à atteindre par un personnage et les différentes actions pour y arriver. Un scénario est représenté par un arbre de séquences d'actions, chaque action pouvant elle-même être divisée en séquences d'actions. Dans notre implémentation, les actions pouvaient être des comportements de l'utilisateur (comme certains mots-clés prononcés ou gestes effectués), ainsi que les actes du langages de l'agent représentés sous la forme d'un fichier FML [Kopp et al., 2006]. Le format FML permet de définir une suite d'intentions communicatives ainsi qu'un texte à prononcer par un agent. Les étiquettes FML sont instanciées en signaux multimodaux nécessaires à la réalisation des intentions définies dans le fichier ; puis sont traduits en animations. Nous avons donc représenté les scénarios de REVERIE sous la forme de fichiers FML tout en créant la structure des scénarios pour Disco4Game. Dans le scénario de la visite du parlement européen, l'agent devait guider les utilisateurs participant à la session dans l'hémicycle. Il se déplaçait dans l'hémicycle tout en donnant une explication sur le fonctionnement du parlement et de l'union européenne. Dans le scénario de la partie de *Jacques-a-dit*, l'agent joue le rôle d'instructeur ; il contrôle le déroulement du jeu en expliquant les règles et en donnant les instructions aux participants. A la fin de la partie, il désigne le vainqueur. Le guide devait avoir une attitude amicale et l'instructeur devait avoir une attitude dominante.

7.4 Conclusion

Notre contribution au projet REVERIE a été significative. Le travail de cette thèse est au cœur du développement de l'architecture d'agent social du système REVERIE. Nous avons aussi fourni des composants qui dépassent le cadre du travail présenté dans ce rapport. La collaboration avec les autres partenaires nous a permis d'enrichir notre architecture d'agent social et de proposer des réflexions sur

l'intégration de la dimension sociale dans un agent [Ravenet et al., 2014]. Ce travail a d'ailleurs servi de modèle pour l'architecture d'un agent membre d'un groupe proposée dans le chapitre 5.

Dans le chapitre suivant, nous concluons, discutons des résultats et présentons les perspectives pour ce travail.

8 Conclusion

Au cours de cette thèse, nous avons proposé un modèle de simulation de groupes conversationnels au sein desquels leurs membres peuvent exprimer différentes attitudes sociales envers les uns et les autres et ce à travers leur comportement non-verbal. En se basant sur les travaux en sciences humaines et sociales dans le domaine de l'expression d'attitudes et des groupes conversationnels, nous avons proposé une nouvelle approche permettant de construire un modèle computationnel pour l'expression d'attitudes sociales d'agents conversationnels. Dans ce chapitre, nous concluons en rappelant les différentes contributions de notre travail dans une première section ; puis nous discutons des limites et des perspectives de nos travaux dans une deuxième section.

8.1 Nos différentes contributions

Dans cette section, nous rappelons les différentes contributions de notre travail.

8.1.1 Modèle computationnel d'expression d'attitudes pour un agent locuteur engagé dans une dyade avec l'utilisateur

Nos travaux ont permis d'apporter :

- Une méthode de collecte de corpus en utilisant une approche de *Crowdsourcing* nous permettant de recueillir des descriptions de comportements non-verbaux d'attitudes sociales.
- Un modèle computationnel d'expression non-verbale d'attitudes sociales.

Nous avons proposé une méthodologie pour la construction d'un modèle computationnel de comportement d'agents virtuels. Cette méthodologie s'appuie sur une approche de *Crowdsourcing*, qui consiste à déléguer la création d'un corpus à des utilisateurs d'une plateforme en ligne, pour ensuite apprendre le modèle computationnel

grâce à l'utilisation d'une méthode d'apprentissage sur les données collectées. Cette méthodologie peut également être réutilisée dans d'autres contextes que le nôtre comme par exemple pour apprendre le comportement associé à différentes personnalités ou intention communicative . Afin de l'utiliser pour construire notre modèle computationnel d'expression non-verbale d'attitudes sociales dans une dyade, nous avons construit notre propre plateforme de collecte de données. Cette plateforme permet à des utilisateurs de manipuler les paramètres du comportement non-verbal d'un agent afin de lui faire exprimer différentes attitudes. Les paramètres du comportement non-verbal manipulables dans notre plateforme ont été choisis à partir d'une étude de la littérature en sciences humaines et sociales sur l'expression d'attitudes.

A partir des données collectées, nous avons proposé d'apprendre un réseau Bayésien pour construire notre modèle computationnel. Nous avons décidé d'utiliser un réseau Bayésien car c'est un modèle qui est facile à interpréter et qui permet de bien représenter les relations de causalité entre les attitudes et les paramètres du comportement non-verbal. Ce modèle a été implémenté dans un prototype de recruteur virtuel afin de réaliser une évaluation de celui-ci. Pour cela, un modèle d'attitude verbale nous a permis de compléter l'architecture du recruteur virtuel pour l'évaluation. Les résultats ont montré que l'expression verbale et non-verbale des attitudes amicale et hostile était reconnue.

Par la suite, nous avons proposé un modèle permettant l'expression d'attitudes dans des situations de groupe.

8.1.2 Modèle computationnel d'agent conversationnel capable d'exprimer différentes attitudes sociales dans un groupe à travers son comportement non-verbal

Nos travaux ont apporté les contributions suivantes :

- Un modèle de génération de tours de parole suivant les attitudes que les agents ont les uns envers les autres.
- Un modèle de génération de formations de groupe suivant les attitudes des agents les uns envers les autres.
- Création d'agents capables de gérer simultanément plusieurs fonctions communicatives telles que la conversation, l'appartenance à un groupe et l'expression d'attitude.

- Proposition d'une architecture d'agent social.

Le modèle de génération de tours de parole se base sur une représentation à partir de machine à états permettant à un agent d'alterner entre les rôles de locuteurs et d'auditeurs en fonction de ses attitudes grâce à un ensemble de transitions. Ces transitions sont construites à partir d'une étude de la littérature. Ce modèle s'accompagne d'un modèle de génération de formations de groupe conçu à partir d'un ensemble de règles permettant de manipuler les distances, les orientations du corps et les directions du regard entre les agents en fonction des attitudes qu'ils expriment les uns envers les autres. Ces deux modèles ont été couplés avec le modèle de génération de comportement dans une dyade afin de proposer un modèle permettant de ne plus considérer uniquement des conversations dyadiques avec l'utilisateur mais aussi des conversations de groupes. L'idée directrice du modèle est que la notion de groupe et de conversation émerge du comportement de chaque individu. C'est l'alternance des activités de locuteur et d'auditeur chez les agents qui fait émerger les tours de parole. De même pour la formation de groupe, c'est le comportement individuel de chaque membre qui fait émerger la notion de groupe. Nous avons implémenté le système complet dans une application de type jeu sérieux, grâce au moteur Unity3D, dans laquelle le joueur peut naviguer dans une scène peuplée de groupes conversationnels.

8.1.3 Évaluations du système avec différents niveaux d'immersions

Nos contributions portent de plus sur l'élaboration de protocoles d'évaluation proposant :

- Une simplification du nombre de configurations possibles afin de tester l'interaction de l'expression d'attitudes entre deux agents et la reconnaissance des attitudes par les participants.
- Différents niveaux d'immersion.

Nous avons réalisé deux évaluations du système. La première évaluation a été réalisée en ligne, en montrant des vidéos de scènes générées par notre modèle. La deuxième évaluation a été réalisé en laboratoire, mettant le participant au cœur de la scène en temps réel dans un environnement immersif de type CAVE. Les résultats ont permis de valider les modèles proposés. Les participants ont bien reconnus les différentes attitudes exprimées par les différents personnages que ce soit dans les vidéos ou dans

l'environnement immersif.

8.2 Limites et Perspectives

Notre contribution principale dans ce travail est un modèle de comportements non-verbaux pour un agent membre d'un groupe conversationnel. Ce modèle permet à l'agent d'adopter différentes stratégies de prise de parole dans une conversation et de générer son comportement non-verbal en fonction des attitudes qu'il exprime envers les autres membres du groupe sans considération pour le contenu verbal. Nous commençons par voir les limites de notre travail puis nous discutons des perspectives.

8.2.1 Limites

8.2.1.1 Le contexte

Nous avons étudié l'expression d'attitudes par des agents conversationnels dans un contexte particulier. Notre étude de la littérature s'est concentrée sur des travaux portant sur la culture occidentale et les participants à nos études venaient également majoritairement de pays issus de cette culture. Nous n'avons pas pris en compte les différences culturelles dans la modélisation des attitudes et dans la perception de celles-ci. Or, par exemple, Une étude menée dans [Keating et al., 1981] a montré des différences dans la perception de la dominance en fonction de la culture. Ils ont trouvé notamment que l'association entre le fait de froncer les sourcils et la perception de la dominance se faisait surtout dans la culture occidentale.

Nous avons également travaillé avec 4 agents ayant chacun une apparence distincte mais cela ne constituait pas un panel représentatif des différentes apparences possibles (pas de différences de tailles, de carrures et d'ethnies par exemple). Or l'apparence peut jouer dans la perception de l'attitude [Argyle, 1988]. Bien que nous ayons cherché à minimiser l'effet de celles-ci dans nos manipulations, il est difficile de généraliser nos résultats pour toutes apparences et cultures.

8.2.1.2 Les comportements

Le nombre de modalités que nous considérons dans l'expression d'attitudes n'est pas exhaustif et il est possible que les attitudes seraient mieux exprimées si d'autres

comportements étaient influencés également comme la hauteur et le volume de la voix par exemple [Sanchez-Cortes et al., 2013]. Pour les paramètres du comportement non-verbal que nous avons considéré, nous avons utilisé des valeurs discrètes. De plus, nos agents ne sont pas non plus capables d'exprimer des *backchannels*, ces comportements qu'une personne fait pour montrer qu'elle écoute et pour inciter l'autre à continuer de parler. Le contenu verbal n'est pas non plus considéré alors que nous savons que les attitudes se manifestent verbalement [Chindamo et al., 2012].

8.2.1.3 L'utilisateur

Dans notre système, nous n'avons pas pu intégrer l'humain dans les conversations des agents du fait qu'ils utilisaient un langage fictif et que la reconnaissance du langage naturelle n'est pas encore fiable. L'impact de notre modèle sur un utilisateur pourrait être très différent s'il se sentait impliqué dans la conversation.

8.2.1.4 Les attitudes

Nous avons utilisé seulement 4 attitudes dans notre évaluation, correspondant aux extrémités des axes Statut et Affiliation. Nous n'avons pas étudié comment ces attitudes peuvent se combiner (par exemple une attitude dominante et amicale) et nous n'avons pas regardé les effets d'un changement d'attitudes durant une interaction. Nous n'avons pas non plus étudié différentes intensités dans l'expression d'attitudes.

8.2.2 Perspectives

Ce travail et ses limites offrent des perspectives d'amélioration pour le modèle proposé et pour explorer de nouvelles voies autour de l'expression d'attitudes.

8.2.2.1 Court terme : les modalités et le groupe d'agents

Tout d'abord, un premier axe de recherche pourrait viser à améliorer la capacité du modèle à mieux exprimer les attitudes en étudiant l'ajout de nouveaux comportements (comme la posture [Ballin et al., 2004] et les *Backchannels*), de nouvelles

modalités (comme la qualité de la voix et le contenu verbal [Gatica-Perez, 2009]) et en augmentant le nombre de valeurs possibles pour celles que nous considérons déjà. Par exemple, nous avons utilisé 3 expressions faciales différentes que nous combinons avec l'animation des lèvres pour exprimer des attitudes. Une expression positive de joie, une expression neutre et une expression négative de colère. Nous pourrions considérer bien plus d'expressions faciales avec différentes intensités. Cela impliquerait alors d'améliorer la méthode de *Crowdsourcing* que nous avons proposée afin de permettre la collecte de données continues et de nouveaux paramètres. Une des problématiques serait de concevoir une interface qui reste simple pour un utilisateur naïf tout en permettant de collecter des données plus riches.

En ce qui concerne la configuration du groupe, nous pourrions chercher à étudier la généralisation aux différentes apparences d'un agent, comme sa taille, son genre, son style vestimentaire et sa carrure. L'apparence physique d'une personne peut renvoyer une impression sur sa personnalité qui peut influencer l'attitude perçue [Argyle, 1988]. Plusieurs axes de recherche sont envisageables. Un premier axe consiste à intégrer l'apparence physique comme un paramètre dans l'expression de l'attitude (une personne grande et musclée peut renvoyer de base une image plus dominante par exemple). Un deuxième axe consiste à proposer un modèle qui ne considère pas l'apparence comme un paramètre mais qui fonctionne pour toutes les apparences. Ensuite, différentes possibilités s'offrent également pour la généralisation à différentes tailles de groupes : prendre en compte les autres membres du groupe comme paramètres pour l'expression d'attitudes (dans un groupe de grande taille, il peut être plus difficile d'exprimer de la dominance, à cause du nombre de personnes et à cause de leurs attitudes par exemple) ou proposer un modèle qui ne considère pas ce paramètre mais fonctionne pour toute taille de groupe.

8.2.2.2 Moyen terme : la conversation et les attitudes

Nous avons considéré une conversation avec une durée indéfinie. Une amélioration consisterait à prendre en compte l'évolution de l'interaction et de ses différentes étapes, en particulier son début et sa fin. Les comportements produits peuvent être différents et l'attitude perçue dans les premières secondes d'une interaction peut avoir un effet sur le reste de celle-ci [Cafaro et al., 2012] de même que ceux produits à la toute fin. Comme nous avons étudié des attitudes figées pour toute la durée

des interactions, cela ouvre également la voie à l'étude de l'évolution de l'attitude au cours de l'interaction. Notre modèle est conçu afin d'exprimer les attitudes avec chaque changement d'action (parler, s'arrêter de parler, vouloir parler, écouter) et ce de façon probabiliste. En effet notre approche dans la conception de ce modèle s'est basée sur la définition de règles probabilistes de comportement, à partir d'un corpus de données collecté et à partir de la littérature en sciences humaines et sociales associant un comportement et un seul sur chaque modalité pour chacune des différentes actions. Une autre approche pourrait être de considérer une expression continue de l'attitude tout au long de l'interaction et non plus juste au moment des changements d'action. Certains chercheurs se sont d'ailleurs intéressés à l'expression d'attitudes dans des séquences de comportements [Chollet et al., 2014].

Une perspective importante est l'intégration de l'utilisateur dans l'interaction avec un groupe d'agents. Pour cela, plusieurs étapes sont nécessaires. Il faut développer un modèle de dialogue permettant d'exprimer différentes attitudes dans le contenu verbal en fonction des attitudes entre les participants. Plus important encore, il est nécessaire de comprendre l'activité de l'utilisateur. Afin de l'intégrer dans notre modèle de tours de parole, nous devons être capable d'identifier dans quel état conversationnel il se trouve, et pour l'intégrer dans le dialogue, nous devons également comprendre ce qu'il dit. Ceci peut être réaliser en utilisant des solutions existantes de reconnaissance de comportements [Wagner et al., 2013] et de la parole [Lamere et al., 2003].

8.2.2.3 Long terme : extension du modèle à différentes interactions

Dans le futur, il sera intéressant de continuer ce travail en cherchant à appliquer le modèle et la méthodologie proposés dans cette thèse à d'autres types d'interactions comme un travail collaboratif ou un jeu par exemple, et non plus juste une conversation. Pour l'instant nous avons évalué le modèle dans des scénarios où un seul utilisateur était face à un groupe d'agents et il nous faudra également vérifier que les agents animés par notre modèle sont capables, à l'inverse, de rejoindre un groupe d'humains et d'interagir avec eux. A terme, le modèle devra permettre à un agent de rejoindre un groupe composé aussi bien d'agents que d'humains et d'interagir avec les uns et les autres en produisant des comportements cohérents vis-à-vis de son rôle dans l'interaction. On peut également imaginer faire fonctionner un tel

modèle avec des robots. Dès lors, on peut imaginer des scénarios où personnages virtuels, robots et humains peuvent se rejoindre et peuvent interagir tous ensemble.

Ce travail nous offre ainsi de nombreuses voies afin de continuer à améliorer le modèle proposé ainsi que l'expression de comportements de personnages virtuels. Bien que nos résultats se soient montrées encourageants, il existe encore de nombreuses pistes de recherche que nous devons mener afin de pouvoir proposer un modèle riche de générations de comportements permettant l'expression d'attitudes sociales.

A

Distributions de probabilités du réseau Bayésien

Nous détaillons ici les distributions de probabilités des différents nœuds de sortie du réseau Bayésien présenté dans le chapitre 4.

Table A.1 : Distribution de probabilités du nœud Amplitude

	Amplitude des gestes		
Attitude	Petite	Normale	Grande
Dominant	0,30	0,27	0,43
Soumis	0,56	0,18	0,26
Amical	0,29	0,34	0,37
Hostile	0,30	0,21	0,49

Table A.2 : Distribution de probabilités du nœud Force

	Force des gestes		
Attitude	Faible	Normale	Forte
Dominant	0,25	0,24	0,51
Soumis	0,53	0,27	0,20
Amical	0,41	0,38	0,21
Hostile	0,21	0,23	0,56

Table A.3 : Distribution de probabilités du nœud Orientation de la tête

	Orientation de la tête			
Attitude	Vers le bas	Sur le côté	Droite	Vers le haut
Dominant	0,22	0,12	0,31	0,35
Soumis	0,60	0,17	0,13	0,10
Amical	0,18	0,40	0,29	0,13
Hostile	0,30	0,24	0,24	0,22

Table A.4 : Distribution de probabilités du nœud Regard

	Direction du regard	
Attitude	Sur le côté	Droit
Dominant	0,15	0,85
Soumis	0,48	0,51
Amical	0,17	0,82
Hostile	0,28	0,72

TABLE A.5 : Distribution de probabilités du nœud Expression Faciale

		Expressions faciales		
Actes Illocutoires	Attitude	Positive	Négative	Neutre
Approuver	Dominant	0,38	0,39	0,23
	Soumis	0,47	0,03	0,50
	Amical	0,92	0,03	0,05
	Hostile	0,13	0,66	0,21
Désapprouver	Dominant	0,12	0,64	0,24
	Soumis	0,24	0,29	0,47
	Amical	0,80	0,13	0,07
	Hostile	0,04	0,93	0,03
Accepter	Dominant	0,23	0,48	0,29
	Soumis	0,37	0,20	0,43
	Amical	0,96	0,03	0,01
	Hostile	0,09	0,89	0,02
Refuser	Dominant	0,33	0,55	0,12
	Soumis	0,19	0,22	0,59
	Amical	0,77	0,13	0,10
	Hostile	0,07	0,78	0,15
Informer	Dominant	0,14	0,31	0,55
	Soumis	0,56	0,07	0,37
	Amical	0,75	0,12	0,13
	Hostile	0,09	0,75	0,16
Demander	Dominant	0,27	0,48	0,25
	Soumis	0,49	0,15	0,36
	Amical	0,78	0,11	0,11
	Hostile	0,10	0,73	0,17

TABLE A.6 : Distribution de probabilités du nœud Mouvements

		Mouvements			
Actes Illocutoires	Attitude	Bras et tête	Tête seule	Bras seuls	Aucun
Approuver	Dominant	0,29	0,39	0,26	0,06
	Soumis	0,20	0,42	0,14	0,24
	Amical	0,40	0,31	0,26	0,03
	Hostile	0,17	0,55	0,12	0,16
Désapprouver	Dominant	0,42	0,28	0,15	0,15
	Soumis	0,22	0,33	0,29	0,16
	Amical	0,33	0,31	0,29	0,07
	Hostile	0,56	0,24	0,19	0,01
Accepter	Dominant	0,28	0,42	0,21	0,09
	Soumis	0,27	0,43	0,15	0,15
	Amical	0,36	0,27	0,35	0,02
	Hostile	0,27	0,29	0,21	0,23
Refuser	Dominant	0,35	0,38	0,25	0,02
	Soumis	0,30	0,44	0,17	0,09
	Amical	0,31	0,46	0,20	0,03
	Hostile	0,44	0,30	0,17	0,09
Informer	Dominant	0,32	0,21	0,42	0,05
	Soumis	0,23	0,19	0,25	0,33
	Amical	0,24	0,15	0,39	0,22
	Hostile	0,19	0,09	0,51	0,21
Demander	Dominant	0,41	0,05	0,30	0,24
	Soumis	0,16	0,26	0,27	0,32
	Amical	0,17	0,21	0,34	0,28
	Hostile	0,29	0,12	0,33	0,26

B

Documents des évaluations du système

Nous présentons ici le détail des deux évaluations que nous avons conduites afin de valider notre modèle (en ligne et en environnement immersif). Nous expliquons comment celles-ci se sont déroulées, leur contenu, leurs différences et nous donnons également les résultats.

B.1 Déroulement de l'évaluation en ligne

Cette évaluation a été conduite en ligne, grâce à un site Internet, composé de plusieurs pages, que nous avons développé nous-mêmes. La figure B.1 présente la première page qui était montrée aux participants afin de recueillir leur accord de participation et bien leur expliquer que bien que cette étude soit parfaitement anonyme, ils pouvaient à tout moment se retirer de celle-ci et demander à ce que leurs informations soient supprimées.

Dans la page suivante (figure B.2), nous leur demandions quelques informations démographiques (âge, genre, niveau d'étude, pays d'origine) puis nous leur expliquions comment répondre à cette étude. Il leur était expliqué comment répondre aux questions et il leur était demandé de regarder les vidéos en plein écran si possible.

Ensuite, 4 pages similaires à celle montrée figure B.3 leur étaient présentées, chaque page correspondant à chaque situation que les participants devaient évaluer. En dessous de chaque vidéo se trouvaient deux questionnaires, un questionnaire comme montré dans la figure B.4 à propos des attitudes des agents et un questionnaire à propos du groupe (figure B.5). A la fin de la session, le participant se voyait montré une page expliquant les objectifs de l'étude à laquelle il venait de participer (figure B.6).

B.2 Déroulement de l'évaluation en environnement immersif

Cette évaluation reprend la même procédure que dans l'évaluation en ligne. Un ordinateur était présent en dehors du CAVE sur lequel une application présentait les mêmes formulaires et questionnaires que dans l'évaluation en ligne. Les participants, entre chacune des quatre scènes, devaient sortir de la CAVE pour répondre aux questions sur l'ordinateur. La seule différence est que nous n'avons pas conservé le questionnaire sur le groupe dans cette seconde évaluation.

B.3 Questionnaire pour mesurer l'attitude d'un agent

Nous avons mesuré les attitudes indépendamment sur chaque dimension de notre représentation des attitudes (Statut et Affiliation). Par conséquent, nous avons conçu deux questionnaires, un pour mesurer l'attitude sur la dimension de Statut et un pour mesurer l'attitude sur la dimension d'Affiliation. Les évaluations étant conduites en anglais, les questionnaires le sont également. Afin de mesurer l'attitude d'un agent, nous posions 4 questions, issues de [Funder et al., 2000] et écrites à l'aide d'adjectifs issus de la classification de Wiggins autour de l'*Interpersonal Circumplex* [Wiggins et al., 1988]. Chaque question devait être répondue suivant une échelle de Likert à 5 points. Les quatre réponses étaient ensuite combinées en une valeur normalisée entre 0 et 1.

B.3.1 Questionnaire pour mesurer le niveau d'Affiliation

Quatre questions étaient posées pour chaque agent. Dans l'exemple suivant, il s'agit des quatre questions pour l'Agent Gauche, mais les mêmes questions respectives étaient posées pour l'Agent Droite.

On a scale from 1 to 5, how much do you agree with the following statements describing the attitude of the Left Agent towards the Right Agent ?

- Left Agent expresses warmth towards the Right Agent.
- Left Agent seems detached from the interaction with the Right Agent.
- Left Agent tries to be likeable when interacting with the Right Agent.
- Left Agent expresses hostility towards the Right Agent.

B.3.2 Questionnaire pour mesurer le niveau de Statut

Quatre questions étaient posées pour chaque agent. Dans l'exemple suivant, il s'agit des quatre questions pour l'Agent Gauche, mais les mêmes questions respectives étaient posées pour l'Agent Droite.

On a scale from 1 to 5, how much do you agree with the following statements describing the attitude of the Left Agent towards the Right Agent ?

- Left Agent tries to control the interaction with the Right Agent.
- Left Agent expresses insecurity (e.g. seems touchy or overly sensitive) towards the Right Agent.
- Left Agent dominates the interaction with the Right Agent.
- Left Agent seeks reassurance from Right Agent (e.g. asks for agreement, fished for praise).

B.4 QUESTIONNAIRE SUPPLÉMENTAIRE SUR LE GROUPE

Dans un but d'explorer la perception du groupe par les participants, nous avons également posé trois questions à propos du groupe après chaque vidéo. Ces questions n'ont pas été posées dans l'évaluation en environnement immersif.

How accurately the following statements describe the group ?

- The group is socially rich (the agents follow social norms typically happening in human-human conversations).
- The group is believable (the agents behave human-like).
- The group appears involved in its conversation (the agents seem engaged in the conversation).

FIGURE B.1 : Accord du participant

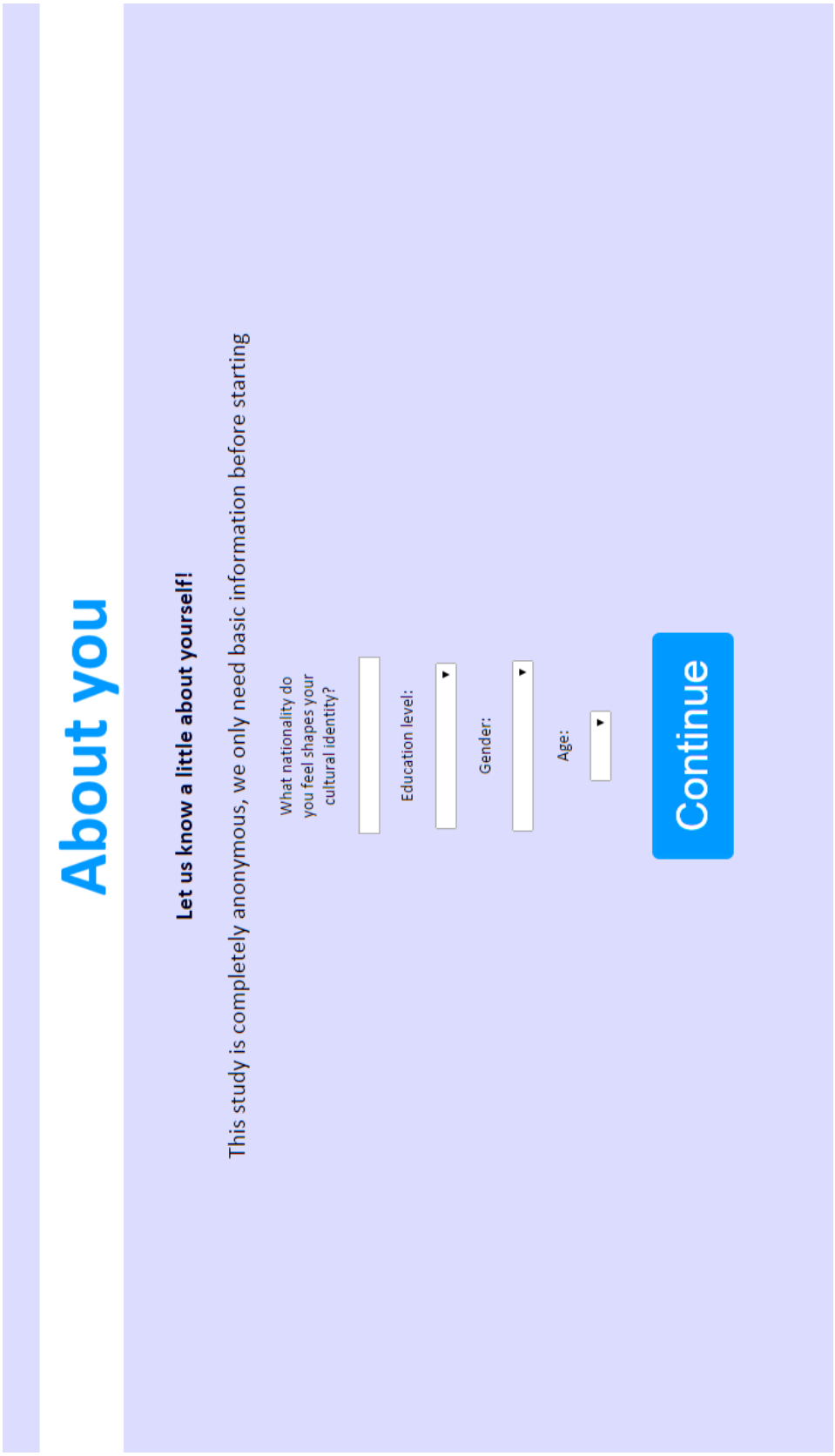


FIGURE B.2 : Questionnaire démographique

FIGURE B.3 : Comment les vidéos étaient montrées aux participants

FIGURE B.4 : Questionnaire sur les attitudes des agents

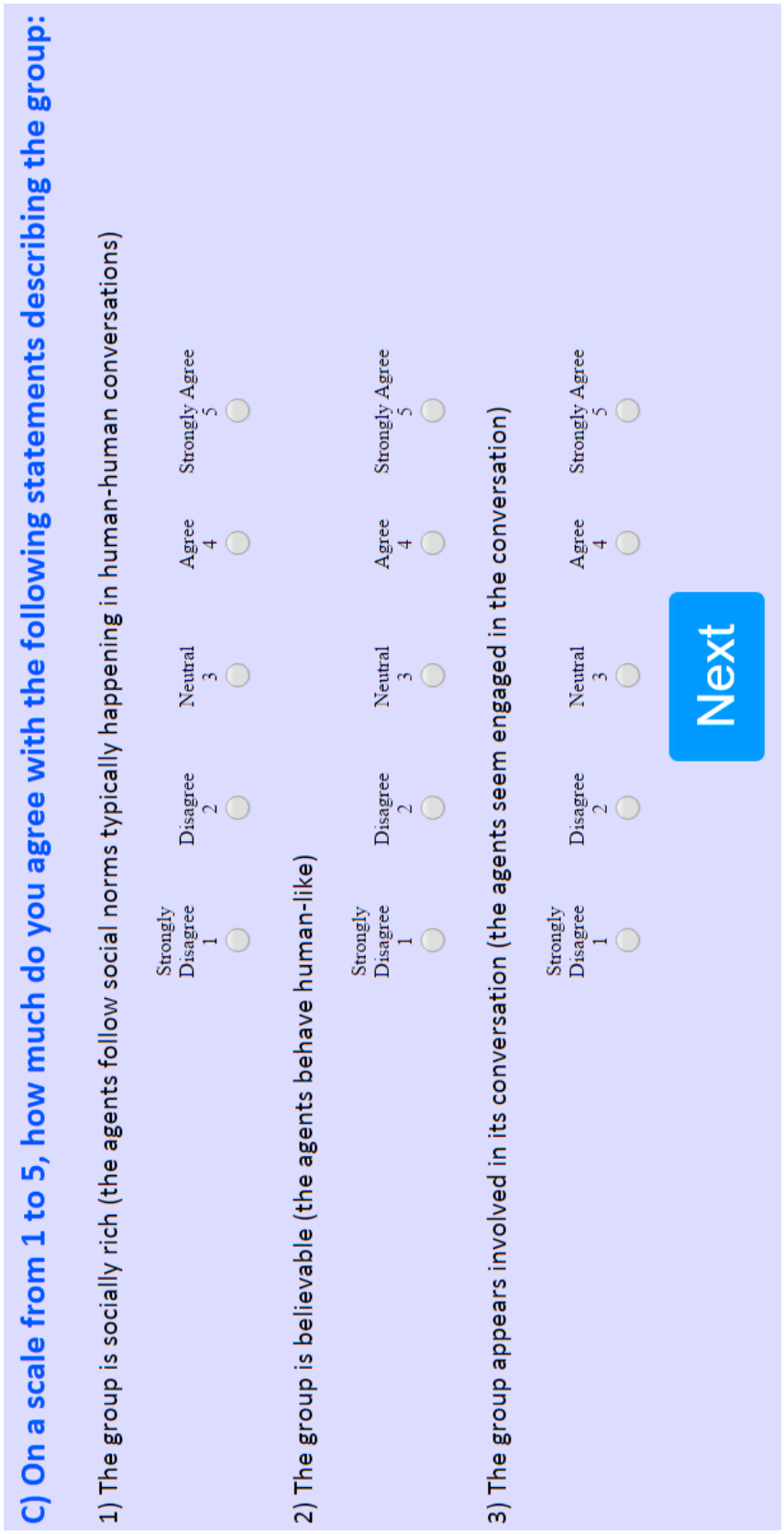


FIGURE B.5 : Questionnaire sur le groupe

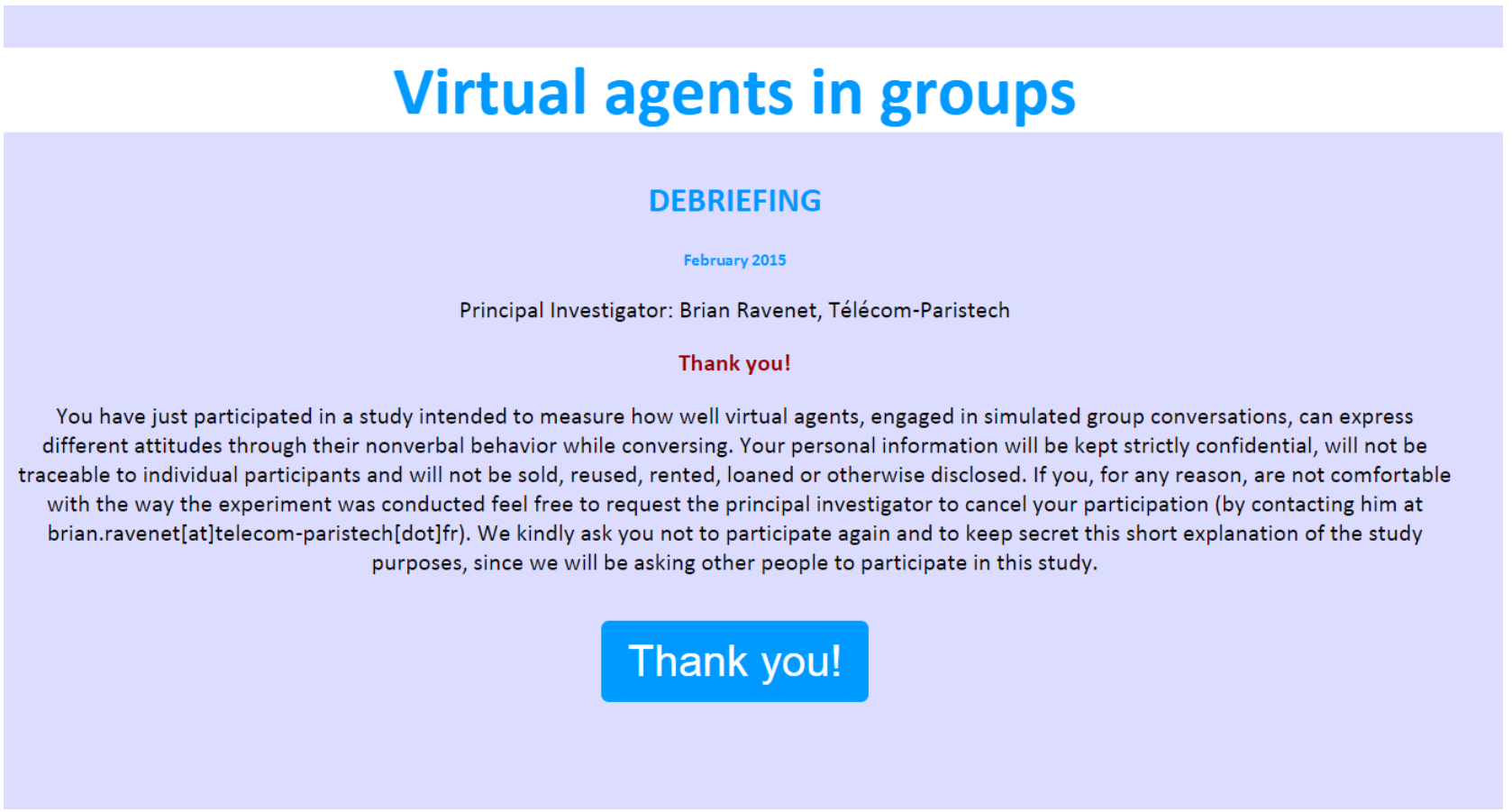


FIGURE B.6 : Page de débriefing

TABLE B.1 : Moyennes des attitudes mesurées par les participants dans l'évaluation en ligne (valeurs normalisées entre 0 et 1)

	Attitudes exprimées			
	Etude Statut		Etude Affiliation	
Attitudes mesurées	Dominante	Soumise	Amicale	Hostile
Agent Gauche	0.59	0.44	0.61	0.37
Agent Droite	0.60	0.46	0.58	0.41

TABLE B.2 : Moyennes des attitudes mesurées par les participants dans l'évaluation en environnement immersif (valeurs normalisées entre 0 et 1)

	Attitudes exprimées			
	Etude Statut		Etude Affiliation	
Attitudes mesurées	Dominante	Soumise	Amicale	Hostile
Agent Gauche	0.67	0.41	0.60	0.40
Agent Droite	0.66	0.42	0.57	0.41

C

Liste des publications

- (Lauréat du meilleur article) **Conversational Behavior Reflecting Interpersonal Attitudes in Small Group Interactions.** *Brian Ravenet, Angelo Cafaro, Beatrice Biancardi, Magalie Ochs, Catherine Pelachaud.* International Conference on Intelligent Virtual Agent (IVA2015), Delft, Pays-Bas, Août 2015.
- **Architecture of a Socio-Conversational Agent in Virtual Worlds.** *Brian Ravenet, Magalie Ochs, Catherine Pelachaud.* International Conference on Image Processing (ICIP2014), Paris, France, Octobre 2014.
- **Interpersonal Attitude of a Speaking Agent in Simulated Group Conversations.** *Brian Ravenet, Angelo Cafaro, Magalie Ochs, Catherine Pelachaud.* International Conference on Intelligent Virtual Agent (IVA2014), Boston, USA, Août 2014.
- **Nonverbal behavior of a virtual agent expressing attitudes in a group.** *Brian Ravenet, Angelo Cafaro, Magalie Ochs, Catherine Pelachaud.* Workshop Affect, Compagnon Artificiel, Interaction (WACAI2014), Rouen, France, Juin 2014.
- **A Model to Generate Adaptive Multimodal Job Interviews with a Virtual Recruiter.** *Zoraida Callejas, Brian Ravenet, Magalie Ochs, Catherine Pelachaud.* Ninth International Conference on Language Resources and Evaluation (LREC2014), Reykjavic, Islande, Mai 2014.
- **A Computational Model of Social Attitudes for a Virtual Recruiter.** *Zoraida Callejas, Brian Ravenet, Magalie Ochs, Catherine Pelachaud.* International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS2014), Paris, France, Mai 2014.
- **Vers des Agents Conversationnels Animés Socio-affectifs.** *Magalie Ochs, Yu Ding, Nesrine Fourati, Mathieu Chollet, Brian Ravenet, Florian Pecune,*

Nadine Glas, Ken Prepin, Chloé Clavel, Catherine Pelachaud. Journal d'Interaction Personne-Système (JIPS), JIPS, 3 (2), pp.1-23, Mars 2014.

- **Vers des Agents Conversationnels Animés Socio-Affectifs.** *Magalie Ochs, Catherine Pelachaud, Yu Ding, Nesrine Fourati, Mathieu Chollet, Brian Ravenet, Florian Pecune, Nadine Glas, Ken Prepin, Chloé Clavel.* 25th french conference on human-machine interaction (IHM2013), Bordeaux, France, Novembre 2013.
- **A Crowdsourcing Method for a User-Perception Based Design of Social Virtual Actors.** *Magalie Ochs, Brian Ravenet, Catherine Pelachaud.* International Workshop Computers are Social Actors (CASA2013), Intelligent Virtual Agent Conference (IVA), Edimbourg, Royaume-Uni Août, 2013.
- **From a User-Created Corpus of Virtual Agent's Non-Verbal Behaviour to a Computational Model of Interpersonal Attitudes.** *Brian Ravenet, Magalie Ochs, Catherine Pelachaud.* International Conference on Intelligent Virtual Agent (IVA2013), Edimbourg, Royaume-Uni, Août 2013.
- **A Framework for Human-like Behavior in an Immersive Virtual World.** *Fons Kuijk, Sigurd Van Broeck, Claude Dareau, Brian Ravenet, Magalie Ochs, Konstantinos Apostolakis, Petros Daras, David Monaghan, Noel E O'Connor, Julie Wall, Ebroul Izquierdo.* International Conference on Digital Signal Processing (DSP2013), Grèce, Juillet 2013.
- **A Computational Model of Social Attitude Effects on the Nonverbal Behavior for a Relational Agent.** *Brian Ravenet, Magalie Ochs, Catherine Pelachaud.* Workshop Affect, Compagnon Artificiel, Interaction (WACAI2012), Research Report, (RR-LIG-039), p. 94-101, LIG, Grenoble, France, 2012.

Bibliographie

[Adam et al., 2009] Adam, C., Herzig, A., and Longin, D. (2009). A logical formalization of the occ theory of emotions. *Synthese*, 168(2) :201–248. 61

[Allbeck and Badler, 2002] Allbeck, J. and Badler, N. (2002). Toward representing agent behaviors modified by personality and emotion. *Embodied Conversational Agents at AAMAS*, 2 :15–19. 12

[Allwood et al., 2012] Allwood, J., Chindamo, M., and Ahlsen, E. (2012). On identifying conflict related stances in political debates. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 918–925. 131

[Argyle, 1988] Argyle, M. (1988). *Bodily Communication*. University paperbacks. Methuen. 21, 23, 25, 46, 61, 66, 68, 103, 152, 154

[Austin, 1975] Austin, J. L. (1975). *How to do things with words*, volume 367. Oxford university press. 67

[Bales, 1950] Bales, R. F. (1950). Interaction process analysis ; a method for the study of small groups. 34, 56

[Ballin et al., 2004] Ballin, D., Gillies, M., Crabtree, B., et al. (2004). A framework for interpersonal attitude and non-verbal communication in improvisational visual media production. 46, 54, 58, 153

[Bates, 1994] Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM*, 37(7) :122–125. 10, 13, 61

[Baur et al., 2013] Baur, T., Damian, I., Gebhard, P., Porayska-Pomsta, K., and André, E. (2013). A job interview simulation : Social cue-based interaction with a virtual character. In *Social Computing (SocialCom), 2013 International Conference on*, pages 220–227. IEEE. 88

[Beck and Clark, 1997] Beck, A. T. and Clark, D. A. (1997). An information processing model of anxiety : Automatic and strategic processes. *Behaviour research and therapy*, 35(1) :49–58. 88

[Becker et al., 2007] Becker, C., Kopp, S., and Wachsmuth, I. (2007). *Why Emotions should be Integrated into Conversational Agents*, pages 49–67. John Wiley and Sons, Ltd. 11

[Bee et al., 2009] Bee, N., Franke, S., and Andrea, E. (2009). Relations between facial display, eye gaze and head tilt : Dominance perception variations of virtual agents. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1 –7. 47, 49

[Bee et al., 2010] Bee, N., Pollock, C., Andrea, E., and Walker, M. (2010). Bossy or wimpy : Expressing social dominance by combining gaze and linguistic behaviors. In Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., and Safonova, A., editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271. Springer Berlin / Heidelberg. 47, 49

[Beebe and Masterson, 2009] Beebe, S. A. and Masterson, J. T. (2009). *Communication in small groups : principles and practices*. Boston : Pearson Education, Inc. 28, 99

[Bergmann, 2012] Bergmann, K. (2012). *The Production of Co-Speech Iconic Gestures : Empirical Study and Computational Simulation with Virtual Agents*. PhD thesis, Bielefeld University. 78

[Bickmore and Cassell, 2001] Bickmore, T. and Cassell, J. (2001). Relational agents : a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '01, pages 396–403, New York, NY, USA. ACM. 10, 42, 46

[Bickmore et al., 2011] Bickmore, T., Pfeifer, L., and Schulman, D. (2011). Relational agents improve engagement and learning in science museum visitors. In *Proceedings of the 10th international conference on Intelligent virtual agents*, IVA'11, pages 55–67, Berlin, Heidelberg. Springer-Verlag. 12, 13

[Bickmore and Picard, 2005] Bickmore, T. W. and Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2) :293–327. 42, 46, 63

[Bowman et al., 2007] Bowman, D., McMahan, R. P., et al. (2007). Virtual reality : how much immersion is enough ? *Computer*, 40(7) :36–43. 117

[Briton and Hall, 1995] Briton, N. J. and Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, 32 :79–90. 25, 27, 65, 69

[Buisine et al., 2006] Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L., and Pelachaud, C. (2006). Perception of blended emotions : From video corpus to expressive agent. In *Intelligent virtual agents*, pages 93–106. Springer. 63

[Burgoon et al., 1984] Burgoon, J. K., Buller, D. B., Hale, J. L., and de Turck, M. A. (1984). Relational Messages Associated with Nonverbal Behaviors. *Human Communication Research*, 10(3) :351–378. 21, 22, 25, 27, 68, 69, 76

[Burgoon and Le Poire, 1999] Burgoon, J. K. and Le Poire, B. A. (1999). Nonverbal cues and interpersonal judgments : Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs*, 66(2) :105–124. 22, 25, 27, 68, 69

[Busso et al., 2007] Busso, C., Deng, Z., Grimm, M., Neumann, U., and Narayanan, S. (2007). Rigid head motion in expressive speech animation : Analysis and synthesis. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3) :1075–1086. 78

[Cafaro et al., 2012] Cafaro, A., Vilhjálmsson, H. H., Bickmore, T., Heylen, D., Jóhannsdóttir, K. R., and Valgardsson, G. S. (2012). First impressions : Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *Intelligent Virtual Agents*, pages 67–80. Springer. 13, 47, 48, 49, 154

[Callejas et al., 2014] Callejas, Z., Ravenet, B., Ochs, M., and Pelachaud, C. (2014). A computational model of social attitudes for a virtual recruiter. In *Autonomous Agent and Multiagent Systems*. 83, 85, 88, 89

[Cappella et al., 1985] Cappella, J. N., Siegman, A. W., and Feldstein, S. (1985). Controlling the floor in conversation. *Multichannel integrations of nonverbal behavior*, pages 69–103. 36, 104, 106

[Carney et al., 2005] Carney, D. R., Hall, J. A., and LeBeau, L. (2005). Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior*, 29 :105–123. 22, 25, 27, 68, 69, 76

[Cassell et al., 1999] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., and Yan, H. (1999). Embodiment in conversational interfaces : Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems : the CHI is the limit*, CHI '99, pages 520–527, New York, NY, USA. 10, 42, 43

[Cassell et al., 2000] Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (2000). *Embodied Conversational Agents*. MIT Press. 9, 41

[Chindamo et al., 2012] Chindamo, M., Allwood, J., and Ahlsén, E. (2012). Some suggestions for the study of stance in communication. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 617–622. IEEE. 20, 153

[Chollet et al., 2014] Chollet, M., Ochs, M., and Pelachaud, C. (2014). From non-verbal signals sequence mining to bayesian networks for interpersonal attitudes expression. In *Intelligent Virtual Agents*, pages 120–133. Springer. 12, 13, 22, 48, 49, 63, 155

[Clark, 1996] Clark, H. H. (1996). *Using language*, volume 1996. Cambridge University Press Cambridge. 35, 53, 101, 106

[Courgeon et al., 2008] Courgeon, M., Martin, J.-C., and Jacquemin, C. (2008). Marc : a multimodal affective and reactive character. In *in Proceedings of the 1st Workshop on AFFective Interaction in Natural Environements*. 12

[Creed et al., 2008] Creed, P. A., Doherty, F., and O'Callaghan, F. (2008). Job-seeking and job-acquisition in high school students. *Journal of Vocational Behavior*, 73(2) :195–202. 85

[Cristani et al., 2011] Cristani, M., Paggetti, G., Vinciarelli, A., Bazzani, L., Menegaz, G., and Murino, V. (2011). Towards computational proxemics : Inferring social relations from interpersonal distances. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 290–297. IEEE. 33, 109, 110

[Damian et al., 2011] Damian, I., Endrass, B., Huber, P., Bee, N., and André, E. (2011). Individualized agent interactions. In *Motion in Games*, pages 15–26. Springer. 57, 58, 110

[Darwin, 2002] Darwin, C. (2002). *The expression of the emotions in man and animals*. Oxford University Press. 25

[De Silva et al., 1997] De Silva, L. C., Miyasato, T., and Nakatsu, R. (1997). Facial emotion recognition using multi-modal information. In *Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on*, volume 1, pages 397–401. IEEE. 13

[DePaulo, 1992] DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological bulletin*, 111(2) :203. 24

[Dias and Paiva, 2005] Dias, J. and Paiva, A. (2005). Feeling and reasoning : A computational model for emotional characters. In *Progress in Artificial Intelligence*, volume 3808 of *Lecture Notes in Computer Science*, pages 127–140. Springer Berlin / Heidelberg. 61, 62

[Dunbar and Burgoon, 2005] Dunbar, N. E. and Burgoon, J. K. (2005). Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships*, 22(2) :207–233. 37

[Duncan, 1972] Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2) :283. 35, 51, 52, 107, 108, 111, 112

[Ekman, 1979] Ekman, P. (1979). About brows : Emotional and conversational signals. *Human ethology*, pages 169–202. 23

[Ekman, 1999] Ekman, P. (1999). Facial expressions. *Handbook of cognition and emotion*, 53 :226–232. 25

[Ekman and Friesen, 1969] Ekman, P. and Friesen, W. (1969). The repertoire of nonverbal behavior : Categories, origins, usage and coding. *Semiotica*. 24, 25

[Ekman and Friesen, 1976] Ekman, P. and Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1) :56–75. 23

[El-Nasr et al., 2000] El-Nasr, M. S., Yen, J., and Ioerger, T. R. (2000). Flame :fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3) :219–257. 11

[Funder et al., 2000] Funder, D. C., Furr, R. M., and Colvin, C. R. (2000). The riverside behavioral q-sort : a tool for the description of social behavior. *Journal of Personality*, 68(3) :451–489. 124, 162

[Gatica-Perez, 2009] Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups : A review. *Image and Vision Computing*, 27(12) :1775–1787. 38, 153

[Gebhard, 2005] Gebhard, P. (2005). Alma : a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pages 29–36, New York, NY, USA. ACM. 11, 61

[Gifford and Hine, 1994] Gifford, R. and Hine, D. W. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispositions. *Journal of Research in Personality*, 28(2) :115–132. 37

[Gillies and Ballin, 2003] Gillies, M. and Ballin, D. (2003). A model of interpersonal attitude and posture generation. In Rist, T., Aylett, R., Ballin, D., and Rickel, J., editors, *Intelligent Virtual Agents*, volume 2792 of *Lecture Notes in Computer Science*, pages 88–92. Springer Berlin Heidelberg. 22

[Gillies et al., 2004] Gillies, M., Crabtree, I. B., and Ballin, D. (2004). Customisation and context for expressive behaviour in the broadband world. *BT Technology Journal*, 22(2) :7–17. 46

[Gobl et al., 2003] Gobl, C., Ni, A., et al. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech communication*, 40(1) :189–212. 68

[Goffman, 1981] Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press. 102

[Goldberg and Cannon-Bowers, 2015] Goldberg, B. and Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game : Pedagogical agents make a difference. *Computers in Human Behavior*, 52 :1–11. 12

[Goldberg, 1990] Goldberg, J. A. (1990). Interrupting the discourse on interruptions : An analysis in terms of relationally neutral, power-and rapport-oriented acts. *Journal of Pragmatics*, 14(6) :883–903. 37, 101, 102, 104, 105, 106, 111, 112

[Gordon and Breazeal, 2015] Gordon, M. and Breazeal, C. (2015). Designing a virtual assistant for in-car child entertainment. In *Proceedings of the 14th International Conference on Interaction Design and Children*, IDC '15, pages 359–362, New York, NY, USA. ACM. 12

[Griol et al., 2014] Griol, D., Callejas, Z., López-Cózar, R., and Riccardi, G. (2014). A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech & Language*, 28(3) :743–768. 14

[Gurtman, 2009] Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex. *Social and Personality Psychology Compass*, 3(4) :601–619. 22

[Hall, 1969] Hall, E. T. (1969). *The hidden dimension*, volume 1990. Anchor Books New York. 30, 48, 54, 56, 109, 112

[Hall et al., 2009] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software : an update. *ACM SIGKDD Explorations Newsletter*, 11(1) :10–18. 79

[Hartholt et al., 2013] Hartholt, A., Traum, D., Marsella, S. C., Shapiro, A., Stratou, G., Leuski, A., Morency, L.-P., and Gratch, J. (2013). All together now. In *Intelligent Virtual Agents*, pages 368–381. Springer. 13

[Heider, 1958] Heider, F. (1958). *The Psychology of Interpersonal Relations*. Psychology Press. 45, 46

[Hess et al., 2005] Hess, U., Adams, R., and Kleck, R. (2005). Who may frown and who should smile ? dominance, affiliation, and the display of happiness and anger. *Cognition and Emotion*, 19(4) :515–536. 27, 68

[Hess and Thibault, 2009] Hess, U. and Thibault, P. (2009). Why the same expression may not mean the same when shown on different faces or seen by different people. In Tao, J. and Tan, T., editors, *Affective Information Processing*, pages 145–158. Springer London. 27

[Heylen et al., 2008] Heylen, D., Kopp, S., Marsella, S. C., Pelachaud, C., and Vilhjálmsson, H. (2008). The next step towards a function markup language. In *Proceedings of the 8th international conference on Intelligent Virtual Agents*, IVA '08, pages 270–280, Berlin, Heidelberg. Springer-Verlag. 88

[Hill et al., 2001] Hill, R., Gratch, J., Johnson, W. L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiébaux, M., Tuch, L., Whitney, R., Douglas, J., and Swartout, W. (2001). Toward the holodeck : Integrating graphics, sound, character and story. In *Proceedings of the Fifth International Conference on Autonomous Agents*, AGENTS '01, pages 409–416, New York, NY, USA. ACM. 51, 54

[Hofmann et al., 2015] Hofmann, J., Platt, T., Ruch, W., Niewiadomski, R., and Urbain, J. (2015). The influence of a virtual companion on amusement when watching funny films. *Motivation and Emotion*, 39(3) :434–447. 12

[Hogg, 2000] Hogg, M. (2000). Social processes and human behavior. In Pawlik, K. and Rosenzweig, M., editors, *The International Handbook of Psychology*. SAGE Publications. 20

[Homans, 2013] Homans, G. C. (2013). *The human group*, volume 7. Routledge. 28, 99

[Huffcutt et al., 2011] Huffcutt, A. I., Van Iddekinge, C. H., and Roth, P. L. (2011). Understanding applicant behavior in employment interviews : A theoretical model of interviewee performance. *Human Resource Management Review*, 21(4) :353–367. 85

[Isbister, 2006] Isbister, K. (2006). *Better Game Characters by Design : A Psychological Approach (The Morgan Kaufmann Series in Interactive 3D Technology)*. Morgan Kaufmann. 25, 27, 89

[Jan and Traum, 2005] Jan, D. and Traum, D. R. (2005). Dialog simulation for background characters. In *Intelligent Virtual Agents*, pages 65–74. Springer. 54

[Jan and Traum, 2007] Jan, D. and Traum, D. R. (2007). Dynamic movement and positioning of embodied agents in multiparty conversations. In *Proceedings of the Workshop on Embodied Language Processing*, pages 59–66. Association for Computational Linguistics. 54, 55, 58

[Kasap et al., 2009] Kasap, Z., Moussa, M. B., Chaudhuri, P., and Magnenat-Thalmann, N. (2009). Making them remember emotional virtual characters with memory. *IEEE Computer Graphics and Applications*, pages 20–29. 44

[Keating et al., 1981] Keating, C. F., Mazur, A., Segall, M. H., Cysneiros, P. G., Kilbride, J. E., Leahy, P., Divale, W. T., Komin, S., Thurman, B., and Wirsing, R. (1981). Culture and the perception of social dominance from facial expression. *Journal of personality and social psychology*, 40(4) :615. 152

[Kendira, 2010] Kendira, A. (2010). *L'animation comportementale d'avatars comme support à la communication dans les environnements virtuels informés*. PhD thesis, Compiègne. 79

[Kendon, 1990] Kendon, A. (1990). *Conducting interaction : Patterns of behavior in focused encounters*, volume 7. CUP Archive. 28, 29, 30, 48, 54, 56, 107, 112

[Kennedy and Camden, 1983] Kennedy, C. W. and Camden, C. (1983). Interruptions and nonverbal gender differences. *Journal of Nonverbal Behavior*, 8(2) :91–108. 37, 65

[Knapp et al., 2013] Knapp, M., Hall, J., and Horgan, T. (2013). *Nonverbal communication in human interaction*. Cengage Learning. 23

[Koller and Friedman, 2009] Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models : Principles and Techniques*. MIT Press. 78

[Kopp et al., 2003] Kopp, S., Jung, B., Leßmann, N., and Wachsmuth, I. (2003). Max - a multimodal assistant in virtual reality construction. *KI*, 17(4) :11. 11

[Kopp et al., 2006] Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Tharisson, K., and Vilhjalmsson, H. (2006). Towards a common framework for multimodal generation : The behavior markup language. In Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors, *Intelligent Virtual Agents*, volume 4133 of *Lecture Notes in Computer Science*, pages 205–217. Springer Berlin / Heidelberg. 12, 81, 86, 145, 147

[Kosinski, 2008] Kosinski, R. J. (2008). A literature review on reaction time. *Clemson University*, 10. 103

[Kotsiantis et al., 2007] Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning : A review of classification techniques. 77

[Krenn et al., 2014] Krenn, B., Endrass, B., Kistler, F., and André, E. (2014). Effects of language variety on personality perception in embodied conversational agents. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, pages 429–439. Springer. 12

[Kuijk et al., 2013] Kuijk, F., Van Broeck, S., Dareau, C., Ravenet, B., Ochs, M., Apostolakis, K., Daras, P., Monaghan, D., O'Connor, N. E., Wall, J., et al. (2013). A framework for human-like behavior in an immersive virtual world. In *Digital Signal Processing (DSP), 2013 18th International Conference on*, pages 1–7. IEEE. 144

[Laird, 2012] Laird, J. (2012). *The Soar cognitive architecture*. MIT Press. 61

[Lamere et al., 2003] Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., Warmuth, M., and Wolf, P. (2003). The cmu sphinx-4 speech recognition system. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong*, volume 1, pages 2–5. Citeseer. 146, 155

[Leary, 1958] Leary, T. (1958). Interpersonal diagnosis of personality. *American Journal of Physical Medicine & Rehabilitation*, 37(6) :331. 22, 66

[Lee and Marsella, 2011] Lee, J. and Marsella, S. (2011). Modeling side participants and bystanders : The importance of being a laugh track. In Vilhjálmsson, H. H., Kopp, S., Marsella, S., and Thórisson, K. R., editors, *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pages 240–247. Springer Berlin Heidelberg. 47, 49, 57, 58

[Leßmann et al., 2004] Leßmann, N., Kranstedt, A., and Wachsmuth, I. (2004). Towards a cognitively motivated processing of turn-taking signals for the embodied conversational agent max. Proceedings Workshop Embodied Conversational Agents : Balanced Perception and Action, pages 57–64. IEEE Computer Society. 51, 52, 53

[Li et al., 2007] Li, T., Qiu, Y., Yue, P., and Zhong, G. (2007). Exploiting model of personality and emotion of learning companion agent. In *Computer Systems and Applications, 2007. AICCSA'07. IEEE/ACS International Conference on*, pages 860–865. IEEE. 61

[Macan, 2009] Macan, T. (2009). The employment interview : A review of current studies and directions for future research. *Human Resource Management Review*, 19(3) :203–218. 85

[Mairesse and Walker, 2011] Mairesse, F. and Walker, M. A. (2011). Controlling user perceptions of linguistic style : Trainable generation of personality traits. *Computational Linguistics*, 37(3) :455–488. 89

[Mancini and Pelachaud, 2007] Mancini, M. and Pelachaud, C. (2007). Dynamic behavior qualifiers for conversational agents. In Pelachaud, C., Martin, J.-C., Andre, E., Chollet, G., Karpouzis, K., and Pele, D., editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, pages 112–124. Springer Berlin / Heidelberg. 81

[Marsella and Gratch, 2009] Marsella, S. C. and Gratch, J. (2009). Ema : A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1) :70 – 90. 11, 13

[Mason and Suri, 2012] Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior research methods*, 44(1) :1–23. 64

[Mast, 2002] Mast, M. S. (2002). Dominance as expressed and inferred through speaking time. *Human Communication Research*, 28(3) :420–450. 38

[McCrae and Costa, 1987] McCrae, R. R. and Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1) :81. 89

[McCrae and John, 1992] McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2) :175–215. 89

[McGrath, 1984] McGrath, J. E. (1984). *Groups : Interaction and performance*, volume 14. Prentice-Hall Englewood Cliffs, NJ. 28, 33, 55, 99

[McNeill, 1992] McNeill, D. (1992). *Hand and Mind : What Gestures Reveal about Thought*. University of Chicago Press. 24

[McTear, 2004] McTear, M. F. (2004). Spoken dialog technology : Toward the conversational user interface. 62

[Mehrabian, 1969] Mehrabian, A. (1969). Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin*, 71(5) :359. 22, 25, 27, 32, 107, 109, 110, 111, 112

[Niewiadomski et al., 2009] Niewiadomski, R., Bevacqua, E., Mancini, M., and Pelachaud, C. (2009). Greta : an interactive expressive eca system. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '09, pages 1399–1400, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems. 12

[Nouri and Traum, 2014] Nouri, E. and Traum, D. (2014). Generative models of cultural decision making for virtual agents based on user's reported values. In *Intelligent Virtual Agents*, pages 310–315. Springer. 62

[Ochs et al., 2012] Ochs, M., Niewiadomski, R., Brunet, P., and Pelachaud, C. (2012). Smiling virtual agent in social context. *Cognitive Processing*, 13(2) :519–532. 64, 70, 72, 77

[Ochs et al., 2009] Ochs, M., Sabouret, N., and Corruble, V. (2009). Simulation of the dynamics of nonplayer characters' emotions and social relations in games. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(4) :281 –297. 44, 46

[O'Connell et al., 1990] O'Connell, D. C., Kowal, S., and Kaltenbacher, E. (1990). Turn-taking : A critical analysis of the research tradition. *Journal of psycholinguistic research*, 19(6) :345–373. 35, 36, 102, 105

[Ortony et al., 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The cognitive structure of emotions*. Campbridge University Press. 44

[Osgood and Tannenbaum, 1955] Osgood, C. E. and Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological review*, 62(1) :42. 43, 46, 56

[Parsons, 2014] Parsons, T. D. (2014). Virtual teacher and classroom for assessment of neurodevelopmental disorders. In *Technologies of Inclusive Well-Being*, pages 121–137. Springer. 12

[Pecune et al., 2014] Pecune, F., Cafaro, A., Chollet, M., Philippe, P., and Pelachaud, C. (2014). Suggestions for extending saiba with the vib platform. In *Proceedings of the Workshop on Architectures and Standards for Intelligent Virtual Agents at IVA 2014*. 70, 112

[Pecune et al., 2013] Pecune, F., Ochs, M., and Pelachaud, C. (2013). A formal model of social relations for artificial companions. In *Proceedings of The European Workshop on Multi-Agent Systems (EUMAS)*. 13

[Pedica and Högni Vilhjálmsson, 2010] Pedica, C. and Högni Vilhjálmsson, H. (2010). Spontaneous avatar behavior for human territoriality. *Applied Artificial Intelligence*, 24(6) :575–593. 56, 57, 58

[Pedica and Vilhjálmsson, 2012] Pedica, C. and Vilhjálmsson, H. H. (2012). Lifelike interactive characters with behavior trees for social territorial intelligence. In *ACM SIGGRAPH 2012 Posters*, page 32. ACM. 113

[Pedica et al., 2010] Pedica, C., Vilhjálmsson, H. H., and Lárusdóttir, M. (2010). Avatars in conversation : the importance of simulating territorial behavior. In *Intelligent Virtual Agents*, pages 336–342. Springer. 56, 110, 112

[Pelachaud and Bilvi, 2003] Pelachaud, C. and Bilvi, M. (2003). Computational model of believable conversational agents. In *Communication in multiagent systems*, pages 300–317. Springer. 11

[Pelachaud and Poggi, 2002] Pelachaud, C. and Poggi, I. (2002). Subtleties of facial expressions in embodied agents. *The Journal of Visualization and Computer Animation*, 13(5) :301–312. 11, 79

[Poggi et al., 2005] Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., and De Carolis, B. (2005). Greta. a believable embodied conversational agent. In *Multimodal intelligent information presentation*, pages 3–25. Springer. 11

[Power and Dal Martello, 1986] Power, R. J. and Dal Martello, M. F. (1986). Some criticisms of sacks, schegloff, and jefferson on turn taking. *Semiotica*, 58(1-2) :29–40. 35

[Prada and Paiva, 2005] Prada, R. and Paiva, A. (2005). Believable groups of synthetic characters. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 37–43. ACM. 55, 56, 58

[Prada and Paiva, 2009] Prada, R. and Paiva, A. (2009). Teaming up humans with autonomous synthetic characters. *Artificial Intelligence*, 173(1) :80–103. 45, 46

[Raux and Eskenazi, 2009] Raux, A. and Eskenazi, M. (2009). A finite-state turn-taking model for spoken dialog systems. In *Proceedings of Human Language Technologies : The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 629–637. Association for Computational Linguistics. 50, 53

[Ravenet et al., 2014] Ravenet, B., Ochs, M., and Pelachaud, C. (2014). Architecture of a socio-conversational agent in virtual worlds. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 3983–3987. IEEE. 147

[Read et al., 2007] Read, S., Miller, L., Kostygina, A., Chopra, G., Christensen, J. L., Corsbie-Massay, C., Zachary, W., LeMentec, J.-C., Iordanov, V., and Rosoff, A. (2007). The personality-enabled architecture for cognition (pac). In *Affective Computing and Intelligent Interaction*, pages 735–736. Springer. 61

[Rehm and Endrass, 2009] Rehm, M. and Endrass, B. (2009). Rapid prototyping of social group dynamics in multiagent systems. *AI and Society*, 24 :13–23. 10.1007/s00146-009-0191-8. 56, 58

[Rehm et al., 2008] Rehm, M., Nakano, Y., André, E., and Nishida, T. (2008). Culture-specific first meeting encounters between virtual agents. In *Intelligent virtual agents*, pages 223–236. Springer. 12

[Reilly, 1996] Reilly, W. S. (1996). Believable social and emotional agents. Technical report, DTIC Document. 61

[Rich and Sidner, 2012] Rich, C. and Sidner, C. L. (2012). Using collaborative discourse theory to partially automate dialogue tree authoring. In *Intelligent Virtual Agents*, pages 327–340. Springer. 147

[Rist et al., 2003] Rist, T., Schmitt, M., Pelachaud, C., and Bilvi, M. (2003). Towards a simulation of conversations with expressive embodied speakers and listeners. In *Computer Animation and Social Agents, 2003. 16th International Conference on*, pages 5–10. IEEE. 43, 46

[Romano and Wong, 2004] Romano, D. M. and Wong, A. K. L. (2004). Personality model of a social character. In *18th British HCI Group Annual Conference*. 61

[Rousseau et al., 1996] Rousseau, D. et al. (1996). Personality in computer characters. In *Proceedings of the 1996 AAAI Workshop on Entertainment and AI/A-Life, AAAI Press, Portland, Oregon*, page 3843. 61

[Sacks et al., 1974] Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735. 34, 50, 51, 52, 53, 100, 101, 106

[Sadler and Woody, 2010] Sadler, P. and Woody, E. (2010). Interpersonal complementarity. *Handbook of interpersonal psychology : Theory, research, assessment, and therapeutic interventions*, page 123. 119

[Sanchez-Cortes et al., 2013] Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., and Gatica-Perez, D. (2013). Emergent leaders through looking and speaking : from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces*, 7(1-2) :39–53. 33, 153

[Scheflen and Ashcraft, 1976] Scheflen, A. E. and Ashcraft, N. (1976). Human territories : How we behave in space-time. 29, 30, 56, 107, 112

[Scherer, 2005] Scherer, K. (2005). What are emotions ? and how can they be measured ? *Social Science Information*. 20, 21, 41, 46

[Schutz, 1958] Schutz, W. C. (1958). Firo : A three-dimensional theory of interpersonal behavior. 21

[Searle, 1976] Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5 :1–23. 67

[Serban and Pauchet, 2013] Serban, O. and Pauchet, A. (2013). Agentslang : A fast and reliable platform for distributed interactive systems. In *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, pages 35–42. IEEE. 12

[Sidnell, 2011] Sidnell, J. (2011). *Conversation analysis : An introduction*, volume 45. John Wiley & Sons. 35

[Siebel, 2003] Siebel, M. (2003). Illocutionary acts and attitude expression. *Linguistics and Philosophy*, 26(3) :351–366. 65, 67

[Smith-Lovin and Brody, 1989] Smith-Lovin, L. and Brody, C. (1989). Interruptions in group discussions : The effects of gender and group composition. *American Sociological Review*, pages 424–435. 37

[Ter Maat et al., 2010] Ter Maat, M., Truong, K. P., and Heylen, D. (2010). How turn-taking strategies influence usersímpressions of an agent. In *Intelligent Virtual Agents*, pages 441–453. Springer. 50

[Thalmann, 2007] Thalmann, D. (2007). *Crowd simulation*. Wiley Online Library. 54

[Thórisson, 2002] Thórisson, K. R. (2002). Natural turn-taking needs no manual : Computational theory and model, from perception to action. In *Multimodality in language and speech systems*, pages 173–207. Springer. 51

[Thórisson et al., 2010] Thórisson, K. R., Gislason, O., Jonsdottir, G. R., and Thorisson, H. T. (2010). A multiparty multimodal architecture for realtime turntaking. In *Intelligent Virtual Agents*, pages 350–356. Springer. 51, 53, 102, 106

[Traum and Rickel, 2002] Traum, D. and Rickel, J. (2002). Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems : part 2*, pages 766–773. ACM. 51

[Tross and Maurer, 2008] Tross, S. A. and Maurer, T. J. (2008). The effect of coaching interviewees on subsequent interview performance in structured experience-based interviews. *Journal of Occupational and Organizational Psychology*, 81(4) :589–605. 85

[Videla and Williams, 2012] Videla, A. and Williams, J. J. (2012). *RabbitMQ in action*. Manning. 146

[Vinciarelli et al., 2009] Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing : Survey of an emerging domain. *Image and Vision Computing*, 27(12) :1743–1759. 25, 26

[Wagner et al., 2013] Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., and André, E. (2013). The social signal interpretation (ssi) framework : multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 831–834. ACM. 88, 155

[Wall et al., 2014] Wall, J., Izquierdo, E., Argyriou, L., Monaghan, D. S., O'Connor, N. E., Poulakos, S., Smolic, A., and Mekuria, R. (2014). Reverie : Natural human interaction in virtual immersive environments. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 2165–2167. IEEE. 10, 143

[Wiggins, 2003] Wiggins, J. S. (2003). *Paradigms of personality assessment*. Guilford Press. 22

[Wiggins et al., 1988] Wiggins, J. S., Trapnell, P., and Phillips, N. (1988). Psychometric and geometric characteristics of the revised interpersonal adjective scales (ias-r). *Multivariate Behavioral Research*, 23(4) :517–530. 124, 162

[Yabar and Hess, 2007] Yabar, Y. and Hess, U. (2007). Display of empathy and perception of out-group members. *New Zealand Journal of Psychology*. 25, 27

[Yuen et al., 2011] Yuen, M.-C., King, I., and Leung, K.-S. (2011). A survey of crowdsourcing systems. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 766–773. IEEE. 64

Modélisation de comportements non-verbaux et d'attitudes sociales dans la simulation de groupes conversationnels

Brian Ravenet

RESUME :

Les Agents Conversationnels Animés sont des personnages virtuels dont la fonction principale est d'interagir avec l'utilisateur. Ils sont utilisés dans différents domaines tels que l'assistance personnelle, l'entrainement social ou les jeux vidéo et afin d'améliorer leur potentiel, il est possible de leur donner la capacité d'exprimer des comportements similaires à ceux des humains. Les utilisateurs, conscient d'interagir avec une machine, sont tout de même capable d'analyser et d'identifier des comportements sociaux à travers les signaux émis par les agents. La recherche en ACA s'est longtemps intéressée aux mécanismes de reproduction et de reconnaissance des émotions au sein de ces personnages virtuels et maintenant l'intérêt se porte sur la capacité d'exprimer différentes attitudes sociales. Ces attitudes reflètent un style comportemental et s'expriment à travers différentes modalités du corps comme les expressions faciales, les regards ou les gestes par exemple.

Nous avons proposé un modèle permettant à un agent de produire différents comportements non-verbaux traduisant l'expression d'attitudes sociales dans une conversation. L'ensemble des comportements générés par notre modèle permettent à un groupe d'agents animés par celui-ci de simuler une conversation, sans tenir compte du contenu verbal. Deux évaluations du modèle ont été conduites, l'une sur Internet et l'autre dans un environnement de réalité virtuelle, afin de vérifier que les attitudes étaient bien reconnues.

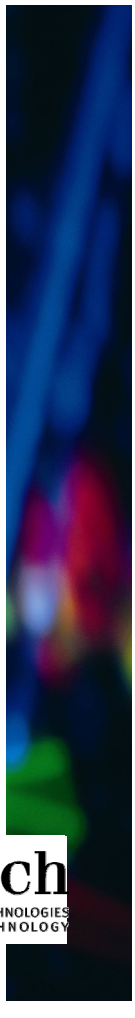
MOTS-CLEFS : Agents Conversationnels Animés - Groupes - Attitudes Sociales - Réalité Virtuelle

ABSTRACT :

Embodied Conversational Agents are virtual characters which main purpose is to interact with a human user. They are used in various domains such as personal assistance, social training or video games for instance. In order to improve their capabilities, it is possible to give them the ability to produce human-like behaviors. The users, even if they are aware that they interact with a machine, are still capable of analyzing and identifying social behaviors through the signals produced by these virtual characters. The research in Embodied Conversational Agents has focused for a long time on the reproduction and recognition of emotions by virtual characters and now the focus is on the ability to express different social attitudes. These attitudes show a behavioral style and are expressed through different modalities of the body, like the facial expressions, the gestures or the gazes for instance.

We proposed a model that allows an agent to produce different nonverbal behaviors expressing different social attitudes in a conversation. The whole set of behaviors produced by our model allows a goup of agents animated by it to simulate a conversation, without any verbal content. Two evaluations of the model were conducted, one on the Internet and one in a Virtual Reality environment, to verify that the attitudes produced are well recognized

KEY-WORDS : Embodied Conversational Agents - Groups - Social Attitudes - Virtual Reality



TELECOM ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY