Modélisation des stratégies verbales d'engagement dans les interactions humain-agent

Nadine Glas

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2016-ENST-0047

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Traitement du Signal et des Images »

présentée et soutenue publiquement par

Nadine GLAS

le 13 Septembre 2016

Modélisation des Stratégies Verbales d'Engagement dans les Interactions Humain-Agent

Directeur de thèse : **Catherine PELACHAUD**

Jury

M. Dirk HEYLEN, Professeur, HMI, University of Twente	Rapporteur
Mme. Sophie ROSSET, Directeur de Recherches, LIMSI, CNRS	Rapporteur
M. Jonathan GINZBURG, Professeur, UFR d'études anglophones, Université Paris-Diderot	Examinateur
M. Frédéric LANDRAGIN, Directeur de Recherches, LATTICE, CNRS	Examinateur
M. Maxime MORGE, Maître de Conférences, SMAC, Université de Lille	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

THÈSE

To my Brother and Parents

Acknowledgements

This thesis was not possible without the help of several people, each of whom I owe my sincere gratitude.

First and foremost I would like to thank my supervisor Catherine Pelachaud. I am grateful for the extraordinary opportunity that she provided me by offering me a PhD position. I would like to thank her very much for her valuable advice along the way, sometimes from the other side of the world or in the middle of the night, which shows her commitment and sincere care for her students. Besides this, I am thankful that she introduced me to the domain of virtual agents and enabled me to go to a summer school, workshops, and conferences, giving me the opportunity to gather knowledge, meet interesting people, and visit amazing places.

My sincerest expression of gratitude also go to my jury members: Dirk Heylen, Sophie Rosset, Frederic Landragin, Jonathan Ginzburg and Maxim Morge. I am thankful for the time they invested in my work, their questions and remarks. Thanks in particular to Jonathan Ginzburg and Sophie Rosset for also providing valuable comments at earlier stages of my work. For the same reason I would like to thank Charles Rich, Candy Sidner, and Andrew Kehler.

My colleagues have contributed enormously to my PhD by always providing help and good atmosphere: Ken was my co-supervisor at the beginning of this thesis. I would like to thank him for interesting discussions and getting me started in the lab. I thank Brice, who was of incredible help on numerous occasions, even in the evenings and weekends, with things such as getting a website running or debugging the SVN. Sabrina I'm grateful for the fun we had and results we achieved working both on project A1:1. Similarly, I enjoyed working with Angelo on interruptions, thanks. I am grateful to David for acting in some video stimuli.

I also shared the lab for a longer period of time with colleagues Florian, Caroline, Nesrine, Yu, Brian, Andre-Marie, Mathieu, Pierre, Jing, Chloé, Beatrice, Valentin, Irina, Soumia, Kevin, and Thomas. I would like to thank them all very much for their help in for example correcting my French, being guinea pigs for experiments, and for their pleasant company. Others I have seen for a shorter period in the lab but have also contributed to this pleasant working atmosphere: Zoraida, Marco, Nicholas, Hector, Amyr, Matthieu, Herwin, Abhishek, Quoc-Ahn, Lucas, Jessica,

Pierre C., Clémence, Rachel, and Radek.

I owe the results that are described in this thesis to the numerous anonymous participants of the empirical studies; I am very grateful for their contribution.

Even when we were most of the time 8937 km apart and in different time zones, Hervé has in many ways travelled the entire road of this PhD by my side. This means a lot to me and I am very thankful for this.

I owe my parents thanks for having built my foundations, and thereby the foundations of this thesis, as they have set me an example of working hard and critical thinking. My brother is a source of comfort as I can always rely on him. My brother and my parents have provided support and encouragement throughout this PhD, even when that meant that I was living abroad. These are some of the reasons why I dedicate this thesis to them.

I am also grateful to my other family members, Hervé's family members, and my friends, you know who you are, for their (digital) presence and the welcome distractions they provided.

Lastly, I would like to thank everyone who was at the defence of this thesis, some taking a day off work and coming from far, including Lennart, Nancy, Mathilde and Nathalie.

Contents

List of Figures

List of Tables

Résumé Long

Introduction

Un aspect essentiel des interactions humain-agent est la mesure dans laquelle l'humain est engagé. L'engagement de l'utilisateur est une variable sociale qui fait référence à la connexion de l'utilisateur humain avec l'agent et l'interaction. Un niveau minimum d'engagement de l'utilisateur est nécessaire dans les interactions humain-agent pour atteindre l'objectif pour lequel l'agent est construit. En raison du nombre croissant d'agents conversationnels et du rôle important que joue l'engagement dans le domaine des interactions humain-agent, au cours des dernières années l'engagement a gagné de plus en plus d'attention.

Les travaux de recherche décrits dans cette thèse visent à contribuer à une meilleure compréhension de l'engagement en explorant les stratégies visant à renforcer l'engagement de l'utilisateur au cours de l'interaction face-à-face avec un agent conversationnel. Cette recherche est menée dans le contexte du Projet A1:1 qui vise à construire un agent virtuel conversationnel à taille humaine dans un musée qui engage l'utilisateur au cours d'une conversation sur le musée et ses œuvres d'art. Ce cadre constitue l'application de cette thèse.

L'engagement peut être exprimé et favorisé à la fois par le comportement verbal et non verbal. Nous concentrons notre recherche sur le comportement verbal de l'agent. En effet, les recherches antérieures sur l'engagement se concentrent souvent sur le comportement non-verbal. Le rôle du comportement verbal, en particulier dans des contextes non orientés vers la tâche, est peu étudié. Selon [Bickmore and Cassell, 2001] "savoir quand et comment utiliser le langage pour atteindre des objectifs sociaux est crucial pour nos agents conversationnels si ils doivent être aussi efficaces que les humains, et si nous voulons que les humains soient en mesure d'utiliser nos agents facilement, efficacement, et en coopération".

Nous essayons de contribuer à la connaissance de l'engagement dans l'interaction humain-agent en tenant compte des stratégies verbales potentiellement engageantes. Étant donné que le comportement verbal d'un agent comprend de nombreux aspects, nous nous limitons à l'étude de trois stratégies différentes : la première est liée à la forme des énoncés de l'agent, la deuxième est liée au timing des énoncés, et la troisième est quant à elle liée au contenu des énoncés. Concrètement, nous regar-

dons respectivement un aspect de la politesse, des interruptions, et des sujets dans l'interaction non-orientée-tâche. Bien que ces aspects du comportement verbal aient tous été étudiés antérieurement, nous les étudions sous un angle différent en considérant si et comment ils impactent et/ou sont dépendants de l'engagement des participants de l'interaction.

Les travaux réalisés dans cette thèse portent sur plusieurs domaines de la recherche, comme les agents conversationnels animés, l'intelligence artificielle, la linguistique computationnelle, l'analyse conversationnelle et les sciences cognitives. Si ces domaines sont généralement étudiés de manière isolée, nous essayons dans cette thèse de les étudier tous ensemble. Par exemple, dans le domaine de la linguistique computationnelle, les systèmes de dialogue sont souvent développés comme systèmes indépendants qui gèrent la parole, sans tenir compte d'autres aspects cognitifs des participants de l'interaction. D'autre part, dans le domaine des interactions humain-agent, les systèmes cognitifs qui dirigent par exemple les émotions et le comportement non verbal sont considérés en détail, mais les systèmes de dialogue (basés sur les tâches) sont souvent employés comme un simple outil pour doter l'agent de la parole sans apporter plus d'attention à ce composant. Nous croyons cependant que si nous voulons qu'un agent dispose d'un véritable comportement humain nous devons traiter les systèmes de dialogue et les systèmes cognitifs comme concepts entrelacés ne pouvant pas être considérés l'un sans l'autre. Dans cette thèse, nous essayons d'apporter une contribution pluridisciplinaire.

CONTEXTE THÉORIQUE

Définitions d'engagement

Dans le chapitre 2 nous donnons un aperçu des différentes définitions de l'engagement et identifions les différences et les points communs entre les définitions. Compte tenu du nombre et de la variation des définitions nous pouvons affirmer que l'engagement dans l'interaction humain-agent est un concept complexe. Compte tenu des différences parmi les définitions, les études peuvent se concentrer sur un aspect ou interprétation particulière de l'engagement sans être en mesure de couvrir toute la gamme des interprétations.

Les différentes définitions impliquent toutes une relation dynamique d'un participant d'une interaction vers/avec un autre participant et/ou l'interaction, en raison d'une connexion et de la coopération entre les participants de l'interaction. Les

parties impliquées dans la relation, les perspectives, et les mesures qui sont sous-entendues par les définitions, font que chaque définition souligne un autre aspect de l'engagement. Les études peuvent exprimer leur point de vue sur l'engagement en choisissant une définition au lieu d'une autre; en fonction de l'objectif de l'étude une certain définition peut être plus approprié que d'autres. Par exemple, pour certaines études, il peut être intéressant de considérer l'engagement de tous les participants à la fois, tandis que pour d'autres études, il peut être utile de se concentrer sur le niveau d'engagement d'un seul participant à la fois. Vu qu'une définition peut indiquer l'objet d'une étude, et si l'étude indique clairement la définition de l'engagement qu'elle emploie, nous ne considérons pas que la coexistence de multiples définitions est problématique.

Dans cette thèse, nous basons tout notre travail sur l'engagement dans l'interaction humain-agent sur la définition de [Poggi, 2007]:

> "La valeur qu'un participant à une interaction attribue à l'objectif d'être ensemble avec l'autre participant(s) et de poursuivre l'interaction."

Cette définition est applicable à tous les types d'interactions, y compris l'interaction humain-humain et décrit l'engagement comme attribué à un participant individuel. Avec cette définition, nous considérons et mesurons l'engagement comme un état, et non pas comme par exemple un processus pour arriver à cet état.

Concepts liées à l'engagement

Afin de favoriser la participation des utilisateurs lors d'interactions humain-agent, nous devons connaître la façon dont l'engagement est manifesté dans une interaction. Par conséquent, dans le chapitre 3 nous donnons un aperçu des concepts et des comportements qui sont associés à l'engagement dans les travaux de recherche antérieurs.

Il existe de nombreux concepts qui sont souvent mentionnés en relation avec l'engagement et qui sont même parfois utilisés de manière interchangeable [Peters et al., 2009]. Nous avons discuté des concepts liés à l'engagement qui sont fréquemment mentionnés dans le contexte de l'interaction humain-agent dans le but de clarifier la façon dont les différents concepts sont liés les uns aux autres. Bien qu'il puisse exister des interprétations diverses pour chacune des ces concepts, nous considérons que *l'attention*, *l'involvement*, *l'intérêt* et *l'empathie* peuvent former des conditions et/ou des conséquences de l'engagement, et le *rapport* constitue une partie de l'engagement. Tous ces concepts sont positivement liés à l'engagement: En

favorisant une de ces concepts, il est probable que l'engagement d'un participant d'une interaction (un ou les deux participants) est également favorisé. Des *stances* particulières peuvent exprimer l'engagement.

Comportements liée à l'engagement

Pour un agent conversationnel engageant il y a trois tâches qui impliquent un comportement associé à l'engagement :

— Détection et interprétation du comportement de l'utilisateur enfin de déduire une perception du niveau d'engagement de l'utilisateur.

— Génération de réactions (engagées) appropriées, exprimant ainsi un niveau d'engagement de l'agent.

— Manifestation du comportement qui influe sur l'engagement de l'utilisateur, créant ainsi un agent engageant.

Dans cette thèse nous nous sommes principalement intéressés au troisième point, mais pour chacune de ces tâches nous devons être en mesure d'associer un comportement à l'engagement et vice-versa. Alors, dans la deuxième partie du chapitre 3 nous regardons des comportements liés à l'engagement dans les travaux antérieurs. Dans ce chapitre nous résumons les types de comportement de haut niveau et de bas niveau qui ont été associés à l'engagement dans l'interaction humain-agent dans des travaux précédents. Le comportement haut niveau peut être réalisé par différents types de comportement de bas niveau: le comportement de bas niveau se réfère à des actions comme des gestes, des mouvements, le regard ou la parole, alors que le comportement de haut niveau se réfère à la fonction communicative, l'émotion ou l'attitude.

Comme nous l'avons vu à partir des définitions d'engagement dans l'interaction humain-agent et les concepts qui sont étroitement liés à l'engagement, l'engagement est caractérisé par une relation et une coordination entre les participants d'une interaction. Par conséquent, il n'est pas surprenant que le comportement qui peut être associé à l'engagement et/ou de ses concepts connexes, est également caractérisé par l'expression d'une connexion et d'une coordination entre les participants de l'interaction.

Les formes de comportement de haut niveau liés à l'engagement dont nous discutons sont la synchronie, l'alignement, le mimétisme, le *feedback* et *backchannels*, la collaboration sur une tâche, et le temps de l'interaction.

Resumé Long

Les formes de comportement de haut niveau qui sont liés à l'engagement peuvent être exprimés par des différents comportements bas niveau (verbal et non verbal). Il y a aussi plusieurs études où un type de comportement de bas niveau est mentionné comme un indicateur de l'engagement direct.

Les comportements non verbaux dont nous parlons sont le regard, les mouvements de tête, les gestes, les postures, les expressions faciales, les caractéristiques temporelles et spatiales, et les réactions physiologiques. L'engagement est également exprimé par le comportement verbal, car cela peut transmettre de la connectivité [Sidner et al., 2005a]. Nous classons les différents aspects du comportement verbal liés à l'engagement selon trois aspects: la forme, le timing et le contenu du comportement verbal.

Avec la forme du comportement verbal nous nous référons à la façon dont quelque chose est verbalement exprimé. La forme d'un énoncé se réfère donc à de multiples aspects tels que les choix prosodiques, lexicaux, et syntaxiques. Le comportement de prise de parole et les *backchannels* verbales forment des exemples de comportements verbaux liés à l'engagement qui sont caractérisés par leur dimension temporelle (timing). Concernant le contenu, plusieurs aspects ont été associés à l'engagement dans les travaux antérieurs. Des exemples sont les *adjacency pairs*, *small-talk* et les expressions des émotions et des attitudes.

La liste des comportements potentiellement engageants est plus longue car potentiellement chaque type de comportement peut avoir une influence sur "la valeur qu'un participant à une interaction attribue à l'objectif d'être ensemble avec l'autre participant et continuer l'interaction " (définition [Poggi, 2007]). Dans cette thèse nous modélisons de telle stratégies de comportement d'agents qui ont de fortes chances de favoriser l'engagement de l'utilisateur. Nous nous concentrons sur les comportements verbaux de l'agent en considérant des stratégies qui réfèrent respectivement aux trois aspects du comportement verbal: la forme (chapitre 5), le timing (Chapitre 6) et le contenu (chapitres 7-11).

État de l'art

Nous visons à développer des stratégies de comportement verbal pour un agent virtuel conversationnel qui joue le rôle d'un visiteur dans un musée, engageant l'utilisateur lors d'une interaction sur les œuvres du musée. Dans le chapitre 4 nous discutons des travaux précédents qui sont liés à cet objectif. Nous abordons respectivement les agents conversationnels animés qui ont été conçus pour une ap-

plication dans un musée, et les agents conversationnels animés qui tentent d'engager leurs utilisateurs à l'aide des stratégies de comportement verbal.

Nous avons trouvé qu'il existe plusieurs agents conversationnels placés dans des musées avec différentes capacités d'interaction. Dans toutes ces applications l'agent joue le rôle d'une sorte de guide de musée, tandis que dans notre travail basé sur le projet A1: 1, l'agent joue le rôle d'un visiteur du musée.

Dans les travaux précédents, des agents conversationnels animés ont adapté leur comportement verbal dans une gamme de différentes façons à l'utilisateur afin d'engager l'utilisateur dans l'interaction. Il existe des comportements d'agent qui sont spécifiquement orientés vers le commencement ou la fin de l'interaction, en commençant et terminant ainsi l'engagement. Les systèmes peuvent, par exemple, décider quand [Sidner et al., 2005b], comment [Bohus and Horvitz, 2014] et avec qui [Bohus and Horvitz, 2009b] commencer et terminer une interaction. D'autres systèmes [Bickmore et al., 2013, Bickmore and Picard, 2005, Bickmore and Cassell, 2005, Sidner et al., 2013] utilisent des comportements verbaux qui peuvent favoriser l'engagement lors de l'interaction; stratégies à maintenir ou à réparer l'engagement. La plupart des systèmes ne considèrent pas un seul comportement subtil pour favoriser l'engagement de l'utilisateur; ils emploient soit une large gamme de comportements en même temps (par exemple [Bickmore et al., 2013, Bickmore and Picard, 2005]), ou un comportement qui s'altère pendant toute la durée de l'interaction en changeant (dans le temps) le sujet ou le but de l'interaction (par exemple [Bickmore and Cassell, 2005, Sidner et al., 2013]). Bien que ces approches puissent entraîner des agents plus engageant que des agents sans ces comportements, l'effet de comportements individuels peut ne pas être toujours clair.

Dans notre travail, nous nous concentrons sur le comportement de l'agent pour favoriser l'engagement de l'utilisateur lors de l'interaction en cours. Nous essayons de contribuer à la compréhension de l'engagement de l'utilisateur en recueillant des indications précises sur la façon dont les comportements verbaux de l'agent pourraient individuellement favoriser l'engagement de l'utilisateur.

DÉCOUVRIR DES STRATÉGIES POUR FAVORISER L'ENGAGEMENT

Engagement et Politesse

Afin de favoriser l'engagement de l'utilisateur autant que possible, dans le chapitre 5 nous essayons d'obtenir des indications sur la coordination optimale de la

politesse. Nous regardons dans ce chapitre (un aspect de) la forme de comportement verbal de l'agent (comment dire quelque chose). Dans le chapitre 6, nous examinons l'influence du timing du comportement verbal de l'agent (quand dire quelque chose) et dans le chapitre 7 le contenu du comportement verbal de l'agent (ce qu'il faut dire).

Afin d'obtenir plus de connaissance sur la coordination optimale de la politesse de l'agent avec l'utilisateur dans le chapitre 5 nous menons une étude perceptive. Dans cette étude nous examinons s'il y a un lien entre la perception du niveau d'engagement de l'interlocuteur, et les stratégies de politesse du locuteur. L'étude examine ce lien dans l'interaction humain-humain parce que nous souhaitons obtenir un comportement d'agent aussi humain que possible. Les résultats de cette étude nous donnent une indication sur la façon de modéliser le comportement de la politesse d'un agent virtuel comme celle d'un humain. En effet, les résultats peuvent nous indiquer que si un agent de type humain veut continuer l'interaction avec son utilisateur (ce qui maintient l'engagement des utilisateurs), il doit parler plus prudemment à quelqu'un qui est moins engagé qu'à quelqu'un qui est très engagé dans l'interaction courante.

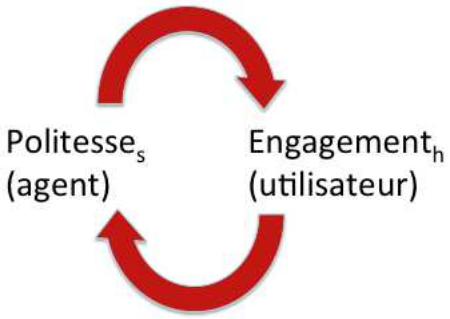


Figure 1 : L'engagement de l'interlocuteur (h, utilisateur) peut influencer la politesse du locuteur (s, l'agent) influençant l'engagement de l'utilisateur (interlocuteur).

Selon [Brown and Levinson, 1987] il y a des actes de dialogue qui intrinsèquement menacent la réputation ("face") de l'interlocuteur, appelé des FTAs. W_x, la valeur numérique qui mesure la pesanteur, à savoir le danger, du FTA x est calculé par [Brown and Levinson, 1987]:

$$W_x = D(S,H) + P(H,S) + R_x$$

où $D(S,H)$ est la distance sociale entre le locuteur et l'interlocuteur, $P(H,S)$ le pouvoir que l'interlocuteur a sur le locuteur, et Rx le degré dans lequel le FTA x

est évalué comme une imposition dans cette culture. Les variables de distance et de pouvoir sont conçues comme des dimensions sociales pan-culturelles très générales [Brown and Levinson, 1987].

Nous émettons l'hypothèse que, tout comme la distance sociale $D(S, H)$, la distance temporelle et dynamique qui est décrite par l'engagement joue également un rôle dans le poids perçu d'un acte menaçant (FTA). Cette hypothèse peut être formulée en ajoutant la variable d'engagement à l'équation de [Brown and Levinson, 1987]:

$$W_x = D(S, H) + P(H, S) + Rx - Eng_h$$

où Eng_h est le niveau de l'auditeur d'engagement.

Depuis la théorie [Brown and Levinson, 1987] on peut déduire qu'on peut estimer le poids menaçant perçue d'un FTA en regardant la stratégie de politesse qui est employée pour placer le FTA. Ainsi, afin de savoir si la menace perçue d'un FTA est en effet dépendante du niveau d'engagement perçu de l'interlocuteur, nous examinons les stratégies qui sont utilisées pour placer la même FTA dans des circonstances où seul le niveau d'engagement perçu de l'interlocuteur diffère. Pour cela, nous créons deux fragments de dialogue de deux humains en interaction (conditions) entre lesquelles les stratégies de politesse utilisées peuvent être comparés; un fragment de dialogue dans lequel un participant apparaît (très) engagé (exprimant le désir de continuer l'interaction et être ensemble avec l'autre participant d'interaction) et un autre dans lequel il apparaît moins engagé (exprimant beaucoup moins envie de continuer l'interaction et être ensemble avec l'autre).

Pour assurer que l'interlocuteur du FTA démontre les niveaux d'engagement souhaités mais toutes les autres variables de l'interaction ont été maintenues aussi constantes que possible, nous scriptons des interactions écrites. Les interactions écrites sont présentées aux observateurs humains. Ils jugent dans les contextes d'interaction quel énoncé (réalisant une stratégie de politesse) ils trouvent la plus appropriée pour le locuteur afin de réaliser le FTA. Nous vérifions cela pour 3 FTAs:

— *Désaccord* sur la préférence pour une peinture.

— *Suggestion* pour aller voir une autre œuvre d'art.

— *Demande* pour des conseils sur quelle autre œuvre il faudra aller voir.

De cette façon, nous avons vérifions s'il y a une différence entre le poids des stratégies de politesse que les locuteurs utiliseraient en interaction avec une personne engagée et en interaction avec une personne moins engagée. Cela nous a dit s'il y a

un lien entre la perception du locuteur du niveau d'engagement de l'interlocuteur, et les stratégies de politesse de locuteur, dans l'interaction humain-humain.

Les résultats nous montrent que dans la création de stimuli représentant les deux conditions, nous avons démontré un modèle de comportement verbal réussi pour transmettre le niveau d'engagement d'un participant. Nous n'avons pas trouvé de différence globale significative entre la recommandation de stratégies de politesse sur les deux conditions. La politesse reste un phénomène très subjectif avec de grandes différences individuelles.

Nous avons cependant trouvé que dans le contexte du FTA *demande*, les participants qui recommandent des stratégies de politesse pesantes (respectivement non pesantes) ont tendance à percevoir un niveau d'engagement inférieur (respectivement supérieur) de la part de l'interlocuteur. Dans ces contextes (pour certains FTAs) un agent de type humain qui veut poursuivre l'interaction avec son utilisateur, et maintenir ainsi l'engagement de l'utilisateur, a besoin de parler plus poliment à quelqu'un qui est moins engagé que à quelqu'un qui est très engagé dans l'interaction.

Engagement et Interruptions

Dans le chapitre 6 nous examinons la coordination optimale de la prise de tour de parole de l'agent dans le but de faire apparaître l'agent engagé tel que le serait un humain, afin de rendre l'agent plus attrayant pour l'utilisateur. Nous considérons une forme de prise de parole en particulier, les interruptions.

L'organisation de la prise de parole est fondamentale dans l'interaction humain-humain [Schegloff, 2007] et est donc également importante pour rendre les agents conversationnels socialement crédibles dans les interaction humain-agent et agent-agent [Crook et al., 2010]. Les interruptions représentent une violation des règles de prise de parole de base et une manière possible de réclamer le tour [Beattie, 1981]. Une interruption est "un démarrage d'une intervention par une seule personne alors qu'un autre tour est en cours" [Schegloff, 2001].

Le comportement lié à la prise des tours de paroles dans une interaction est lié à l'engagement. On pourra penser par exemple au fait que ne pas prendre un tour de parole attendu peut signaler désengagement [Sidner et al., 2003]. La relation exacte entre les interruptions et l'engagement n'est cependant pas tout à fait claire. À notre connaissance, aucun travail précédent n'a encore proposé une analyse systématique de la perception des interruptions qui tient compte à la fois de la quantité de paroles qui se chevauchent et le contenu des interruptions, afin de comprendre les effets des

interruptions plus en détail. Dans le chapitre 6 nous présentons donc une telle analyse systématique des interactions agent-agent, et ce dans le but de clarifier les effets des interruptions sur la perception de l'engagement des agents et des attitudes interpersonnelles.

Nous considérons des interruptions avec différentes longueurs de chevauchement, que nous appelons les *types* d'interruption:

— *Interruption silencieuse*: où le locuteur interrompu arrête son énoncé dès qu'il est interrompu.

— *Interruption simple*: où le locuteur interrompu poursuit son énoncé pendant un certain temps quand il est interrompu, mais ne termine pas son énonciation.

— *Superposition*: où le locuteur interrompu poursuit son énoncé quand il est interrompu jusqu'à ce que son tour de parole soit complété.

Et nous considérons 2 *stratégies* d'interruptions, qui sont caractérisés par leur contenu:

— "*Interruptions coopératives* sont destinées à aider le locuteur en coordonnant le processus et/ou le contenu de la conversation en cours." [Li, 2001]

— "*Interruptions intrusives* constituent une menace pour le territoire du locuteur à cause d'interrompre le processus et/ou le contenu de la conversation en cours." [Li, 2001, Murata, 1994, Goldberg, 1990].

Sur la base de la littérature et une étude de validation, nous créons une série d'interruptions qui incorpore les différents types d'interruptions (quantités de chevauchement) et les stratégies (contenu) d'interruption.

Pour évaluer les effets des différents types d'interruptions et les stratégies d'interruptions sur la perception de l'humain par rapport aux agents interrupteurs et interrompus en ce qui concerne l'engagement (engagement et implication) et l'attitude interpersonnelle (dominance et amicalité), nous concevons une expérience empirique. Au moyen d'un sondage en ligne, nous exposons les participants humains à une série de fragments de vidéo de courte durée où deux agents interagissent. Les types et les stratégies d'interruption varient systématiquement dans une sous-catégorie particulière où toutes les interruptions ont la même fonction communicative. Nous demandons ensuite aux participants humains d'évaluer chaque fragment sur le niveau perçu de l'engagement, l'implication, la domination et l'amicalité des deux agents.

Comme prévu, pour la perception de l'engagement et l'implication de l'interrupteur, c'est la stratégie d'interruption qui conduit à des effets significat-

ifs : stratégies coopératives sont signe de plus grands engagement et implication que les stratégies disruptives. Les interruptions sont sans effet sur la perception de l'engagement et pratiquement sans effet sur le niveau de l'implication de l'agent interrompu.

En ce qui concerne la perception qu'ont les humains sur les attitudes interpersonnelles des agents, c'est d'abord le type d'interruption qui importe le plus : plus il y a de de chevauchement, plus l'agent sera perçu comme dominant et hostile. La stratégie d'interruption peut également influencer la perception des attitudes relationnelles des agents même si cela dépend du contenu des interruptions individuelles. Lorsque le type d'interruption n'a pas de chevauchement, l'effet de la stratégie d'interruption peut parfois augmenter.

Engagement et Préférences

Nous avons d'abord examiné dans le chapitre 5 les aspects du comportement de la politesse des agents, ce qui fait référence à la forme du comportement verbal de l'agent (comment dire quelque chose), puis nous avons considéré dans le chapitre 6 des interruptions qui se réfèrent principalement à la synchronisation du comportement verbal de l'agent (quand dire quelque chose), le chapitre 7 est quant à lui principalement orientée vers le contenu du comportement verbal de l'agent (ce qu'il faut dire). Concrètement, dans le chapitre 7 nous vérifions si et comment une adaptation des sujets dans les interaction humain-agent non-orientée-tâche pourrait favoriser l'engagement de l'utilisateur. Nous explorons si dans l'interaction humain-agent non-orientée-tâche, la personnalisation d'un agent selon les préférences de l'utilisateur peut influer sur l'engagement de l'utilisateur. Plus précisément, nous vérifions si la préférence de l'utilisateur pour un objet physique (œuvre d'art) joue un rôle dans le niveau d'engagement de l'utilisateur lors de la discussion de cet œuvre avec un agent virtuel. Nous interprétons une préférence en fonction de la définition de Scherer [Scherer, 2005] comme "un jugement évaluatif relativement stable dans le sens d'aimer ou non un stimulus".

Afin de savoir si la préférence de l'utilisateur pour un objet dans un musée (œuvre d'art) joue un rôle dans le niveau d'engagement de l'utilisateur lorsqu'il en discute avec un agent virtuel, nous effectuons une étude perceptive. Dans cette étude, nous recueillons des données sur les utilisateurs concernant leurs préférences pour des œuvres d'art. Cela afin de les comparer avec les estimations par les utilisateurs de leur propre engagement au cours des différentes phases de discussion de leur interaction

avec un agent virtuel (car les préférences et l'engagement font partie de l'état mental de l'utilisateur). L'étude consiste, pour chaque participant de l'évaluation, d'une visite d'un petit musée improvisé, une conversation avec un agent virtuel appelé Leonard, et d'un questionnaire.

Figure 2 : Agent Virtuel Leonard @Cantoche

Les résultats de l'étude montrent que le degré d'appréciation (préférence) et l'intérêt de l'utilisateur pour une œuvre d'art du musée est en effet corrélé, de manière significative, avec l'engagement et l'intérêt de l'utilisateur lors de la discussion de cette œuvre avec un agent virtuel. Nous pouvons ainsi déduire que l'une des stratégies de dialogue de l'agent lorsqu'il donne des informations, probable de favoriser plus d'engagement chez les utilisateurs lors de l'interaction, est l'adaptation des sujets aux préférences de l'utilisateur pour les différents objets physiques.

Réciproquement, les corrélations indiquent également que la détection d'un niveau d'engagement de l'utilisateur lors de l'interaction humain-agent (en supposant que l'utilisateur communique son niveau d'engagement par son comportement), permet d'obtenir des indications par rapport à la préférence de l'utilisateur sur le sujet en cours et son objet physique sous-jacent. Cela permettrait à un système d'estimer les préférences d'un utilisateur grâce à son engagement, sans avoir besoin de les demander explicitement.

Ces deux conclusions nous ont conduit à développer un modèle d'agent qui tente de renforcer l'engagement de l'utilisateur en personnalisant le contenu (sujets) de l'interaction humain-agent. Nous décrivons ce modèle dans la partie V de cette thèse.

MODÉLISATION DES STRATÉGIES POUR FAVORISER L'ENGAGEMENT

Modèle de Sélection de Sujet

Dans la partie IV, nous avons exploré comment des aspects du comportement verbal peuvent influer ou refléter l'engagement ("la valeur qu'un participant d'une interaction attribue à l'objectif d'être ensemble avec l'autre participant (s) et de poursuivre l'interaction" [Poggi, 2007]). Nous avons examiné les stratégies liées à la politesse, les interruptions et les préférences de l'utilisateur. Dans le chapitre 8, en nous basant sur les dernières observations décrites dans le chapitre 7, nous renforçons le modèle avec la possibilité pour l'agent d'initier des sujets de discussion adaptés aux préférences de l'utilisateur et/ou aux préférences propres à l'agent.

Les agents conversationnels emploient souvent une structure de dialogue "orienté-tâche" où la réalisation de sous-tâches précises permet de compléter la tâche principale, le but, pour laquelle ils ont été développés. D'autre part, les systèmes type *chat* permettent une interaction moins rigide mais l'agent a moins de contrôle sur le thème de l'interaction. Nous nous sommes intéressés par un système dialogue à mi-chemin entre ces catégories : il n'y a pas une tâche claire à atteindre et l'interaction n'est pas complètement ouverte non plus, mais il y a un panel de choix dans le sujet sur un domaine défini. L'interaction n'est donc pas "orientée-tâche", mais "orientée-engagement". Pour ce type d'interaction nous présentons dans ce chapitre un modèle de sélection de sujet pour un agent conversationnel qui essaye de favoriser l'engagement. Le modèle donne à l'agent la possibilité de sélectionner des sujets de dialogue qui sont adaptés à l'utilisateur.

En prenant en compte 1) l'objectif de vouloir favoriser l'engagement de l'utilisateur, 2) la perception de l'agent de l'engagement de l'utilisateur, et 3) l'état mental propre de l'agent, y compris les préférences et les associations de l'agent, le dialogue n'est plus simplement orienté que pour l'utilisateur, mais prend en compte également l'agent en tant que participant de l'interaction avec ses avis propres. Ce composant de sélection de sujet (partie du modèle de gestionnaire de sujet) décide quel sujet choisir à chaque fois que l'agent a besoin d'en introduire un nouveau.

Alors que le modèle de sélection des sujets présenté dans le chapitre 8 peut être appliqué à plusieurs domaines, nous l'illustrons avec le domaine pour lequel il a été développé à l'origine: un agent qui essaie d'engager les utilisateurs humains dans une interaction dyadique face-à-face dans une musée et sur certaines œuvres afin

d'en informer les visiteurs (projet A1:1).

Préférences et Engagement de l'Utilisateur : Le modèle considère plusieurs variables. Tout d'abord, afin de sélectionner des sujets de discussion intéressants, l'agent doit être en mesure de prédire le niveau d'engagement de l'utilisateur sur des sujets jusqu'ici non adressés (objets). L'étude d'évaluation dans le chapitre 7 a confirmé que la préférence de l'utilisateur pour une œuvre d'art (objet de musée) a une corrélation significative avec l'engagement de l'utilisateur lors de la discussion de cet objet avec un agent virtuel. De cette constatation, nous pouvons déduire que les caractéristiques de cet objet physique permettent à l'agent de faire des prédictions des préférences de l'utilisateur pour les autres objets physiques (œuvres d'art) qui n'ont pas encore été évoqués. L'estimation de ces préférences peuvent donner une indication de l'engagement futur de l'utilisateur lors de la discussion de ces objets. Ainsi, la perception de l'agent de l'engagement de l'utilisateur lors de la discussion d'un objet, via ses caractéristiques, permettent à l'agent de faire des prédictions de l'engagement futur de l'utilisateur pour d'autres objets non discutés. L'agent peut alors utiliser cette prédiction pour sélectionner un sujet d'engagement de conversation.

Préférences et Engagement de l'Agent : Pour simuler des caractéristiques humaines à un agent qui joue un visiteur d'un musée, l'agent doit avoir ses propres préférences pour les œuvres d'art. Représenter les préférences de l'agent est fondamental pour tout modèle d'agent [Casali et al., 2011]. Après la corrélation que nous avons trouvé dans le chapitre 7 un agent préfère parler de ses sujets préférés car ils favorisent son engagement. Cependant, l'agent que nous modélisons veut aussi engager l'utilisateur. L'agent tente ainsi d'optimiser l'engagement de l'utilisateur et le sien (ce que nous appellerons maintenant l'engagement combiné). Pour ce faire, pour chaque (sous-)sujet (objet et caractéristique) qui peut être adressé, l'agent calcule un niveau (estimé) d'engagement combiné prévu et sélectionne le (sous-)sujet avec le score le plus élevé (optimisation) en suivant le (sous-)sujet de l'interaction. De cette façon, l'agent sélectionne un nouveau (sous-)sujet de conversation basée sur une combinaison de ses propres préférences pour les œuvres d'art et de sa prévision du niveau d'engagement de l'utilisateur lors de la discussion sur les œuvres d'art.

Associations de l'Agent : Une dernière variable humaine qui doit être représentée lorsque l'agent sélectionne un sujet de conversation sont ses propres associations entre sujets. Nous intégrons ce concept dans le composant de sélection de sujet (du gestionnaire de sujet) car, dans l'esprit humain, des événements qui partagent sens ou de la similitude physique sont associés [Dellarosa, 1988]. Pour

imiter ce processus humain, nous représentons les associations de l'agent par des scores de similarité entre chaque paire d'objets physiques correspondant à deux sujets. De cette façon, l'agent a, par exemple, une forte association entre deux sujets portant sur la peinture abstraite. A contrario, son association entre la discussion d'une peinture abstraite et la discussion d'une statue réaliste est beaucoup plus faible, ce qui rend alors l'évocation du deuxième sujet déclenché par la première moins probable.

Les associations (en fonction des similitudes d'objet) permettent à l'agent de faire des prédictions sur l'engagement de l'utilisateur pour des sujets jusqu'ici non adressés. Lorsque l'utilisateur a un certain niveau d'engagement au cours de la discussion sur un sujet (et donc une préférence liée à l'objet de la discussion), les sujets de conversation proches devraient avoir des niveaux similaires de préférence de l'utilisateur et donc conduire à des niveaux d'engagement équivalent des utilisateurs. Le modèle de sélection des sujets garantit que lorsque l'agent prédit un niveau d'engagement de l'utilisateur pour un sujet associé suffisamment élevé (en combinaison avec ses propres préférences de l'agent), ce sujet est un nouveau sujet potentiel de la conversation, déclenchée par les associations de l'agent.

Le modèle de sélection de sujet essaie de sélectionner pour l'interaction les (sous-)sujets qui optimisent l'engagement combiné (engagement de l'agent et l'utilisateur). Pour cela, le modèle calcule en permanence pour chaque potentiel (sous-)sujet, l'engagement attendu (prévu) de l'agent et de l'utilisateur ensemble (engagement combiné) au cours d'une discussion potentielle des (sous-)sujets. Pour calculer ces niveaux d'engagement combinés prévus pour chaque sujet et sous-sujet potentiel, dans l'esprit de [Stede and Schlangen, 2004], nous définissons un modèle représentant les savoirs conceptuel (base de connaissances) et l'historique du dialogue. Le modèle fait partie de la connaissance de l'agent et est enrichi dynamiquement avec des informations sur l'interaction. L'engagement détecté de l'utilisateur sert à décider quand changer un sujet et à mettre à jour les variables qui jouent un rôle dans la sélection des sujets.

Le modèle permet d'éviter d'avoir à fournir des informations à priori sur l'utilisateur et permet l'adaptation de l'interaction à tout utilisateur. Les configurations des préférences de l'agent, des associations de l'agent (scores de similarité), et de l'orientation de l'agent envers soi et/ou envers l'utilisateur permettent la modélisations de différents types d'agents.

L'initialisation de la base de connaissances de l'agent peut être simplifié à l'aide d'un catalogue de musée qui liste les objets et leurs caractéristiques. Le modèle

de sélection de sujets peut être étendu facilement à d'autres domaines si ces sujets peuvent être structurés de la même manière que des objets de musée.

Modèle de Transition de Sujet

Afin d'arriver à un discours d'agent semblable à celui d'un humain les (sous-)sujets choisis par le modèle de sélection de sujet afin d'engager l'utilisateur (partie du gestionnaire de sujet, Chapitre 8) doivent être initiés par l'agent d'une façon naturelle et cohérente dans l'interaction en cours. Pour les sous-sujets cela est simple; tous les sous-sujets au sein d'un sujet traitent évidemment un aspect de la même œuvre. Ils font tous partie de l'ensemble de tours qui constituent un sujet clairement délimité.

Sur un plan du sujet cependant, il est moins facile d'initier un sujet d'une manière cohérente car les sujets sont clairement délimités. Chacun des sujets parle d'une œuvre différente, avec laquelle le (l'œuvre du) sujet précédent peut avoir peu en commun avec le (l'œuvre du) sujet suivant. Ceci s'explique par le fait que les sujets sont principalement choisis en fonction de leur potentiel d'engager l'utilisateur au lieu de leur cohérence par rapport au sujet précédent. Ce que les différents sujets tentent de réaliser ensemble peut ne pas être évident, réalisant une simple concaténation de sujets peut être insuffisamment cohérente.

Afin de résoudre une éventuelle incohérence des sujets suivants, et ainsi rendre le discours de l'agent plus semblable à l'humain dans le chapitre 9 examinons les énoncés qui peuvent être utilisés pour initier une transition d'un sujet à l'autre et les stratégies de transition qui sont réalisées avec ces énoncés. Nous y décrivons une étude évaluative vers des stratégies de transition afin de tenter de répondre à deux questions: 1) Quelles stratégies ont le potentiel de maintenir/rendre le dialogue cohérent? Et 2) Quels sont les effets de l'utilisation des différentes stratégies de transition sur la perception de l'agent conversationnel? Les réponses à ces questions servent alors à générer automatiquement des stratégies d'agent pour connecter un sujet à un autre dans un dialogue non-orienté tâche. Dans le contexte que nous examinons, chaque sujet correspond à la discussion d'une œuvre d'un musée.

Basé sur la littérature des sciences sociales et de l'analytique conversationnelle nous construisons d'abord un ensemble de stratégies de transition potentielles.

Afin de vérifier si chacune de ces stratégies peut être générée par l'agent pour introduire un sujet dans une interaction non-orientée tâche, nous effectuons une étude empirique. Au moyen d'un questionnaire en ligne, nous testons comment et si les

différentes stratégies de transition influent sur la perception humaine du dialogue et de l'agent. Nous montrons à chaque participant à l'étude en ligne 2 fragments de dialogue, composés par des énoncés d'agent et des entrées d'utilisateur simulées. Un tel fragment forme le contexte des stratégies de transition. Chaque scénario (fragment) est suivi de 3 énoncés de transition répartis de façon aléatoire (les réalisations des stratégies de transition), affichées les unes à côté des autres. Nous ne montrons pas les énoncés qui peuvent suivre les stratégies de transition. De cette façon, nous ne montrons pas l'acceptation ou le rejet du sujet par l'utilisateur ([Clark, 1996, Svennevig, 2000]). Directement après chacune des 3 stratégies de transition, une liste de questions demande aux participants leur perception du dialogue et de l'agent.

Nous constatons que les stratégies qui obtiennent de bons résultats sur toutes les dimensions et toutes les circonstances testées sont celles qui font appel à l'expérience de l'utilisateur (si l'utilisateur *a vu* une œuvre), et celles qui se réfèrent aux préférences des participants de l'interaction. Ces stratégies maintiennent la cohérence du dialogue tout en maintenant/établissant une perception positive de l'agent, même si les sujets précédents et suivants n'ont rien en commun mis à part le fait qu'ils concernent tous les deux des œuvres d'art. Cela montre que la composante de sélection du sujet du gestionnaire de sujet de l'agent peut en effet être autorisé à sélectionner n'importe quel sujet nécessaire pour favoriser l'engagement à tout moment dans la conversation.

Nous utilisons les résultats de l'étude empirique pour modéliser un composant du gestionnaire de sujet qui sélectionne automatiquement les stratégies de transition appropriées pour l'agent conversationnel à chaque fois que le modèle de sélection de sujet initie une changement de sujet. Ce modèle de transition de sujet génère des énoncés de transition en tenant en compte de

— La relation entre le sujet précédent et suivant (caractéristiques d'objet commun ou pas);

— L'orientation de l'agent (par rapport à l'agent utilisateur);

— Les variables dans le processus de raisonnement de l'agent pour sélectionner un sujet;

— L'historique de stratégies de transitions dans le dialogue;

— L'importances des caractéristiques communes entre le sujet précédent et le suivant;

— L'historique de sous-sujets pour le sujet précédent.

L'Implémentation du Gestionnaire du Sujet

Dans le chapitre 10, nous proposons une implémentation du gestionnaire de sujet dont les modèles ont été décrits dans les chapitres 8 (sélection des sujets) et 9 (transitions de sujet). Plus précisément, nous décrivons l'intégration d'un module de gestionnaire de sujet dans la plateforme VIB et une architecture qui permet au gestionnaire de sujet de l'agent de gérer le timing, la sélection, l'initiation et l'abandon des sujets dans l'interaction humain-agent.

Dans la plateforme VIB le module de gestionnaire de dialogue gère le comportement conversationnel de l'agent par le réseau de tâches hiérarchiques Disco [Rich and Sidner, 2012]. Disco offre une approche strictement orientée-tâche, où les tâches de l'interaction sont exécutées dans un ordre prédéfini. Alors pour faire en sorte que l'agent soit capable de gérer des interactions dont les sujets sont adaptés en continu à l'interaction en cours, nous avons intégré un module de gestionnaire de sujet qui est relié au module de gestionnaire de dialogue. Le module de gestion de sujet ajoute de la flexibilité à l'ordre fixe d'exécution de tâches en exploitant et en modifiant en continu l'arbre de tâches qui est exécuté par Disco lors de l'interaction.

De cette façon, nous continuons à utiliser un ordre prédéfini d'exécution de la tâche lorsque cela est possible (dans les sous-sujets), mais ajoutons de la flexibilité lorsque c'est nécessaire (parmi les (sous-)sujets). Nous montrons comment le gestionnaire de sujet effectue des modifications à l'arbre de dialogue à chaque fois que 1) la structure de l'interaction l'exige, et 2) à chaque fois que l'engagement de l'utilisateur nécessite un changement de sujet. De cette façon, le gestionnaire de sujet gère les sujets de l'interaction en temps réel, ce qui réalise une interaction flexible adaptée à l'utilisateur.

L'Evaluation du Gestionnaire du Sujet

Dans les chapitre 8, 9 et 10 nous avons développé et implémenté un gestionnaire de sujet pour un agent conversationnel qui tente d'engager l'utilisateur tout en lui fournissant des informations. Le gestionnaire de sujet est conçu de manière à adapter les sujets de l'interaction en temps réel en décidant quels (sous-)sujets il faut introduire, et quand et comment introduire les sujets dans l'interaction en cours.

Dans le chapitre 11, nous évaluons le gestionnaire de sujet dans une application où un agent virtuel joue un visiteur dans un musée qui tente de donner un maximum d'information culturelle à l'utilisateur. Dans le chapitre 7, nous avons vu

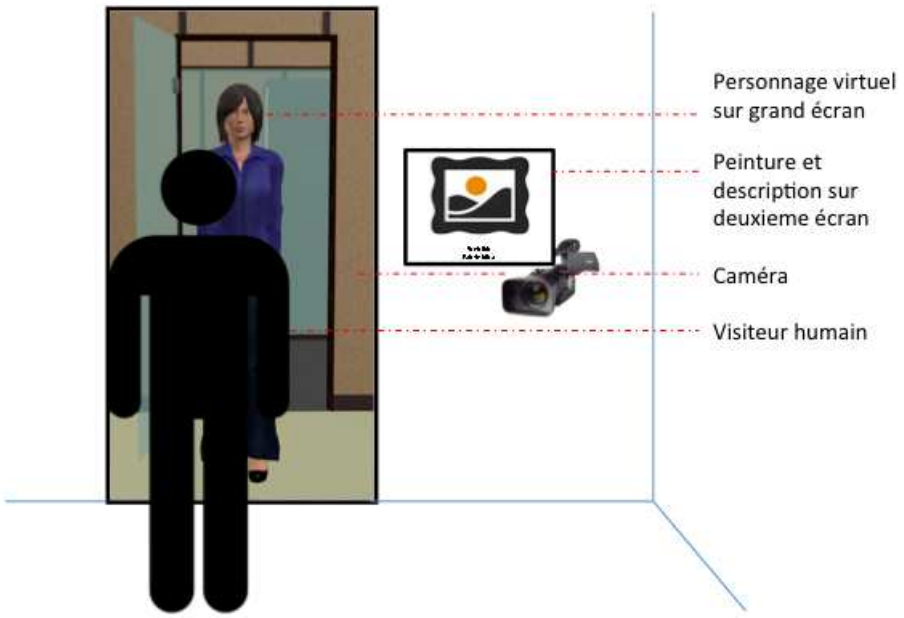


Figure 3 : Mise en place des interactions humain-agent

que l'engagement de l'utilisateur est favorisé si un agent parle d'œuvres d'art que l'utilisateur préfère. Dans l'évaluation du chapitre 11, nous cherchons donc à savoir si les actions du gestionnaire de sujets sont bien perçues, à savoir l'adaptation des sujets (à différents degrés) à l'engagement (et donc préférences) de l'utilisateur et les propres préférences de l'agent en associant les sujets d'interaction. Deuxièmement, nous vérifions si le gestionnaire de sujet a une influence sur la perception humaine de l'agent et de son dialogue.

Afin d'évaluer les effets du gestionnaire de sujet nous considérons plusieurs interactions humain-agent où dans chaque interaction l'agent gère les sujets de l'interaction différemment. Pour évaluer les effets du gestionnaire de sujets et contrôler les autres variables dans les interactions nous assemblons des fragments vidéo des contributions de l'agent virtuel calculés et des contributions de l'utilisateur humain actés, en créant des vidéos d'interactions humain-agent présumées. Ces vidéos forment les stimuli de l'étude d'évaluation. Les stimuli sont présentés à des observateurs humains auxquels nous demandons leur avis sur ces interactions. L'agent virtuel et l'utilisateur humain sont modélisés comme des visiteurs d'un musée. Le contexte des interactions est montré dans Figure 3.

Afin d'évaluer les effets du gestionnaire de sujet et de ses configurations, nous considérons 4 conditions différentes (interactions humain-agents). Dans chaque con-

dition l'agent a une façon différente de gérer les sujets. Dans 3 conditions l'agent gère les sujets avec différentes configurations du gestionnaire de sujet, et 1 autre condition forme la condition de contrôle sans le gestionnaire de sujet. Le tableau 1 montre les différents modes de gestion des sujets pour chaque condition.

Nom	**Condition**	**Selection (Sous-)Sujet** (Quoi)	**Abandonner des sujets** (Quand)	**Strategies de Transition** (Comment)
`TMw=0`	Avec gestionnaire, $w = 0$	Orienté utilisateur	Si l'engagement de l'utilisateur est trop bas	Validés, Orienté utilisateur
`TMw=0.5`	Avec gestionnaire, $w = 0.5$	Orienté utilisateur-agent	Si l'engagement de l'utilisateur est trop bas	Validés, Orienté utilisateur-agent
`TMw=1`	Avec gestionnaire, $w = 1$	Orienté agent	Si l'engagement de l'utilisateur est trop bas	Validés, Orienté agent
`WithoutTM`	Sans gestionnaire	Selection des sujets aléatoire, Sous-sujets fixés	Non	Non validés, Orienté objet

Table 1 : Les conditions dans l'évaluation du Gestionnaire du Sujet.

Les résultats de l'évaluation montrent que les actions du gestionnaire de sujet sont en effet perçus dans les interactions : le gestionnaire de sujet fait en sorte que l'agent soit perçu comme associant les sujets les uns avec les autres, ne sélectionnant pas des sujets au hasard, et adaptant les sujets de l'interaction à l'engagement et aux préférences de l'utilisateur. Différentes orientations du gestionnaire de sujet (orientée vers l'agent et/ou l'utilisateur) permettent d'assurer des différences dans la manifestation des préférences de l'agent et de l'adaptation des sujets aux préférences de l'utilisateur.

Le gestionnaire de sujet et ses configurations n'influencent pas de manière significative la perception par des observateurs humains du dialogue de l'agent et la perception de l'agent lui-même en ce qui concerne des aspects tels que la chaleur et la compétence. Pour ces dimensions, des adaptations à d'autres niveaux que

(seulement) au niveau de sujet sont susceptibles d'avoir plus d'impact. Par exemple, des adaptations au niveau de l'énoncé, ou en ce qui concerne des formes de comportement non-verbal.

PART I : INTRODUCTION

TELECOM ParisTech

1 Introduction

An essential aspect of human-agent interaction is the extent to which the human is engaged. *Engagement* of the user is a social variable that refers to the connection of the human user with the agent and the interaction. A minimum level of user engagement is necessary in any human-agent interaction to achieve the goal for which the agent is built. Due to the growing number of conversational agents and the important role that engagement plays in human-agent interaction, over the last years engagement has gained increasing attention.

The research that is described in this thesis aims at contributing to a better understanding of engagement by exploring strategies to enhance the user's engagement during face-to-face interaction with a conversational agent. This research is conducted in the context of project A1:1 that aims for a human-sized conversational virtual agent in a museum that engages the user during conversation about the museum and its artworks. This setting forms the application of this thesis.

1.1 Focus and Domains

As we will point out in the first chapters of this thesis, engagement can be expressed and favoured by both verbal and non-verbal behaviour. We focus our research on the agent's verbal behaviour. This is because previous research on engagement focuses often on non-verbal behaviour while with respect to verbal behaviour, especially in non-task oriented contexts (see next section), still much is left unclear. According to [Bickmore and Cassell, 2001] "knowledge of when and how to use language to achieve social goals is crucial for our computational agents if they are to be as effective as people, and if we want people to be able to use our agents easily, efficiently, and cooperatively".

We try to contribute to the knowledge of engagement in human-agent interaction by considering potentially engaging verbal agent strategies. Since the verbal behaviour of an agent includes numerous aspects we limit the number of strategies

that we consider to one that considers the *form* of the agent's utterances, one that relates mostly to the *timing* of its utterances, and one that looks at the *content* of its utterances. Concretely, we look respectively at an aspect of politeness, interrupting utterances, and the topics of non-task oriented interaction. While these aspects of verbal behaviour have all been studied previously, we approach them from a different angle by considering if and how they impact and/or are dependent on the engagement of the interaction participants.

The research in this thesis is based on various domains of research, among which embodied agents, artificial intelligence, computational linguistics, conversational analysis, and cognitive science. Where these domains are often considered in isolation, for this thesis we try to bring them together.

For example, in the field of computational linguistics, dialogue systems are often developed as independent systems that manage speech, without considering other cognitive aspects of the interaction participants. On the other hand, in the field of human-agent interaction, cognitive systems that direct for example emotions and non-verbal behaviour are considered in detail, but (task-based) dialogue systems are often employed as merely a tool to endow the agent with speech without giving it further attention. We believe however, that if we want an agent to dispose of true human-like behaviour, we need to treat dialogue systems and cognitive systems as intertwined concepts where the one cannot be considered without the other. In this thesis we try to contribute to such an interdisciplinary view.

1.2 Application

As mentioned before, the application of this thesis is formed by project A1:1 that aims at developing a virtual agent that interacts with human visitors in a museum. The agent has the appearance of a human-size cartoon-like man and interacts face-to-face with human visitors by means of a set of human-like postures, gestures, and facial expressions, as well as spoken natural language. The user interacts with the agent by means of natural language speech while his/her non-verbal behaviour is detected as well. Since the user should be able to interact with the agent in such a natural human-like way (using human interaction modalities) we aim to realise human-agent interaction that is as close to face-to-face human-human interaction as possible.

The agent's task is to engage human users during one-to-one conversation about the museum and its art objects in order to transfer a maximum amount of cultural

information. An important difference with some other virtual museum agents is that the agent is not playing the role of a museum guide, but that of a museum visitor. This makes that the interaction is not strictly task-oriented. The agent is not created, for example, to answer questions about the museum or to show the user around, but is just like the user a visitor of the museum itself. Because the agent is a very frequent visitor it does have lots of knowledge about the artworks, which it can share with the user. Instead of an interaction that is task-driven, the interaction is thereby driven by a social variable of the interaction, namely the engagement of the user.

1.3 Contents and Contributions

This document is divided in 7 parts. In the following part, Part II, we give insights into what engagement exactly is. In the first Chapter (2) of this part we offer an overview and analysis of the definitions of engagement where we identify their implications, commonalities, and differences. In Chapter 3 we give an account of how engagement is manifested in an interaction by giving an overview of the concepts and behaviours that have been associated with engagement in previous research.

To position our work, in Part III (Chapter 4) we address previous efforts that are related to ours, by discussing embodied conversational agents (ECA's) in museum contexts, and ECA's that consider verbal engagement strategies.

In Part IV then, we present three studies to explore potentially engagement-favouring verbal behaviours. We look respectively at aspects of verbal behaviour that consider the *form*, *timing*, and *content* of utterances: 1) In Chapter 5 we present an empirical study to verify the existence of a link between the speaker's politeness and his/her perceived engagement level of the hearer. 2) Chapter 6 presents a systematic empirical analysis of interruptions that considers both the amount of overlapping speech and the content of the interruptions in order to reveal their effects on the perception of the interrupter and the interruptee. 3) And by means of a perceptive study, in Chapter 7 we show the existence of a positive correlation between the user's preferences for an artwork (a physical object) and the user's engagement during the discussion of this object with an ECA.

Based on the outcomes of Part IV, in Part V we propose an engagement-driven topic manager that personalises the topic of conversation in human-agent information-giving chat. The topic manager decides *what* to talk about, *when*, and

how to introduce a new topic in the ongoing interaction: Chapter 8 of Part V presents the topic selection component of the topic manager that takes into account the agent's dynamically updated perception of the user and its own mental state. In Chapter 9, we have checked with an empirical study how the selected topics can be introduced on the dialogue level without loosing the coherence of the interaction, leading to the computational topic transition model for the topic manager. Chapter 10 proposes an implementation of the topic manager. The architecture allows the agent to manage on the fly the *timing*, *selection*, *initiation* and *abandonment* of the topics of the interaction. Chapter 11 presents an evaluation study of the topic manager that shows that the agent's adaptation and personalisation of the topics of the interaction, as enabled by the topic manager, are indeed perceived in the interaction.

In Part VI we conclude our findings and contributions, and give directions for future work. The last part, Part VII, contains the appendices that illustrate previous Chapters. Appendix F describes project A1:1 that forms the application of this thesis.

PART II : THEORETICAL BACKGROUND

2 Definitions of Engagement

In order to favour the engagement level of the user in human-agent interaction we need to know first what engagement actually means. *Engagement* is a term that occurs in a variety of domains and has a wide range of definitions and uses. In this chapter we give an overview of its most relevant definitions and interpretations.We explore the use of the term *engagement* respectively from a layman perspective (Section 2.1), in more expert domains (Section 2.2), and in the specific context of human-agent interaction (Section 2.3). In the next Chapter 3 we then discuss to engagement related concepts and behaviour.

2.1 Engagement in Everyday Use

According to [Harper, 2013] the oldest notion of *engage* in the literature goes back to the beginning of the 15th century in the sense of *to pledge*, coming from the old French *en gage* (respectively *make* and *pledge*). The noun *engagement* was first used around 1600 in the meaning of a *formal promise*. In the 17th century *engage* described senses of *attract the attention of*, *employ* and the specific *promise to marry*. *Engagement* referred to a *promise of marriage* as well but was also used in the sense of *battle* or *fight*. As from 1806 the term *engagement* is used to indicate an appointment [Harper, 2013].

Modern, everyday uses of the term engagement are similar to its original uses [Oxford-University-Press, 2013]:

1. (a) "A formal agreement to get married"
   - "The duration of an agreement to get married"

   (b) "An arrangement to do something or go somewhere at a fixed time"

   (c) "The action of engaging or being engaged"
   ("*Britain's continued engagement in open trading*")

   (d) "A fight or battle between armed forces"

Definitions (1.1a), (1.1b) and (1.1d) do not seem to be related in any way to engagement in interactions, the context we are eventually interested in. Explanation (1.1c) may be more relevant for our framework but its interpretation is less straightforward. Therefore, to see what is exactly meant by "the action of *engaging* or *being engaged*" we have a look at the definition of the verb *to engage* [Oxford-University-Press, 2013]:

2. (a) "Occupy or attract (someone's interest or attention)"
   (b) "Involve someone in (a conversation or discussion)"
   (c) "Participate or become involved in"
   (d) "Establish a meaningful contact or connection with"
   (e) "Arrange to employ or hire (someone)"
   (f) "Pledge or enter into a contract to do something"
   (g) "Dated reserve (accommodation, a place, etc.) in advance"
   (h) "Move into position so as to come into operation"
   (i) "Bring (weapons) together preparatory to fighting"

It can be noticed that definitions (Def.) 2.2e till 2.2i refer to very specific actions which again do not seem compatible with an interpretation of engaging or being engaged in an interaction. Others however, Def. 2.2a till 2.2d, might be compatible with such a context. From these descriptions rise however many questions: For example, when is someone actually involved (in a conversation or discussion) and when exactly do you speak of a meaningful contact? The answer to these questions may depend on the domain we are looking at. We therefore explore the more detailed definitions of engagement in specific fields.

2.2 Engagement in Specific Domains

Specific specialist use of the term engagement occurs at least in the domains of marketing (e.g. [Burns, 2013]), economics (e.g. [Rose and Spiegel, 2009]), diplomacy and politics (e.g. [Welsh and Fearn, 2008]), linguistics (e.g. [Hyland, 2005]), psychology (e.g. [Schaufeli et al., 2002]), philosophy (e.g. [McMahon and Portelli, 2004]), pedagogy (e.g. [Harris, 2011]), psychiatry (e.g. [Cohen-Mansfield et al., 2009]), social sciences (e.g. [Andrew and Sofian, 2012]), information science (e.g. [O'Brien and Toms, 2008]), artificial intelligence (e.g. [Sidner et al., 2005b]) and medicines (e.g. [Dietz and Lanzarone, 2004]). This shows that there are plenty of

definitions, many of which not relevant for our focus. Because we considered that Def. 2a till 2.2d are compatible with an interpretation of engagement in an interaction, we have a look at some examples of definitions in specific domains that go further on these interpretations. In Section 2.3 we subsequently discuss the definitions that are actually used for the specific application of human-agent interaction, the domain of this thesis.

We show that the examples in this Section (related to Def. 2a–2.2d and outside the domain of human-agent interactions) turn out to have all in common that they imply some positive relation of a (group of) person(s) towards a certain task, entity, or another (group of) person(s).

For example, [Schaufeli et al., 2002] refer with engagement to a positive relation of a person with its work:

3. "Engagement is defined as a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption."

An interpretation that is also referred to as *work/employee engagement* (e.g. [Alarcon and Lyons, 2011]). [Macey and Schneider, 2008] give an extensive account of the various uses of *employee engagement*.

Another example is *social engagement* that considers a positive relation of an individual or a member towards society. It has been defined amongst others as [Prohaska et al., 2012]:

4. "The extent to which an individual participates in a broad range of social roles and relationships."

by [Zhang et al., 2011] as :

5. "The commitment of a member to stay in the group and interact with other members."

and by [Millican, 2013] as:

6. "The ability to work constructively within and between social groups to create more resilient and sustainable communities."

The term *community engagement* can refer to a concept close to *social engagement* when used in the sense of [Tindana et al., 2007]:

7. "The process of working collaboratively with relevant partners who share common goals and interests."

The term is however more commonly used for explanations of engagement that point to a relation of an entire community towards some other entity, such as the government [Head, 2007] or research [Ahmed and Palermo, 2010]. The same holds for *public*

engagement, which refers to the relation of an entire public towards for example science [Nisbet, 2009b] or climate change [Nisbet, 2009a]. Similarly used concepts are *civic engagement* (e.g. [Skocpol and Fiorina, 1999]) and *citizen engagement* (e.g. [Powell and Colin, 2008].

A different type of example definition comes from [Martin and White, 2005] who describe engagement as a positive relation between two entities occurring within a text:

8. "Groups together all those locutions which provide the means for the authorial voice to position itself with respect to, and hence to 'engage' with, the other voices and alternative positions construed as being in play in the current communicative context."

In a similar context, [McGrath and Kuteeva, 2011, Ansarin and Aliabdi, 2011] employ [Hyland, 2005]'s definition of engagement as a relation between writers and readers in definition 9:

9. "Writers who relate to their readers with respect to the positions advanced in the text. This is an alignment dimension where writers acknowledge and connect to others, recognizing the presence of their readers, pulling them along with their argument, focusing their attention, acknowledging their uncertainties, including them as discourse participants, and guiding them to interpretations."

There are many more interpretations of engagement that are related to Def. 2a till 2.2d but often they do not provide an explicit definition. Examples are engagement between student readers and text [Gambrell et al., 2004], engagement between genres and society [Swales, 1993], engagement with beauty [Diessner et al., 2008], and engagement with music [Zentner and Eerola, 2010].

The few examples that we mention above already show that there exists lots of variation between the definitions of engagement. However, as we have seen, they all have in common that they imply some positive relation of a (group of) person(s) towards a certain task, entity, or another (group of) person(s). This observation makes that these interpretations of engagement are in line with an interpretation of engagement that we would be interested in for the application of human-agent interaction: Describing some positive relation of a person (the human user) towards another entity (the agent or the interaction). In the following section we try to specify this overall idea of engagement into a working definition in face-to-face human-agent interaction by considering interpretations of engagement that are used

in previous studies on human-agent interaction.

2.3 Engagement in Human-Agent Interaction

Over the last years researchers in face-to-face human-agent interaction have used different interpretations of engagement leading to different definitions and meanings that coexist. In multiple human-agent studies the word engagement is used without providing an explicit definition of the term in that particular study (e.g. [Baker et al., 2014, Bednarik et al., 2012, Bonin et al., 2012, Leite et al., 2013, Mower et al., 2007]). [Bednarik et al., 2012] for example, define six levels of engagement in conversation ranging from "no interest" to "governing/managing discussion" but without giving an explicit definition of what engagement actually is. In these studies without explicit definition, we have to infer the interpretation of the concept from the way in which the concept is used (measured/annotated/simulated/...). This makes such implicit interpretations of engagement less straightforward to contribute to an overall understanding of the concept and to be reused in future studies.

In other studies a clear definition of engagement has been given but exclusively for the purpose of that particular study. These definitions only cover the particular effect of engagement chosen to be measured in that study. [Michalowski et al., 2006] for example, express engagement in terms of spatial distance. In their work they characterise an engaged person as:

10. "Mostly stationary near the booth, perhaps with a detected face."

As such definitions are too context specific to serve as a description of the entire concept we no further elaborate on these. In the rest of this section we identify existing generic definitions of engagement in face-to-face human-agent interaction (including human-robot interaction) in order to obtain an overview of existing interpretations that could contribute to our the understanding of the concept and the positioning of our research. We do not discuss existing interpretations of engagement that *could* be used in the context of human-agent interaction, but limit this account to definitions that have actually been used in this context.

2.3.1 Definitions

In this section we first identify the definitions used in human-agent interaction. Table 2.1 and the following sections subsequently discuss the differences and commonalities among them.

[Goffman, 1966] formulated a definition of engagement in 1966 that is still frequently used in human-robot interaction [Le Maitre and Chetouani, 2013, Couture-Beil et al., 2010]. Speaking about *face engagements* he claims that:

11. "Face engagements comprise all those instances of two or more participants in a situation joining each other openly in maintaining a single focus of cognitive and visual attention – what is sensed as a single mutual activity, entailing preferential communication rights."

[Goffman, 1966] thus describes engagements as instances of the interaction, while implying a relation between participants of a situation (i.e. interaction). This interpretation is often used to describe the process through which people employ eye-contact, gaze and facial gestures to interact with each other [Le Maitre and Chetouani, 2013] as these cues may signal a "focus of cognitive and visual attention".

Another definition where engagement is described as a relation between participants of an interaction comes from [Sidner and Dzikovska, 2002] who define engagement as:

12. "The process by which two (or more) participants establish, maintain and end their perceived connection. This process includes: initial contact, negotiating a collaboration, checking that other is still taking part in interaction, evaluating staying involved, and deciding when to end connection."

Contrary to 11 however, here engagement is not an instance of the interaction, but a process within the interaction. Def. 12 has been employed in multiple studies, e.g. in [Castellano et al., 2012, Holroyd et al., 2011, Nakano and Ishii, 2010].

[Bohus and Horvitz, 2009b] formulated a variation to [Sidner and Dzikovska, 2002]'s definition, in order to fit their interaction model:

13. "The process subsuming the joint, coordinated activities by which participants initiate, maintain, join, abandon, suspend, resume or terminate an interaction."

The only difference with Def. 12 is that more activities of the participants are covered. This interpretation in its turn is reused by [Klotz et al., 2011].

[Poggi, 2007]'s interpretation of engagement does not describe engagement as specific instances (Def. 11) or processes (Def. 12 and 13) of the interaction, but as a concept that refers to the (mental state) of a participant in the interaction. [Poggi, 2007]'s definition of engagement is:

14. "The value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction."

This definition has also been used in for example [Castellano et al., 2009] (as *user engagement*), [Peters et al., 2005b, Sanghvi et al., 2011].

[Hall et al., 2005] formulated a definition of *empathic engagement*. Their definition implies that it is formulated for particular human-agent interaction only as it describes a relation between a "user" and a "synthetic character" instead of participants in an interaction in general:

15. "Empathic engagement is the fostering of emotional involvement intending to create a coherent cognitive and emotional experience which results in empathic relations between a user and a synthetic character."

Here, engagement is not described as an instance (Def. 11) or a process (Def. 12 and 13) of the interaction, and not as a mental state (Def. 14) either. Instead, it is considered as an action ("fostering of emotional involvement").

Other definitions of engagement in human-agent interaction refer to a second concept, such as interest and attention in [Yu et al., 2004]:

16. "User engagement describes how much a participant is interested in and attentive to a conversation."

This definition does not present engagement as a concept on its own but describes it as equal to interest and attention. The definition was developed in the context of a voice communication system and therefore considers engagement of the "user" instead of engagement of interaction "participants". The definition has subsequently been referred to in human-agent interaction [Novielli et al., 2010].

Lastly, as [Yu et al., 2004], [Bickmore et al., 2010] formulated a definition where engagement is being described as being equal as another concept, here involvement.

17. "The degree of involvement a user chooses to have with a system over time."

In contrast to the other definitions, Def. 17 consider engagement between users and systems over multiple interactions at once. Because Def. 16 and 17 point towards other concepts, respectively interest, attention, and involvement these descriptions are not very informative about what engagement is, especially in comparison to other concepts in the interaction. It does not explain why the term engagement exists besides the terms interest, attention and/or involvement. Besides, these latter concepts can be interpreted in many different ways as well.

Table 2.1 : Decomposed definitions of engagement.

Def.	Source	Type	Measure	Perspective	Relation	(Doing) What
11	[Goffman, 1966]	Face	Instances	Situation	Between participants	*Joining each other* openly in maintaining a single focus of cognitive and visual *attention*
12	[Sidner and Dzikovska, 2002]	-	Process	Participants	Between participants	Establishing, maintaining and ending their *perceived connection*
13	[Bohus and Horvitz, 2009b]	-	Process	Participants	Participants	*Joint, coordinated activities* to initiate, maintain, join, abandon, suspend, resume or terminate an interaction
14	[Poggi, 2007]	-	Value	Participant	Of one participant to another + Of one participant towards the interaction	*Being together* and continuing the interaction
15	[Hall et al., 2005]	Empathic	Activity	User	Of a user towards a synthetic character	Fostering emotional involvement
16	[Yu et al., 2004]	User	Description of how much	User	Of a participant to a conversation	Being interested in and attentive to a conversation
17	[Bickmore et al., 2010]	-	Degree	User	Of a user towards the system	Involvement a user chooses to have with a system

2.3.2 Commonalities

Looking at the definitions of engagement that are used for human-agent interaction (Section 2.3.1) there are several commonalities and differences to notice. Table

2.1 summarises the characteristics of each of these definitions. In this Section we further describe the commonalities, and in the following Section 2.3.3 the differences among the definitions.

2.3.2.1 Relation, Connection, and Cooperation

Definitions of engagement outside the scope of human-agent interactions described in Section 2.2 all had in common that they describe some relation of a person (or persons) towards some other entity. The definitions in human-agent interaction all seem to imply a relation between some person and another entity as well. Column 6 of Table 2.1 lists for each definition the relation that it implies. Def. 12 implies for example a relation between participants, and for example Def. 15 refers to a relation between a user and a synthetic character. In Section 2.3.3 we discuss individual differences of the implied relations.

The relations are all expressed by a connection and cooperation between the participants of the interaction. Human-agent interaction requires namely a connection and cooperation between its participant since participants collaborate together in performing the joint action of interacting [Clark, 1996]. It seems that in all above definitions engagement points to this cooperation and connection necessary for interaction. In many of the definitions the connection and cooperation is made explicit. For example, by speaking in terms of *joining each other in* (Def. 11), a *perceived connection* (Def. 12), *joint coordinated activities*, and a *goal of being together* (Def. 14) (Def. 13) (listed in Column 7 of Table 2.1). If the cooperation and connection between participants is not explicit, the definitions refer to a state that is the result of or requirement necessary for a connection and/or collaboration, meaning that this state is the result or condition for the connection or cooperation of the participants. These states are *involvement* (Def. 15 and 17) and *being interested in attentive to a conversation* (Def. 16). In all the definitions the level of connection and cooperation is positively related to the presence or level of engagement.

2.3.2.2 Dynamic

Another commonality that we can derive from the above observations is that engagement is a concept whose occurrence is dynamic. This is because, since the interaction evolves over time the connection and cooperation that are implied by the interactions (see above) can also change over time. The dynamic nature of engage-

ment is implied by the definitions: If engagement is defined in terms of *instances* (Def. 16) or *processes* (Def. 12 and 13) engagement occurs only at certain moments in the interaction. If engagement is described by a *value* (Def. 14), *description of how much* (Def. 16), or *degree* (Def. 17) it indicates possible variation of the engagement intensity during the interaction. And finally, if engagement is described by *activities* (Def. 15) this means that engagement can occur both at specific moments and to different extents (see also Column 4 Table 2.1). This means that engagement can vary during the course of the interaction.

Summarising the findings above, engagement thus evolves around a relation of a participant towards another entity of the interaction and is dependent on the varying connection and cooperation that is established in the interaction.

2

2.3.3 Differences

2.3.3.1 Relations

While the definitions of engagement used in human-agent interaction all imply a relation of an interaction participant (or the "user") with another entity of the interaction (Section 2.3.2.1), the second entity differs among the definitions. In some definitions (see Column 6 of Table 2.1) the relation is with the second participant or the system (e.g. Def. 13 and Def. 17), while in others it is with the interaction itself (Def. 16), or with both the second participant and the interaction (Def. 14).

What also differs among the interpretations is that some definitions refer to a one-sided relation of a person/user towards another person or the interaction (Def. 14, 15–17), while other definitions consider multi-sided (i.e. mutual) relations; instead of focusing on the relation of one participant/user towards another or the interaction, the latter definitions consider the relations of all the interaction participants towards each other (Def. 11, 12, 13, see column 6, Table 2.1).

This means that in studies on engagement in human-agent interaction the selected definition should reflect the type of relations that are taken into account: what parties are involved in the interaction and if the relation is one-sided or multi-sided.

2.3.3.2 Entities and Perspectives

As we have seen in Section 2.3, another difference among the definitions of engagement in human-agent interaction is the nature of engagement: the entity in which it is expressed. We have seen (Section 2.3 and Column 4 of Table 2.1) that

engagement can be defined as instances of a situation (i.e. interaction, Def. 16), processes by participants (Def. 12 and 13), values attributed by participants (Def. 14), fostering of emotional involvement (of a user, Def. 15), a quantity of interest and attention of a user (Def. 16), and the degree of involvement of a user (Def. 17).

These entities reflect different perspectives of the definitions: Engagement is either attributed to the interaction itself (Def. 11), belonging to all the participants of the interaction (Def. 12, 13), or attributed to a single participant (i.e. user) in the interaction (Def. 14, 15–17) (listed in Column 5, Table 2.1).

Again, the definition that is selected for a study can thereby reflect the perspective of that study: either an interaction perspective, a participant/user perspective, or a perspective from all the participants together.

2.3.3.3 Implementation Measures

Another distinction that can be derived from the different ways in which engagement is expressed (e.g. as instance, value, process, etc., see Section 2.3.3.2 and Column 4, Table 2.1) is the way in which engagement could be estimated; it has consequences on the way engagement could be measured.

For example, entities like a *value*, or *instances* describe engagement as a certain state (resulting from a connection and cooperation between interaction participants) that can be estimated at any moment in time. Other definitions focus on the process to get to this state by calling engagement *processes* (Def. 12, 13) or *activities* (Def. 15).

Within the cases where engagement relates to a certain state we can distinct two subcategories: Those that consider engagement as something that is merely present or absence because it is an *instance* (Def. 11) and those that allow a continuous notion of engagement as it is expressed as a *value* (Def. 14) or *degree* (Def. 17).

The definition of engagement in a particular study can thus also reflects the way engagement is measured.

2.3.4 Conclusion

We have given an overview of the different definitions of the term engagement in human-agent interaction and identified the differences and commonalities between the definitions. Given the number and variation of the definitions we can state that engagement in human-agent interaction is a complex concept. Seen from the differ-

ences among the definitions, studies can focus on a particular aspect or interpretation of engagement without being able to cover the entire range of interpretations.

The different definitions all imply a dynamic relation of an interaction participant towards/with another participant and/or the interaction, due to a connection and cooperation between the interaction participants. The different parties involved in the relation, the different perspectives, and the different measures which are implied by the definitions, make that each definition expresses another focus of engagement. Studies may express their focus of engagement by choosing one definition over another; depending on the goal of the study some definitions may be more suitable in certain circumstances than others. For example, for some studies it may be interesting to consider the engagement of all participants at once (as in e.g. Def. 12 [Sidner and Dzikovska, 2002]) while for other studies it may be useful to focus on the engagement level of only one participant at a time (as in e.g. Def. 14 [Poggi, 2007]). Since a definition can indicate the focus of a study, and as long as a study clearly indicates what definition of engagement it employs, we do not consider that multiple definitions coexisting is problematic. In the following Section we motivate our interpretation of engagement in this thesis.

2.4 Engagement in Our Work

In this thesis we choose to work with a definition that is applicable to all types of interactions, including human-human interaction. This is because we aim at making human-agent interactions as believable and natural as human-human interaction. Using a definition that considers an interaction "participant" in general instead of a "user" would simplify a comparison of engagement between both interaction types. We therefore eliminate Def. 15 until 17 from our options.

Regarding the differences in perspective (Section 2.3.3.2) we choose to use a definition that describes engagement as attributed to an individual participant. This choice is based on the fact that describing engagement for every participant separately allows us to specify the contribution of every participant to the overall engagement. Moreover, the participants may experience similar events differently, which naturally leads to different simultaneous levels of engagement. Attributing engagement to a single participant allows us to use from hereon the terms *user engagement* to refer to engagement from the side of a user, and *agent engagement* to refer to engagement from the side of the agent.

The relation of the participant which is implied by the term engagement, can

either be oriented towards another participant or to the interaction/situation, or both (Section 2.3.3.1). We chose the latter interpretation, as this is the richest interpretation covering both aspects.

The only definition that meets all the described choices regarding the implied relations and perspective is the definition by [Poggi, 2007]. We thereby consider and measure engagement as certain state, and not as for example a process to get to this state (Section 2.3.3.3). A practical advantage of using an existing definition is that it contributes to the consistency and possibility of comparison between studies.

In conclusion, we will base all the work on engagement in face-to-face human-agent interaction in this thesis on [Poggi, 2007]'s interpretation:

18. "The value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction."

3 Engagement Related Concepts and Behaviour

In order to favour user engagement in human-agent interaction we need to know not only how to interpret the concept (previous Chapter) but also how it is manifested in an interaction. Therefore, in this Chapter we give an overview of respectively concepts and forms of behaviour that are associated with engagement in previous research.

3.1 Engagement Related Concepts

There exist numerous concepts that are often mentioned in relation with engagement and that are sometimes even used interchangeably [Peters et al., 2009]. In this section we try to explore how these concepts are related. In this way we hope not only to gain a clearer idea of engagement, but also to clarify the existing terminology around engagement. We briefly discuss the concepts *attention*, *involvement*, *interest*, *immersion*, *rapport*, *empathy*, and *stance* to see how they relate to engagement, without detailing all the coexisting definitions of each term.

3.1.1 Attention

According to [Peters et al., 2009] in most studies relating to engagement two underlying fundamentals are apparent: attentional and emotional involvement. It depends on the definition of engagement how attention relates to engagement. For example according to [Yu et al., 2004] attention is one of the two aspects (together with interest) that constitute engagement (Def. 16, Section 2.3.1). Other interpretations consider attention as a vital aspect of engagement but not as a part that constitutes engagement. Attention controls namely the orientation of the senses towards stimuli that are of relevance to engagement [Peters et al., 2005b]. Selective attention to a stimulus seems necessary for a basic form of engagement [Peters et al., 2009] and is thereby a condition for engagement. For interpretations of engagement

where attention does not constitute engagement, attention is a condition for engagement, meaning that attention and engagement may co-occur, but are not the same.

3.1.2 Involvement

Involvement can be interpreted in different ways and can relate to engagement in different ways, depending also on the interpretation of engagement. [Sidner et al., 2005b] claims that involvement concerns being captured by the experience. This interpretation is referred to by [Lombard and Ditton, 1997] and [Lombard et al., 2000] as engagement. [Bickmore et al., 2010] also defines engagement directly in terms of involvement (Def. 17, Section 2.3.1).

[Peters et al., 2005b] talks about cognitive and emotional involvement and commitment and claim that they are key factors that only underlie engagement. Here involvement is not the same as engagement but could, as attention, be interpreted as a condition for engagement.

[Bamoallem et al., 2016] defines six aspects of user involvement: immediacy (physical proximity), expressiveness (energy, activity, enthusiasm), altercentrism (focus on the conversation partner), interaction management (smooth flow), composure dimension (body movement, confidence), and positive affect (good feelings). Here involvement is thus expressed in terms of behaviours. (Some of) these behaviours may be associated with engagement (see also Chapter 3).

While the exact relation of involvement and engagement is thus disputable, it is clear that both concepts are closely related. Because we use the definition of [Poggi, 2007] (Def. 18, Section 2.4) we consider involvement as not being equal to engagement but as only an aspect that underlies (i.e. is a condition for) engagement.

3.1.3 Interest

According to 16's definition of engagement (Def. 16, Section 2.3.1) being interested is one of two concepts (together with attention, see Section 3.1.1 above) that constitute engagement.

According to [Peters et al., 2005a], engagement is however only linked to and possibly caused by interest, while both interest and engagement are caused by attention. Interest can be seen as "an emotional state linked to the participant's goal of receiving and elaborating new and potentially useful knowledge" [Peters et al.,

2005a]. Interest is often used to describe the motivation or goal towards opening and maintaining engagement [Peters et al., 2009].

As we follow [Poggi, 2007]'s definition of engagement (Def. 18, Section 2.4) we do not define engagement in terms of interest (and attention), but follow [Peters et al., 2005a] in considering that engagement is only linked to and can be caused by interest.

3.1.4 Immersion

Immersion is a term that is particularly in use in the area of virtual reality. It is often defined [McMahan, 2003] as "the experience of being transported to an elaborately simulated place" or "the sensation of being surrounded by a completely other reality" [Murray, 1997]. Although there are many different interpretations of the term [Brown and Cairns, 2004] immersion tends to refer to a broader concept than engagement. [Brown and Cairns, 2004] claim in the context of game immersion that engagement can be seen as a part of immersion, namely as the first in three levels of immersion (engagement, moving on to greater involvement in engrossment, and total immersion).

3.1.5 Rapport

Gratch and colleagues [Huang et al., 2011] describe rapport in a broad manner as "the feeling of being in sync with your conversational partner". [Tickle-Degnen and Rosenthal, 1990] specify that rapport only exists in interaction between individuals and that it consists of three essential components: mutual attentiveness, positivity and coordination. Engagement is also directly related to attentiveness (Section 3.1.1) and coordination (Sections 2.3.2.1, 3.2.1.1, 3.2.1.2) but not necessarily to positivity; one can be engaged and angry at the same time.

[Gratch et al., 2007] designed a virtual agent to elicit rapport from human participants in a dyadic narrative task. They found that this agent produced increased speaker engagement [Gratch et al., 2007], which shows a positive correlation between rapport and engagement.

Using [Poggi, 2007]'s definition (Def. 18, Section 2.4) we consider that rapport refers to the first of the two aspects that engagement covers: the value attributed to being together with the other participants.

3.1.6 Empathy

Empathy refers to the capability of sharing and interpreting correctly another's emotions and feelings [Decety and Jackson, 2004]. Literally "it accounts for the naturally occurring subjective experience of similarity between the feelings expressed by self and others without loosing sight of whose feelings belong to whom" [Decety and Jackson, 2004]. Two kinds of empathy have been distinguished: a basic emotional contagion system where "one feels what the other feels" and a more advanced cognitive perspective-taking system where "one understands what the other feels" [Shamay-Tsoory et al., 2009].

Central to both types of empathy is the sharing of one's feelings, which requires a connection between the participants. Engagement also implies a connection between participants (Section 2.3.2.1). This makes that if a participant displays a high level of empathy this person also appears to be engaged. The contrary is however not necessary: A connection between participants leading to engagement does not necessarily mean that the participants share or interpret each others' feelings; one can be highly engaged in a hostile discussion without empathising with the interaction partner(s).

3.1.7 Stance

Stance is another term that is related to engagement and has many definitions. The overview of interpretations provided by [Chindamo et al., 2012] have led them to suggest that stance is "an attitude which, for some time, is expressed and sustained interactively in communication in a unimodal or multimodal manner". Given this description of stance we consider that engagement can be signalled by a stance. A shared stance can depend on the engagement of the agents [Prepin et al., 2012]. [Allwood et al., 2012] mention that aggressive and provocative stances can be considered as special cases of engagement and disagreement.

3.1.8 Conclusion

We listed to engagement related concepts that are frequently mentioned in the context of human-agent interaction, with the goal of clarifying how the different concepts relate to each other. While there may exist various interpretations for each of these concepts, we consider that attention, involvement, interest, and empathy may form conditions and/or consequences of engagement, and rapport forms a part of

engagement. These concepts are positively related to engagement: By favouring one of these concepts it is likely that the engagement of an interaction participant (one or both participants) is also favoured. Particular stances can express engagement.

3.2 Engagement Related Behaviour

All aspects of the agent may potentially influence the user's engagement and/or the related concepts that are discussed in the previous Section. For example, the agent's appearance [Zimmerman et al., 2005], the pleasantness of the agent's speech (synthesised by a Text To Speech tool) [Niculescu et al., 2010], and the agent's capability to understand the user (Speech Recognition, Natural Language Understanding) [Schmitt and Minker, 2012] can play a large role in the course of the interaction and the user's perception of the interaction. However, for our work we consider given (fixed) agent graphics, TTS and SR tools. We therefore do not look further into these aspects. Instead we focus on the behaviour of the interaction participants that may be associated to engagement.

For an engaging conversational agent there are three tasks that involve engagement associated behaviour:

— Detecting and interpreting the user's behaviour to infer a perceived level of *user engagement*.

— Generating appropriate (engaged) reactions, thereby expressing a level of *agent engagement*.

— Demonstrating behaviour that influences the engagement of the user, thereby creating an *engaging agent*.

While in this thesis we are primarily interested in the third point, for each of these tasks we would need to be able to associate behaviour to engagement and vice-versa. In this section we look at what behaviour is related to engagement in previous work. As we have seen from the definitions of engagement in human-agent interaction (Section 2.3) and the concepts that are closely related to engagement (Section 3.1), engagement is characterised by a connection and coordination between the interaction participants. Therefore, it is no surprise that behaviour that can be associated to engagement and/or its related concepts is also characterised by expressing a connection and coordination between the interaction participants.

Below we give an overview by summarising the types of *high-level* (Section 3.2.1) and *low-level behaviour* (Sections 3.2.2 and 3.2.3) that have been associated with

engagement in human-agent interaction. High-level behaviour can be performed by different types of *low level* behaviour: Low-level behaviour refers to actions as gestures, movements, gaze or speech where high-level behaviour refers to the communicative function, emotion or attitude.

3.2.1 High level behaviour

3.2.1.1 Synchrony

A first type of high level behaviour associated to engagement is synchrony. According to the overview of [Delaherche et al., 2012] synchrony "refers to individuals' temporal coordination during social interactions". More specifically, it is "the dynamic and reciprocal adaptation of the temporal structure of behaviours between interactive partners" [Delaherche et al., 2012]. This is illustrated by an example of two people who cross or uncross their legs at the same time or gaze in the same direction simultaneously during an interaction. The behaviours are multimodal as different (verbal and non-verbal) modalities intervene at the same time. But more important than the nature of the behaviours is the timing of the behaviours in comparison with the other participants' behaviours [Delaherche et al., 2012].

Synchrony has multiple effects in interaction between adults [Delaherche et al., 2012]. For instance, it contributes to smoothing social interactions and to achieve a coordination of expectancies among participants [Kendon, 1970]. Most relevant to engagement is that synchrony fosters cooperation [Wiltermuth and Heath, 2009] and empathy [Decety and Jackson, 2004] between individuals. This shows that synchrony is related to engagement.

3.2.1.2 Alignment

Very similar to synchrony is alignment. Alignment also expresses a coordination between interaction participants but this time at a deeper level. [Pickering et al., 2004] refer to alignment as coordination that occurs when interlocutors share the same representation at some level [Branigan et al., 2000, Garrod and Anderson, 1987]. [Pickering et al., 2004] have considered alignment at verbal levels, such as on a lexical and syntactic level, but for example [Bergmann and Kopp, 2012] has demonstrated that the concept can be expanded to non-verbal behaviour. In contrast to synchrony which is multimodal, alignment always occurs in the same modality [Delaherche et al., 2012]. Just as synchrony, alignment is a measure of

coordination which makes it a possible symptom of engagement. Besides, failure of alignment may cause failure to communicate successfully [Branigan et al., 2010] influencing the participants' engagement directly.

3.2.1.3 Mimicry

Mimicry is the behaviour of a participant who does what another person does. Mimicry refers to "an automatic tendency to imitate others" [Van Baaren et al., 2003]. Where in synchrony the important element is the timing, in mimicry the nature of the behaviour is important [Delaherche et al., 2012]. It contributes to the connection between the interaction participants, and favours empathy, liking and the smoothness of the interaction [Chartrand et al., 2005]. [Van Baaren et al., 2003] agrees that mimicry enhances pro-social behaviour and helps to create bonds between individuals. Given that mimicry favours the connection between the participants it is another possible symptom of engagement.

3.2.1.4 Feedback and Backchannels

According to [Allwood and Cerrato, 2003] several definitions of feedback have been proposed that all "agree on the fact that feedback strategies are used as a cooperative way of exchanging information about the successfulness of communication". This indicates that feedback can reveal something about the engagement of the participants: it can indicate that the communication is more or less successful which can be a sign of more or less engagement.

Backchannels can be considered as a type of feedback [Allwood and Cerrato, 2003]. Backchannels are "non-intrusive acoustic and visual signals provided by the listener during the speaker's turn" [Yngve, 1970] and provide information about the listeners engagement and processes which are necessary for communicative interaction: attention, perception, comprehension and internal reactions [Peters et al., 2005a, Poggi, 2007].

3.2.1.5 Collaboration on a Task

Other high-level behaviour that may be a cue for engagement can be found in the form of collaborating on a task. If participants cooperate and coordinate their actions they imply a perceived connection between them. [Sidner et al., 2005a]

claims that engagement is supported by this ability to collaborate on a task, the use of conversation, and gestural behaviour.

3.2.1.6 Interaction Time

In a study by [Sidner et al., 2004] the length of interaction time is considered as one of the measures of how coordinated two participants are. It has been used (in combination with eye gaze and directed talk) to estimate a participant's engagement in interaction with a robot [Sidner et al., 2005a]. The interaction time gives of course a clear indication to what extent the participants value continuing the interaction, which is a key aspect of engagement in [Poggi, 2007]'s interpretation.

3.2.2 Non-verbal behaviour

The forms of high-level behaviour that are related to engagement can be expressed by different forms of low-level (verbal and non-verbal) behaviour. Besides, there are multiple studies where a particular type of low-level behaviour is mentioned as an indicator of engagement directly. As for high-level behaviour, engagement related low-level behaviour is often associated to engagement related concepts such as attention, involvement and interest as well (Section 3.1). Usually different types of verbal and non-verbal behaviour occur together. Here we first mention the non-verbal behaviours before mentioning verbal behaviours in Section 3.2.3.

3.2.2.1 Eye Gaze and Head Movements

In many studies eye gaze is seen as a cue of engagement. [Sidner et al., 2003] for example, specified in the context of a collaborative task that "looking at the speaking collaborative partner is evidence of engagement, while looking around the room for more than very brief moments, is evidence of disinterest in the interaction and possibly the intention to disengage". "Looking at objects relevant to the conversation" is however "not evidence of disengagement" [Sidner et al., 2003]. On the contrary, looking together at the same object as the collaborative partner is an indication of how coordinated two participants are in their interaction [Sidner et al., 2004]. Sequences of looking behaviours such as looking back at the conversational partner after looking at some object also indicate a level of engagement [Sidner et al., 2004].

Looking behaviours can also play a role in grounding; indicating that what has been said is understood [Sidner et al., 2004].

[Peters et al., 2005a] mention that gaze is an important way of providing feedback and subtle signalling through which a listener can show his level of interest and engagement.

Other examples where eye gaze has been used as a cue for engagement are [Castellano et al., 2009] who considers the amount of time the user looks at a robot, and [Nakano and Ishii, 2010] who observed sequences of gaze-behaviours.

Often eye gaze (i.e. looking behaviours) is accompanied by head movements, making this a cue of engagement as well.

Head movements in the form of nods can form feedback providing information about the participant's engagement [Allwood and Cerrato, 2003].

3.2.2.2 Gestures and Postures

Gestures can be clear indications of engagement as they can serve as "a means to indicate the desire to further or discontinue the collaboration" [Sidner et al., 2003]. Engagement gestures are however "sensitive to the conversational and collaborative context of use" [Sidner et al., 2003]. Besides, gestures that are linked to engagement are culturally determined. Nevertheless, "every culture has some set of behaviours to accomplish the engagement task" [Sidner et al., 2003]. [Sidner et al., 2005a] mentions that such conversational gestures generally concern, besides gaze, "pointing behaviours, bodily addressing the conversational participant and other persons/objects in the environment, and various hand signs, all with appropriate synchronisation with the conversational collaborative behaviour". In short, engagement is supported by all gestural behaviour that conveys a connection between the participants [Sidner and Lee, 2003]. Consequently, greeting and goodbye gestures can serve as cues of respectively initiating and ending the interaction and thereby the engagement.

Specific gestures of turn taking are indicators of engagement as well since "the overall choice to take the turn is indicative of continuing the interaction" [Sidner et al., 2003]. An example of a turn taking gesture is returning to face the conversational partner when finishing a conversational turn [Sidner et al., 2005b].

[Peters et al., 2005b] assign for their attention metric a high value to body parts (not only the head) that are oriented towards the viewer, implying that it is also more likely that a person is engaged if he is oriented towards the viewer than if he

is oriented elsewhere.

[Mota and Picard, 2003] have found a correlation between a presumed level of interest and several postures and posture sequences. [Sanghvi et al., 2011] have found a similar correlation between engagement and postures such as the body lean angle, the curvature of the back and a contraction index of the upper body.

Lastly, [Oertel et al., 2011] have conducted a pilot study that concentrates on the quantity of movement. They found indications that the more a person is involved, the more he moves. As involvement is a condition for engagement it is probable that it is also the case that the more a person is engaged, the more he moves. [Sanghvi et al., 2011] confirm that quantity of movement seems to be a key feature for recognition of user engagement with a robot.

3.2.2.3 Facial Expressions

Facial expressions can have a range of different meanings [Niewiadomski et al., 2010]. Smiling can for example, realise mimicry and communicate amusement, embarrassment or politeness [Niewiadomski et al., 2010]. Facial expressions can thus reveal information about the person's engagement in many different ways.

In some contexts smiling (in combination with eye gaze) is seen as a direct indicator of engagement [Castellano et al., 2009]. [Castellano et al., 2009] observed that when users are more engaged with a robot they tend to smile more than when they are not engaged with it.

In short, facial expressions can give information about the performer's attitude and emotions. Since attitudes can convey more or less engagement and emotional involvement is a condition for engagement, facial expressions can also form a cue of the participant's engagement.

3.2.2.4 Temporal and Spatial Features

[Michalowski et al., 2006] claim that "the distance between interactors determines the relative salience of attentional visual cues, shaping our perception of attention and therefore of engagement". Besides, they notice the possibility that people at a certain distance are more likely to be paying attention to a robot. For their particular study, [Michalowski et al., 2006] consider people that are near to the robot and passively observing the robot's behaviour as engaged, while people that are further away are considered to have lower levels of engagement.

We agree that for engagement a minimal distance is required as face-to-face interaction requires a minimum distance between the participants. This makes that there is a higher chance for someone to be engaged if this person is located within a certain distance of the interaction partner. However, for our view of engagement proximity and passive observation alone are not sufficient to talk about engagement. We think these factors are only prerequisites of engagement as they are prerequisites for an interaction to take place. In order to speak of engagement in human-agent interaction, we have to be able to talk about a human-agent interaction, meaning that the agent and user are not simply nearby observing each other, but together contribute to the joint action of interacting (Section 2.3.2.1) [Clark, 1996].

3.2.2.5 Physiological Reactions

Other, but less frequently considered cues of engagement are physiological reactions. Such reactions can give information about one's engagement as they can provide a baseline for determining the subjects' emotions during interactions [Choi et al., 2012]. However, virtual agents cannot express physiological reactions. Therefore physiological cues can only be used as an indication of user engagement but not as a way to simulate agent engagement.

[Mower et al., 2007] for example, found that it is possible to estimate user engagement by physiological cues such as galvanic skin response and skin temperature. [Choi et al., 2012] looked at heart rate and electro-dermal activity.

3.2.3 Verbal behaviour

Engagement is also supported by the use of verbal behaviour as it can convey connectedness [Sidner et al., 2005a]. Conveying connectedness, possibly by realising one of the high-level engagement-related behaviours, can be realised by multiple aspects of verbal behaviour. Where we could specify non-verbal behaviour by its different modalities, we categorise the different aspects of verbal behaviour into three aspects of speech: form, timing, and content.

3.2.3.1 Form

With the form of verbal behaviour we refer to *how* something is verbally expressed. The form of an utterance thus refers to multiple aspects, such as prosodic,

lexical, and syntactic choices. The way (form) in which something is said can reveal something about the engagement of the speakers.

[Yu et al., 2004] for example, links prosodic features directly to engagement. They do this by considering the expression of emotion in someone's voice. This makes sense as, as we have seen earlier, emotional involvement is a condition for engagement. Similarly, [Oertel et al., 2011] found a relationship between prosodic features (level, span and intensity of voice) in relation to their measure of involvement.

When the form of verbal behaviour conveys connectedness it can be a sign of engagement [Sidner et al., 2005a]. Lexical and syntactic choices in human-agent interaction for example, can express, among others, verbal alignment [Pickering et al., 2004]. Also politeness and formality can be aligned among the interaction participants [De Jong et al., 2008]. As we have seen above (Section 3.2.1.2) verbal alignment behaviours can reveal something about the participant's engagement as it expresses a coordination between the participants.

3.2.3.2 Timing

As we have seen already for non-verbal behaviour, turn-taking behaviour is related to engagement. Turn-taking is namely a key characteristic of synchrony between the interaction participants [Harrist and Waugh, 2002] and therefore an expression of engagement (see Section 3.2.1.1). In a study by [Sidner et al., 2003] disengagement from the interaction is understood when a collaborative partner fails to take an expected turn (in combination with loss of the face of the human).

Verbal backchannels (a form of feedback) also provide information about the listeners engagement [Rich et al., 2010] and form another verbal behaviour that is characterised by its timing dimension, since backchannels are inserted at a specific moments within the turn of another participant.

3.2.3.3 Content

On a content (pragmatic) level several aspects have been associated with engagement in previous work.

Initiating an interaction for example, implies (a minimum level of) engagement which makes that conversational greetings can be seen as a cue for initiating an interaction and the possibility of engagement [Sidner et al., 2003, Bohus and Horvitz,

2009a]. Also calling or other opening dialogue moves have been related to engagement [Bohus and Horvitz, 2009b]. Disengagement can be achieved by closing comments [Sidner et al., 2005b].

High-level engagement related behaviour of collaborating on a task may be realised by verbal behaviour (as well as non-verbal behaviour). [Le Maitre and Chetouani, 2013] have looked at two different interaction styles (detected by prosodic features), so-called "system directed speech" and "self-speech" that indicates out-of-task situations. They see these interaction styles as indicators for respectively engagement and disengagement (with a task).

The occurrence of adjacency pairs can be seen as an engagement cue as it contributes to the perceived connection between the interaction participants [Rich et al., 2010]. A second pair part of an adjacency pair is namely adjusted to the first part [Schegloff, 2007], reflecting a coordination between the interaction participants. This concept can be generalized to include not only verbal but also non-verbal communication acts [Rich et al., 2010].

Small talk (i.e. social talk) can serve to build rapport [Bickmore and Cassell, 2005]. As mentioned in Section 3.1.5, rapport refers to "the feeling of being in sync with your conversational partner". As we consider rapport to be a part of engagement (Section 3.1.5) small talk can contribute to the rapport related dimension of engagement.

Further, verbal (as well as non-verbal) expressions of emotions or attitudes can reveal information about the participant's engagement, expressing for example empathy. There are certain words that represent affective concepts [Strapparava and Valitutti, 2004], which may give an indication as whether or not we are dealing with an emotional utterance. Reciprocal self-disclosure is mentioned as resulting in increased engagement [Bickmore et al., 2013]. As it increases the intimacy of the interaction participants [Bickmore et al., 2013] it refers just as small talk, to the rapport dimension of engagement, the connection between the interaction participants. Expressing empathy by expressions of agreement with the user has lead to better closeness and user satisfaction in human-computer dialogue [Higashinaka et al., 2008].

3.2.4 Conclusion

We have given an overview of high-level and low-level behaviours that are in particular associated to engagement in previous research. The behaviours may signal

engagement and in their turn engage the other participant(s) as they generally imply a connectedness between the interaction participants. The list of potentially engaging forms of behaviour is however much longer as potentially each form of behaviour may have an influence on "the value that a participant in an interaction attributes to the goal of being together with the other participant and continuing the interaction" (definition [Poggi, 2007]). In this thesis we aim at forming such agent behaviour strategies that are likely to favour the user's engagement. We thereby focus on the agent's verbal behaviour, considering strategies that regard respectively all three aspects of verbal behaviour: form (Chapter 5), timing (Chapter 6) and content (Chapters 7–11).

PART III : RELATED WORK

TELECOM ParisTech

4 Related Work

We are aiming at developing verbal behaviour strategies for an embodied conversational virtual agent that plays the role of a visitor in a museum, to engage the user during interaction about the museum artworks. In this chapter we mention previous work that is related to this objective. We address respectively embodied conversational agents (ECA) that are built for the application of a museum, and ECAs that especially try to engage their users using verbal behaviour strategies.

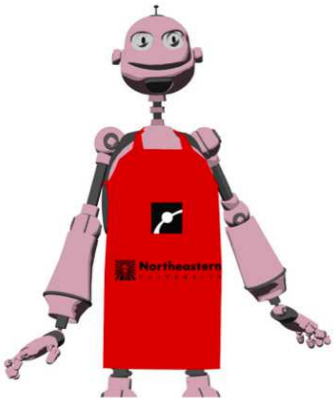
4.1 ECAs for Museums

Nowadays, museums form a popular application for conversational agents. One of the first conversational virtual agents that is successfully employed in a museum, and in a real-world application in general, is Max [Kopp et al., 2005]. Max is a human-sized conversational agent playing the role of a guide in a museum (Figure 4.1a). Its goal is to engage human visitors in conversations in which it provides the users with information about the museum or exhibition, or it conducts small talk. It can also play a guessing animal game with the user. Users give natural language input through a keyboard while Max produces synthesised speech and non-verbal behaviour. Both the agent and the user can take the initiative (i.e. bring up goals that the discourse pursues, such as the goal of coming to know the interlocutor's name). The agent has internal goals, intentions and an emotion system that runs a dynamic simulation to model the agent's emotional state. By evaluating the users' utterances in reaction to the agent they found that users seem willing to interact with the agent and try to be cooperative in answering its questions [Kopp et al., 2005].

An agent that resembles our project because it is both situated in a museum and focuses on the user's engagement, is Tinker [Bickmore et al., 2013]. Tinker, shown in Figure 4.1b, is a human-sized anthropomorphic robot that displays verbal and non-verbal behaviours. It describes exhibits in the museum, gives directions, and

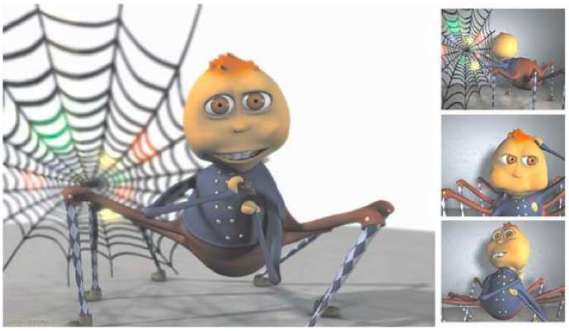
discusses technical aspects of its implementation. It uses a model of the user-agent relationship and tries to establish social bonds with users, continued interaction and repeated visits by using relational behaviour [Bickmore et al., 2013]. This relational behaviour consists of expressions of empathy, questions to the user for acquaintance, reciprocal self-disclosure, agreement to the user's (dis)likes, humour, addressing the user by its name, and expressions of liking the user and the interaction. User utterances are selected through a multiple-choice touch screen. The use of relational behaviour has lead to significantly more engagement of museum visitors (measured as total interaction time and the number of conversations) and greater learning gains of museum visitors (retention of information that Tinker gave the visitors) [Bickmore et al., 2011, Bickmore et al., 2013].

(a) Max [Kopp et al., 2005]



(b) Tinker [Bickmore et al., 2011]

(c) Ada & Grace [Swartout et al., 2010]



(d) Carletto [Damiano et al., 2008]

Figure 4.1 : Museum Agents

[Swartout et al., 2010] developed 2 virtual museum guides, twins called Ada and Grace. The agents have an educational goal and are able to answer a variety of spoken questions about science and technology, museum exhibits, and themselves [Traum et al., 2012]. The agents' verbal behaviour is based on a large collection of human visitor questions and human responses, as well as previous experiences with virtual characters [Swartout et al., 2010]. A voice actor recorded the agents' utterances that were used as the basis for the animations [Swartout et al., 2010]. An evaluation study showed that human participants give an overall rating of 3 ("pretty good") on a scale of 4 from "boring" to "exiting", with respect to "interacting with the exhibit" and "learning more about computers by interacting with the twins" [Traum et al., 2012].

Besides non-mobile agents there are also multiple museum guides that literally guide the human visitors to sites in the exhibition. An example is Carletto, a virtual spider with an anthropomorphic aspect that is used for guided tours on mobile devices [Damiano et al., 2008]. Because of the focus on its mobility the agent differs much from the projects above; Carletto adapts the type and quantity of informative items to the user's location, the duration of the stay in each location, and the interaction history, but the agent does not handle speech by the user. Carletto itself uses dramatic monologue and audiovisual output [Damiano et al., 2008].

A similar project is the Fun Robotic Outdoor Guide, FROG [Evers et al., 2014]. This robot reacts to the affective states of the users and offers location-based services. FROG disposes of audio, video display, movement, projection, possible light and gestural expressions. [Evers et al., 2014] planned to implement displaying content in response to the user's engagement. User engagement is measured by affect detection through facial feature extraction [Evers et al., 2014]. However, while in this way the user is taken into account, to our knowledge the user cannot actively contribute to the interaction, for example by means of speech.

Another example of a guiding robot is described in [Shiomi et al., 2006]. They developed multiple robots that guide people to exhibits randomly and are able to say a few things to the user, though without hearing what the user says. In order to interest the user in exhibits that he/she has not yet visited, the robots have conversations with each other about these exhibits which the user can overhear.

There is also a mobile art eating virtual monster that enhances a museum exploration game for children [Rehm and Jensen, 2015]. The agent-based version of the game made children spend more time with the artworks (what [Rehm and Jensen, 2015] refer to as engagement) than the paper-based version. They made two ver-

sions of the agent, which they plan to evaluate in a future study: one with only task-related verbal behaviours, and one that also has relational verbal behaviours such as self-disclosure, reference to user disclosure, empathy and humour [Rehm and Jensen, 2015].

We have seen in this section that there exist numerous conversational agents that are placed in museums, with varying interaction capabilities. In all these applications the agent plays a type of guide, while in our work based on project A1:1, the agent plays the role of a museum *visitor* (Section 1.2).

4.2 VERBALLY ENGAGING ECAS

While all ECAs try to engage the user in some way to achieve the goal for which they are built, fewer agents are built with favouring the user's engagement during the interaction as (one of its) main purpose(s). Besides, often work in human-agent interaction where engagement is taken into account, is oriented towards the *detection* of user engagement, such as [Nakano and Ishii, 2010, Forbes-Riley et al., 2012, Le Maitre and Chetouani, 2013, DeFalco and Baker, 2013, Castellano et al., 2014, Corrigan et al.,]. We are however particularly interested in the *generation* of verbal behaviour strategies by an agent to appear engaged and engage the user. We do not refer here to engagement with a task but engagement in the human-agent interaction itself. Here we mention ECAs that are related to this objective.

As mentioned above, the museum agent Tinker [Bickmore et al., 2013] (Figure 4.1b) uses what they call relational behaviour consisting of expressions of empathy, questions to the user for acquaintance, reciprocal self-disclosure, agreement to the user's (dis)likes, humour (scripted jokes), addressing the user by its name, and expressions of liking the user and the interaction [Bickmore et al., 2013]. Use of these relational behaviours that are all manifested by the content (Section 3.2.3.3) of the agent's verbal behaviour has lead to more user engagement in sense of total interaction time and the number of conversations.

REA (Figure 4.2a) acts as a real estate salesperson, answering user questions about properties in its database and showing users around the virtual houses [Cassell et al., 1999, Bickmore and Cassell, 2005]. In [Cassell et al., 1999], with respect to the agent's verbal behaviour REAs responses are generated from an engine that mirrors features of the user's last utterance, it can give feedback, allows interruptions by the user and has greeting and farewell functions [Cassell et al., 1999]. It can also initiate error correction when it misunderstands the user in [Bickmore and Cassell, 2005].

These features may favour the user's engagement as described in Section 3.2.3.3, but while engagement has not yet been taken explicitly into account in [Cassell et al., 1999], in [Bickmore and Cassell, 2005] this goal becomes more prominent in REA:

In [Bickmore and Cassell, 2005] REA's dialogue planner represents the relationship between the agent and the user by using a model of interpersonal relationship, based of the dimensions familiarity, solidarity and affect. Based on this relational model the discourse planner can interleave social talk (small talk) and task talk. Social dialogue (small talk) is talk in which interpersonal goals are more important than task goals, if existent, and serves to build rapport and trust [Bickmore and Cassell, 2005]. REA selects a dialogue move (engage in small talk and what kind of small talk) based on the interpersonal closeness with the user, topic coherence, the topics it thinks the user knows about, task goals, and logical preconditions. In this way REA decides to do small talk whenever closeness with the user needs to be increased. It adapts the content of the interaction to favour the user's engagement (Section 3.2.3.3).

[Bickmore and Picard, 2005] developed a FitTrack exercise advisor agent called Laura. Its verbal behaviour includes social dialogue, empathy dialogue, meta-relational communication, humour, continuity behaviours to bridge the time between multiple interactions with the same user, form of address (as a friend or stranger) and politeness strategies (indicative of small or large social distance). Use of these relational behaviours regarding form, timing and content levels (Section 3.2.3.3) resulted in increased quality of the working relationship and higher desires to continue interacting with the agent [Bickmore and Picard, 2005], which is an element of user engagement according to Poggi's definition (Section 2.3 Def. 18).

Favouring rapport is also likely to favour engagement (see Section 3.1.5). [Papangelis et al., 2014] developed a dyadic architectural model of rapport between virtual agents and human users. The model enables an ECA to build, maintain, and end rapport over multiple interactions with the same user. A rapport strategy is selected by taking into account the dyadic state and a model of the user and is based on reinforcement learning. In [Papangelis et al., 2014] no further details have been given about the nature of these rapport strategies.

[Sidner et al., 2005a] have developed Mel (Figure 4.2b), a conversational penguin robot that acts as a host to a human visitor participating in a demo in a laboratory. It is endowed with gestures, gazing behaviour, and verbal behaviour. The robot adapts its behaviour to the user as it interrupts the conversation about the demo when the user fails to take a turn as this is considered as an indication of the desire

(a) REA [Bickmore and Cassell, 2005]

(b) Mel [Sidner et al., 2005a]

Figure 4.2 : Engaging ECAs

to disengage. In this case the robot asks the user if he/she wants to continue. If not, the demo will end by a closing sequence. (To further measure the user's engagement they used interaction time, amount of mutual gaze, talk directed to the robot, overall looking back to the robot, and how closely in time the user tracked the robot's pointing [Sidner et al., 2005a].)

As in [Sidner et al., 2005a], in [Sidner et al., 2005b] engagement refers to processes by which participants establish, maintain, and end the perceived connection (Section 2.3 Def. 12). [Sidner et al., 2005b] therefore describes verbal engaging behaviour of Mel that is handled by the Collagen tool deciding when to begin engagement (i.e. the interaction), whether engagement is succeeding or failing, and when to end engagement (i.e. the interaction). Besides, engagement rules can introduce new agent goals. For example, when the user is disengaging Collagen can introduce a new goal to re-engage the user.

In the Always-On project a relational robotic agent (Reeti) is developed to reduce the isolation of isolated older adults [Sidner et al., 2013]. It tracks the engagement of the user (in the sense of initiating and concluding the interaction). Based on the closeness between the agent and the user, the agent determines which activities are appropriate to suggest to the user each time the agent and user interact, thereby adapting the content of the agent's verbal behaviour. Closeness increases when the agent and user do activities together and decreases when they have not interacted for a period of time [Sidner et al., 2013].

[Bohus and Horvitz, 2009b] describe a system for an embodied agent that senses

and reasons about the engagement, actions and intentions of multiple users, decides whom to (dis)engage with and when, and executes and signals these decisions to the interaction participants. The agent's engagement actions consist of engage, disengage, maintain and no-action. The system realises these actions by head gestures, gaze, facial expressions, salutations, and interjections. For example the engage action can consist of the sequence: gaze towards the users, trigger an interjection like "excuse me" and a greet. By means of an observational study they have found that the system can provide support for effectively managing engagement [Bohus and Horvitz, 2009b].

[Bohus and Horvitz, 2014] developed a model that forecasts when a user is about to terminate his/her interaction (disengage) that then guides a disengagement policy that uses linguistic hesitation actions such as filled and non-filled pauses. The hesitations give the system more time to collect disengagement evidence and leave the option open for continuing or finishing engagement. They found that the joint use of forecasts and hesitations can help manage disengagement and that hesitations policies help agent systems to engage in more seamless and natural interactions [Bohus and Horvitz, 2014].

In conclusion, ECAs have adapted their verbal behaviour in a range of different ways to the user in order to engage the user in the interaction. Some agent behaviours are specifically oriented towards initiating or ending the interaction, thereby starting and ending engagement. Systems can, for example, decide when [Sidner et al., 2005b] and how [Bohus and Horvitz, 2014] and with whom [Bohus and Horvitz, 2009b] to start and end an interaction. Other systems [Bickmore et al., 2013, Bickmore and Picard, 2005, Bickmore and Cassell, 2005, Sidner et al., 2013] use verbal behaviours that can favour engagement during the interaction; strategies to maintain or repair engagement. Most systems do not consider a single subtle behaviour to achieve the favouring of the user's engagement; they employ either a wide range of behaviours (e.g. [Bickmore et al., 2013, Bickmore and Picard, 2005]), or behaviour that modifies the entire course of the interaction by changing (temporally) the topic or goal of the interaction (e.g. [Bickmore and Cassell, 2005, Sidner et al., 2013]). While such approaches can result in more engaging agents than agents without these behaviours, the effect of individual behaviours may not always be clear.

In our work we focus on agent behaviour to favour the user's engagement during the ongoing interaction. We try to contribute to the existing knowledge about user engagement by exploring possible new strategies of engaging verbal agent behaviour.

We will do this by respectively considering an aspect of: the form of the agent's verbal behaviour, namely the politeness of the agent (Chapter 5); the timing of the agent's verbal behaviour by looking at interruptions (Chapter 6); and the content of the agent's verbal behaviour by considering the topics of the interaction and the user's preferences (Chapter 7).

PART IV : DISCOVERING STRATEGIES TO FAVOUR ENGAGEMENT

TELECOM ParisTech

5 Engagement and Politeness

5.1 Introduction

For a range of applications we would like embodied conversational agents to engage their users. For reminder, we consider engagement as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction" [Poggi, 2007]. As we have set out in the previous chapters, numerous recent studies describe how an agent can influence user engagement by coordinating and synchronising its behaviour to that of its user. Such engagement related behaviour includes non-verbal behaviours such as gaze [Sidner et al., 2005b], gestures, postures, facial expressions [Delaherche et al., 2012], and various verbal behaviours [Pickering et al., 2004] (Chapter 3). One of the verbal aspects that can be coordinated with the user is the degree of expressed politeness [De Jong et al., 2008] (Section 3.2.3.1). [En and Lan, 2012] state that a successful implementation of politeness maxims is likely to improve human-agent engagement. In [Bickmore and Picard, 2005] relational behaviour including politeness strategies (indicative of either small or large social distance) resulted indeed in higher desires to continue interacting with the agent [Bickmore and Picard, 2005].

In order to favour the user's engagement as much as possible, in this chapter we try to gain more insight into the optimal coordination of politeness. Where we look in this chapter at (an aspect of) the form of the agent's verbal behaviour (how to say something), in the chapter 6 we will look at the influence of timing of the agent's verbal behaviour (when to say something) and in chapter 7 at the content of the agent's verbal behaviour (what to say).

In order to gain more insight into the optimal coordination of the agent's politeness with the user we have conducted a perceptive study. In this study we examine if there is a link between the speaker's perceived engagement level of the hearer, and the speaker's politeness strategies. The study examines this link in human-human interaction because we aim at modelling agent behaviour that is as human-like as

possible (Chapter 1). The outcomes of this study provide us an indication about how to model the politeness behaviour of a human-like virtual agent. They tell us namely, whether or not a human-like agent who wants to continue the interaction with its user (thereby maintaining user engagement) needs to speak with more caution to someone who is less engaged than to someone who is very engaged in the ongoing interaction.

Figure 5.1 shows this hypothesis: we verify if the speaker's perceived engagement level of the hearer (agent's perception of user engagement) influences the politeness strategies of the speaker (agent). The politeness strategies that the speaker uses (agent) may in their turn influence the engagement level of the hearer (user engagement).

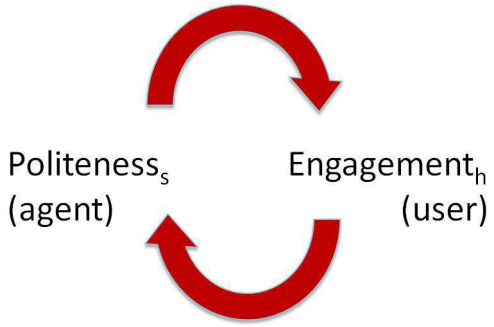


Figure 5.1 : Engagement of the hearer (h, user) may influence the politeness of the speaker (s, agent) influencing the user's engagement.

Seen from the perspective of Brown and Levinson's politeness theory [Brown and Levinson, 1987] on which this study is based, we have tested the hypothesis that the speaker's assessment of the hearer's level of engagement has an impact on the politeness level that the speaker employs in addressing the hearer.

Below we first explain Brown and Levinson's politeness theory. In Section 5.3 we then formulate our hypothesis in more detail, and in Section 5.5 we describe the methodology. This is followed by the results (Section 5.6), the discussion (Section 5.7) and the conclusion (Section 5.8) of this study.

5.2 Politeness Theory

[Brown and Levinson, 1987]'s politeness theory is about saving "the public self-image that every member" in an interaction "wants to claim for himself", which is referred to as this person's 'face'. The concept of 'face' consists of a negative face

which is "the want of every competent adult member that his actions be unimpeded by others"; and a positive face which is "the want of every member that his wants be desirable to at least some others".

According to [Brown and Levinson, 1987] some acts intrinsically threaten face, referred to as Face-Threatening Acts (FTAs). FTAs can be categorised into threats to the addressee's positive face, such as expressions of disapproval, criticism, and disagreements, and threats to the addressee's negative face, such as orders, requests and suggestions.

The speaker of an FTA can try to minimise the face-threat by employing a set of strategies [Brown and Levinson, 1987]. These politeness strategies describe different ways of how the speaker can place the FTA, and are the following:

1. Without redressive action, baldly
2. With positive politeness (i.e. oriented toward the positive face of the hearer)
3. With negative politeness (i.e. oriented toward the negative face of the hearer)
4. Off record (i.e. indirect hints as to what the speaker means)
5. Don't do the FTA.

[Brown and Levinson, 1987] state that "roughly, the more dangerous the particular FTA x is, in the speaker's assessment, the more he will tend to choose the higher numbered strategy". The other way around, this implies that the higher the number of the chosen strategy that is used to place an FTA, the more dangerous the FTA x is, in the speaker's assessment.

W_x, the numerical value that measures the weightiness, i.e. danger, of the FTA x is calculated by [Brown and Levinson, 1987]:

$$W_x = D(S, H) + P(H, S) + R_x \tag{5.1}$$

where $D(S, H)$ is the social distance between the speaker and the hearer, $P(H, S)$ the power that the hearer has over the speaker, and R_x the degree to which the FTA x is rated an imposition in that culture. The distance and power variables are intended as very general pan-cultural social dimensions [Brown and Levinson, 1987]. So when the social distance between the speaker and the hearer, the power that the hearer has over the speaker, and/or the degree of the imposition increase, the FTA is more dangerous, and the speaker will tend to use a higher numbered politeness strategy.

5.3 Adding to the FTA-weight Formula

Besides a very general pan-cultural distance between the participants in an interaction as described by $D(S, H)$ in Equation 5.1, in our view, the level of engagement can be seen as a measure of distance in the interaction as well. This is because, considering the definition of engagement we use [Poggi, 2007], a low level of engagement implies a temporally small value to *continue the interaction and be together with the other interaction participant(s)* and vice versa. This distance that is captured by engagement may be comparable with [Brown and Levinson, 1987]'s distance variable $D(S, H)$, only this time the distance measure has a more temporal and dynamic nature, instead of describing a general pan-cultural social dimension. We therefore hypothesise that, just as the social distance $D(S, H)$, the temporal and dynamic distance that is described by engagement also plays a role in the perceived weight of an FTA. This hypothesis can be formulated by adding the engagement variable to [Brown and Levinson, 1987]'s equation 5.1:

$$W_x = D(S, H) + P(H, S) + R_x - Eng_h \tag{5.2}$$

where Eng_h is the engagement level of the hearer. As for the other variables in the equation, we do not talk about an absolute level of the dimension, but of the *speaker's perception* of the hearer's engagement at the moment the FTA x is placed. Where the power and distances variables are unlikely to vary much during an interaction, the engagement of the hearer can vary much among different FTAs in the same interaction which reflects the temporal and dynamic distance that is described by engagement.

5.4 Related Work

To our knowledge no previous work has looked into the possibility of taking into account the hearer's engagement in estimating the speaker's perceived weight of a face-threatening act, nor in the modelling of the agent's politeness behaviour according to the hearer's engagement.

[Andre et al., 2004] modelled an agent that instead of the perceived engagement of the user takes into account the perceived emotions of the user in adapting its politeness strategy. They added the perceived emotions of the user as a variable in [Brown and Levinson, 1987]'s equation 5.1 as follows:

$$\Theta_E = E(H) * \frac{1}{3}(D(S, H) + P(H, S) + R\alpha) \tag{5.3}$$

where Θ_E is the threat (what [Brown and Levinson, 1987] refer to as the Weight of the threat W_x), $E(H)$ is the emotional state of the hearer, and α is the speech act (what [Brown and Levinson, 1987] refer to as the face-threatening act (FTA) x). In this way [Andre et al., 2004] consider the hearer's emotional state as a situational factor that influences the other variables of the equation. They however did not evaluate this assumption nor the model itself. In our hypothesis (Eq. 5.2) we do not consider the hearer's engagement as a variable that modifies the very general, rigid social distance ($D(S, H)$) and power ($P(H, S)$) dimensions, nor the stable degree to which FTA x is rated an imposition in that culture (R_x), but consider the speaker's perception of the hearer's engagement (Eng_h) as an additional dimension that may play a role in the perceived weight of an FTA.

Other work on politeness in human-agent interaction includes [De Jong et al., 2008] who described a model for the alignment of formality and politeness in a virtual guide. The model generates agent utterances whose politeness is (in adjustable extents) aligned with the user's expressed politeness. An evaluation of written dialogue fragments with different agent alignment settings did not show overall preferences for a particular politeness alignment setting. "[...] some participants liked the way the guide mirrored the user in the strong alignment settings, whereas others liked the fact that in the weak alignment versions, the guide stayed formal and polite even when the user was not" [De Jong et al., 2008].

As mentioned in Section 4.2, the FitTrack ECA [Bickmore and Picard, 2005] incorporates a range of relational behaviours, including politeness strategies. It uses politeness strategies that are indicative of either small or large social distance ($D(S, H)$ in Eq. 5.1) without providing further details about the way these strategies are modelled. Use of the entire range of relational behaviours has lead to increased quality of the working relationship and higher desires to continue interacting with the agent [Bickmore and Picard, 2005].

5.5 Methodology

From [Brown and Levinson, 1987]'s theory, and as explained in Section 5.2, we can derive that a way to infer the perceived threat of an FTA is by looking at the politeness strategy that is employed to place the FTA. So in order to find out if the perceived threat of an FTA is indeed dependent on the perceived engagement level of the hearer as described in Section 5.3, we can look at the strategies that are used to place the same FTA in circumstances where only the perceived engagement level

of the hearer differs.

For this we create two dialogue fragments of two humans interacting (conditions) between which the employed politeness strategies could be compared; one dialogue fragment in which a participant appears (highly) engaged (expressing the desire to continue the interaction and be together with the other interaction participant) and another in which he appears less engaged (expressing much less desire to continue the interaction and be together with the other).

To control the conditions, i.e. to ensure that the 'hearer' of the FTA demonstrated the desired levels of engagement but all other variables of the interaction were kept as constant as possible, we scripted written interactions. Human third party observers are presented with the written interactions (as also in e.g. [De Jong et al., 2008, Mayer et al., 2006]). They judge for the interaction contexts which utterance (realising a politeness strategy) they find most appropriate for the speaker to place the FTA.

We have modelled an engaged and less engaged condition for each of three different FTAs. We looked at only three FTAs in order to limit the length of the perceptive study for the evaluation participants. The FTA types are chosen according to the context of this research: Building a conversational virtual agent that represents a visitor in a French museum. We looked at the FTAs: disagreement, in the preference for a painting; suggestion, to have a look at some other object; and request, for advice about what to see next.

The design of our experiment consisted of two steps which we shall discuss respectively in the sections below: 1) The design of a collection of utterances that represent the politeness strategies among which human judges could choose the most appropriate (Section 5.5.1 and 5.5.2); and 2) the design of the two different conditions (dialogue fragments) in which the strategies needed to be chosen (Section 5.5.3).

5.5.1 Utterances with Politeness Strategies

In order to obtain a set of utterances that represent the politeness strategies that were to be used in the final evaluation study we created a first set based on the literature and selected a subset of these by means of a validation study.

So first of all, following [Brown and Levinson, 1987]'s tactics of formulating politeness strategies and taking inspiration from [De Jong et al., 2008], we constructed a maximal set of French formulations (utterances) for each FTA. See for an example

Table 5.1 for the FTA 'disagreement', where the first three columns from left to right show respectively the politeness strategy that is underlying the utterance, the tactic that is used to realise this strategy, and the formulation to realise this tactic.

The pronoun to address the hearer was kept constant over all utterances to the less formal version of *you* (*tu*). We did not design utterances with a mixture of strategies as this is a delicate matter that may even cause "painful jerks" instead of an accumulating politeness [Brown and Levinson, 1987].

Table 5.1 : Utterances constructed to place the FTA 'disagreement', using different politeness strategies (Section 5.5.1). Column 4 shows the results of the validation procedure (Section 5.5.1.1), listing the average scores of overall politeness (q2 and q5). Column 5 indicates the utterances that are selected for the final experiment and their ranking within this final experiment (Section 5.5.2).

Politeness Strategy	**Tactic**	**Sentence**	**Average politeness (Q2,5)**	**Rank nr.**
Bald on record	Bald on record	**1.** Moi je n'aime pas cette peinture. *I don't like this painting.*	3.73 2.82	1
Positive politeness	Attend to the hearer's interests, wants, needs	**2.** Je comprends pourquoi tu aimes cette peinture, mais moi je ne l'aime pas. *I understand why you like this painting but I don't like it.*	5.73 5.36	-
	Be optimistic	**3.** Je pense que ça ne te dérange pas si je te dis que je n'aime pas cette peinture. *I think you won't mind if I tell you I don't like this painting.*	4.45 4.45	-
	Give reasons	**4.** Je n'aime pas cette peinture car l'art abstrait n'est pas pour moi. *I don't like this painting because abract art is not for me.*	4.45 3.82	-
Negative politeness	Nominalise	**5.** Le fait est que je n'aime pas cette peinture. *The thing is that I do not like this painting.*	4.45 3.45	2

	Give deference, humble oneself	**6.** Je ne suis pas expert en art abstrait mais je n'aime pas cette peinture. *I am not an expert of abstract art but I do not like this painting.*	5.09 4.64	-
	Be conventionally indirect	**7.** Je crois que je n'aime pas cette peinture. *I think I do not like this painting.*	4.82 5.27	3
	Minimise imposition	**8.** Je ne suis pas sûre d'aimer cette peinture tant que ça. *I am not sure I like this painting that much.*	5.73 5.73	5
	Apologise	**9.** Je suis désolée mais je n'aime pas cette peinture. *I am sorry but I do not like this painting.*	5.73 5.09	-
Off Record	Over generalise	**10.** En général l'art abstrait n'est pas pour moi. *In general, abstract art is not for me.*	4.91 4.64	-
	Understate	**11.** Cette peinture n'est pas laide. *This painting is not ugly.*	4.82 4.73	-
	Contradictions	**12.** J'aime bien et je n'aime pas à la fois. *I like and I don't like it.*	4.91 4.91	4

While theoretically the utterances we constructed in this way can be ranked according to their potential of minimising the FTA's risk in the way [Brown and Levinson, 1987] proposed (see Section 5.2 and the order of strategies in Table 5.1 Column 1), in practice [Brown and Levinson, 1987]'s proposed hierarchy is not always entirely respected [De Jong et al., 2008, Andre et al., 2004]. To deal with this issue we validated the utterances, as we shall describe in the two following paragraphs. Only when we know certain how to rank the strategies we are able to compare the use of these strategies between interactions with an engaged hearer and a less engaged hearer.

5.5.1.1 Politeness Strategy Validation: Methodology

To validate the potential of minimising the face threat (i.e. politeness) of each formulated utterance that could be used in our final experiment (e.g. Table 5.1) we performed a questionnaire-based survey.

This questionnaire consisted of three parts, corresponding to the three different FTAs. Appendix A Figure A.1 shows a fragment of one of these parts. Every part first introduced the context in which the formulations were supposed to be uttered (*Two young women in a museum meet thanks to a mutual friend.*), as well as the intention of the speaker which describes the FTA: expressing a contradictory opinion, making a suggestion to have a look at a painting, or asking for advice about what to see next. By specifying in this way the setting of the interaction, the relation between the interaction participants, and the intention of the speaker, we avoided any speculation regarding the distance ($D(S,H)$) and power ($P(H,S)$) between the interaction participants, as well as the (ranking of) the intended face threat (R_x); all the variables from [Brown and Levinson, 1987]'s formula (Equation 5.1).

Below each introduction we listed all the corresponding utterances. For the FTA 'disagreement' this means all the utterances from Table 5.1, Column 3.

Every utterance in the questionnaire was followed by a set of 6 questions regarding their plausibility (q1., yes or no) and politeness level. For the latter we asked respectively 3 questions that have been used in previous work to estimate a degree of politeness, and 1 question that is derived from [Brown and Levinson, 1987]'s description of politeness strategies as minimising the threat to the addressee's face (Section 5.2): We asked for rankings on a 7 point scale regarding respectively:

q2. The utterances' politeness level directly [De Jong et al., 2008];

q3. The degree to which the speaker allowed the hearer to make his own decision [Mayer et al., 2006] (negative politeness);

q4. The degree to which the speaker wanted to work with and appreciated the hearer [Mayer et al., 2006] (positive politeness);

q5. The degree to which the speaker spared the needs or face of the hearer. [1]

We made sure that the order in which the FTAs were listed fluctuated between participants and that each subject was presented with a unique order of utterances. Also repetition of utterance sequences and positions was avoided.

1. Literally *ménager la susceptibilité.*

5.5.1.2 Politeness Strategy Validation: Results

13 subjects participated in the validation study: 5 female, all native speakers of French, aged 23-40. ANOVAs for repeated measures conducted on the ranking of overall politeness (resp. q2 and q5) showed that for each FTA the utterances differ significantly from each other (disagreement $F = 4.07$, $F = 5.84$; suggestion $F = 7.19$, $F = 8.01$; request $F = 14.38$, $F = 13.32$; $p < 0.01$). Our results confirm other studies [De Jong et al., 2008, Andre et al., 2004] that [Brown and Levinson, 1987]'s ranking of politeness according to their strategies is not completely respected. Similar to observations by [De Jong et al., 2008] indirect (off-record) strategies were rated much less polite than expected. The answers to validation questions q3 and q4 did not play a role in the estimation of every utterance's overall politeness level but might serve to provide more insight into the outcomes of the final experiment.

5

5.5.2 Utterance Selection

From the utterances tested in the validation experiment a subset for each FTA was selected for the final experiment. For reminder, in the final evaluation, participants will recommend utterances from this subset to place an FTA. From the strategies that are recommended we then derive the comparative perceived weight of the FTA.

The first step in the selection process of utterances for the final evaluation is eliminating those utterances that were judged more than once as implausible in the sense of an infrequent oral formulation (q1). Among the remaining utterances we selected one utterance for every clearly distinguishable observed level of politeness. We do this because, in the final study, in order to estimate if the weight of the FTA is different among the conditions, the utterances that are used to indicate these weights should have different rankings of politeness. To select the utterances that convey different levels of politeness we looked primarily at the mean and standard deviation of the utterances' scoring on perceived overall politeness (q2, for FTA 'disagreement' averages are listed in Column 4 of Table 5.1). However, which utterances are selected exactly for the final evaluation, does not matter, as long as they convey relatively different levels of politeness.

Because in the final evaluation we verify only if the perceived weight of the FTA is different among the conditions and are not calculating exact perceived weights of the FTA, we do not further require the exact politeness level of the utterances

obtained by the validation study. Instead, for the final evaluation study, we only use the ranking of the selected utterances that represent the politeness levels relative to one another. Table 5.1, Column 5 shows the utterances and rankings that are selected for the final evaluation for the FTA 'disagreement'.

The performed validation procedure could not take into account [Brown and Levinson, 1987]'s heaviest risk minimising strategy of not doing the FTA at all. However, considering the nature of this strategy as complete avoidance of the FTA, we indeed assumed that there is no strategy heavier than this one and added it as such to the strategies that are used for the final experiment. E.g. for the FTA 'disagreement' (Table 5.1) this means that the strategy *not expressing his opinion about the painting* was added with ranking number 6.

5.5.3 Engagement Conditions

The procedure described in the two sections above had provided the politeness strategies (utterances) that were to be tested for appropriateness in interactions with an engaged and a less engaged hearer. As a second step we constructed for each FTA the two versions of dialogue (i.e. conditions, interactions) that provided the contexts for these strategies.

The contributions of the speaker of the FTA, say Person A, stayed constant over both conditions, while the contributions of the hearer, say Person B, differed considerably in order to communicate a high or low level of engagement. Table 5.2 shows the two conditions (contexts) we designed to place the FTA 'disagreement'.

Table 5.2 : The two scenarios preceding the FTA 'disagreement'. Pauline's (Person A) utterances are the same among the two conditions. In one of the two conditions Charlotte (Person B) shows more engagement (utterances following '+'), and in the other she shows less engagement (utterances following '-').

Person	**Eng. level**	**Utterance**	**Engagement strategy**
Pauline:		Bonjour. Je suis Pauline. Ravie de te rencontrer ! *Hello. I am Pauline. Nice to meet you!*	
Charlotte :	-	Bonjour, Charlotte. *Hello, Charlotte.*	

	+	Bonjour. Je suis Charlotte. Ravie de te rencontrer aussi. *Hello. I am Charlotte. Nice to meet you too.*	Extend reaction (Bickmore et al., 2013), expression of liking the other (Bickmore et al., 2013).
Pauline:		Heureusement vous avez réussi à venir aujourd'hui. Caroline m'a dit que vous avez eu des problèmes de transport. *Luckily you were able to make it today. Caroline told me you had some transport problems.*	
Charlotte:	-	Oui. *Yes*	
(Pauline):		Qu'est-ce qui s'est passé ? *What happened?*	
(Charlotte):		Le train avait du retard. *The train was late*	
Charlotte:	+	Oui, c'est vrai ! Notre train avait une heure de retard. *Yes, that's true! Our train was one hour late.*	Extend reaction (Bickmore et al., 2013).
Pauline:		Oh c'est gênant ! Mais, aujourd'hui le musée est ouvert jusqu'à 22 heures donc il n'y a pas de problème. *Oh that's annoying! But today the museum is open until ten o'clock so no problem.*	
Charlotte:	-	Oui. Yes.	
	+	Oui, on a de la chance. *Yes, we are lucky.*	Extend reaction (Bickmore et al., 2013), expression of emotion (Peters et al., 2005).
Pauline:		C'est la première fois que tu viens ? *Is this the first time you come here?*	
Charlotte:	-	La deuxième. *Second.*	

	+	Non, c'est la deuxième fois et toi ? *No, this is my second time, and you?*	Extend reaction (Bickmore et al., 2013), show interest in the other (Peters et al., 2005).
Pauline:		Moi c'est ma première visite. *For me it's my first visit.*	
Charlotte:	-	-	
	+	Ah oui ? J'adore ce musée : je l'ai déjà visité l'année dernière. *Really? I love this museum. I've already visited it last year.*	Add feedback (Gratch et al., 2006), expression of emotion (Peters et al., 2005).
(Pauline):		Ah c'est super ! *That's great!*	
Pauline:		Pour l'instant je n'ai vu qu'une seule partie du musée, mais j'ai beaucoup aimé. *At the moment I've only seen part of the museum, but I like it a lot.*	
Charlotte:	-	-	
	+	Ah bon ? *Really?*	Add feedback (Gratch et al., 2006).
(Pauline):		Oui ! *Yes!*	
Pauline:		Et toi tu as déjà tout vu ? *And you, have you seen everything already?*	
Charlotte:	-	Oui. *Yes.*	
	+	Oui. Moi, j'ai tout vu. *Yes. I have seen everything.*	Extend reaction (Bickmore et al., 2013).
Pauline:		Ah c'est bien. Et qu'est-ce que tu as aimé le plus ? *Ah that's nice. And what did you like most?*	
Charlotte:	-	Guernica. *Guernica.*	

+	Guernica, la grande peinture de Picasso. *Guernica, the big painting by Picasso.*	Extend reaction (Bickmore et al., 2013).

For the context in which Person B was minimally engaged her utterances were kept as brief and uninterested as possible. Her engagement level was just high enough to participate in the interaction so far [Poggi, 2007]. In the interactions where Person B was to demonstrate a high level of engagement we used cues that have been linked to engagement in former studies (Chapters 3 and 4) and which can be expressed in written text, as shown in Table 5.2 Column 4): We made Person's B reactions longer so as to extend the interaction time [Bickmore et al., 2013] (see Section 3.2.1.6), we added more feedback [Gratch et al., 2006] (see Section 3.2.1.4 and 3.2.3.2), added expressions of emotion [Peters et al., 2005b] (see Section 3.2.3.3) and of liking their interaction partner [Bickmore et al., 2013] (see Section 4.2), and showed interest in Person A [Peters et al., 2005b] (see Section 3.1.3).

Having created the utterances that convey different politeness strategies, as well as the conditions with an engaged and a less engaged hearer, we completed the design of the stimuli for the overall evaluation study.

5.5.4 Questionnaire

We presented human subjects with the stimuli by means of an online questionnaire. Each participant is given one interaction condition (engaged or less engaged Person B) for each of the three FTAs. We did not show the participants both conditions of one FTA because we did not want the participants to reflect about the differences between the conditions and try to compare and reason their answers. Instead, we tried to collect answers that are given as intuitively as possible, as would be the case in spoken face-to-face interaction as well.

Appendix A Figure A.2 shows a fragment of the online questionnaire containing the questions for one FTA. For each FTA, first the context was introduced by the same context description that was used for the validation procedure: *two young women meet in a museum through a mutual friend and start to talk.* Again, this description serves to fix the 'distance' ($D(S,H)$), 'power' ($P(H,S)$) and 'ranking' (R_x) variables from [Brown and Levinson, 1987]'s formula (Equation 5.1).

Below each context description followed the written dialogues with either an engaged or less engaged Person B (see Table 5.2). At the end of this dialogue fragment the participants were asked to recommend one of the utterances (each representing a politeness strategy) to Person A, under the instruction that this participant wanted to place the FTA (communicate his disagreeing opinion, do a suggestion, or ask for advice) but also wanted to absolutely continue the conversation with his interaction partner, Person B. Concretely, for the FTA 'disagreement' this question looks as follows (translated):

Q1. *Pauline does not like the Guernica (a painting by Picasso). But she absolutely wants to continue the conversation with Charlotte. Which of the following options would you advice Pauline at this moment in the conversation?*

(a) *Alright. I don't like this painting.*

(b) *Alright. The thing is that I do not like this painting.*

(c) *Alright. I think I do not like this painting.*

(d) *Alright. I like it and I don't like it.*

(e) *Alright. I am not sure I like this painting that much.*

(f) *Alright.*

Sentence (f) represents the politeness strategy of not doing the FTA at all.

Following Q1 we verified whether or not we had actually successfully communicated the high or low engagement levels of person B by asking the observers to indicate on a scale from 1 to 7:

Q2. The value that B attributed to the goal of being together with A;

Q3. The value that B attributed to the goal of continuing the interaction.

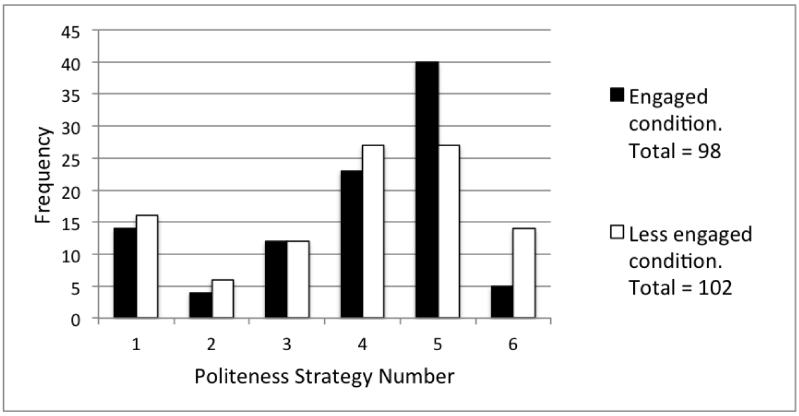
These are the direct measurements for engagement derived from the definition of engagement by [Poggi, 2007]. We also asked respectively for Person B's level of involvement (as interpreted by [Sidner et al., 2005b], referred to as 'engagement' by [Lombard and Ditton, 1997, Lombard et al., 2000]), rapport, and interest as these concepts are closely related to engagement (see Section 3.1):

Q4. To what extent the interaction seemed engaging for Person B [Sidner et al., 2005b];

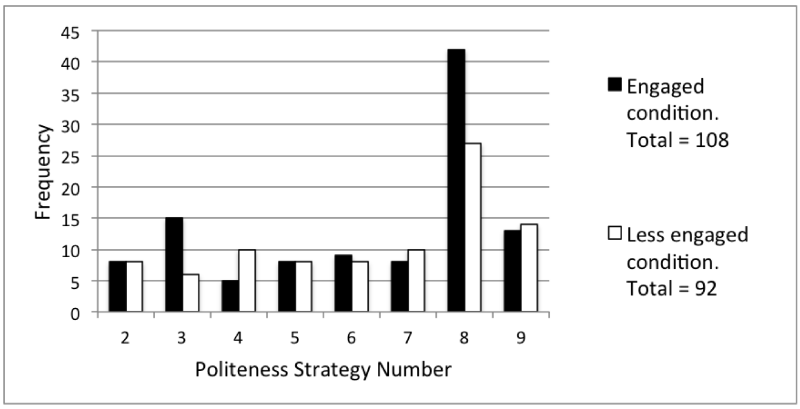
Q5. If Person A and Person B wanted to become friends [Ringeval et al., 2013];

Q6. If Person B seemed interested in the interaction (Section 3.1.3).

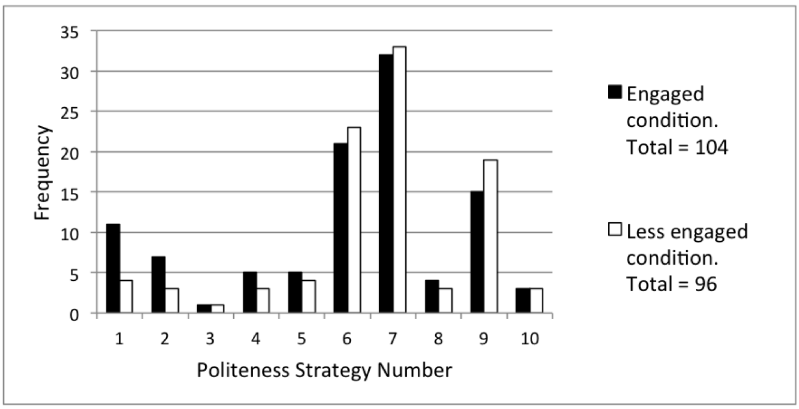
5.6 Results



(a) FTA disagreement



(b) FTA suggestion



(c) FTA request

Figure 5.2 : Distributions of the recommended politeness strategies (Q1).

200 subjects filled out our online questionnaire: 68.5% female, 100% native speakers of French (required for participation), aged 16-75.

For every FTA the participants perceived the hearer's engagement, involvement, rapport and interest levels significantly higher in the engaged condition than in the less engaged condition (t-tests $p < 0.01$, Q2 till Q6). Mann-Whitney U tests on each FTA separately, show however no significant overall differences with respect to the distributions of the recommended politeness strategies between the two conditions, which are shown in Figure 5.2 (Q2).

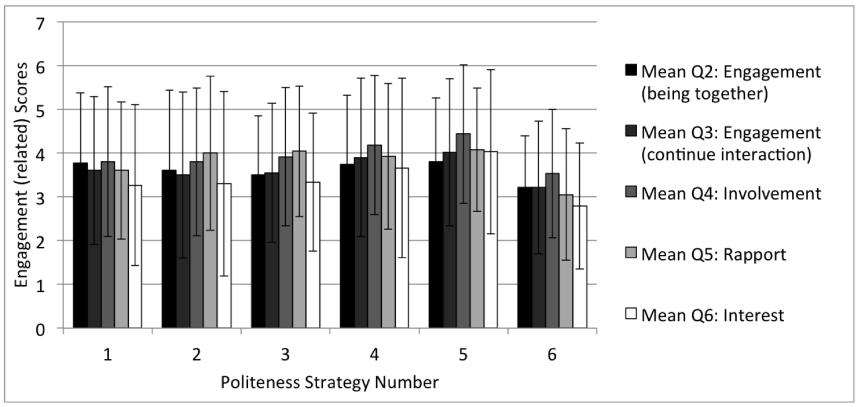
For an inter-participant analysis we took the data for both conditions of one FTA together and compared the rank of the participants' selected politeness strategies with the scores they attributed to Person B's engagement and related concepts (see Figure 5.3). Kendall Tau tests show significant negative correlations ($p < 0.05$), for the FTA 'request', between the rank of the chosen politeness strategy and the level of engagement ($\tau = -0.127$, Q2; $\tau = -0.111$, Q3), involvement ($\tau = -0.110$, Q4) and interest ($\tau = -0.107$, Q6). Similarly, the FTA 'suggestion' holds a significant negative correlation regarding the perceived level of involvement ($\tau = -0.109$, Q4).

5.7 Discussion

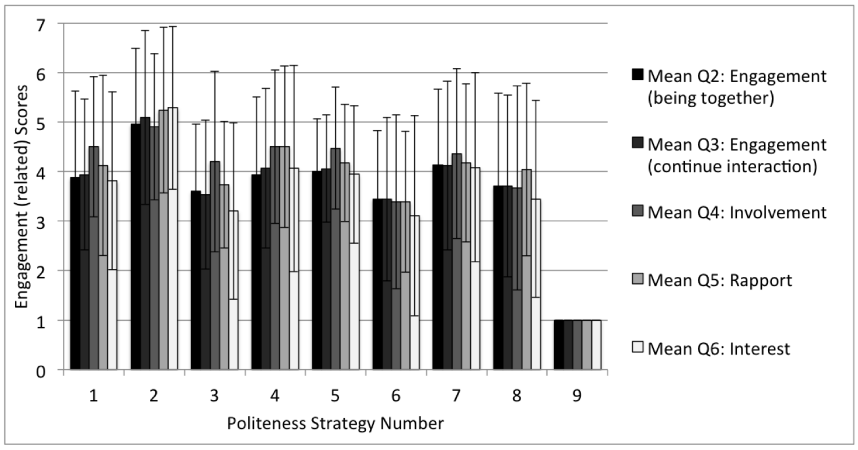
The fact that for every FTA, observers perceived the hearer's engagement, involvement, rapport and interest levels significantly higher in the engaged condition than in the less engaged condition, shows that we have successfully created two conditions for each FTA between which the perceived engagement levels of Person B are different. A pretest had already ensured that also the politeness utterances do what they should, namely representing different levels of politeness. This means that the overall stimuli indeed convey what we want them to convey.

The results we obtained with these stimuli however, do not demonstrate that the recommendation of politeness strategies differs overall between both conditions. The lack of such a clear overall difference confirms the idea that politeness is a highly subjective phenomenon [Danescu-Niculescu-Mizil et al., 2013].

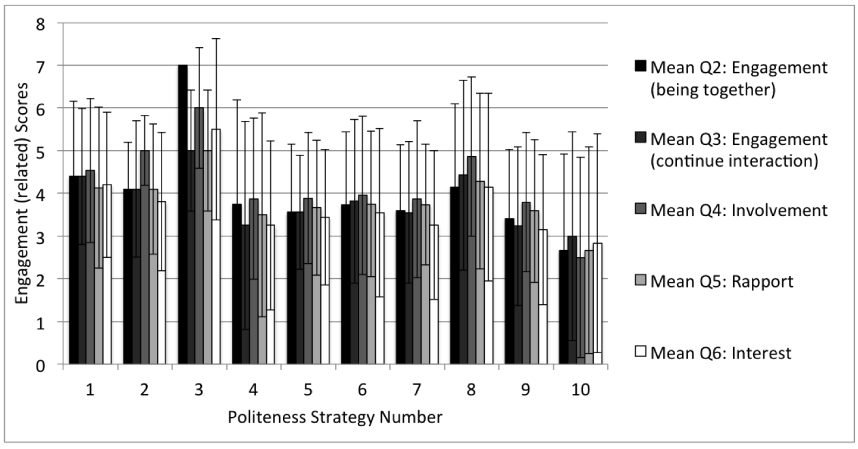
We therefore performed a finer grained, inter-participant analysis of the results. More specifically, we compared the ranking of a participant's recommended politeness strategy with the score this participant gave to person's B level of engagement and engagement related concepts. It must be noted that the data points are not equally distributed as not every politeness strategy was chosen with the same frequency. In this way significant negative correlations were revealed in the contexts



(a) FTA disagreement



(b) FTA suggestion



(c) FTA request

Figure 5.3 : Average engagement (related) scores per selected politeness strategy.

of the negative FTAs 'request' and 'suggestion'.

The only FTA in our experiment that is a threat to the addressee's positive face, namely 'disagreement', does not show such significant negative correlations. A possible explanation for this is that here such a tendency interferes with a preference for alignment. This is because a low level of engagement is conveyed by features that overlap with features that indicate positive impoliteness (e.g. "ignore or fail to attend to the addressee's interests" and "being disinterested and unconcerned" [Rehm and Krogsager, 2013]). Thus, the condition where Person B has a lower level of engagement may not only convey a lower level of engagement, but also convey that Person B is not attentive to Person A's positive face (Section 5.2), which implies positive impoliteness. Some people prefer strong alignment settings [De Jong et al., 2008] and may therefore be inclined to answer positive impoliteness (by Person B) with less caution for the addressee's positive face as well (by Person A). This tendency might interfere with an opposing preference to employ stronger politeness strategies when the addressee seems to have a lower level of engagement. This interference might explain why no significant results were found regarding a preference for alignment of politeness in [De Jong et al., 2008], nor for engagement adapted politeness in the present study.

For the negative FTA 'request' we have found negative correlations between the ranking of politeness strategies and the level of engagement, involvement and interest. Specific research towards 'rapport' will be needed to explain why rapport (as in [Ringeval et al., 2013]) does not show a similar correlation.

The other negative FTA 'suggestion' shows only one such negative correlation, namely between the politeness strategies and the perceived levels of involvement. This may be due to the fact that the FTA can be interpreted as not really face-threatening. While [Brown and Levinson, 1987] categorise suggestions as FTAs, they may be interpreted as a remark that is placed purely in the interest of the hearer, which removes its face-threatening aspect. To a smaller extent the correlations found for the FTA 'request' might be weakened due to the fact that asking someone for his advice on what to see next implies an interest of the speaker in the addressee's values and knowledge, thereby diminishing the threat.

Further it must be noted that we only considered verbal behaviour while non-verbal behaviour such as prosody, gaze and gestures can influence the way in which verbal behaviour is interpreted and can reveal a range of information about the person's attitude and perceptions [Sidner et al., 2005b, Peters et al., 2005b]. Non-verbal expressions of feedback and mimicry for example, can play a large role in

building and creating rapport [Gratch et al., 2006].

5.8 Conclusion

In this chapter we have verified the existence of a link between the speaker's perceived engagement level of the hearer, and the speaker's politeness strategies, in human-human interaction. We have done this by verifying if there is a difference between the weight of politeness strategies that speakers would use in interaction with an engaged person and in interaction with a less engaged person.

In the creation of stimuli representing the two conditions we have demonstrated a successful verbal behaviour model to convey a participant's engagement level. We have not found a significant overall difference between the recommendation of politeness strategies over both conditions. Politeness remains a highly subjective phenomenon with large individual differences.

We did, however, find that in the context of a particular negative FTA participants who recommend weightier politeness strategies, tend to perceive a lower level of the addressee's engagement level, and vice versa. In these contexts, our hypothesis

$$W_x = D(S,H) + P(H,S) + R_x - Eng(H) \tag{5.4}$$

seems confirmed. The diagram in Figure 5.1 is in these contexts indeed confirmed as the speaker adapts his/her politeness strategies to the engagement level of the hearer.

The outcomes of this study in human-human interaction give us an indication for the optimal coordination of the agent's politeness in human-agent interaction: It predicts that in some contexts (for some FTAs) a human-like agent that wants to continue the interaction with its user, and thereby maintaining the user's engagement, needs to speak more politely to someone who is less engaged than to someone who is very engaged in the ongoing interaction.

6 Engagement and Interruptions

6.1 Introduction

In order to favour the user's engagement, in the previous chapter we looked into the optimal coordination of the agent's politeness. In this chapter [1] we look at the optimal coordination of the agent's turn-taking behaviour with the goal of making the agent appear engaged and human-like, and therefore more engaging for the user. We consider one form of turn-taking in particular, namely interruptions.

The organisation of turn-taking is fundamental in human-human interaction [Schegloff, 2007] and is therefore also important in making socially believable conversational agents in human-agent and/or agent-agent interactions [Crook et al., 2010]. Interruptions represent a violation of the basic turn-taking rules and a possible way of claiming the turn [Beattie, 1981]. An interruption is "a starting up of some intervention by one person while another's turn is in progress" [Schegloff, 2001].

Turn-taking behaviour is related to engagement as for example, a failure to take an expected turn can signal disengagement (see Section 3.2.3.2) [Sidner et al., 2003]. The exact relation between interruptions and engagement is however not completely clear. For example, [Maat et al., 2010] has shown that leaving pauses between turns creates the feeling of having more rapport, which is an aspect of engagement (see Section 3.1.5), while other research in human-human interaction has shown that interruptions can create high involvement, and promote interest (and enthusiasm, and solidarity) [Tannen, 1981], which are also aspects of engagement (see Section 3.1.2 and 3.1.3). Previous findings have also showed that an agent's turn-taking behaviour can influence the human's impressions of the agent in terms of interpersonal attitude and personality [Maat et al., 2010].

To our knowledge no previous work has however offered a systematic analysis of the perception of interruptions, that considers both the amount of overlapping

1. performed jointly with Angelo Cafaro

speech and the content of interruptions, in order to understand the effects of interruptions in more detail. In this chapter we therefore present such a systematic analysis in agent-agent interaction with the goal of clarifying the effects of interruptions on the perception of the agents' engagement and interpersonal attitudes.

By considering both a classification based on the amount of overlapping speech (i.e. 'interruption types') and a classification based on the content of the interrupting utterance (i.e. 'interruption strategies'), we try to obtain a finer-grained understanding of the impact of interruptions, and answer questions such as: Is an interrupting agent always considered as dominant? And does the amount of overlap influence the perceived engagement level of the interruptee? Or is the content of the interruption more important?

The study is based on agent-agent interactions as this allows complete control and evaluation of both the interrupter's and interruptee's spoken behaviour. The outcomes of the study will then give us indications of how to model the turn-taking behaviour of a conversational agent in human-agent interaction.

Below we first explain the categorisations of interruptions in respectively interruption types (Section 6.2.1) and strategies (Section 6.2.2). In Section 6.3 we discuss related work, Section 6.4 describes the methodology, and 6.6 the results. In Section 6.7 we discuss and conclude the study.

6.2 Theoretical Background

6.2.1 Interruption Types

As mentioned above, an interruption is defined as a starting up of some intervention by one person while another's turn is in progress, often including "not letting them finish" [Schegloff, 2001]. An overlap refers to the fact of more than one person talking at a time [Schegloff, 2001], which means that an interruption does not necessarily implies overlap.

There are multiple ways in which interruptions can be further subdivided based on the temporal aspects of the speaker's switch (e.g. [Li, 2001], [Ferguson, 1975], [Schegloff, 2001]). We use the taxonomy of attempted speaker-switches by [Ferguson, 1975] that was later modified by [Beattie, 1981]. Figure 6.1 shows this taxonomy (everything in black and white). As we will explain below, we only focus on a subset of the interruption types that are identified in this taxonomy, indicated by red squares in Figure 6.1, which makes that the first classification of interruptions

that we take into account, is based on the amount of overlap of the interrupter and interruptee's turns.

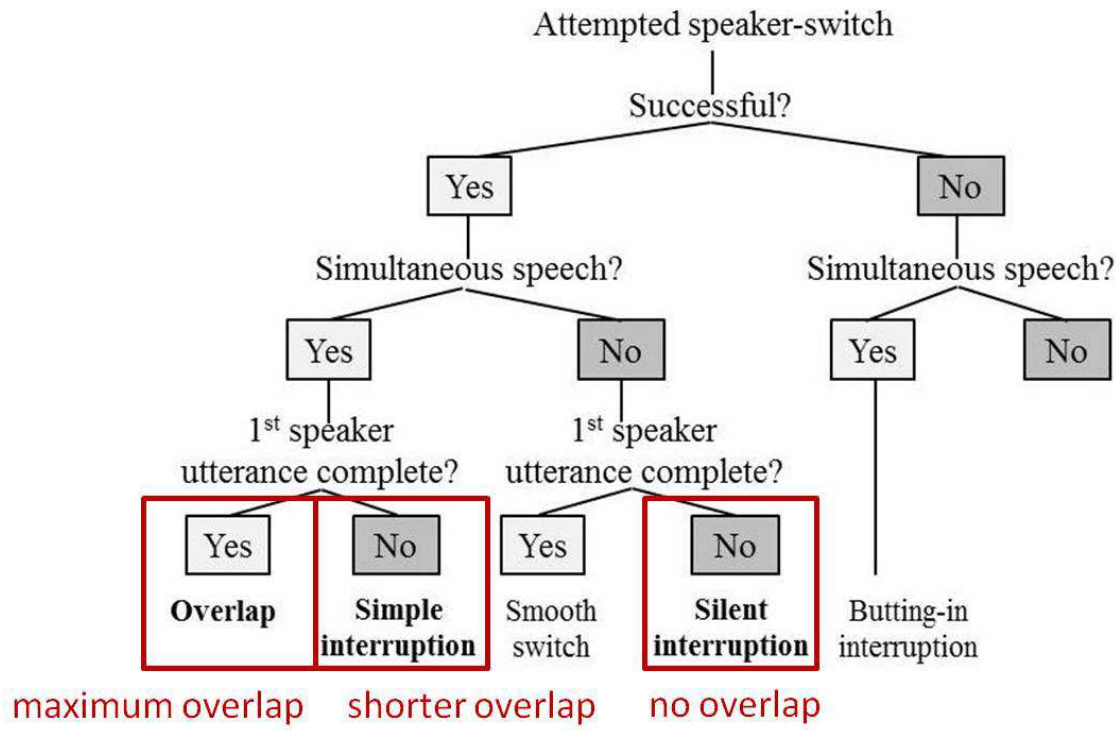


Figure 6.1 : [Beattie, 1981]'s taxonomy of interruption types (everything in black and white). The red squares indicate the interruptions that we consider.

[Beattie, 1981]'s taxonomy makes a first division according to whether or not the speaker-switch is successful, meaning that the initiator of the attempted interruption takes indeed the floor [Beattie, 1981] (top level of Figure). Further division is based respectively on the presence of simultaneous speech (second level), and first speaker utterance completeness (third level). In this taxonomy backchannels (such as *yeah, hmm, exactly*) are not considered as interruptions [Beattie, 1981]. Backchannels are not placed as an attempt of speaker-switching but are non-intrusive signals provided by the listener during the speaker's turn [Yngve, 1970] (Section 3.2.1.4).

In our study we focus on successful interruptions, where the interrupter takes indeed the floor. This is because we are also interested in the content of the interruption (see next Section, 6.2.2), which can only be transmitted if the interrupter is able to utter its turn. Further, as we are interested in the effect of interruptions only, we do not consider the smooth speaker switch as this represents the 'standard' smooth way of taking the floor without interruption.

This leaves us three different interruptions with different lengths of overlap, which we will refer to as *interruption types* (in contrast to *interruption strategies* in Section 6.2.2). From right to left in Figure 6.1 we consider the types:

— *Silent interruption*, where the interruptee stops its utterance as soon as it is interrupted.

— *Simple interruption*, where the interruptee continues its utterance for a while when it is interrupted but does not complete its utterance.

— *Overlap*, where the interruptee continues its utterance when it is interrupted until its utterance is completed.

6.2.2 Interruption Strategies

Interruptions have long been associated with interpersonal dominance [Youngquist, 2009] and are often considered power displays [Li, 2001], but recently a more balanced view has emerged [Li, 2001]. This view has led to the identification of two broad categories of interruptions that are based on the content of the interruptions, often referred to as intrusive and cooperative interruptions [Li, 2001, Murata, 1994].

According to [Murata, 1994]

— "*cooperative interruptions* are intended to help the speaker by coordinating the process and/or content of the ongoing conversation" [Li, 2001];

while

— "*intrusive interruptions* pose threats to the current speaker's territory by disrupting the process and/or content of the ongoing conversation" [Li, 2001, Murata, 1994, Goldberg, 1990].

We will take the difference between intrusive and cooperate interruptions into account as a second way of classifying and analysing interruptions (next to interruption types Section 6.2.1). We refer to these categories as *interruption strategies* (as opposed to *interruption types* in Section 6.2.1).

In previous work different terms have been used to refer to intrusive and cooperative interuptions. For example 'power' and 'non power' interruptions [Goldberg, 1990], or 'disconfirming' and 'confirming' interruptions [Kennedy and Camden, 1983] We use the terms *disruptive* [Ng et al., 1995] and *cooperative* interruptions as these terms refer to the interaction without already implying an attitude or stance of the interaction participants.

There is no consensus about the classification of possible cooperative and disruptive sub-strategies. For example, [Murata, 1994] does not mention any cooperative sub-strategy, while [Li, 2001, Kennedy and Camden, 1983] mention three:

— *Agreement*, showing concurrence, compliance, understanding or support. Sometimes also serving as an extension or elaboration of the idea presented by the speaker;

— *Assistance*, providing the current speaker with a word, phrase, sentence or idea in order to help completing his/her utterance;

— *Clarification*, having the current speaker clarify or explaining previously elicited piece of information that the listener is unclear about.

[Li, 2001] combines four disruptive sub-strategies from [Murata, 1994] and [Kennedy and Camden, 1983]:

— *Disagreement*, disagreeing with what the current speaker is saying and wanting to voice his or her opinion immediately;

— *Floor taking*, developing the topic of the current speaker by taking over the floor from the current speaker;

— *Topic change*, changing the topic of the conversation;

— *Tangentialisation*, reflecting awareness, usually by summarisation, of the information being sent by the current speaker. It prevents the interrupter from listening to an unwanted piece of information either because it has been already presented previously or because it is already known to the listener.

When we look into the effects of cooperative and disruptive interruption strategies on the perception of interrupting and interrupted agents we will also consider the sequence organisation of the interactions (conversational analysis) [Schegloff, 2007], as this may (also) influence how the agents are perceived. More precisely, we will consider the adjacency pair of which the interruption might be a part. An adjacency pair consists of two speaker turns where the second part of a pair is responsive to the action of the first part [Schegloff, 2007]; for example an offer followed by an acceptance or declination, and a question followed by an answer to this question [Schegloff, 2007]. Because a first utterance makes different responses relevant [Schegloff, 2007] the sequence assessment–(dis)agreement for example [Levinson, 1983], may be perceived differently than the sequence assessment–question.

Therefore, in order to compare merely the effects of cooperative versus disruptive interruption strategies, we will consider pairs of interruption strategies where

the cooperative and disruptive interruption have the same sequential *function* with respect to the first part of the adjacency pair. We will for example, compare a cooperative interruption that agrees with the preceding assessment, with a disruptive interruption that disagrees with the preceding assessment. Both interruptions have the same function of expressing their agreement with the previous assessment, which allows for a more balanced comparison of strategies.

Section 6.4.1 describes how the above theory about the functions, strategies and types of the interruptions are used in the design of the experimental stimuli.

6.2.3 Engagement

In Chapters 2 and 3 we have given an account of engagement. As in the rest of this thesis, for the present study we use [Poggi, 2007]'s interpretation as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction". Engagement refers to the coordination and cooperation of participants in the interaction (Section 2.2). Because turn-taking is a key characteristic of synchrony [Harrist and Waugh, 2002] (Section 3.2.3.2) turn-taking is also related to engagement.

In this study we try to find out what the consequences are of coordinating (i.e. synchronising) turns in such a way that they form interruptions, on the perception of the interrupting and interrupted agents' engagement (and interpersonal attitudes discussed in the following section). Previous work provide opposing expectations with respect to the effect of interruptions on engagement as [Maat et al., 2010] showed that leaving pauses between turns creates the feeling of having more rapport (Section 3.1.5), while other research has shown that simultaneous speech can be a way to create involvement (Section 3.1.2) and promote interest (Section 3.1.3) [Tannen, 1981].

Besides the perception of the agents' engagement we will look at the agents' perceived involvement (Section 3.1.2). Involvement is only one aspect underlying engagement (Section 3.1.2) and may therefore give us more fine-grained results. We use [Sidner et al., 2005b]'s interpretation of involvement as "being captured by the experience".

6.2.4 Interpersonal Attitudes

Next to the effects on the perception of the agents' engagement and involvement we will verify the effects of interruptions on the interrupting and interrupted agents' perceived interpersonal attitudes. This is because the role of interruptions in interpersonal attitudes such as dominance has been subject of discussion in previous work (as mentioned in Section 6.2.2) [Li, 2001].

Interpersonal attitudes can be defined as an individual's conscious or unconscious evaluation of how he/she feels about and relates to another person [Argyle, 1988, p. 85]. Several researchers attempted to identify the dimensions that can best represent the different interpersonal attitudes that are expressed during social interaction. For example, [Schutz, 1958] proposed the dimensions of 'inclusion', 'control' and 'affect'. [Burgoon and Hale, 1984] identified twelve dimensions, defining different communication styles, such as 'dominance', 'intimacy', and 'confidence'. [Argyle, 1988] proposed a two-dimensional representation composed of 'affiliation' and 'status' where:

— *affiliation* is characterised as the degree of liking or friendliness, ranging from unfriendly to friendly;

— *status* is related to power and assertiveness during the interaction, ranging from submissive to dominant.

In this study we chose [Argyle, 1988]'s representation because it offers an intrinsic simplicity (two dimensions only) and the best compromise of explanatory power and parsimony.

6.3 Related Work

Related research with respect to the perception of turn-taking in interaction with an agent includes [Maat et al., 2010] who showed how different ways of turn-taking in user-agent interaction, including interruptions, can influence the user's impression of the agent's personality (agreeableness), emotion and social attitudes. Their analysis did however not take into account the speech content nor the interruption strategy. [Thórisson et al., 2010] evaluated a turn-taking model (YTTM) in terms of scalability and believability, though they focused on timing and prosody features for autonomous turn assigning in multi-party agent interaction. [Ravenet et al., 2015] presented a multi-agent model for reflecting interpersonal attitudes in conversing groups. Their emphasis was on nonverbal behaviour (e.g. interpersonal

space, gestures) and ways of turn-taking but they did not consider verbal behaviour and different interruption types and/or strategies.

Work related specifically to interruptions includes [Crook et al., 2010, Crook et al., 2012] who presented a mechanism for handling "barge-in" interruptions from a user interacting with an embodied conversational agent. Their agent was able to detect and respond to user interruptions. The handling process consisted of an address phase where the agent addressed the particular interruption that occurred, and a resumption phase where the agent implemented context-sensitive strategies for continuing or aborting the current conversational plan. While they modelled a handling mechanism that considered the user's perception of the interrupted agent, they did not consider the effects of the interrupter's (user's) interruption strategy on the perceived engagement level and interpersonal attitude in interaction with the agent.

Further, [Gravano and Hirschberg, 2012] examined interruptions in a corpus of spontaneous task oriented dialogue. They looked at timing and acoustic/prosodic features that predict interruptions, where we investigate the effects of different interruption types and strategies by keeping the prosody and timing of the interruptions constant.

In sum, previous work focused on the impact of different ways of turn-taking on the perceived agent's personality and attitude [Maat et al., 2010], on handling interruptions [Crook et al., 2010, Crook et al., 2012], and on predicting interruptions [Gravano and Hirschberg, 2012]. To our knowledge no work examined the effects of different interruption types (Section 6.2.1) and strategies (Section 6.2.2) on the perceived engagement (Section 6.2.3) and interpersonal attitudes (Section 6.2.4) of both an interrupting and an interrupted agent.

6.4 Methodology

To evaluate the effects of different interruption types and strategies on the human's perception of interrupting and interrupted agents' engagement (engagement, involvement) and interpersonal attitude (dominance and friendliness) we designed an empirical experiment. By means of an online survey we exposed human participants to a series of short video fragments where two agents interact (as further detailed below). The interruption types and strategies vary systematically within a particular subcategory where all interruptions have the same communicative function (see Section 6.2.2 and 6.4.1.2). After each fragment participants were asked to

rate the perceived level of engagement, involvement, dominance, and friendliness of the two agents. For clarity we here onwards refer to the interrupted agent as 'A' and the interrupting agent as 'B'.

In the following subsection we describe the details of this methodology: respectively the design of the evaluation stimuli and the questionnaire.

6.4.1 Stimuli Design and Production

6.4.1.1 Interruption Types

We based the generation of interruptions on existing literature in human-human interaction that we described in Sections 6.2.1 and 6.2.2. As mentioned before, with respect to the interruption types we focused on successful speaker-switches, as shown in Figure 6.1, where the interrupter takes indeed the floor. This makes it possible for the interrupter to utter a complete utterance (in the other cases the sentence is partially uttered) and thus deploy an interruption strategy as well. Of the remaining speaker-switches we only consider the interruptions (see Figure 6.1).

In [Beattie, 1981] each of the interruptions is illustrated with an example, giving us further details with respect to for example the possibility of pauses in between the turns. According to the characteristics of the interruption types (Figure 6.1) and its examples provided in [Beattie, 1981], we defined conversational fragments with different amounts of overlap that represent the three interruption types, as shown in Table 6.1, Column 1. The design of the content of the utterances is described in the following subsection; here we only define the sequential organisation of the utterances.

While [Beattie, 1981]'s classification of interruptions indicates how an interruption can be modelled with respect to its overlap, no explicit information is given about when, during the first speaker's utterance, the interruption should start. We noticed that in [Beattie, 1981]'s examples interruptions start at *possible completion points*. A possible completion point is a point during the first speaker's turn where a speaker-switch could take place [Mazeland, 2003]). Following [Beattie, 1981]'s examples we fixed the onset (starting point) of the interruptions [Schegloff, 2001] in our study (Table 6.1) at a possible completion point as well. There are three aspects that participants in an interaction use to determine possible completion points: syntax, prosody and pragmatics of the first speaker's utterance [Mazeland, 2003]. In the conversation fragments we designed (Table 6.1) the second speaker interrupts at the

moment where the first speaker's turn can indeed be judged syntactically and pragmatically complete. As detailed in Section 6.4.1.3, because the first speaker's turn is not really completed when it is interrupted, the first turn shall not be produced with prosodic completion.

To compare the effect of the three interruption types with each other, each interruption type is modelled with the same set of utterances (see Table 6.1) to keep all other variables as constant as possible and reduce possible confounding effects. This means that:

— In each conversation fragment the first speaker (agent A) is doing the same utterance when it is interrupted.

— The onset of the interruptions [Schegloff, 2001] is as equal as possible among the different interruption types, meaning that in all interruptions the interrupter (B) begins speaking at the same moment relative to the interruptee's current turn (A). Only in the *silent interruption type* a small pause is added within the interruptee's turn, following the examples of silent interruptions in [Beattie, 1981].

— The surface forms (designed in the next subsection) of the interrupting utterances (second speaker utterances) are the same among the interruption types.

The only element that differs among the interruption types is the amount of overlap. In Table 6.1 synchronous speech (i.e. overlap) of the interruptee (A) and the interrupter (B) is indicated between squared brackets. The length of synchronous speech is not only the result of theoretical considerations but depends further on the speech duration of the agents' synthesised speech (see Section 6.4.1.3).

6.4.1.2 Interruption Strategies

We represented the interruption strategies by designing the content of the interruptions according to the descriptions of disruptive and cooperative strategies, and their examples in the form of sub-strategies (described in Section 6.2.2). Besides, according to theory about adjacency pairs and motivated in Section 6.2.2, in order to compare the effects of cooperative and disruptive strategies we need to formulate pairs of strategies where the cooperative strategy and disruptive strategy have a similar function in relation to the preceding turn.

Above considerations have lead us to design the following 4 pairs of comparable interruption strategies, respectively disruptive and cooperative:

Table 6.1 : Interruption types and strategies for the function categories investigated in our study. The Interruptee (A) is the first speaker, the Interrupter (B) is the second speaker. In the silent interruption (0.2) refers to a silent pause in seconds.

Interruption Type	Function	Interruption Strategy	Agent	Utterances	
Overlap			A	You know I've read the story Alice in Wonderland. It tells an amazing story [about a little girl.]	(1)
	Understanding	Disruptive	B	[When were you in] the Wonderland theme park?	(2)
	Question	Cooperative	B	[Do you mean the book writ]ten by Lewis Carrol?	(3)
	Opinion	Disruptive	B	[No, it does] not tell an amazing story at all.	(4)
		Cooperative	B	[Great! I] love it and I've read it several times.	(5)
	Partner Communication	Disruptive	B	[I've read the book s]o I know the story.	(6)
		Cooperative	B	[About a busy rab]bit and a smiling cat.	(7)
	Topic	Disruptive	B	[I borrowed the book B]eauty and the Beast.	(8)
		Cooperative	B	[It was first rele]ased in 1865.	(9)
Simple Interruption			A	You know I've read the book Alice in Wonderland. It tells an amazing story [about]	(10)
	Understanding	Disruptive	B	[When] were you in the Wonderland theme park?	(11)
	Question	Cooperative	B	[Do you] mean the book written by Lewis Carrol?	(12)
	Opinion	Disruptive	B	[No,] it does not tell an amazing story at all.	(13)
		Cooperative	B	[Great]! I love it and I've read it several times.	(14)
	Partner Communication	Disruptive	B	[I've re]ad the book so I know the story.	(15)
		Cooperative	B	[About] a busy rabbit and a smiling cat.	(16)
	Topic	Disruptive	B	[I bo]rrowed the book Beauty and the Beast.	(17)
		Cooperative	B	[It was] first released in 1865.	(18)
Silent Interruption			A	You know I've read the book Alice in Wonderland. It tells an amazing story (0.2)	(19)
	Understanding	Disruptive	B	When were you in the Wonderland theme park?	(20)
	Question	Cooperative	B	Do you mean the book written by Lewis Carrol?	(21)
	Opinion	Disruptive	B	No, it does not tell an amazing story at all.	(22)
		Cooperative	B	Great! I love it and I've read it several times.	(23)
	Partner Communication	Disruptive	B	I've read the book so I know the story.	(24)
		Cooperative	B	About a busy rabbit and a smiling cat.	(25)
	Topic	Disruptive	B	I borrowed the book Beauty and the Beast.	(26)
		Cooperative	B	It was first released in 1865.	(27)

— *Question implying misunderstanding – Clarification question*;

— *Disagreement – Agreement*;

— *Tangentialisation (see Section 6.2.2) – Completion*;

— *New topic introduction – Topic elaboration.*

These pairs do not represent all possible interruption sub-strategies, but evaluating 4 pairs only allows us to obtain sufficient data per interruption while still exploring some variation in interruption sub-strategies. For practical reasons we respectively name each of the interruption pairs by the *function* that both interruptions have in common:

— *Understanding question*: both sub-strategies are consequences of the interrupter's (mis)understanding of the interruptee's turn;

— *Opinion*: both express the interrupter's (dis)agreement with the interruptee;

— *Partner Communication*: both try to adjust (future) utterances of the interruptee;

— *Topic*: both strategies contribute to the topic management of the interaction.

Table 6.1 gives an overview of the strategies and utterances that we formulated in combination with the interruption types defined in Section 6.4.1.2.

The surface form of the first speaker's utterance (interruptee) is designed to introduce the book Alice in Wonderland because this study was primarily performed for another project (European project H2020 *ARIA-VALUSPA*[2]) than our museum agent project (*A1:1*, Section 1.2), namely for a virtual agent that enables retrieval of information about a book (Alice in Wonderland). A's utterance is fixed among the strategies and varies as little as possible among the interruption types in order to allow for comparison of interruptions with respect to their types and strategies only.

The surface forms of the second speaker's utterances are modelled as a reaction to this first utterance, and as mentioned above, follow the descriptions of disruptive and cooperative strategies and the descriptions of the selected sub-strategies (functions) (described in Section 6.2.2). For example, to realise the function *understanding question* we formulated one question utterance that *implies misunderstanding* the first speaker's utterance, which is a disruptive strategy, and one question utterance that *asks for clarification* of the first speaker's utterance, which is a cooperative strategy. We modelled the question that implies misunderstanding (disruptive strategy) as

2. http://aria-agent.eu

asking more information about a reference from A's utterance (Wonderland) that shows that B had misunderstood this reference (B thinks that the first speaker has read the story Alice in a theme park named Wonderland, while Wonderland is actually part of the book title Alice in Wonderland, see Table 6.1). The clarification question (cooperative strategy) is modelled as asking more information about a reference from A's utterance as well, but this time by checking if B has well understood this reference (B checks if A indeed refers to the book B thinks A refers to, Table 6.1).

In order to verify that for each function the instances of the interruption strategies (i.e. surface forms of second speaker turns) were indeed perceived as disruptive/cooperative strategies, we had first performed a validation study (manipulation check) that has lead to the final interruptions of Table 6.1.

Interruption Strategy Validation : We had performed a validation study of the interruption strategies by means of an online questionnaire. Appendix B Figure B.1 shows a fragment of this online questionnaire. To disregard the interruption types, we showed written fragments of conversation between two interaction participants where we placed the interrupting utterance (as designed above) directly after the first speaker's utterance is completed. In this way no interruption takes place but we present the utterance, that will later be used to interrupt, in the context of a smooth speaker-switch (Section 6.2.1), that thereby serves as a baseline for measuring the utterance's disruptive/cooperative potential. Participants were asked to rate on a 5-point Likert scale to what extent they found the second utterance in the conversation fragments (the later interrupting utterance) disruptive/cooperative (with anchors *"very disruptive"* and *"very cooperative"*). We provided definitions of "disruptive" and "cooperative" (see Section 6.2.2).

Results from 11 participants (7 males) indicated that for all functions the strategies were correctly recognised in the utterances, except for the disruptive utterance belonging to the *opinion* function (sub-strategy *disagreement*): *"Well, that's debatable, it's not my favorite one."* On average this utterance was judged more cooperative than expected ($M = 3, S.E. = 1.3$). This can be explained by the fact that the disagreement was presented in a mitigated way with respect to the first speaker's assessment (Table 6.1) (politeness theory [Brown and Levinson, 1987]). We therefore formulated multiple alternative utterances for the disagreement sub-strategy, namely direct non-mitigated disagreements, and performed a second round of validation (see Appendix B Figure B.2 for a fragment). Based on the results of this second round of validation we selected an utterance for the final evaluation study

from the set of alternatives that was clearly judged as disruptive: *"No, it does not tell an amazing story at all."* Table 6.1 shows the validated list of utterances that is used in the final evaluation study.

6.4.1.3 Stimuli Production

Now that we have designed the conversational fragments that represent the interruptions for the final evaluation study, we needed to create video fragments of two agents (agent-agent interaction, see Section 6.1) acting out these conversational fragments. These videos (stimuli) then allow us to obtain human perceptions of the interrupter and interruptee agents in agent-agent interaction.

We thus produced a total number of 24 conversation fragments (6 for each function/trial) corresponding to all the 24 cases in Table 6.1. We were aware of possible gender influences on the interactants' behaviour [Zimmermann and West, 1996, Beattie, 1981, Youngquist, 2009] and the perception of their attitudes [Beattie, 1981, Youngquist, 2009, Goldberg, 1990]. Considering that our primary interests concern interruption type and strategy effects, we did not variate the gender of the agents in this study but started out with two male gendered agents for all interruptions.

The agents were represented by two still male head silhouettes facing each other on a black background in order to avoid any biases from non-verbal behaviour or appearances of the agents. The silhouettes were white grey and were indicated as "Left Agent" and "Right Agent". Along with the agents' speech subtitles progressively appeared below the corresponding silhouettes. Figure 6.2 shows a still image of one of the video fragments.

The agents' speech was produced with a male synthesised voice from the Cereproc text-to-speech (TTS) tool [Aylett and Pidcock, 2007]. We used synthesised voice instead of human pre-recorded speech to keep the prosody of the utterances constant among all interruptions. A human speaker might pronounce the same sentence with different intonations adding possible biases. Besides, in final applications agent voices are produced with synthesised speech from TTS tools as well. We used the same synthesised voice for both the interruptee and the interrupter to avoid a possible imbalance between the perception of the interpersonal attitudes (e.g. dominance) of both speakers due to their voices. Some voices may for example sound more dominant than others, which could lead to a bias of the results. We made sure that in the video fragments both voices were played in stereo sound to give

Figure 6.2 : Still image of one of the agent-agent video fragments

the impression that the left agent's speech is really coming from the left agent and the right agent's speech really from the right agent, enabling better discrimination of the agents' voices.

Subtitles were added to compensate any possible miscomprehension of the agents' speech due to unnatural effects of the synthesised voices, thereby ensuring that the participants were correctly exposed to the interruption stimuli. Together with the stereo sound the subtitles further allowed us to use the same voice for both agents while still being able to discriminate between the two speech sources (agents). Participants of a small pilot test (2 male and 2 female) did indeed think that we used different voices "since the speech came from different audio sources".

As explained in Section 6.2.1, the silent interruptions contain a small pause in the interruptee's turn following the examples in [Beattie, 1981]. To produce these pauses in the video fragments we added 0.2 seconds of white noise at an amplitude of 0.001. This keeps the left channel (of the interruptee) open and avoids an abrupt termination that might have sounded as a recording error.

Except for the small pause, the first (left) speaker is always doing the same utterance when it is interrupted. We therefore did not change the prosody of the first speaker's turn among the conversation fragments (see Section 6.2.1). To produce this effect we synthesised speech (with the TTS tool) for the completed first utterance, with accompanying prosody, and cut the fragment to select for each of the interruption types only the part of the first utterance that is pronounced. In this way an interrupted utterance also sounds prosodically as interrupted. The speech of the second speaker's turn was synthesised and inserted in its entirety within the

first speaker's turn, forming the conversational fragments of Table 6.1.

6.4.2 Questionnaire

6.4.2.1 Website

We presented the stimuli to evaluation participants by means of an online website. The website started of with a consent page, a sound check page, and a tutorial page. Participants were then shown 6 pages with on each a stimuli (video fragment) and questions asking how the human participants perceived the agents. Figure B.3 in Appendix B shows a printscreen (only a part) of such a page. After evaluating all 6 video fragments participants were asked for some demographic information, after which the questionnaire was completed.

Participants were shown the 6 stimuli that belong to the trial to which they were assigned (randomly but balancing the number of participants per trial). We created a total of 4 trials corresponding to the 4 functions of the interruptions (see Column 2 of Table 6.1). For example all the video fragments where the interruptions have the function of an 'understanding question' form the stimuli of one trial. No more stimuli are shown to the same participant in order to limit the length of the evaluation for each participant and because we are only interested in the effects of interruption types and strategies within the same function (as explained in Sections 6.4.1.2 and 6.2.2).

6.4.2.2 Questions

For each stimulus, we asked the participants to rate the perceived level of engagement and involvement of both agents, and the agents' attitudes (dominance and friendliness) towards each other.

For assessing the agents' level of engagement we asked two questions per agent that are based on [Poggi, 2007]'s definition (Chapter 2) and that we also use in other studies towards engagement such as in Chapter 5. The questions for the interrupting agent were: (1) *"What value does the Right Agent attribute to being together with the Left Agent?"* and *"What value does the Right Agent attribute to continuing the interaction?"*. For the interrupted agent "left" and "right" are switched. Answers were collected on a 5-points labelled Likert scale ranging from *"No value"* to *"A maximum Value"*.

We also assessed the agents' level of involvement as interpreted in [Sidner et al., 2005b] (explained in Section 6.2.3). Similar as in Chapter 5 we asked *"How engaging was the interaction for the Right* (or Left) *Agent?"* [Sidner et al., 2005b] on a 5 point labelled Likert scale ranging from *"Not at all"* to *"Very much"*. This question is chosen from [Sidner et al., 2005b] because it fits our design and is directly linked to engagement [Lombard and Ditton, 1997, Lombard et al., 2000] (Section 3.1.2).

For measuring the agents interpersonal attitudes we adapted 2 statements for each attitude dimension from the Riverside Q-Sort inventory [Funder et al., 2000]. We selected the two most reliable items (adjectives or key terms) of positive and negative valence defined by [Wiggins et al., 1988]'s interpersonal circumplex inventory. The items we adopted to assess dominance were *"tries to control"* and *"seems insecure"*, and for friendliness they were *"tries to be likable"* and *"expresses hostility"*. For example, for estimating the interrupter's dominance (right agent), participants were asked to what extent they agreed with 2 statements including "*The Right Agent tries to control the interaction with the Left Agent*". Answers were given on a 5-points labelled Likert scale ranging from *"Completely disagree"* to *"Completely agree"*.

An overview of the dependent variables (the variables we measured) is shown in Table 6.2.

Table 6.2 : Summary of the evaluated measures and their dependent variable abbreviations. *A* refers to the interrupted agent (left agent) and *B* to the interrupting agent (right agent).

Measure	DV names
Engagement	`A-Eng, B-Eng`
Involvement	`A-Inv, B-Inv`
Attitude: Dominance	`A-Dom, B-Dom`
Attitude: Friendliness	`A-Friend, B-Friend`

6.5 Hypotheses

6.5.1 Engagement and Involvement

In this evaluation study we verified the effects of interruptions on the perceived engagement, involvement and interpersonal attitudes of an interrupted and inter-

rupting agent in agent-agent interaction. We hypothesised some outcomes of this study.

First of all, we expected that when an agent is interrupting with a cooperative strategy it shows more engagement and involvement compared to when it interrupts with a disruptive strategy. Cooperative interruptions coordinate the process and/or content of the conversation, while disruptive ones disrupt the process and/or content of the ongoing interaction (Section 6.2.2) [Murata, 1994]. Given that engagement is characterised by a coordination and cooperation between the interaction participants (Section 2.3.2.1) we expect that a cooperative interruption is a sign of more engagement and involvement of the interrupting agent than a disruptive strategy (main effect).

With respect to the interruption types previous research has shown contradictory effects on engagement related measures of the interrupter (see Section 6.2.3) preventing us from predicting a specific outcome.

To our knowledge no previous research has considered effects of interruptions, any type or strategy, on the perception of the interruptee's (interrupted agent) engagement and/or involvement. We therefore do not hypothesise any effects of the interruption strategies and types on the perception of the interruptee. After all, the interruption strategy is only decided by the interrupting agent; and with respect to the interruption types we believe that both a continuation of the first speaker's utterance as a sudden break-off of the utterance could be interpreted as manifestations of engagement and involvement of the first speaker (interrupted agent).

In summary, we formulated two hypotheses regarding the engagement and involvement dimensions; we hypothesised for each trial/function (where the interrupter is indicated as `B`, see Table 6.2) that:

— `H.B-Eng`: The interruption *strategy* has a main effect on the perceived interrupter's engagement: The cooperative strategy leads to *higher* `B-Eng` compared to the disruptive strategy.

— `H.B-Inv`: The interruption *strategy* has a main effect on the perceived interrupter's involvement: The cooperative strategy leads to *higher* `B-Inv` compared to the disruptive strategy.

6.5.2 Interpersonal Attitudes

Previous work has shown that disruptive interruption strategies in human-human interaction increase the perceived dominance of an interrupter [Goldberg,

1990, Youngquist, 2009], whereas cooperative strategies increase the interrupter's affiliation (i.e. liking or friendliness) [Goldberg, 1990]. We therefore hypothesised that the perceived dominance and friendliness of an interrupting agent (in agent-agent interaction) would also be affected by the interruption strategy that it uses (main effect).

As for the interruption types, their difference is expressed by the amount of overlap of the first and second turns (in Table 6.1 respectively long overlap, short overlap, and no overlap). In human-human interaction a speaker that holds the turn (i.e. trying to end his/her utterance, resulting in a long overlap) is perceived as more dominant [Beattie, 1981, Goldberg, 1990] and less likable [Goldberg, 1990]. Considering that the interrupting agent always completes its utterance (in order to deploy a specific strategy) while the interrupted agent varies the amount of produced overlap among the interruption types, we hypothesised that the perceived dominance and friendliness of the interrupted agent would be affected by the interruption type (main effect).

In summary we hypothesised for each trial/function (where A is the interrupted agent and B the interrupting agent):

- `H.A-Dom`: The interruption *type* has a main effect on the perceived interruptee's dominance: The *longer* the overlap the *higher* `A-Dom`.
- `H.A-Friend`: The interruption *type* has a main effect on the perceived interruptee's friendliness: The *longer* the overlap the *lower* `A-Friend`.
- `H.B-Dom`: The interruption *strategy* has a main effect on the perceived interrupter's dominance: The disruptive strategy leads to *higher* `B-Dom` compared to the cooperative strategy.
- `H.B-Friend`: The interruption *strategy* has a main effect on the perceived interrupter's friendliness: The cooperative strategy will lead to *higher* `B-Friend` compared to the disruptive strategy.

6.6 Results

We recruited a total of 72 participants via mailing lists (18 in each trial). 46% of the participants was between 18 and 30 years old, and 34% between 31 and 40. 57% of the participants was male, 39% female, and 4% preferred not to tell their gender. 82% of the participants was well educated (master level and further). Participants

had different cultural backgrounds[3], where France (20.8%) and the USA (14%) formed the two largest groups.

We conducted 4 statistical analyses on the results of each function (trial). On each trial we conducted 3x2 repeated measures MANOVAs (Multivariate Analysis of Variance) on the interrupted (A) and interrupting (B) agents' *engagement* and *involvement* (`A-Eng` and `A-Inv`, `B-Eng` and `B-Inv`), *dominance* and *friendliness* (`A-Dom` and `A-Friend`, `B-Dom` and `B-Friend`), with within-subject factors interruption *type* (3) and interruption *strategy* (2). Prior to the analyses we normalised all dependent variables to values within the range $[0-1]$.

Below we report per function the significant main effects and interactions of follow-up univariate analyses (i.e. analyses per independent variable) (the sphericity assumption was not violated). Simple main effects of interactions between factors were tested (post-hoc analyses) with Bonferroni adjustments for multiple comparisons. Effect sizes (η_p^2) for all comparisons ranged from 0.18 to 0.76. Figure 6.3 provides a summary of all quantitative findings.

6.6.1 Function: Question

Interruptee's Engagement and Involvement : No significant effect is found with respect to `A-Eng`. With respect to `A-Inv` there is a significant interaction effect between *type* and *strategy* ($F(2,34) = 4.4$, $p < .05$): For the silent interruption (i.e. interruption without overlap, Section 6.2.1), A is considered significantly more involved in the interaction when B's interruption strategy is cooperative ($M = .60, SE = .04$) compared to when B's strategy is disruptive ($M = .70, SE = .03$).

Interrupter's Engagement and Involvement : With respect to `B-Eng`, main effects of *type* ($F(2,34) = 5.5$, $p < .05$) and *strategy* ($F(1,17) = 17.0$, $p < .005$) have been found but no significant interaction effects. `H.B-Eng` is partially confirmed. A decomposition of these effects revealed that overall B was perceived as more engaged when it used a cooperative strategy ($M = .61, SE = .05$) compared to when it uses a disruptive strategy ($M = .45, SE = .05$). But when it uses a disruptive strategy, it was perceived significantly more engaged when it interrupts in silence (i.e. no overlap) ($M = .54, SE = .05$) in comparison with the two other interruption types (short overlap: $M = .41, SE = .05$; maximum overlap: $M = .40, SE = .06$).

Strategy had a main effect on `B-Inv` ($F(1,17) = 12.5$, $p = .003$): Cooperative

3. As part of the demographic information we collected, we asked participants to indicate the nationality that most represented their cultural identity.

Interrupted Agent (A)			
Function	**Type** (↗ *overlap*)	**Strategy** (*disr. to coop.*)	**Type AND Strategy** (*disr. to coop.*)
QUESTION	↗ Dominance ↘ Friendliness	*n.s.*	↗ Involvement (when type: *no overlap*)
OPINION	*n.s.*	*n.s.*	↗ Friendliness (when type: *no overlap*)
PARTNER	*n.s.*	↗ Dominance	*n.s.*
TOPIC	*n.s.*	*n.s.*	*n.s.*

(a) Results Interrupted Agent (interruptee)

Interrupting Agent (B)			
Function	**Type** (↗ *overlap*)	**Strategy** (*disr. to coop.*)	**Type OR Strategy** (*disr. to coop.*)
QUESTION	↗ Dominance	↗ Involvement	↗ Friendliness ↗ Engagement (particular cases)
OPINION	*n.s.*	↗ Friendliness ↗ Engagement ↗ Involvement	*n.s.*
PARTNER	↗ Dominance	↗ Engagement ↗ Involvement	↗ Friendliness (more remarked when type: *no overlap*)
TOPIC	↘ Friendliness	*n.s.*	*n.s.*

(b) Results Interrupting Agent (interrupter)

Figure 6.3 : Overview of the results. Inclined arrows indicate the effect directions of the dependent variables. For example, in the *question* trial, changes in interruption *type* (increasing ↗ overlap) decreased ↘ A's perceived level of *friendliness*. Another example, in the *opinion* trial, changing the interruption *strategy* from *disruptive* to *cooperative* increased ↗ B's perceived *engagement*.

interruptions made that B was perceived more involved ($M = .59, SE = .04$) in comparison to when B uses disruptive strategies ($M = .40, SE = .06$). `H.B-Inv` (see Section 6.5) is therefore supported.

Interruptee's Dominance and Friendliness : The analyses revealed a main effect of the interruption *type* on `A-Dom` ($F(2, 34) = 9.1$, $p < .001$). As the amount of overlap increased, A's perceived dominance increased (no overlap: $M = .48, SE = .04$; short overlap: $M = .53 SE = .03$; and maximum overlap: $M = .64 SE = .04$). This supports hypothesis `H.A-Dom`.

Type had an opposite main effect on `A-Friend` ($F(2, 34) = 3.6$, $p < .05$): An increasing overlap decreased A's perceived friendliness (no overlap: $M = .74, SE = .04$; short overlap: $M = .71 SE = .03$; and maximum overlap: $M = .66 SE = .04$), supporting `H.A-Friend`.

Interrupter's Dominance and Friendliness : Only the interruption *type* has a main effect on `B-Dom` ($F(2, 34) = 4.0$, $p < .05$), which makes that `H.B-Dom` (hypothesising an effect of *strategy*) is rejected: Silent interruptions (without overlap) led to significantly less dominance of the interrupter ($M = .70, SE = .04$) compared to the two other interruption types (short overlap: $M = .78, SE = .04$; maximum overlap: $M = .78, SE = .03$).

For `B-Friend` we found main effects of *type* ($F(2, 34) = 7.9$, $p < .005$) and *strategy* ($F(1, 17) = 11.9$, $p < .005$) but no significant interaction effects, confirming `H.B-Friend` partially. We decomposed the effects and found that overall a cooperative strategy accounts for higher friendliness of the interrupter ($M = .53, SE = .04$) compared to a disruptive strategy ($M = .42, SE = .04$). When the interrupting agent used a disruptive strategy the interrupter was perceived significantly more friendly if the interruption was silent ($M = .53, SE = .05$) in comparison with the other two interruption types (short overlap: $M = .32, SE = .05$; maximum overlap: $M = .39, SE = .04$).

6.6.2 Function: Opinion

Interruptee's Engagement and Involvement : In this trial no significant effects were found regarding the engagement and involvement levels of the interruptee.

Interrupter's Engagement and Involvement : The interruption *strategy* has significant main effects on `B-Eng` ($F(1, 17) = 26.0$, $p < .001$) and `B-Inv` ($F(1, 17) = 23.8$, $p < .001$), supporting thereby `H.B-Eng` and `H.B-Inv`. Interrupting with a cooperative strategy led to higher perceived engagement (M=.46, SE=.04) and involvement levels (M=.46, SE=.04) of the interrupter compared to the disruptive strategy ($M = .22, SE = .02 and M = .23, SE = .04$).

Interruptee's Dominance and Friendliness : No significant effects have

been found regarding `A-Dom`, which rejects `H.A-Dom`. There is a significant interaction effect on `A-Friend` ($F(2, 34) = 3.8$, $p < .05$), which partially confirms `H.A-Friend`: When we deal with a silent interruption *type* (without overlap), if B interrupts with a cooperative *strategy*, A was perceived as less friendly ($M = .66, SE = .05$) compared to when B interrupted with a disruptive strategy ($M = .74, SE = .03$).

Interrupter's Dominance and Friendliness : With respect to `B-Dom` no significant effects have been found (`H.B-Dom rejected`). *Strategy* has a main effect on `B-Friend` ($F(1, 17) = 54.8$, $p < .001$), which supports `H.B-Friend`. B was more perceived as more friendly if it used a cooperative interruption ($M = .52, SE = .04$) compared to a disruptive strategy ($M = .26, SE = .03$).

6.6.3 Function: Partner

Interruptee's Engagement and Involvement : There are no significant effects on the interruptee's engagement and involvement in this trial.

Interrupter's Engagement and Involvement : The interruption *strategy* has significant main effects on `B-Eng` ($F(1, 17) = 30.7$, $p < .005$) and `B-Inv` ($F(1, 17) = 15.5$, $p < .001$). `H.B-Eng` and `H.B-Inv` are therefore confirmed. In both cases a cooperative strategy led to higher perceived engagement ($M = .42, SE = .04$) and involvement of B ($M = .43, SE = .03$) than a disruptive strategy ($M = .18, SE = .04$, and $M = .21, SE = .04$).

Interruptee's Dominance and Friendliness : With respect to `A-Dom` *strategy* has a main effect ($F(1, 17) = 12.8$, $p < .005$). When B used a cooperative interruption strategy, A was perceived as significantly more dominant ($M = .60, SE = .03$) than when B interrupted with a disruptive strategy ($M = .55, SE = .03$). `H.A-Dom` that predicted an effect of the type of the interruption is rejected. No significant main effects or interactions have been found for `A-Friend`, rejecting `H.A-Friend`.

Interrupter's Dominance and Friendliness : *Type* has a main effect on `B-Dom` ($F(2, 34) = 5.3$, $p < .05$) instead of strategy, which rejects `H.B-Dom`. B's dominance was significantly higher when A stopped its utterance after a short overlap with B's interruption (simple interruption: $M = .75, SE = .03$) compared to when A did not continue its utterance after it was interrupted in the small pause of the utterance (silent interruption: $M = .67, SE = .03$).

Regarding the interrupter's friendliness (`B-Friend`) we found main effects of both *type* ($F(2, 34) = 8.6$, $p < .005$) and *strategy* ($F(1, 17) = 30.8$, $p < .001$) but no

significant interaction effects. `H.B-Friend` that hypothesised only a main effect of strategy is partially supported. A decomposition of the effects reveals that for all interruption types a cooperative strategy accounts for higher interrupter friendliness ($M = .52, SE = .04$) compared to a disruptive strategy ($M = .27, SE = .03$). This effect is stronger when the interruption type is silent (no overlap, cooperative: $M = .62, SE = .05$; disruptive: $M = .33, SE = .04$).

6.6.4 Function: Topic

The only effect we found in this trial is a significant main effect of the interruption *type* on `B-Friend` ($F(2, 34) = 7.9$, $p < .005$): The interrupter's (B) perceived friendliness was significantly higher when A did not continue its utterance after B interrupted in the short pause of A's utterance (silent interruption, no overlap: $M = .47, SE = .05$) compared to the two other interruption types (maximum overlap: $M = .38, SE = .03$, short overlap: $M = .35, SE = .03$). Hypotheses `H.A-Dom`, `H.A-Friend`, `H.B-Dom`, and `H.B-Friend` are not confirmed in this function (trial).

6

6.7 Discussion

The results have shown that overall agent A's (the interruptee's) perceived engagement was never influenced by B's (the interrupter's) interrupting behaviour, even if A decided to stop or continue its utterance in reaction to B's interruption (different interruption types).

The perception of A's involvement however, was influenced by the interruption type and strategy in the *question* trial. A possible explanation for this is that a disruptive misunderstanding question is the only interruption in this study that can reveal something about A's quality of interacting; B wouldn't need to interrupt with this strategy if A's utterance was completely clear for B. B's misunderstanding could in this way be a sign of A's lack of involvement.

As hypothesised, in the *question*, *opinion*, and *partner communication* trials cooperative strategies led to higher perceived engagement and involvement levels in B compared to disruptive strategies. This reflects the nature of the interruption strategies: Disruptive strategies "disrupt the process and/or content of the ongoing conversation" (Section 6.2.2) [Murata, 1994, Goldberg, 1990, Li, 2001], which turn out to convey indeed a low level of "wanting to be together with the other participant and continuing the interaction" (definition of engagement [Poggi, 2007]). Coopera-

tive interruptions on the other hand, convey a cooperation between the participants which is a key aspect of engagement (Section 2.3.2.1), and indeed turn out to signal engagement.

No significant effects on engagement and involvement have been found in the *topic* trial. This may be due to the fact that participants in this trial particularly, found it difficult to detect differences among the stimuli, as also reported by comments of 5 out of 18 participants in this trial (e.g. *"I was unable to detect significant differences between the examples aside from timing." "The differences may be too subtle for me to notice."*). Further research would be necessary to find out why this is.

With respect to the agents' mutual interpersonal attitudes, both *type* and *strategy* led to main and interaction effects. Contrary to our expectations, the interruption type was of greater importance than the interruption strategy for both A and B's perceived dominance and friendliness.

In the *question* trial (function) the perceived dominance of A increased and its friendliness decreased as the amount of overlap increased (*type*). B was also perceived less dominant during silent interruption types (no overlap). This suggests that no matter how disruptive or cooperative an interruptive question is, the amount of overlap of the interactants' utterances has a higher impact on the humans' perception of the agents' dominance and friendliness.

In the *opinion* trial (disagreement vs. agreement) *strategy* did have an effect, but only on friendliness (no dominance): A cooperative strategy increases B's friendliness, while this effect for A is also dependent on the type of interruption (silent, no overlap). When B used a disruptive strategy in a silent interruption (no overlap), A' friendliness increased. The fact that strategy does play a role in this trial may be due to the fact that in this trial B's disruptive strategy can be perceived as very hostile (strong disagreement, face threatening) which can make A look much more friendly in comparison. The lack of differences in dominance levels caused either by *type* or *strategy* may be explained by the fact that this is the only function where the interruption forms an adjacency pair with the first speaker's utterance. Both an agreement and a disagreement form an adjacency pair 'assessment-(dis)agreement' (see Section 6.2.2 [Levinson, 1983]) with the first speaker's utterance ("It tells an amazing story"). The first turn makes both an agreement and a disagreement relevant (an assessment asks for a (dis)agreement) [Schegloff, 2007] which can cause that none of both strategies is perceived as much more or less dominant than the other.

In the *partner communication* trial, next to the interruption *type*, B's interruption *strategy* had effects on B's friendliness and A's dominance: A disruptive strategy led to less friendliness of B (more pronounced in silent interruption) and lower dominance of A compared to a cooperative strategy. This may be due to the fact that B's disruptive strategy in this function has a strong interrupting power: It rejects the course of the interaction that A introduced (*"I've read the book so I know the story."*). This could possibly be interpreted as not so much a sign of dominance of B (only of less friendliness), but as a lack of dominance of A. This effect can be explained by the notion of complementarity of interpersonal attitudes in dyads [Markey et al., 2003].

As in the former trials in the *partner communication* trial the interruption *type* had an impact on B's dominance: more overlap, more dominant. Similarly in the *topic* trial the interruption *type* has an opposite effect on B's friendliness: more overlap, less friendly.

In short, the interruption *strategy* is especially important in how human observers (third parties) perceive the interrupter's *engagement* and *involvement*. Primarily the interruption *type* has an impact on the users' perception of *interpersonal attitudes* of both agents (the more overlap the more dominant and less friendly), though in some trials the strategy did have an influence on the interrupter's (and rarely the interruptee's) friendliness and dominance. In silent interruption types (no overlap) the content of the interrupting utterance may have become more noticeable which may be the reason of an increased effect of the interruption *strategy* for this interruption *type*.

We should however be careful in generalising these results outside the conditions of the study; The interrupting utterances that we evaluated are not representative for their respective functions and interruption strategies, and the amount of overlap in the simple interruption is not the only amount of overlap possible for this interruption type.

We also kept the onset of the interruptions (Section 6.4.1.1) constant across conditions to focus on the effects of the interruption types and strategies. Though, manipulating the moment when the interruption occurs during the interruptee's turn may also influence the perception of both agents.

It should be noticed further that in this study we only looked at interruptions in a very small conversation fragment. Some participants found it difficult to judge the agents' by a small fragment (e.g. *"It was not clear how to rate the short segment in the questions."*) while others noticed effects easily (e.g. *"It never occurred to me*

so directly that the timing and content of a sentence completion could have such a profound effect on perception of character."). Further research would be necessary to see if we find the same results when the interruptions are presented in a longer context.

Lastly, in human-human interaction personal characteristics such as the gender and status of the interaction participants can play a role in their interruption behaviour and on the way the participants are perceived [Zimmermann and West, 1996, Beattie, 1981]. We did not vary these variables in our agent-agent study in order to focus solely on the effect of the interruption types and strategies.

6.8 Conclusion

In this chapter we have performed an evaluation study in agent-agent interaction to understand the effects of interruptions on the perceived engagement, involvement, and interpersonal attitudes of the interaction participants. This study has the goal of trying to obtain indications of the optimal coordination of the agent's turn-taking behaviour in human-agent interaction to make the agent appear more engaged and human-like, and thereby engaging the user.

Based on the literature and a validation study we created a set of interruptions that incorporated different interruption types (i.e. amounts of overlap) and strategies (i.e. contents). The interruptions were acted out by two agents. Human observers (third party) of these short interaction fragments judged the agents' engagement, involvement and interpersonal attitudes.

As expected, it turned out that for the perception of the interrupter's *engagement* and *involvement*, it is the interruption *strategy* that leads to significant effects: cooperative strategies are sign of more engagement and involvement than disruptive ones. The interruptions have no effect on the engagement and hardly any effect on the involvement levels of the interrupted agent.

With respect to the humans' perception of the agents' *interpersonal attitudes* it is primarily the interruption *type* that matters: the more overlap the more dominant and less friendly the agents are perceived. The interruption strategy can also influence the perception of the agents interpersonal attitudes but this is dependent on the content of the individual interruptions. When the interruption type has no overlap, the effect of the interruption strategy can increase.

The results can be used in the design of conversational agents in human-agent interaction. For example, the interruption behaviour of the user can update the

agent's mental model of the user ('Theory of Mind') with respect to its engagement and interpersonal attitudes. But also the agent can adapt its interrupting behaviour to convey more or less engagement and dominance for example. We thereby achieved our goal of finding the optimal coordination of the agent's turn-taking behaviour; agent interruption behaviour that conveys the most engagement consists of interruptions with cooperative strategies.

7 Engagement and Preferences

7.1 Introduction

After considering aspects of the agents' politeness behaviour referring to the *form* of the agent's verbal behaviour in Chapter 5 (*how* to say something), and considering interrupting behaviour referring primarily to the *timing* of the agent's verbal behaviour in Chapter 6 (*when* to say something), the present chapter is primarily oriented towards the *content* of the agent's verbal behaviour (*what* to say). Concretely, in this chapter we verify if and how an adaptation of the *topics* in non-task oriented human-agent interaction could favour the user's engagement.

As mentioned before, engagement is crucial in human-agent interaction as it is a prerequisite for the interaction to go on, and thus a prerequisite for the agent to deliver its messages and/or to complete the objectives of the interaction [Bickmore et al., 2010]. As pointed out in Section 2.4, we consider engagement as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction" [Poggi, 2007].

Previous research has demonstrated that a personalisation according to user preferences may contribute to, amongst others, an optimisation of user experience in game playing [Yannakakis and Togelius, 2011], an improvement in customer relationships [Jiang et al., 2008], and an enhancement of learning efficiency and experience (intelligent tutoring systems) [Cha et al., 2006]. In this chapter we explore if in non-task oriented human-agent interaction, an agent's personalisation according to the user's preferences can influence the user's engagement. Specifically, we will verify if the user's preference for a physical object (artwork) plays a role in the user's level of engagement during the discussion of this object with a virtual agent. The outcome of this perceptive study will be used to develop dialogue strategies for the virtual agent aiming at enhancing user engagement in human-agent interaction (Part V).

In the following two sections we specify the type of interaction we look at and

introduce the notion of preferences. In Section 7.4 we present our methodology and in Section 7.5 our results. In Section 7.7 we conclude and discuss our findings.

7.2 Information-Giving Chat

As mentioned in Section 1.2 this thesis is conducted in the context of the project 'A1:1' that aims at developing a human-sized virtual agent playing the role of a visitor in a museum. The agent's task is to engage human users in one-to-one face-to-face interaction about the museum and some of its artworks with the objective of providing the visitor information about these subjects. The choice of the exact subject is secondary: what matters is that some amount of cultural information is transferred. We shall refer to this type of interaction as an *information-giving chat* (as opposed to *information-seeking chat* [Stede and Schlangen, 2004]). Like information-seeking chat [Stede and Schlangen, 2004], information-giving chat has a more exploratory and less task-oriented nature but is more structured than general free conversation.

Our aim is to explore if a personalisation of the topic of conversation based on the user's preferences (explained in Section 7.3) for the artworks under discussion is likely to enhance the user's engagement, thereby augmenting the interaction time and thus the agent's opportunities to transfer cultural information.

7.3 Preferences and Related Work

We interpret a preference according to the definition of Scherer [Scherer, 2005] as "a relatively stable evaluative judgement in the sense of liking or disliking a stimulus". Preferences are everywhere in our daily lives [Kießling, 2002] and the development of personalised content based on preferences increases in multiple domains of human-computer interaction, such as e-commerce, news reading and computer games [Yannakakis and Togelius, 2011].

In the domain of natural language generation user preferences have been taken into account as well, for example, the user's preferences for a communication channel [Landragin, 2013], preferences for the level of detail of an explanation [Wallis and Shortliffe, 1984], and preferences for houses [Carenini and Moore, 2006], restaurants [Walker et al., 2004] and airline flights [Moore et al., 2004] in task-oriented, recommender-style systems [Foster and Oberlander, 2010].

[Garber-Barron and Si, 2013] considered an interactive digital storyteller and

found that subjects younger than 40 prefer stories that are adapted to the user's preferences for certain events in the story.

In the domain we focus on, interactions between a human user and an ECA, user personalisation is also increasing. Variables taken into account include the user's affective state [Prendinger and Ishizuka, 2005, Castellano et al., 2013, Leite et al., 2014, Paiva et al., 2004], appearance [Paiva et al., 2004], expressions of appreciation [Campano et al., 2014] and politeness and formality [De Jong et al., 2008].

In the present work we will verify if we should augment this list of personalisation variables in non-task oriented human-agent interaction with user preferences that are directly linked to the topic of conversation. Our goal is namely, to explore if a personalised topic of conversation according to the user's preferences regarding physical objects is likely to enhance the user's engagement during the interaction.

Some previously built virtual agent systems give their users the opportunity to directly select or reject the topics of interaction [Bickmore et al., 2011][Kopp et al., 2005], thereby adapting the interaction to a choice of the user for certain information. However, the systems do not take into account the user preferences that may underlie the user's choice. In our work we will verify if the user's preferences towards the physical objects under discussion play a role in the user's interest for the discussion and his/her engagement. This will give us indications for the development and usefulness of dialogue strategies aiming at agent initiated topic selection (described in Part V).

7.4 METHODOLOGY

In order to find out if the user's preference for a museum object (artwork) plays a role in the user's level of engagement during the discussion of this artwork with a virtual agent we performed a perceptive study. In this study we collect data about the users' self-reported preferences for artworks and compare them with the users' own estimations of their engagement (because preferences and engagement are part of the mental state of the user) during the different discussion phases of their interaction with a virtual agent. The study consisted respectively, for each evaluation participant, of a visit of a small improvised museum, a conversation with a virtual agent called Leonard, and the completion of a questionnaire. Below we briefly discuss each of these steps.

Figure 7.1 : The improvised museum.

7.4.1 Museum

Since at the time of this study virtual agent Leonard that is used in project A1:1 (detailed in Chapter F) is not yet installed in the museum, we simulated a small museum in our laboratory; We exhibited 4 pictures of existing artworks in a first room and gave each participant as much time as he/she needed to observe the artworks, just as they would do in a regular museum (Figure 7.1). The artworks are shown in Appendix C Figure C.1 and were chosen as to vary in style: a photo of the exhibition of *Balloon Dog* by Jeff Koons, and printed images of the paintings *The Kiss* by Gustav Klimt, *Composition A* by Piet Mondrian, and *The Anatomy Lesson of Dr. Frederick Ruysch* by Jan Van Neck. When the participant indicated that he/she finished looking at the artworks we explained that the visit would continue in the next room and that there he/she will talk with Leonard, a virtual character who also visits the museum. We placed another artwork between the screen of the virtual agent and the user that serves as a first conversation topic (Figure 7.2): a picture of a statue named *Soldier drawing his Bow*, by Jacques Bousseau (Appendix C, Figure C.1a).

7.4.2 Virtual Agent and Interaction

A technical limit for interacting with the agent (Leonard) is that at the moment of this study we did not dispose of reliable speech recognition and natural language understanding modules. To resolve these issues we used a Wizard of Oz: as shown in Figure 7.3 we predefined keywords with which the participants needed to formulate

Figure 7.2 : The 'statue' between the agent and the user.

Figure 7.3 : Leonard and the keywords that are needed to talk with it.

their reaction and then transmitted the user's choice for that keyword to the agent. The keywords represent branches in the agent's predefined dialogue tree, meaning that the use of each keyword can lead to a particular predefined agent response. The dialogue tree is developed by using the hierarchical task network Disco for Games [Rich and Sidner, 2012]. (More details about this tool will be given in Chapter 10.)

As soon as the human participant entered the room where the agent was located, the agent started the interaction. The agent Leonard has the appearance of a cartoon-like version of a man of about 70 years old, is displayed on a 75-inch vertically placed screen giving it the size of a human, speaks French, and does gestures. The user was recorded with two kinects and one camera (see Figure 7.4).

Figure 7.4 : The setting of the interaction.

In the first couple of turns Leonard presented itself and asked for the name and region of the user. This small talk phase (in the sense of [Bickmore and Cassell, 1999]) serves to let the user get used to the character and the way of interacting [Bickmore and Cassell, 1999].

After the small talk phase Leonard started talking about the artwork in front of it (placed between the agent and user, Figure 7.2) as a way to open the conversation about the artworks. Then, the agent switched to discuss the other artworks of the improvised museum. For every artwork Leonard provided some information about the object, asked what the participant thinks of it, and optionally expressed its own opinion (as described in [Campano et al., 2015b]). The order in which the 4 artworks from the first room were discussed was random and changed among the participants. In this way possible effects of novelty [Leite et al., 2014] on the user's engagement during these discussion phases are outbalanced. After the discussion of all artworks Leonard ended the interaction. Depending on the user utterance lengths the entire interaction took between 6 and 10 minutes.

7.4.3 Questionnaire

Directly after the interaction the evaluation participants were presented a questionnaire. Similarly to [Garber-Barron and Si, 2013] we posed several questions to acquire self-report measurements of the user's engagement during the different phases of the preceding interaction. For this dimension we used the same (here French) questions that measured engagement in the previous evaluation studies

(Chapter 5 and 6) which are based on the definition of [Poggi, 2007] (Section 2.3):

Q1. to what extent the user wanted to be together with Leonard during the different discussion phases;

Q2. to what extent the user wants to continue the interaction during the different discussion phases;

The different discussion phases for which these questions were asked were: the small talk phase (name and region), and each separate discussion around an artwork. All the questions needed to be answered on a 7-point scale ranging from *not at all* to *extremely*. We also asked:

Q3. to what extent the user was *interested* in the discussion during the different phases;

In order to estimate the *preferences* of the users we asked the participants

Q4. to what extent the user likes the different artworks (degree of liking following Scherer's definition of preferences [Scherer, 2005], Section 7.3);

Q5. and to what extent the user finds the different artworks interesting.

7.5 Results

33 participants took part in the study (13 female, aged 19-58, all proficient in French). Analyses of the data show that the participants' *degrees of liking* (i.e. preference for) an artwork (Q4) is significantly, positively correlated to the users' self-estimated *engagement* during the discussion of that artwork with Leonard (Kendall Tau tests). This is the case for both of the aspects of engagement we looked at: wanting to be together with Leonard (Q1, $p < 0.001$, $\tau = 0.50$) and wanting to continue the interaction (Q2, $p < 0.001$, $\tau = 0.52$). These results are obtained by taking, for all the participants, and all the 4 (in random order discussed) artworks from the first room, the participants' degree of liking the artwork (Q4), and comparing this with the scores that the participants attributed to their engagement during the corresponding discussion phases of all these artworks.

In the same way as above, a positive correlation is found between the extent to which the users found an *artwork interesting* (Q5) and their level of both engagement measurements during the discussion of this artwork (Q1, $p < 0.001$, $\tau = 0.45$; Q2, $p < 0.001$, $\tau = 0.54$). The users' liking (Q4) and interest (Q5) for an artwork are also positively correlated to the extent to which the participants found the *discussion of the artwork interesting* (Q3) (Q4, $p < 0.001$, $\tau = 0.56$; Q5, $p < 0.001$, $\tau = 0.49$).

For all of the above results, possible effects of novelty [Leite et al., 2014] are outbalanced by the random order in which the artworks were discussed. The discussion of the artwork that is located between the agent and the user (Figure 7.2) and that is always discussed first, before the in random order discussed museum objects (Figure 7.1), does not lead to a significant different level of engagement or interest in the discussion than the other artworks (Kruskal-Wallis).

7.6 Discussion

The results of the study confirm our hypothesis that the user's preference for the physical object (artwork) under discussion plays a role in the user's level of engagement in non-task oriented human-agent interaction: the higher the user's preference for the object, the higher the user's engagement during the discussion of this object with a virtual agent. This conclusion is confirmed by comments from the participants: *"This (answer) shows that wanting to stay with Leonard depends on how much I find the topic interesting." "Talking about a work that I don't like is not pleasant and makes me less involved in the interaction".*

We found no significantly different level of engagement or interest during the discussion of the object that was physically present during the interaction in comparison with the objects that were located in another room. This means that this study does not give us reasons to suspect that the physical presence of an object during the interaction is required in order to engage the user. It also does not give us reasons to expect (in interactions of 6-10 minutes) that the user is much more engaged during a first object discussion (at the beginning of the interaction), than during object discussions later on in the interaction.

It should be noticed that in this study we only considered a link of the user's preferences and the user's temporal engagement during the discussion phase of the object under discussion. Further work would be necessary to verify the consequences of adapting the topic of conversation to the user preferences on the long time; during the entire interaction and the overall engagement level of the user.

Besides, we only looked at estimations of the users' own engagement. The present work does not tell if the users also communicate their engagement through their behaviour, which would make it possible for external observers and/or systems to recognise the levels of user engagement (and thereby infer the user's preference towards the object under discussion).

We also did not consider here to what extent, and how exactly, a personalisation

of content *within* the discussion phase of an artwork (object) may influence the user's engagement further.

7.7 Conclusion

In this chapter we have described a perceptive study to find out if the user's engagement in human-agent interaction is correlated with the user's preferences for the artworks that are discussed in the interaction. For this, human participants respectively visited an improvised museum, interacted face-to-face with a human-size agent using natural language speech (Wizard of Oz), and filled in a questionnaire.

The results of the study show that the user's degree of liking of (i.e. preference, Section 7.3) and interest in a museum artwork is indeed significantly, positively correlated with the user's engagement and interest during the discussion of this artwork with a virtual agent.

From this finding we can derive that one of the agent's dialogue strategies that is likely to favour user engagement in human-agent interaction is personalising the topics in information-giving chat to the user's preferences for the different physical objects. The other way around, the revealed positive correlations also indicate that by detecting a level of user engagement during human-agent interaction (assuming that the user communicates its engagement level through its behaviour) we can obtain indications with respect to the user's preference towards the current topic of conversation and its underlying object. This would allow a system to infer user preferences from the user's engagement, without needing to ask for the user's preferences explicitly (as opposed to e.g. [Edwards and Barron, 1994]).

Both of these conclusions have lead us to develop an agent model that tries to enhance the user's engagement by personalising the content (i.e. topics) of the human-agent interaction. We describe this model in Part V of this thesis.

PART V : MODELLING STRATEGIES TO FAVOUR ENGAGEMENT: TOPIC MANAGER

TELECOM ParisTech

8 Topic Selection Model

8.1 Introduction

In order to favour the user's engagement level previous research manipulated the agent's non-verbal behaviour including gaze [Peters, 2005], gestures [Sidner and Lee, 2003], postures [Peters, 2005] and facial displays [Bohus and Horvitz, 2009b], as well as some aspects of the agent's verbal behaviour, such as expressions of empathy [Bickmore et al., 2013] and prosody [Foster, 2007] (see Chapters 3 and 4). In Part IV we explored how other aspects of verbal behaviour may have an effect on, or be a sign of, engagement (i.e. 'the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction" [Poggi, 2007]): We looked at politeness strategies, interruptions and user preferences. In this chapter we build further on the latter observations described in Chapter 7, by proposing a model where the agent initiates discussion topics that are adapted to the user's preferences and/or the agent's own preferences.

Conversational agents often employ a strict task-oriented dialogue structure in order to achieve the particular task for which they are built. Chat-based systems on the other hand, allow for less rigid interaction but the agent has less control of the topic of the interaction. We are interested in dialogue that falls in between these categories: Where there is not a clear task to achieve and where the interaction is not completely open either, but where there is freedom of topic choice within a certain domain. The interaction is not task-oriented but engagement oriented. For this type of interaction we present in this chapter a topic selection model for a conversational agent that tries to favour engagement. The model gives the agent the ability to select dialogue topics that are adapted to the user.

By taking into account 1) the goal of wanting to favour the user's engagement, 2) the agent's perception of the user's engagement, and 3) the agent's own mental state including the agent's preferences and associations, we do not consider the dialogue merely from a user-oriented system point of view, but model the agent as a human-

like interaction participant that contributes to the interaction from its own point of view. In Section 8.4 we will detail the exact interpretation of the variables that the agent takes into account.

In the following section we will first further specify the type of interaction and topics we are looking at. In Section 8.3 we present related work and in Section 8.4 we introduce the variables that will be taken into account in the topic selection model. In Section 8.5 we present the topic selection model itself, and in Section 8.6 we concludes our findings.

In this chapter we focus only on the agent's selection of the topics of interaction. This topic selection component (of the topic manager) decides whenever the agent needs to introduce a new topic in the interaction, which one to choose. In the following chapters of this thesis we will further propose a way of introducing the selected dialogue topics on a dialogue level and implementing the corresponding agent behaviour in a virtual agent platform. All aspects (chapters of Part V) together form a topic manager.

8.2 Context

While the topic selection model presented in this chapter can be applied to multiple domains we will illustrate it for the domain for which we developed it originally: an agent that tries to engage human users in one-to-one face-to-face interaction about a museum and some of its art objects with the objective of giving the visitors information about these subjects (project 'A1:1'). The choice of the exact subject is secondary: what matters is that some amount of cultural information is transferred. As mentioned before (Section 7.2), we refer to this type of interaction as an *information-giving chat* (as opposed to information-seeking chat [Stede and Schlangen, 2004]). The freedom that the broad interaction task of information-giving chat provides makes that locally the interaction is not strictly task-driven but can be driven by social variables of the interaction. The topic manager we propose (in the entire Part V) allows an agent to adapt the topics of the interaction on the fly to favour the user's engagement.

The information-giving chat we consider for the topic selection model is agent-initiated in order to increase the likelihood of understanding the user's contributions. The agent cannot anticipate all possible user contributions. By directing the interaction however, it can anticipate a list of user reaction types that are most probable at that moment in time. Due to limitations of speech recognition and natural language

understanding modules it is therefore also the agent who introduces (initiates) the topics in the interaction.

The notion of *topic* in interactions can mean different things [Brown and Yule, 1983]. We define topic from a discourse perspective as "what is being talked about in a conversation" [Brown and Yule, 1983]. In the context of the information-giving chat defined above, each topic refers to the discussion phase of an artwork in the museum [O'Donnell et al., 2001]. In the model each topic is thus associated to a fragment of conversation (similar to [Macias-Galindo et al., 2012]) consisting of at least 1 pair of agent-user turns. Subtopics are subfragments of these larger conversation fragments and discuss a particular aspect of the artwork. For example, the artist of the artwork or the historical period during which it was created.

8.3 Related Work

Outside the domain of human-agent interaction, for example in automatic description generation systems, content has been adapted to the user's preferences by means of recommendation-like systems [Carenini and Moore, 2006, Walker et al., 2004, Moore et al., 2004]. We take some inspiration from recommendation systems, such as the accumulative nature of the preferences for an object's characteristics to determine the overall preference for the object (Multiattribute Utility Theory [Clemen and Reilly, 2013]). However, our model is not only based on the presence or absence of an object's characteristics, but will also take into account the agent's associations between the characteristics and objects, and considers the preferences of both the agent and the user, as we will explain further in Section 8.4.

[Garber-Barron and Si, 2013] modelled an interactive digital storyteller consisting of text and pictures that adapts the events of the story to the user's preferences derived from selected user responses in reaction to events of the story. It selects subsequent events of the story by balancing the presentation of new information with information that the user has previously expressed an interest in. Again, our model differs because our interaction is not solely user oriented, but also takes into account an agent that has its own preferences and associations. Moreover the application and goal of the interactions differ: linear digital story telling versus more flexible engagement-driven information-giving chat by an ECA. In information-giving chat all the presented information is new compared to the previous events of the interaction since providing new information is the goal of the interaction.

As pointed out in Chapter 7 Section 7.3, in human-agent interaction with an ECA

different efforts are oriented towards a personalisation of the agent's behaviour to the user (e.g. [Leite et al., 2014, Paiva et al., 2004, De Jong et al., 2008]). To our knowledge, so far no model exists that makes the agent adapt the topics of non task-oriented interaction to the user's preferences for physical objects in order to favour the user's engagement during the interaction, as motivated in Chapter 7.

We have seen that some previously built virtual agent systems did give their users the opportunity to directly select or reject the topics of interaction [Bickmore et al., 2011, Kopp et al., 2005], thereby adapting the content of the interaction to the user. However, these systems only offer the user a choice for certain information. They do not present a conversational virtual agent than can select interaction topics itself based on dynamic social variables in the interaction.

In order for the agent to be able to select appropriate interaction topics itself, it needs to dispose of a domain knowledge representation mapping to the possible discussion topics. Several dialogue systems dispose of some kind of representation of domain knowledge, developed for various modules such as natural language understanding [Milward and Beveridge, 2003], topic tracking [Carlson and Hunnicutt, 1996, Jokinen and Wilcock, 2012], question-answering [Agostaro et al., 2005], response generation [Pilato et al., 2011], surface realization [Milward and Beveridge, 2003], and the selection or generation of dialogue topics [Chakraborty et al., 2007, Macias-Galindo et al., 2012, Stede and Schlangen, 2004]. We are interested in the latter where domain knowledge is organised in such a way that it represents (potential) interaction topics.

The topic representations can be divided in specific task-oriented models [Chakraborty et al., 2007] and non task-oriented models. As information-giving chat has a less task-oriented structure (Section 8.2) we focus on the latter category. In this category [Macias-Galindo et al., 2012] use a semantic relatedness mechanism to transition between conversational snippets in an agent that engages in chatty dialogue, and [Stede and Schlangen, 2004] use an ontology-like topic structure that makes the agent produce coherent topic follow-ups in information-seeking chat. However, these systems do not take into account the user's engagement during the different discussion phases (topics). They are merely oriented towards dialogue coherence and are therefore not sufficient for an optimisation of engagement by topic selection.

[Song et al., 2009] do take into account the user's interests in that they decide when the agent should switch topic. The system decides about the timing of a topic change, and then extracts a new topic from the web. For the selection of the topics

themselves the user's preferences are however not taken into account.

By using some concepts of the models described above, we aim at building a topic structure in the agent's mind to retrieve dynamically, during human-agent information-giving chat, engaging interaction topics. Opposite to existing topic selection systems that have focused exclusively on dialogue coherence, our topics will be generated from an agent perspective: The topic structure is representing a part of the agent's knowledge, which is located within the agent's mind, and the agent's objective is to constantly favour engagement. As such the topic selection will include human-like features by taking into account the agent's dynamically updated perception of the user's engagement, the agent's preferences and the agent's associations with respect to the current topic of conversation. In the section below we define these variables.

8.4 Variables for Topic Selection

8.4.1 User Preferences and Engagement

In order to select engaging discussion topics, the agent needs to be able to *predict the user's engagement* level during the discussion of so far unaddressed topics (objects). For this we needed to know if there are any underlying observable *preferences* that can help the agent collect indications with regard to its prediction of the user's engagement. We interpret a preference as "a relatively stable evaluative judgement in the sense of liking or disliking a stimulus" [Scherer, 2005]. Since a topic of conversation in our interaction setting corresponds to the discussion of a particular artwork, we verified in the previous Chapter (7) by means of a perceptive study if there exists a relation between the user's engagement level during the discussion of an artwork with a virtual agent, and the user's preference for the physical artwork that is discussed.

The evaluation study in Chapter 7 has confirmed that the user's preference for an artwork (museum object) is significantly, positively correlated with the user's engagement during the discussion of this object with a virtual agent (schematically shown in Figure 8.1). From this finding we can derive that the agent can obtain an indication of the user's preferences for a physical object from the user's engagement at the moment this object is addressed in the interaction. The characteristics of this physical object allow the agent to make predictions of the user's preferences for other, so far non-addressed, physical objects (artworks). These predicted preferences

can give an indication of the user's future engagement during the discussion of these objects. Cutting this process short, the agent's perception of the user's engagement during the discussion of an object, and the characteristics of this physical object, allow the agent to make *predictions of the user's future engagement* for so far unaddressed objects. The agent can then use this prediction to select an engaging topic of conversation.

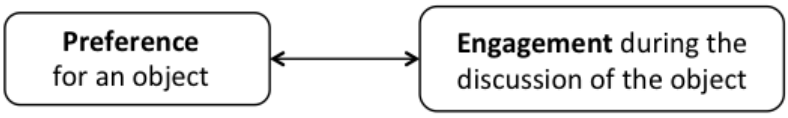


Figure 8.1 : A human's preference and engagement.

8.4.2 Agent Preferences and Engagement

To represent human-like features in an agent that plays a museum visitor, the agent needs to have its own preferences for the artworks as well, as representing *agent preferences* is fundamental for any agent model [Casali et al., 2011]. Besides, the preference representation of the agent can be used to express (consistent) agent appreciations, which has shown to significantly favour the user's engagement [Campano et al., 2015a].

Following the correlation we found above (Figure 8.1) an agent likes to talk most about its preferred topics as those maximise its own engagement. However, the agent we model also wants to engage the user. The agent thus tries to optimise the engagement level of both the user and the agent itself (from here onwards indicated as *combined engagement*). To achieve this, for each (sub)topic (object and characteristic, Section 8.2) that can be addressed the agent calculates an *expected (predicted) level of combined engagement* and selects the (sub)topic with the highest score (optimisation) as next (sub)topic of the interaction. In this way the agent selects a new (sub)topic of conversation based on a combination of the agent's own preferences for the artworks and its prediction of the user's level of engagement during the discussion of the artworks. Figure 8.2 shows this relation.

8.4.3 Associations between Topics

A last human-like variable that needs to be represented when the agent selects a topic of conversation in information-giving chat are its own associations between

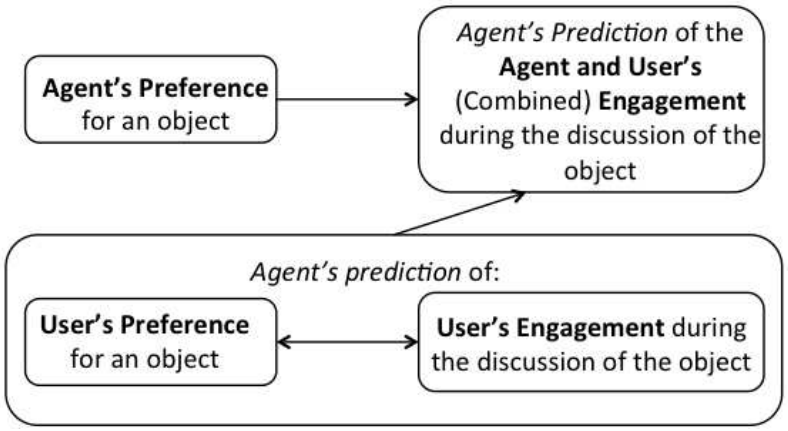


Figure 8.2 : The agent's prediction for the level of combined engagement during the discussion of an object depends on several variables.

topics. We integrated this concept in the topic selection component (of the topic manager) because in a human mind "events that share meaning or physical similarity become associated" [Dellarosa, 1988]. "Activation of one unit activates others to which it is linked, the degree of activation depending on the strength of association" [Dellarosa, 1988]. To imitate this human process the topic selection component ensures that the discussion of one topic can be associated with other topics in the agent's mind. The "degree of activation depending on the strength of the association" is represented by the similarities of the physical artworks that underlie the various topics. In this way the agent has a strong association between for example two topics that both discuss an abstract painting. Its association between the discussion of an abstract painting and the discussion of a realistic statue for example, is much weaker, which makes that here the activation of the second topic triggered by the first is very unlikely. In the topic selection model we will thus represent the agent's associations by similarity scores between every pair of physical objects underlying two topics (detailed in Section 8.5.1).

The associations (based on object similarities) allow the agent to make predictions about the user's engagement during so far unaddressed topics: When the user has a certain engagement level during the discussion of the current topic (and thereby a related preference towards the current object under discussion (Section 8.4.1)), similar conversation topics are expected to have similar levels of user preference and are thus expected to lead to similar levels of user engagement (Figure 8.1). The topic selection model detailed in the following section ensures that when the agent's predicted user engagement level for an associated topic is high enough (in combi-

nation with the agent's own preferences) it is a potential new topic of conversation, triggered by the agent's associations.

8.5 Topic Selection Model

Incorporating the variables described above (Section 8.4), in this section we present the topic selection model of the topic manager. We first define the model's knowledge base that contains all the agent's knowledge that is required to select (sub)topics for the interaction. In Subsection 8.5.2 we describe the initial state of the knowledge base that leads to the first (sub)topic of the interaction, and in Section 8.5.3 we describe how the knowledge base is updated and subsequent (sub)topics emerge. In Subsection 8.5.4 we explain how a detected engagement measurement is processed, and in Subsection 8.5.5 we give an example of the topic selection procedure.

Table 8.1 gives an overview of all the variables of the topic selection component, which will be defined throughout this section.

Table 8.1 : Overview of the variables in the topic selection model.

Variable	Description
	Variables of the knowledge base with fixed values:
G	Knowledge base graph
O	Set of all objects (nodes, mapping to topics)
S	Set of all similarity scores between objects (topics)
o_i	Object (node, mapping to a topic)
o_i^c	The c^{th} characteristic of object o_i (mapping to a subtopic)
$Char(c)$	Set of all possibles values for a given characteristic of index c (e.g. *artist*)
$\texttt{SIM}(o_i, o_j)$	Similarity score between objects o_i and o_j
$\underset{c}{\texttt{sim}}(o_i, o_j)$	Similarity score between the values of the c^{th} characteristic (e.g. *artist*) of objects o_i and o_j (e.g. between *Mondrian* and *Van Gogh*)
$\texttt{PREF}_a(o_i)$	Agent's ($_a$) preference for object o_i
$\underset{c}{\texttt{pref}}_a(o_i)$	Agent's preference for the value of the c^{th} characteristic of o_i (e.g. preference for *Mondrian*)

Variables of the knowledge base with continuously updated values:

$\texttt{ENG}^{*}_{u}(t+1, o_i)$	Predicted (*) level of user ($_u$) engagement during a potential discussion of object o_i (topic o_i) at time $t+1$
$\underset{c}{\texttt{eng}}^{*}_{u}(t+1, o_i)$	Predicted level of user engagement during a potential discussion of the c^{th} characteristic of o_i (subtopic c of topic o_i) at time $t+1$
$\texttt{ENG}^{*}_{a+u}(t+1, o_i)$	Predicted level of combined engagement (agent and user engagement $_{a+u}$) during a potential discussion of object o_i (topic o_i) at time $t+1$
$\underset{c}{\texttt{eng}}^{*}_{a+u}(t+1, o_i)$	Predicted level of combined engagement (agent and user engagement $_{a+u}$) during a potential discussion of the c^{th} characteristic of o_i (subtopic c of topic o_i) at time $t+1$
$w(t)$	Ratio (weight) of the agent's orientation towards itself (agent preferences) versus the user (user engagement) at time t
$\mathit{Recent}\texttt{ENG}^{obs}_{u}(t, o_i)$	Observed recent engagement of the user (during the current topic, i.e. currently addressed object o_i) at time t
$\texttt{ENG}^{obs}_{u}(t, o_i)$	Observed overall user engagement level during the current topic (currently addressed object o_i) at time t
$\underset{c}{\texttt{eng}}^{obs}_{u}(t, o_i)$	Observed overall user engagement during subtopic c of the current topic (currently addressed object o_i) at time t

Input:

$\overset{meas}{\texttt{Eng}}_u\ (t)$	User engagement measurement (detection) at time t
$\mathit{Cert}(\overset{meas}{\texttt{Eng}}_u\ (t))$	Certainty (i.e. confidence) of the user engagement measurement $\overset{meas}{\texttt{Eng}}_u\ (t)$

Output:

Topic	The topic (i.e. object o_i) that is selected to be addressed in the interaction

$Subtopics(o_i)$	The ordered subtopic(s) of a selected topic o_i (i.e. characteristics $(o_i^{\sigma(1)}, \ldots, o_i^{\sigma(k)})$) that are selected to be addressed in the interaction

8.5.1 Knowledge Base

The topic selection component of the topic manager selects the topic and subtopics that are most likely to optimise the combined engagement level of the user and the agent at that moment in the interaction. As mentioned before, in our application, the topic corresponds to the discussion of an artwork, and a subtopic of a discussion to the artwork's characteristic.

As explained in Section 8.4.2 the topic selection component tries to select (sub)topics for the interaction that optimise the combined engagement (agent and user engagement). For this, the topic selection component continuously calculates for each potential (sub)topic, the expected (predicted) engagement of the agent and user together (combined engagement) during the potential future discussion of the (sub)topics. To calculate these predicted combined engagement levels for each potential topic and subtopic, in the spirit of [Stede and Schlangen, 2004] we define an ontology-like model of domain knowledge holding the conceptual knowledge and dialogue history (knowledge base). The model is part of the agent's knowledge and dynamically enriched with information representing the variables described above (Section 8.4).

The topics all consist of artwork discussions and are therefore not hierarchically ordered in the knowledge base but are represented in a weighted non-directed graph $G = (O, S)$ (see Figure 8.3 for an example) where:

$O = \{o_i, i \in \{1, \ldots, N\}\}$ is the set of all N nodes (objects that map to topics).

$S = \{\texttt{SIM}(o_i, o_j), i, j \in \{1, \ldots, N\}, i \neq j\}$ contains the weights (i.e. similarity scores) attributed to the vertices that connect all objects to each other (detailed below).

Each node o_i is a k-components vector, describing k characteristics (attributes) of the corresponding object (artwork). For instance, if $k = 4$, $o_i = (o_i^1, o_i^2, o_i^3, o_i^4)$, where each entry describes a characteristic, for example the object's *artist: "Van Gogh"*; *period: "19th century"*; *style: "post-impressionist"*; and *genre: "portrait"*.

Each of these object characteristics map in the topic manager to a potential subtopic (objects map to topics).

We denote $Char(c)$ as the set of all possible values for a given characteristic of index c (corresponding to e.g. *artist*):

$$o_i^c \in Char(c), c \in \{1, \ldots, k\}, i \in \{1, \ldots, N\}$$

where k is the number of characteristics for any object o_i.

The weights of the edges that link all objects (nodes) to each other represent object similarities in the agent's mind and are expressed by similarity scores $\texttt{SIM}(o_i, o_j)$ with range [0,1]. The similarity scores are responsible for the possible *associations of the agent* (explained in Section 8.5.3).

Besides object similarities, characteristics are also linked in the agent's mind by means of similarity scores, to other characteristics of the same type (e.g. *artist: "Mondrian" – "Van Gogh"*). We denote $\underset{c}{\texttt{sim}}$ (with range $[0, 1]$) as the constant similarity function between the c^{th} characteristic of two objects o_i and o_j:

$$\underset{c}{\texttt{sim}}(o_i, o_j) = f_c(o_i^c, o_j^c)$$

where f_c is an arbitrarily defined function, independent of the objects o_i, o_j, meaning that the similarity score between two characteristics is the same independently of which objects have these characteristics.

The similarity scores between the objects (topics) are derived from all the separate similarity scores assigned to the objects' characteristics (subtopics). Similarity scores between objects (with range $[0, 1]$) are thus a constant function of the similarity scores between the objects' characteristics:

$$\texttt{SIM}(o_i, o_j) = F(\underset{1}{\texttt{sim}}(o_i, o_j), \ldots, \underset{k}{\texttt{sim}}(o_i, o_j))$$

where F is also an arbitrarily defined function. The freedom of setting the similarity scores between characteristics ($\underset{c}{\texttt{sim}}(o_i, o_j)$) to any value desired (between 0 to 1) allows for different agent associations and thereby different agent types. The function that describes the similarity scores between the artworks ($\texttt{SIM}(o_i, o_j)$) can also reflect different agent types. For example, to model an agent that is particularly focused on historic aspects of the artworks, the similarity weights of the objects' characteristic *period* can be assigned a larger impact on the similarities between the objects than the other characteristics of the objects.

For every object (node, topic) and characteristic (subtopic) the agent has its own preferences, which are fixed during the entire interaction (i.e. do not change over

the course of the interaction). The *agent's preference* for a characteristic is defined as:

$$\underset{c}{\texttt{pref}}_a(o_i) = g_c(o_i^c),$$

where g_c is a function independent of the object o_i, whose output lies between 0 and 1. Following the definition of preferences (Section 8.4), 0 corresponds to no liking and 1 to a maximum liking of the characteristic.

Optionally, the agent's preference for an object (topic) can be directly derived from the agent's preferences for the artwork's characteristics:

$$\texttt{PREF}_a(o_i) = \begin{cases} G\left(\underset{1}{\texttt{pref}}_a(o_i), \ldots, \underset{k}{\texttt{pref}}_a(o_i)\right), \text{ if accumulative preferences for } o_i \\ \tilde{G}_{o_i} \text{otherwise} \end{cases}$$

where G and $\tilde{G}$ are arbitrarily defined functions of respectively the agent preferences for the object's subtopics and for the object itself. Besides the similarity scores (agents associations), the agent's preferences for characteristics and objects can thus also reflect different types of agents (as recommended by [Amgoud and Parsons, 2002]). Agent preferences can also be set for example, to values that are close to the users' preferences in previous interactions. As shown above, the agent's preferences for the objects can be directly related to the sum of its preferences for the characteristics of the object (as for example in Multiattribute Utility Theory [Clemen and Reilly, 2013]) or not. Attributing more importance to the preference for one characteristic than for another can reflect a particular focus of the agent towards a certain characteristic.

All above values in the knowledge base are constant. Further variables of the knowledge base are continuously updated during the ongoing interaction (see also Table 8.1). Firstly, the agent's knowledge base holds for every object and characteristic a continuously updated *predicted level of the user's engagement* during the potential future discussion (at time $t+1$) of these objects (i.e. topics) and characteristics (i.e. subtopics):

For each object o_i: $\texttt{ENG}_u^*(t+1, o_i)$

For each characteristic c of the object o_i: $\underset{c}{\texttt{eng}}_u^*(t+1, o_i)$ (independent of the object o_i, meaning that the predicted user engagement levels for all subtopics of topics that address the same characteristic value (e.g. *Van Gogh*) are equal)

ranging from 0 to 1, where 0 refers to the minimum level of engagement to continue the interaction and 1 refers to the maximum level of engagement.

Finally, the agent's knowledge base holds continuously updated predicted future $(t+1)$ levels of *combined* (user and agent) *engagement* for the discussion of:

Each object o_i: $\texttt{ENG}^*_{a+u}(t+1, o_i)$

Each characteristic c of the object o_i: $\texttt{eng}^*_{c\ a+u}(t+1, o_i)$ (independent of the object o_i, meaning that the predicted combined engagement levels for all subtopics of topics that address the same characteristic value (e.g. *Van Gogh*) are equal)

ranging from 0 to 1. Figure 8.3 shows an example of the knowledge base graph with all the variables described in this Section.

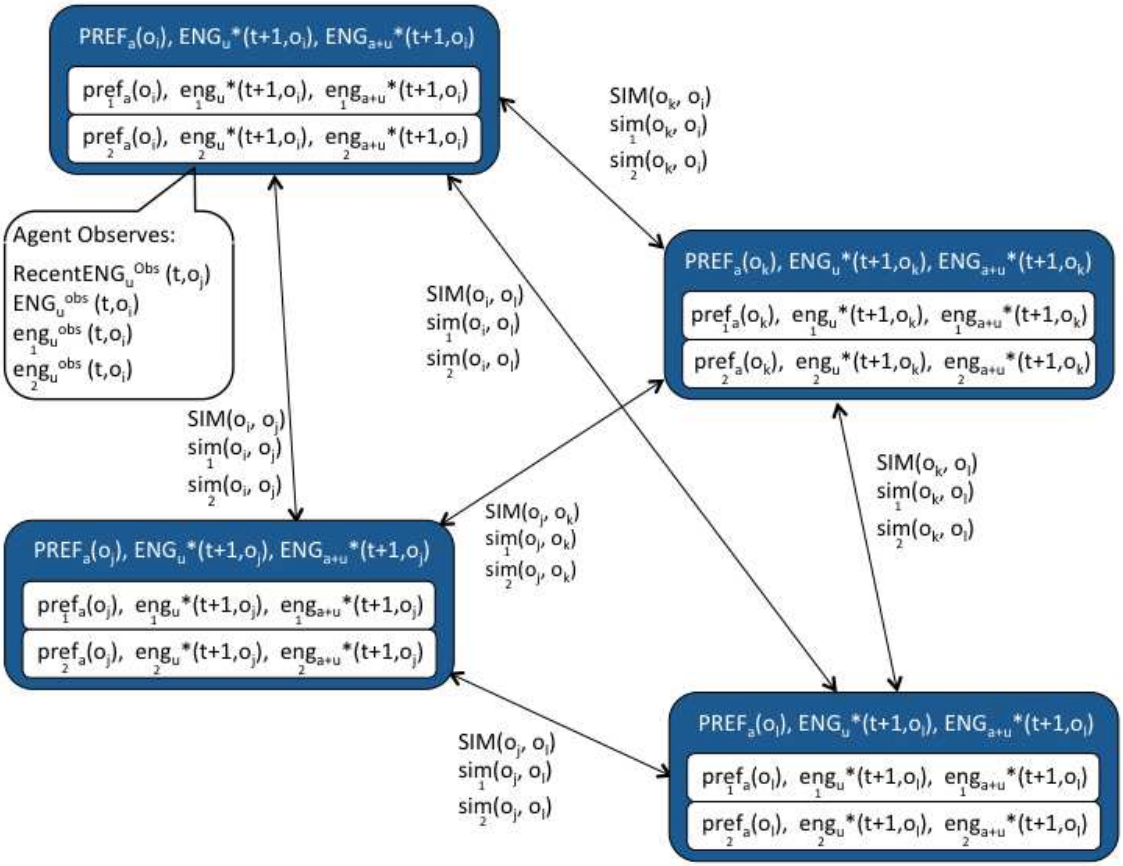


Figure 8.3 : The graph representing an example of the knowledge base in the agent's mind at a time t. Each blue box represents an object (topic) and each white box within a blue box a characteristic (subtopic). o_j is the current object under discussion.

At any time t, the agent's *predicted level of combined engagement* for time $t+1$ during the potential discussion of any o_i is a trade-off between the *agent's preferences*

for the object and the agent's *predicted user engagement* for the discussion of the object (as explained in Section 8.4.2):

$$\texttt{ENG}^*_{a+u}(t+1, o_i) = w(t) \cdot \texttt{PREF}_a(o_i) + (1 - w(t)) \cdot \texttt{ENG}^*_u(t+1, o_i) \tag{8.1}$$

where $w(t)$ is the ratio indicating to what extent the agent values its own preferences in comparison to the user's engagement at moment t. A similar equation holds for the predicted engagement of any subtopic c:

$$\underset{c}{\texttt{eng}}^*_{a+u}(t+1, o_i) = w(t) \cdot \underset{c}{\texttt{pref}}_a(o_i) + (1 - w(t)) \cdot \underset{c}{\texttt{eng}}^*_u(t+1, o_i) \tag{8.2}$$

8.5.2 Initial State: First (Sub)Topic(s)

The initial state of the knowledge base used for the agent's topic selection contains all the objects and characteristics that are known to the agent; for these objects (with at least one characteristic), dialogue fragments have been created.

For all $i, j \in 1, \ldots, N$ and for all $c \in 1, \ldots, k$: *the agent's preferences*, $\underset{c}{\texttt{pref}}_a(o_i)$, $\texttt{PREF}_a(o_i)$, and *similarity scores*, $\underset{c}{\texttt{sim}}(o_i, o_j)$, $\texttt{SIM}(o_i, o_j)$, can be initialized at any value between 0 and 1 (as defined in the subsection above).

At the beginning of the interaction (time t_0) the *prediction of the user's engagement level* during the potential discussion of respectively objects and characteristics at time t_0+1, $\texttt{ENG}^*_u(t_0+1, o_i)$ and $\underset{c}{\texttt{eng}}^*_u(t_0+1, o_i)$, can be set to any value, leading to different types of agents and interactions. If the agent has no additional clues about the user's preferences before the interaction starts we propose to set the predictions of the user's future (time $t_0 + 1$) engagement level for all (sub)topics to 0.5, which represents equal, intermediate engagement values.

With the initial user engagement predictions and the agent's preferences for the underlying objects, the topic selection component uses Equation 8.1 to calculate a *predicted combined engagement* level for each topic $\texttt{ENG}^*_{a+u}(t_0 + 1, o_i)$ and subtopic $\underset{c}{\texttt{eng}}^*_{a+u}(t_0 + 1, o_i)$). For this, $w(t_0)$ can be set to any value between 0 and 1, representing the trade-off that the agent makes at time 0 between its own preferences and the user's engagement.

The topic selection component then selects the first topic of the interaction in the same way as the subsequent topics: out of the set of topics that are equal to or surpass a threshold e of predicted combined engagement it selects the topic that is most likely to optimise the combined engagement level at that moment in the

interaction:

$$Topic = \begin{cases} \arg\max_{i \in \{1,\ldots,N\}} \left(\texttt{ENG}^*_{a+u}(t_0+1, o_i) \right), \text{ if } \texttt{ENG}^*_{a+u}(t_0+1, o_i) \geq e \\ \text{no engaging topic, otherwise} \end{cases} \tag{8.3}$$

where e is a threshold representing the minimum level of predicted mutual engagement level that the agent finds acceptable for the interaction. For example, the agent can decide to only talk about a topic if it is predicted to lead to half the maximum level of engagement, setting e to 0.5.

A subtopic of a selected topic, o_i^c, is selected for discussion when its predicted combined engagement level is equal to or falls above the same threshold e. The subtopics are ordered to be introduced in descending order of predicted combined engagement to maximize local engagement:

$$Subtopics(o_i) = \left(o_i^{\sigma(1)}, \ldots, o_i^{\sigma(k)} \right), \tag{8.4}$$

where σ is a permutation of $(1, \ldots, k)$, such that $\forall p < q, \texttt{eng}^*_{\substack{a+u \\ \sigma(p)}}(t+1, o_i) \geq \texttt{eng}^*_{\substack{a+u \\ \sigma(q)}}(t+1, o_i)$. But besides meeting the engagement criterion, the subtopics also have to be compatible with other steps of the dialogue manager: The topic manager will only actually add the selected subtopic to the interaction if the corresponding characteristic has not already been addressed in the introduction of the topic (described in Chapter 9), and if indeed agent behaviours are defined to realise the subtopic (see Chapter 10).

8.5.3 Updating: Subsequent (Sub)Topic(s)

There are two situations where the topic manager decides to introduce a new topic: when the current topic has been discussed completely, and when the recently (detected) engagement level of the user during the current topic ($Recent\texttt{ENG}_u^{obs}(t, o_i)$) is too low, meaning that it falls below the same threshold e that was used for the topic selection (in Equations 8.3 and 8.4):

$$Recent\texttt{ENG}_u^{obs}(t, o_i) < e \tag{8.5}$$

In the following Section 8.5.4 we describe how we can obtain this value $Recent\texttt{ENG}_u^{obs}(t, o_i)$ by detecting the engagement of the user. Here we first continue the explanation of topic selection itself.

Just before a next topic is selected (because the previous topic has been discussed completely, or the user's recent engagement is too low) the values in the knowledge base are updated with information that is gathered during the previous discussion phases. Figure 8.4 shows the chain of updates that are performed at this moment, which we explain below.

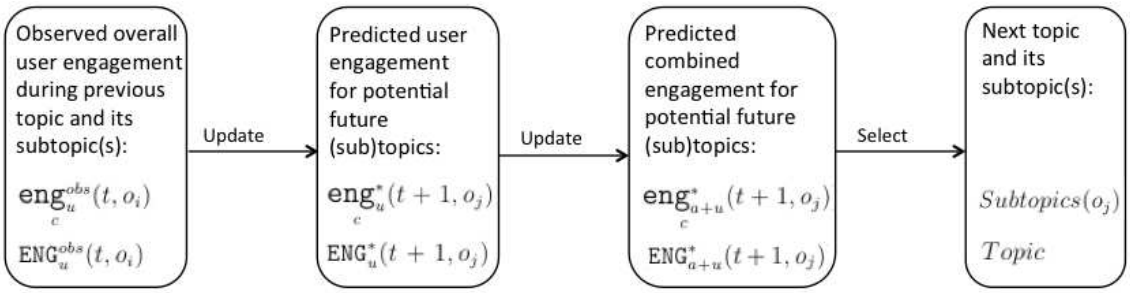


Figure 8.4 : Updating process

Concretely, just before a next topic is selected, the knowledge base is updated with the agent's detected (observed) level of user engagement during the discussion of the preceding topic $\mathtt{ENG}_u^{obs}(t, o_i)$ and subtopic(s) $\underset{c}{\mathbf{eng}}_u^{obs}(t, o_i)$. (In Section 8.5.4 we propose a way of obtaining these values.)

By means of the agent's associations that are represented by similarity scores between the topics, the detected user engagement values update, in combination with previous engagement predictions, the expected user engagements ($\mathtt{ENG}_u^{*}(t+1, o_j)$) for topics that have not yet been addressed:

$$\mathtt{ENG}_u^{*}(t+1, o_j) = \mathtt{ENG}_u^{obs}(t, o_i) \cdot \mathtt{SIM}(o_i, o_j) + \mathtt{ENG}_u^{*}(t, o_j) \cdot (1 - \mathtt{SIM}(o_i, o_j)) \quad (8.6)$$

The same update holds for all subtopics with indexes c between 1 and k:

$$\underset{c}{\mathbf{eng}}_u^{*}(t+1, o_j) = \underset{c}{\mathbf{eng}}_u^{obs}(t, o_i) \cdot \underset{c}{\mathtt{sim}}(o_i, o_j) + \underset{c}{\mathbf{eng}}_u^{*}(t, o_j) \cdot (1 - \underset{c}{\mathtt{sim}}(o_i, o_j)) \quad (8.7)$$

These updates reflect the agent's expectation that the user has similar preferences for artworks that the agent associates with each other (i.e. that are similar in the agent's knowledge base). (Sub)topics that discuss similar artworks/characteristics are therefore expected to lead to similar levels of user engagement (see Section 8.4).

For all objects o_i, with $i \in 1, \cdots, N$, that are part of the dialogue history (already discussed topics) we set:

$$\begin{aligned} \mathtt{ENG}_u^{*}(t+1, o_i) &= 0 \\ \text{and } w(t) &= 0 \end{aligned} \quad (8.8)$$

This implies that the agent makes the assumption that once a topic has been addressed the user does not want to address it again. The agent finds this more important than its own preferences ($w(t) = 0$). This simplification makes that the system shall not discuss a topic twice.

The knowledge base does not keep a history for the subtopics as every topic can contain similarly named subtopics. For example, the discussion of an object's artist (e.g. subtopic *artist*: *"Mondrian"*) does not prevent the discussion of the same artist in the context of another object (e.g. subtopic *artist*: *"Mondrian"*). On the contrary, given that characteristic values (e.g. *"Mondrian"*) are connected to themselves with a similarity score of 1, the predicted user engagement value of another object's subtopic (e.g. that talks about *"Mondrian"*) obtains by Equation 8.7 the value of the actually detected user engagement level during the discussion of the subtopic of the previous topic that discussed this characteristic ($\underset{c}{\mathtt{eng}}{}^{obs}_{u}(t, o_i)$).

With the updated values of predicted user engagement for each topic ($\mathtt{ENG}^*_u(t+1, o_j)$) and the fixed agent preferences for the objects ($\mathtt{PREF}_a(o_j)$), Equation 8.1 subsequently updates the predicted combined engagement levels ($\mathtt{ENG}^*_{a+u}(t+1, o_j)$). (Topics that have already been discussed obtain in this way $\underset{c}{\mathtt{ENG}}{}^*_{a+u}(t+1, o_i) = 0$.)

Similarly, with the updated values of predicted user engagement for each subtopic ($\underset{c}{\mathtt{eng}}{}^*_u(t+1, o_j)$) and the fixed agent preferences for the objects' characteristics ($\underset{c}{\mathtt{pref}}{}_a(o_j)$), Equation 8.2 updates the predicted combined engagement levels ($\underset{c}{\mathtt{eng}}{}^*_{a+u}(t+1, o_j)$).

These outcomes are then used in Equations 8.3 and 8.4 to select the next topic and subtopics of the conversation, triggered by the agent's associations.

No more topics are selected (i.e. the discussion of artworks (objects) ends) at the moment when no more topics fulfil the minimum engagement threshold (Equation 8.3) that make them worth addressing, or when a predefined requirement has been fulfilled. For example, when a predefined number of objects has been discussed, or when a minimum interaction time has passed.

8.5.4 Engagement Detection

As mentioned above, the detected engagement of the user serves to decide when to change a topic (Equation 8.5), and to update variables that play a role in the selection of the topics (Equations 8.6 and 8.7). In this section we describe how the topic manager calculates the values that are used for these procedures.

We assume that the topic manager receives measurements of detected user engagement with a value between 0 and 1 ($\overset{meas}{\texttt{Eng}}_u\ (t)$), accompanied by a certainty score (between 0 and 1) of the measurement ($Cert(\overset{meas}{\texttt{Eng}}_u\ (t))$). How the detection of the user's verbal and non-verbal behaviour leads to a user engagement measurement is out of the scope of this thesis.

During the discussion of the artwork topics and subtopics, meaning *not* during the opening and closing phases of an interaction, the topic manager integrates in real-time each user engagement measurement in 3 scores: the user's recent engagement during the current topic ($Recent\texttt{ENG}_u^{obs}(t, o_i)$), the overall user engagement during the current topic ($\texttt{ENG}_u^{obs}(t, o_i)$), and the overall user engagement during the current subtopic ($\underset{c}{\texttt{eng}}_u^{obs}(t, o_i)$).

The first score serves to decide if the current topic sufficiently engages the user at that moment in the interaction or if a topic change is necessary (Equation 8.5). It balances the most recent measurement with its certainty score and, in case of uncertainty, all previous measurements during the current topic:

$$Recent\texttt{ENG}_u^{obs}(t, o_i) = \begin{cases} \overset{meas}{\texttt{Eng}}_u\ (t) \cdot Cert\left(\overset{meas}{\texttt{Eng}}_u\ (t)\right)) + \\ 0.5 \cdot (1 - Cert\left(\overset{meas}{\texttt{ENG}}_u\ (t)\right), \text{ if } t = t_{o_i} \\ \\ \overset{meas}{\texttt{Eng}}_u\ (t) \cdot Cert\left(\overset{meas}{\texttt{Eng}}_u\ (t)\right) + \\ Recent\texttt{ENG}_u^{obs}(t-1, o_i) \cdot \\ (1 - Cert\left(\overset{meas}{\texttt{Eng}}_u\ (t)\right), \text{ otherwise} \end{cases} \tag{8.9}$$

where o_i is the current topic, and t_{o_i} is the time when the first measurement of o_i is received.

The overall detected user engagement level during a topic serves to update the expected engagement levels for so far unaddressed topics (Equation 8.6). Where for the recent user's engagement (Equation 8.9) we prioritise the last user engagement measurement, here we take all the measurements that are received during the current topic and balance them with their respective certainties:

$$\texttt{ENG}_u^{obs}(t, o_i) = \frac{\sum_{n=t_{o_i}}^{t} \overset{meas}{\texttt{Eng}}_u\ (n) \cdot Cert\left(\overset{meas}{\texttt{Eng}}_u\ (n)\right)}{\sum_{n=t_{o_i}}^{t} Cert\left(\overset{meas}{\texttt{Eng}}_u\ (n)\right)} \tag{8.10}$$

where o_i is the current topic, and t_{o_i} is the time when the first measurement of o_i is received. t is the present time. Similarly, we obtain the overall detected user engagement during a subtopic ($\underset{c}{\mathtt{eng}}{}_u^{obs}(t, o_i)$) by defining $t_{o_i^c}$ as the time where the first measurement of the current subtopic (o_i^c) is received:

$$\underset{c}{\mathtt{eng}}{}_u^{obs}(t, o_i) = \frac{\sum_{n=t_{o_i^c}}^{t} \mathtt{Eng}_u^{meas}(n) \cdot Cert\left(\mathtt{Eng}_u^{meas}(n)\right)}{\sum_{n=t_{o_i^c}}^{t} Cert\left(\mathtt{Eng}_u^{meas}(n)\right)} \tag{8.11}$$

The benefit of using detected measurements of user engagement is that the agent does not need to ask the user for his/her engagement or preferences explicitly (see also Section 7.7). However, in circumstances where a detection of the user's engagement is not possible, Equations 8.6 and 8.7 (and subsequently Equations 8.1 and 8.2) can be updated by entering a value that represents the user's explicitly uttered preferences for the lastly discussed object and characteristics at the place of respectively $\mathtt{ENG}_u^{obs}(t, o_i)$ and $\underset{c}{\mathtt{eng}}{}_u^{obs}(t, o_i)$. We allow this as a user's preference is directly related to the user's engagement (Chapter 7).

8

8.5.5 Example

For the sake of clarity, in this section we demonstrate the working of the topic selection component of the topic manager with a small example for which the knowledge base graph is shown in Figure 8.5. In this example the agent knows about 4 topics (objects) with each 2 characteristics (*type* and *period*) that are listed in Table 8.2.

Object	Type	Period
o_i	Statue	Antiquity
o_j	Statue	17^{th} century
o_k	Painting	17^{th} century
o_l	Painting	18^{th} century

Table 8.2 : The objects of the example topic structure of Figure 8.5.

Figure 8.5 shows how the variables for each topic can evolve over time $t_0 - t_2$ during an interaction. We limit this example to the calculation and selection of topics. The calculation of the weights of the subtopics occurs, with exception of already discussed subtopics (see Section 8.5.3), in the same manner.

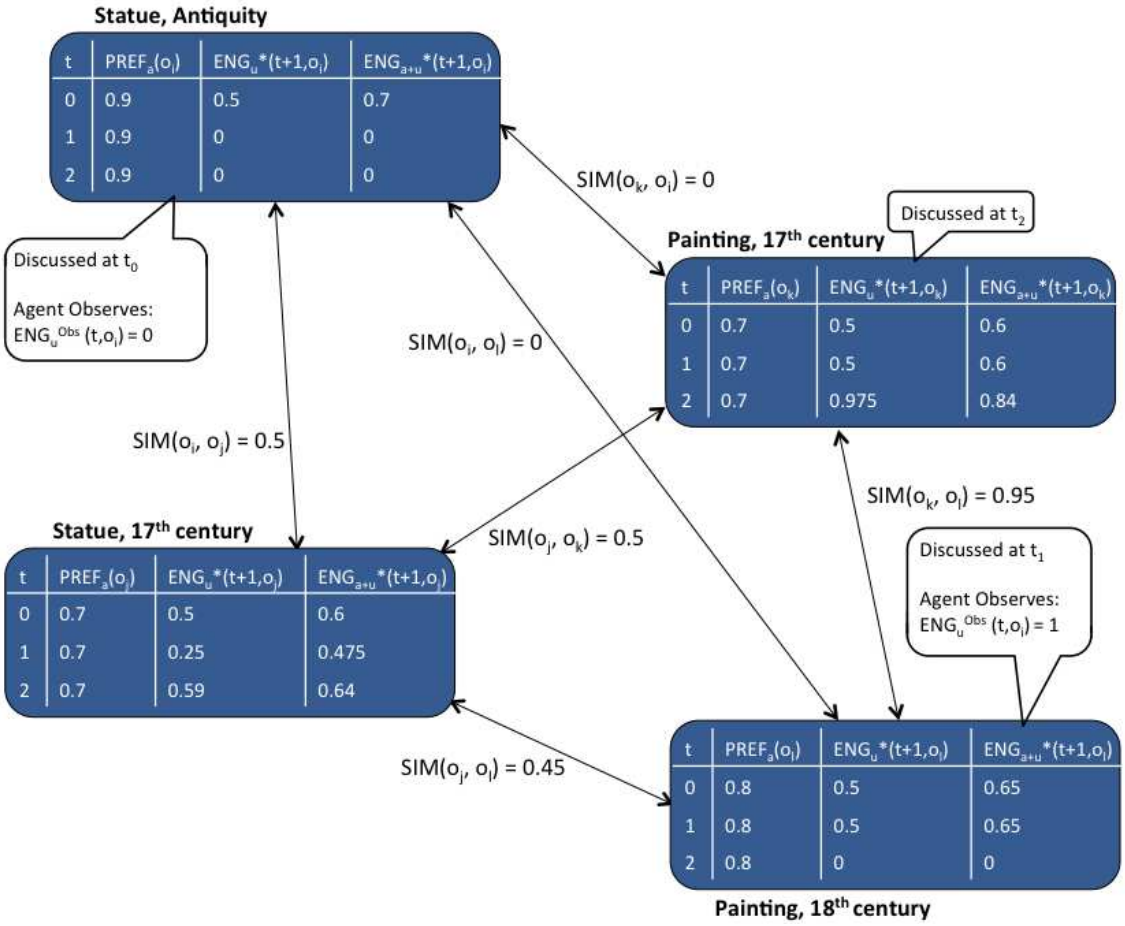


Figure 8.5 : Example of the evolution of the weights in the topic structure over time $t_0 - t_2$ for the topics/objects (no characteristics) only. In this example, at each t, $w(t) = 0.5$.

The agent's associations (similarity scores) between the objects, shown as the edges of the knowledge base graph, are the average of the similarity scores for the objects' characteristics: Equal characteristics are connected to themselves with similarity scores of 1, and *periods: 17th century – 18th century* are assigned a similarity of 0.9. All other characteristic pairs are assigned a similarity score of 0.

The values of the variables for each topic at the beginning of the interaction (time t_0) represent the initial state of these variables (Figure 8.5). Given that at this moment the predicted combined engagement for potential future discussion ($\texttt{ENG}^*_{a+u}(t_0, .)$) is the highest for object (topic) o_i, this object is the first to be selected for discussion (an object with characteristics *type: statue* and *period: antiquity*). When the agent then, when a topic switch is required (see Section 8.5.3), observed

a minimum level of overall user engagement during the discussion of this previous topic ($\texttt{ENG}_u^{obs}(t, o_i) = 0$) the updated variables (see t_1, by Equations 8.6 and 8.1) lead to the selection of object o_l as next topic (Equation 8.3), which has nothing in common with the former object (*type: painting*, *period: 18th century*). To show the opposite extreme situation, after the discussion of object o_l the agent receives a maximum level of overall observed user engagement during this o_l ($\texttt{ENG}_u^{obs}(t, o_l) = 1$). This leads to the selection of o_k as third topic, staying close to the characteristics of the former object (*type: painting*, *period: 17th century*). Of course the knowledge base of Figure 8.5 is only a limited example and not sufficient to illustrate the full potential of the trade-off between agent and user oriented variables in the selection of a topic.

8.6 Conclusion

In this chapter we have proposed an engagement driven topic selection model for an information-giving agent. By taking into account the agent's perceived (detected) levels of user engagement, as well as the agent's own preferences, and its associations, the agent is endowed with human-like features, selecting the topics of the interaction from its own point of view.

The model avoids the need for any pre-entered information of the user, but allows the interaction to be adapted to any user. Configurations of the agent's preferences, the agent's associations (similarity scores), and the agent's orientation towards itself and/or the user ($w(t)$) allow for different types of agents.

The initialisation of the agent's knowledge base (Section 8.5.1) can be simplified with the help of a museum catalogue that already lists the objects and their characteristics. The topic selection model can also be easily extended to other domains that can be structured similarly as museum objects. This means that the agent needs to have its preferences for the different topics and can associate the topics to each other by means of similarity scores. Selecting subtopics is only possible if the topics' characteristics (attributes) can be defined.

In this chapter we focused on the agent's selection of the topics of the interaction. In Chapter 11 we perform a perceptive study to evaluate the entire topic manager, including this topic selection component, in human-agent interaction. For this, in the following chapters we first complete the description of the topic manager:

— In the following chapter (9) we study ways of switching topics on a dialogue level. This is necessary since even when two consecutive topics have no charac-

teristics in common, the agent needs to present the new topic in a natural way in the conversation without loosing the dialogue coherence [Levinson, 1983].

— In Chapter 10 we propose an implementation of the topic manager in a virtual agent platform, leading to a dynamic interaction where the agent pastes in real time dialogue parts to the ongoing conversation to realise the selected topics and subtopics.

9 Topic Manager: Topic Transition Model

9.1 Introduction

The choice of the dialogue topics that an agent initiates in non-task oriented human-agent interaction is important for several aspects of the interaction such as the coherence of the interaction [Macias-Galindo et al., 2012] and the user's engagement (Chapters 7 and 8). Some efforts are oriented towards the selection of the appropriate topic at a specific point in an interaction [Macias-Galindo et al., 2012, Stede and Schlangen, 2004] (Section 8.3). However, *how* the selected topic can or should be introduced by the agent has not been given much consideration in non-task oriented dialogue. In this chapter we will consider the latter aspect.

Concretely, in order to obtain human-like agent discourse the (sub)topics that are selected by the topic selection component of the topic manager in order to engage the user (presented in the previous chapter) have to be initiated by the agent in a natural and coherent way in the ongoing interaction. On a subtopic level this is straightforward; all subtopics within a topic talk evidently about an aspect of the same artwork. They all form part of the ensemble of turns that constitute a clearly delimited topic.

This makes that the subtopics are naturally connected to each other, permitting direct concatenation of the subtopic dialogue fragments. Besides, on an utterance level the subtopics can be fairly similar. For example, a subtopic that describes the *style* of the artwork may also make reference to (i.e. talk shallowly about) the *period* in which the artwork was realised, thereby favouring the coherence of the interaction.

On a topic level however, it is less straightforward to establish a topic swift in a coherent manner as the topics are clearly delimited. Each of the topics addresses a different artwork, where the (artwork of the) preceding topic may have little in common with the (artwork of the) subsequent topic. This is because the topics are primarily selected according to their potential of engaging the user in-

stead of their coherence with respect to the previous topic (Chapter 8). Where the subtopics together try to achieve an apparent goal of discussing an artwork, it may not be apparent what the different topics try to achieve together, making a simple concatenation of topics possibly insufficiently coherent.

In order to resolve a possible incoherence of subsequent topics, and thereby making the agent's discourse more human-like, in this chapter we will look into utterances that may be used to initiate a transition from one topic to another and the *transition strategies* that are executed with (can underlie) such a *transition utterance*. In the first part of this chapter we describe an evaluative study towards transition strategies in order to try to answer two questions: 1) What strategies have the potential of keeping/making the dialogue coherent? And 2) what effect do the use of the different transition strategies have on the perception of the conversational agent? The answers to these questions then serve in the second part of this chapter (Section 9.6) to automatically generate agent strategies to connect one topic to another in non-task oriented dialogue.

So where other dialogue systems look at what topic is coherent at a specific point in the interaction (e.g. [Macias-Galindo et al., 2012, Breuing and Wachsmuth, 2012, Wong et al., 2012]), we are looking at possible strategies to introduce a topic coherently.

Below we first discuss respectively related work (Section 9.2) and the methodology of the perceptive study (Section 9.3). In Section 9.4 we describe the results of this study followed by their discussion in Section 9.5. Section 9.6 describes the generation of transition utterances in the agent's topic manager, and Section 9.7 concludes our findings.

9.2 Related Work

9.2.1 Transition Strategies in Theory

Literature about transition strategies outside task-oriented applications can be found in the domains of conversational analysis and social sciences, where they are studied from an observational (detection) point of view. [Downing, 2000] distinguishes two forms of introducing a topic: by means of an *informative statement*, and by *asking a question*. "By informing the speaker assigns him/herself the role of topic supplier, whereas by questioning or more exactly by eliciting a topic this role is offered to an interlocutor in the discourse."

Similarly, [Button and Casey, 1985] define two global ways of introducing a topic that is not related to the prior topic in a conversation: by a *topic initial elicitor* that is used "to elicit a candidate topic from the next speaker whilst being mute with respect to what that topic may be", and by *topic nominations* that "are oriented to particular newsworthy items". Two sequence types that may be used for topic nomination are *itemised news enquires* and *news announcements*. An itemised news inquiry is oriented to a recipient's newsworthy item where a news announcement is oriented to a speaker's newsworthy item.

[Maynard and Zimmerman, 1984] identified four topic initiation strategies in dyadic human-human conversations. For acquainted parties: *displaying prior experience* and *using setting talk*, and for unacquainted parties: *categorisation question-answer pairs* (e.g. year in school, academic major, etc.), and *question-answer pairs involving activities* that are related to the categories.

[Hobbs, 1990] focuses on three coherence relations that he claims are responsible for most of the so-called topic drift in dialogue: *parallelism*, *explanation* and *metatalk*. "*Parallelism* between two segments occur when the two segments assert propositions from which we can infer that identical properties hold of similar entities", or "when, from the assertions of the two fragments we can infer similar properties holding for identical entities". An *explanation* occurs when the situation or event described in the second segment can serve as a cause for the situation or event described in the previous segment. A segment is *metatalk* if it asserts a relation between some second segment and the goals of the conversation [Hobbs, 1990].

9.2.2 Transition Strategies in Dialogue Systems

To our knowledge, existing dialogue systems that explicitly consider different strategies to introduce a particular topic have been developed exclusively for task oriented interaction, in particular in the form of task interruption strategies. In this context [McFarlane, 2002] defines four primary methods: *immediate*, *negotiated*, *mediated*, and *scheduled interruption*. [Yang et al., 2008] found out that dialogue partners usually use *discourse markers* and *prosody cues* to signal task switching. Guided by these works [Heinroth et al., 2011] looked at 4 different task switching strategies: *unassisted immediate topic shift*, *discourse markers combined with prosody cues*, and two full sentence initialising topic shifts "to produce a more natural dialogue flow and to increase the timespan the user has for task switching": *explanation* and *negotiation* strategies. The explanation strategy explains what task is about to be

started and the negotiation strategy asks for permission to switch a task. They evaluated the use of these four strategies on several dimensions and found that the explanation strategy showed high scores regarding efficiency and user-friendless and supports the user to memorise the tasks. Other strategies showed advantages such as being less irritating.

9.2.3 Guidelines for Topic Transitions

The above mentioned research demonstrates that there does not exist one overall taxonomy of transition strategies that can be used as a recipe for transition strategy generation in non-task oriented dialogue. This lack shows the need of our own research towards transition strategies and makes us fall back to the following generally accepted ideas about topic switching: According to [Clark, 1996] a topic can be described as a joint project as it is jointly established during ongoing conversations. [Svennevig, 2000] adds that every spoken contribution may raise new potential topics whose actual realisation depends on the co-participant's acceptance by picking up one of these topics within his or her reply. To conclude, Sacks (1971, April 5 in: [Levinson, 1983, p. 313]) made an overall remark that

> "What seems to be preferred is that, if A has been talking about X, B should find a way to talk about Z (if Z is the subject he wants to introduce) such that X and Z can be found to be 'natural' fellow members of some category Y. However it should not be thought from this that such co-class membership is somehow antecedently given; rather it is something that is actually achieved in conversation."

No instructions have been given of how such a presentation could be generated.

Based upon these general ideas of topic switching, below we try to collect more precise indications about how to generate transition strategies in non-task oriented dialogue by means of a perceptive study.

9.3 Methodology

In order to find out what strategies a conversational agent can use to initiate topic transitions in non-task oriented dialogue we follow [Heinroth et al., 2011] (Section 9.2.2) by testing a set of potential transition strategies with respect to their effects on the perception of the dialogue and the agent. In the subsections below we respectively discuss the steps to achieve this: the specification of the context

of the transition strategies (Section 9.3.1), the design of the transition strategies themselves (Section 9.3.2), the setup of the experiment to test the set of strategies (Section 9.3.3), and the questionnaire that will be used for this (Section 9.3.4).

9.3.1 Context of the Transition Strategies

The strategies that are mentioned in previous work vary with respect to the context. Some strategies work for topics that are interesting for the listener and others for those that are interesting for the speaker [Button and Casey, 1985]. Some strategies are used by acquainted parties and others by unacquainted parties [Maynard and Zimmerman, 1984]. Explanation strategies in the sense of Hobbs (1990), as well as metatalk, only work for a specific set of topics.

These constraints imply that the strategies that can be used to introduce a topic in a conversation depend on the relation between the current topic of the dialogue and the new topic that is to be introduced. The first step in generating transition strategies is thus to define this relation. In the context of project Avatar 1:1 we are looking at strategies that an agent can employ in interaction with an unacquainted user to make the transition between two discussion phases about two different artworks. In the current work we will focus on what seems the most extreme case, namely the transition between discussion phases of two very different artworks: Artworks that have nothing in common except from the fact that they are both artworks in the same museum. In this way we test if the agent's topic manager can indeed be allowed the flexibility to select any given artwork of the museum as next topic of the discussion (as presented in Chapter 8). Such flexibility helps finding (initiating) the topic that engages the user most (Chapter 7).

To be more precise, in Table 9.1 we give an example of a dialogue fragment that precedes the moment at which the new topic, corresponding to a very different artwork than the one discussed, is to be introduced. As the timing of introducing a new topic may have an influence on the perception of the topic switch [Clark, 1996] we limit this research to a topic switch that occurs after the conversation has addressed respectively the subtopics *artist* and *period* of the former discussed artwork.

Speaker	Dialogue about "Luncheon on the Grass" by Claude Monet	Subtopic
	[...]	
Agent:	Claude Monet was a French painter. He lived his entire life at Giverny, a beautiful village north of Paris.	Artist
User:	Yes I know. I visited Giverny last year.	
Agent:	This painting was made around 1865.	Period
User:	Yes, I've read so too.	

Table 9.1 : An example of a dialogue fragment preceding a topic switch initiated by a transition strategy. In the evaluation study this dialogue fragment (translated) serves as the context of scenario 1 (Section 9.3.3).

9.3.2 Design of Potential Transition Strategies

Due to the nature of the context we are dealing with, the potential transition strategies to introduce a discussion phase of another artwork are limited to the following categories from the literature: explanations in the sense of [Heinroth et al., 2011], informative statements [Downing, 2000], itemised news enquires and news announcements [Button and Casey, 1985], categorisation question-answer pairs and question-answer pairs involving activities [Maynard and Zimmerman, 1984], and parallelism [Hobbs, 1990]. It is however not prescribed how we could generate formulations for each of these detection-based categories for the context we are looking at. We thus base the manual creation of a set of potential transition strategies that belong to one or multiple of these categories, on the general guideline by Sacks [Levinson, 1983] (Section 9.2.3).

According to Sacks [Levinson, 1983] we need to find a way to let the former (current) and the next (selected) topic be members of some category Y. We try to do this by (indirectly) referring to an element that is used in the agent's reasoning process to talk about the next topic. As described in Section 8.5.1 from the previous chapter, the agent disposes of a knowledge base that holds information about certain artworks from the museum. From this set of artworks it selects dynamically a new topic of discussion with the goal of maximising the combined (agent and user) engagement level, taking into account (the characteristics of) the artworks,

the user's engagement (an indication of the user's preferences for the artworks), the agent's preferences for artworks, and the agent's associations (Chapter 8). The set of potential transition strategies that we created by referring to these elements is listed in Table 9.2. For each of the strategies we formulated an agent utterance to realise the strategy.

Strategies 9 and 10 that insist on the (in this case contrasting) characteristics of the artworks are added as a reference to the strategies that we would use for the transition between artworks that have characteristics in common (the category Y).

Table 9.2 : Potential transition strategies to connect the discussion phases of two very different artworks (translated). The strategies (Column 3) are based on concepts of the topic selection component (Column 2) but do not reflect its calculations (Chapter 8). KB = Knowledge Base, o_i is the current topic of the interaction and o_j is the one to be introduced, $_a$ = Agent, $_u$ = User. The example utterances (column 3, translated) serve as the transition utterances in scenario 1 of the evaluation study (Section 9.3.3).

Orientation	**Reasoning Variable**	**Transition Strategy**	**No.**
Agent	Agent Preferences	$(Preference_a(o_i) > e) \wedge (Preference_a(o_j) > e)$ E.g. "I also like the Balloon Dog by Jeff Koons."	(1)
Agent	Agent Preferences	$Preference_a(o_j) > Preference_a(o_i)$ E.g. "Personally, I prefer the Balloon Dog by Jeff Koons."	(2)
Agent	Agent Associations	$Association_a(o_i, o_j)$ E.g. "This work reminds me of another work, the Balloon Dog by Jeff Koons."	(3)
User	User Preferences (Engagement)	$(Preference_u(o_i)) > e) \wedge (Preference_u(o_j) > e)$? E.g. "If you like this work, maybe you also like the Balloon Dog by Jeff Koons."	(4)
User	User Preferences (Engagement)	$Preference_u(o_j) > Preference_u(o_i)$? E.g. "Maybe you prefer the Balloon Dog by Jeff Koons?"	(5)
Agent	Objects in KB	$Experience_a(o_i) \wedge Experience_a(o_j)$ E.g. "I've also seen the Balloon Dog by Jeff Koons."	(6)
User	Objects	$Experience_u(o_i) \wedge Experience_u(o_j)$?	(7)

	in KB	E.g."Have you also seen the Balloon Dog by Jeff Koons?"	
Object	Objects in KB	$\exists(o_j) \wedge (o_j \neq o_i)$ E.g. "Another artwork is the Balloon Dog by Jeff Koons."	(8)
Objects	Characteristic of Objects	$\exists(o_j) \wedge (o_j^{artist} \neq o_i^{artist})$ E.g. "An artwork from another artist is the Balloon Dog by Jeff Koons."	(9)
Object	Characteristic of Objects	$\exists(o_j) \wedge (o_j^{period} \neq o_i^{period})$ E.g. "An artwork from another period is the Balloon Dog by Jeff Koons."	(10)

9.3.3 Experimental Setup

Inspired by the existing literature, above we have created a set of potential transition strategies for the context we are looking at. In order to verify if each of these strategies is suitable to be generated by the agent to switch the topic in the information-giving chat with the user we perform an empirical study. By means of an online questionnaire, shown partially in Appendix D Figure D.1, we test if and how the different transition strategies affect the human perception of the dialogue and the agent.

To this end we present each participant of the online study with 2 different dialogue fragments consisting of French agent utterances and simulated user inputs (as in [Macias-Galindo et al., 2012]), such as the translated fragment in Table 9.1, which form the contexts of the transition strategies. Each scenario is followed by 3 randomly assigned transition utterances (realisations of strategies), displayed next to each other (see Appendix D Figure D.1). We do not show the utterances that may follow the transition strategies. In this way we do not show an acceptance or rejection of the topic by the user ([Clark, 1996, Svennevig, 2000]). Directly after each of the 3 transition strategies we ask the participants to answer several questions, which we point out in Section 9.3.4.

We use a written setup to allow the participants to consider multiple strategies at the same time in the same context, enabling cross-comparison and rereading as much as desired. Besides, in this way the judgements are not disturbed by unnatural text-to-speech realisations.

As mentioned before, the dialogue fragment that represents the former topic in the context and the topic that is addressed in the transition strategies (next topic) are about very different artworks. We test 2 topic pairs for each participant (2 different scenarios) to anticipate possible effects that are due to individual characteristics of a particular context. Scenario 1 consists of the discussion of a painting by Monet (translation shown in table 9.1) followed by transition utterances (representing transition strategies) introducing a statue by Jeff Koons (translations listed in Table 9.2). Scenario 2 consists of the discussion of a painting by Mondrian followed by transition utterances introducing David, the statue by Michelangelo. The alternation of agent-user utterances, the number of utterances and the order of the subtopics are the same in both context fragments. The order in which the scenarios are presented to the participants is random. Pictures of the artworks next to the questionnaire make sure that all the participants know what the artworks look like (Appendix D Figure D.1).

9.3.4 Questionnaire

For each of the 3 randomly assigned transition strategies we ask questions on a scale from 1-9 (Appendix D Figure D.1) (following [Bickmore and Cassell, 2005]). The first 3 questions relate to the perception of the dialogue and serve to answer the first question we try to answer (Section 9.1): What strategies have the potential of keeping/making the dialogue coherent? We ask respectively if the participant finds the dialogue *natural* [Nakano and Ishii, 2010, Bickmore and Cassell, 2005]), *coherent* [Macias-Galindo et al., 2012], and *smooth* [Higashinaka et al., 2008, Nakano and Ishii, 2010].

The following 5 questions serve to answer our second question (Section 9.1): What effect do the use of the different transition strategies have on the perception of the conversational agent? We ask respectively to what extent the participants find the agent *friendly*, *warm*, *fun*, *competent*, and *informed* [Bickmore and Cassell, 2005]. These measures are related to 2 important social aspects, warmth and competence [Fiske et al., 2007].

9.4 Results

83 subjects filled out the online questionnaire: 56 female (67%), all native speakers of French as this was a requirement for participation, aged 19-69. In the sub-

sections below we show the results of the experiment specified for the two issues we are looking at: the perception of the dialogue and the perception of the agent.

9.4.1 Dialogue Perception

For each strategy, the perception of the dialogue has been questioned for the two scenarios and on three dimensions: *naturalness*, *coherence*, and *smoothness*. For each of these dimensions the results show no significant difference between the two scenarios (Kruskal-Wallis). This means that we can take the data for both scenarios together, as shown in Figures 9.1, 9.2, and 9.3.

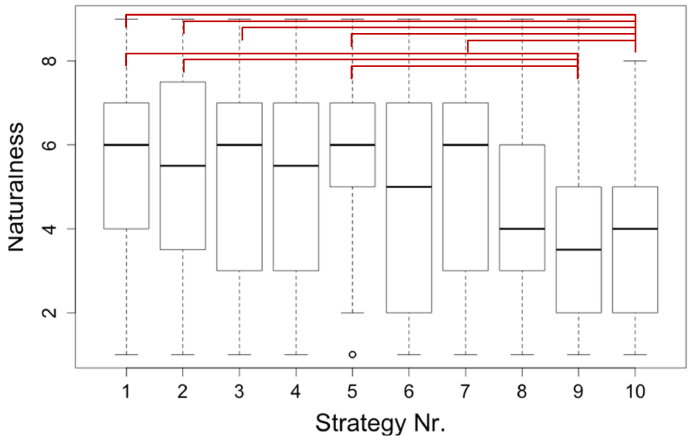


Figure 9.1 : Naturalness for each strategy, $p < 0.01$. The numbers of the strategies correspond to the numbering in Table 9.2.

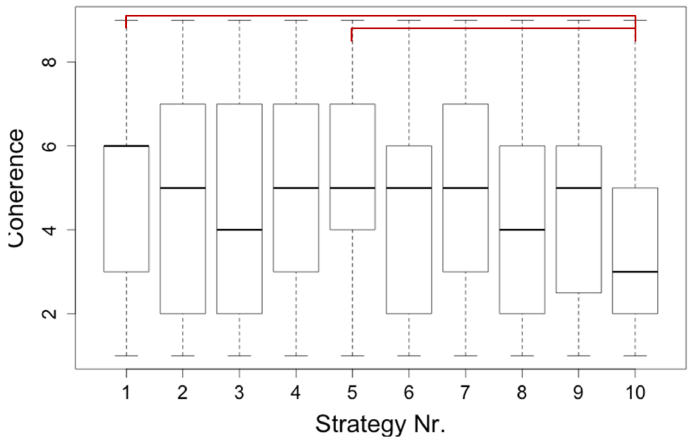


Figure 9.2 : Coherence for each strategy, $p < 0.01$.

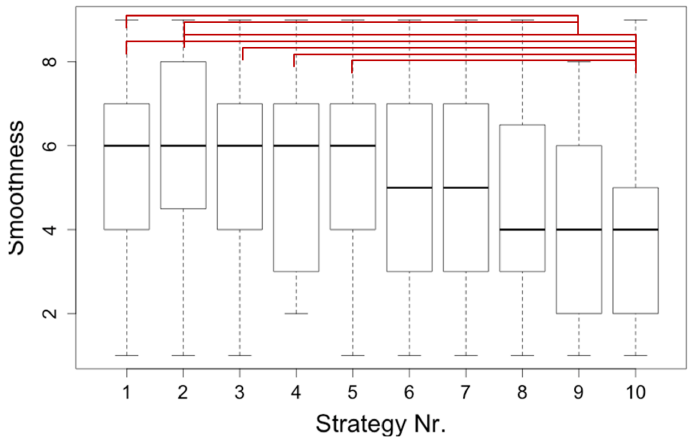


Figure 9.3 : Smoothness for each strategy, $p < 0.01$.

For all three dimensions the scores differ significantly among the strategies (Kruskal-Wallis $p < 0.01$). Regarding the level of *naturalness* and *smoothness*, Kruskal-Wallis multiple comparisons show that the significant differences are due to the strategies 9 and 10 that score significantly lower than some others, indicated by the horizontal brackets in the graphs. Regarding the *coherence* of the dialogue, strategy number 10 leads to a significantly lower level than other strategies.

Only strategies 1 to 7 show a higher mean than the average level (> 4.5 on a scale of 9) on the dimensions of *naturalness* and *smoothness*. With respect to the level of *coherence*, except from strategies 8, 9 and 10, strategy 3 also scores lower than average (mean).

As mentioned in Section 9.3.2 the strategies are either oriented towards the agent, the user, or the object (artwork) (Table 9.2, Column 2). The strategies from the latter group lead to significantly lower levels of *naturalness*, *coherence* and *smoothness* in comparison to the strategies with another orientation (both scenarios, Kruskal-Wallis, $p < 0.01$). There is no significant difference in the scoring of the strategies that are agent oriented versus the ones that are user oriented with respect to the perception of the dialogue.

9.4.2 Agent Perception

Questions 4 to 8 are about the way the participants of the experiment perceive the social competence [Fiske et al., 2007] of an agent that would use the transition strategies in the context in which they are presented. The results show that between the two scenarios, the participants find the agent not significantly different with

respect to its level of *friendliness* and *knowledge* ("informed") (Kruskal-Wallis). For these dimensions we can thus analyse the data for both scenarios together. Figure 9.4 and 9.5 specify the distributions of these dimensions for every strategy.

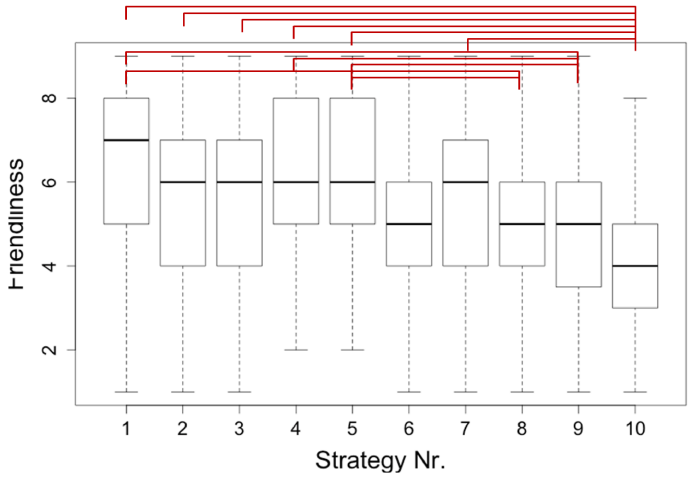


Figure 9.4 : Friendliness for each strategy, $p < 0.01$.

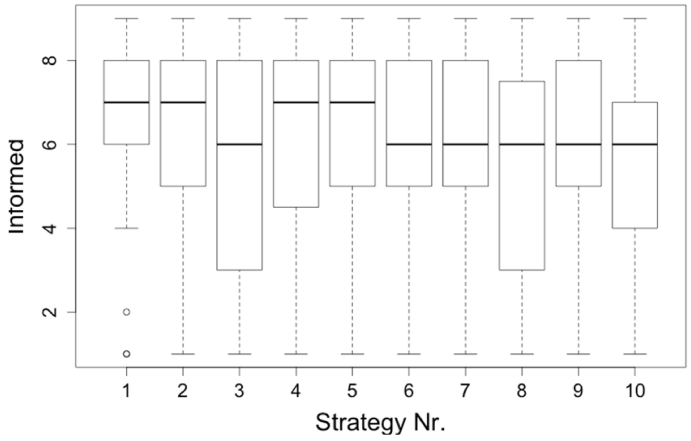


Figure 9.5 : Knowledge (i.e. agent being informed) for each strategy.

The level of *friendliness* differs significantly among the strategies (Kruskal-Wallis $p < 0.01$), which is due to strategies 8, 9, and 10 (Kruskal-Wallis multiple comparisons). However, only strategy 10 scores below average for the level of *friendliness* (< 4.5).

For all strategies the agent is not perceived significantly different with respect to its *knowledge* ("informed") and all of the strategies score above average on this dimension (> 4.5).

In contrast to the agent's level of *friendliness* and *knowledge* ("informed"), for the dimensions of *warmth*, *fun* and *competence*, some strategies are significantly differently judged among both scenarios (Kruskal-Wallis $p < 0.05$). Figure 9.6, 9.7 and 9.8 show the distribution of the results specified for both scenarios. The circled numbers indicate the strategies that are judged differently between both scenarios.

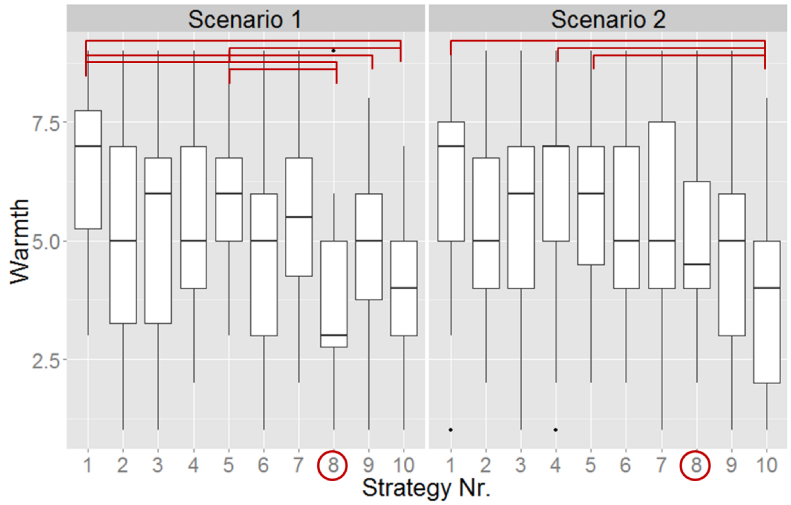


Figure 9.6 : Warmth for each strategy, $p < 0.01$. Between scenarios strategy 8 differs $p < 0.05$.

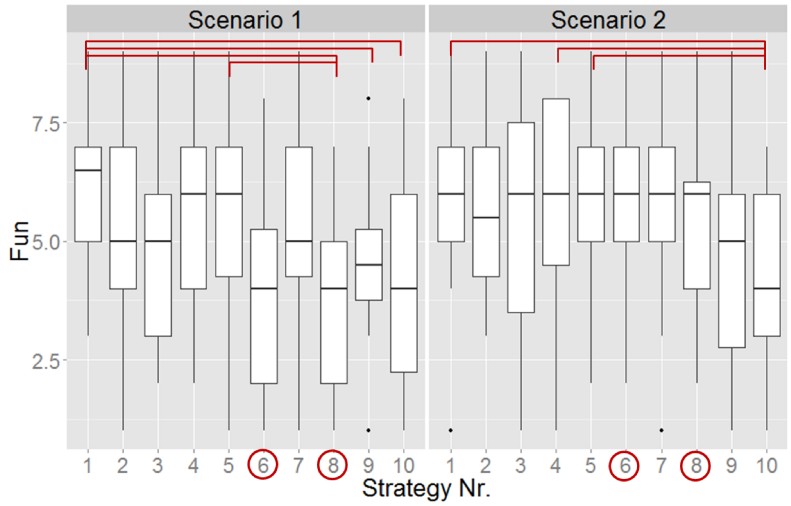


Figure 9.7 : Fun for each strategy, $p < 0.01$. Between scenarios strategies 6, $p < 0.05$, and 8, $p < 0.01$, differ.

On the dimension of *warmth*, strategy number 8 scores significantly higher in the

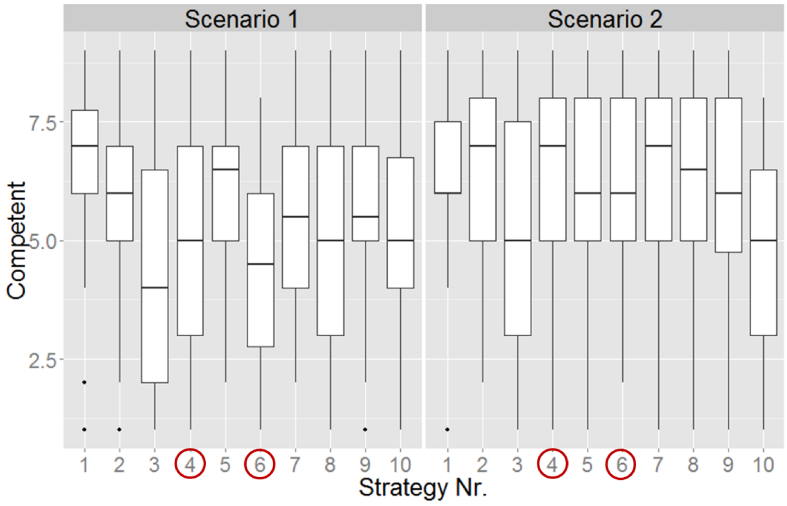


Figure 9.8 : Competence for each strategy, $p < 0.01$ for scenario 1. Between scenarios strategies 4, $p < 0.05$, and 6, $p < 0.01$, differ.

second scenario (Monet-Koons) than in the first (Mondrian-Michelangeloo). Together with strategy 6, strategy 8 also scores higher in the second scenario with respect to the level of *fun* that the participants perceived in the agent. Further, in the second scenario strategies 4 and 6 score higher on the dimension of *competence* than in the first scenario.

With respect to the agent's perceived level of *warmth* as well as *fun*, in scenario 1, strategies 8, 9 and 10 score significantly lower than other strategies (Kruskal-Wallis $p < 0.01$). In this scenario, strategies 6, 8 and 10 also score below average (< 4.5). For scenario 2 strategy 10 scores significantly lower than other strategies (Kruskal-Wallis $p < 0.01$) and falls below average.

Scenario 1 shows significant differences between the scorings of the agent's perceived level of *competence* (Kruskal-Wallis $p < 0.01$). Multiple comparisons (Kruskal-Wallis) do not indicate a specific pair of strategies that is responsible for this difference. Strategy 3 is the only strategy that scores below average (< 4.5). In scenario 2 the strategies show no significant differences or scorings below average.

Comparing the strategies that are oriented towards the agent with those that oriented towards the user (Table 9.2) does not lead to a significant difference with respect to the perception of the agent (both scenarios, Kruskal-Wallis). The strategies that are not oriented towards the agent or user, but refer to the (characteristics) of the object (artwork) lead to significantly lower levels of *friendliness*, *warmth* and *fun* in comparison with the strategies that are oriented towards the interaction par-

ticipants (both scenarios, Kruskal-Wallis $p < 0.01$).

Within the group of strategies that are agent or user oriented, we can make another grouping according to the element of the agent's reasoning process that is referred to (Table 9.2, Column 1): *preferences* (1,2,4,5), *associations* (3) and the presence of an *artwork* in the agent's knowledge base (6,7). A comparison between these groups leads to one significant result: in scenario 1, strategies that use *preferences* score significantly higher on competence than the strategies from the other groups (Kruskal-Wallis $p < 0.01$, Kruskal-Wallis multiple comparisons).

9.5 Discussion

The results show that the transition strategies that insist on contrasting `characteristics of the objects` (artworks) such as a different artist (strategy 9) or period (strategy 10) lead to scores below average (< 4.5 on a scale of 9) with respect to the level of *naturalness*, *coherence* and *smoothness*. On the level of *naturalness* and *smoothness* the difference with the other strategies is significant. This demonstrates that even when the artist and period of the former artwork have been discussed just before the transition strategies, making transition strategies that are based on (referring back to) the earlier discussed characteristics (subtopics), does not guarantee a *natural*, *coherent* and *smooth* dialogue.

The fact that strategy 8, that presents another artwork as just being another artwork (`object in KB`), scores below average on *naturalness* and *coherence*, shows that "being an artwork" is not a category that can sufficiently bind both topics ("category Y", see Section 9.2.3 [Levinson, 1983]). This can have several reasons: The transition strategy may not succeed in presenting the former and latter artwork in being natural fellow members of some category Y, in this case being an artwork; Category Y may need to be a more restrictive (distinctive) category than one to which all topics belong (all topics are artworks) in order to bind two specific topics; Or, making both topics natural fellow members of some category Y may not be sufficient in general to establish a natural and coherent dialogue.

A reason why strategy 3 that is based on the `associations of the agent`, scores bad on *coherence* but well on *naturalness* and *smoothness* may be due to the fact that the participants can find the association itself incoherent. Due to the contrasting characteristics of the artworks the participants may find it incoherent that the first artwork reminds the agent of the second. However, given that the strategy is considered *natural* and *smooth* implies that it might be a suitable strategy

to connect the discussions of two similar artworks.

The same explanation can be given for the low scoring of this strategy (3) (referring to the `agent's associations`) with respect to the perception of the agent's level of *competence* (scenario 1). When the agent associates two artworks that do not seem alike the agent is perceived less *competent* than average.

The strategies that lead to low scores on *naturalness*, *coherence* and *smoothness* of the dialogue (8, 9 and 10) also score relatively low with respect to the perception of *friendliness*, *fun* and *warmth* of the agent. This gives us reasons to suspect that both aspects are related: when a dialogue is not considered *natural*, *coherent* or *smooth*, the agent is not considered as very *friendly*, *fun* and *warm*.

The participants do not perceive the agent significantly more or less *informed* when it uses certain transition strategies instead of others. This shows that referring explicitly to the `characteristics of the objects` such as its artist (9) or period (10) does not make the agent look more *informed* than when the strategies refer to more subjective aspects of the agent's reasoning process, such as its preferences or associations.

On the contrary, strategies that refer to the preferences of the interaction participants (`agent preferences` or `user preferences`) score significantly higher with respect to the agent's level of *competence* than the strategies that use other variables from the agent's reasoning process.

With respect to the consequences of the transition strategies on the perception of the dialogue the results have shown no significant difference between both scenarios. The effects on the perception of the dialogue that are discussed in this Section seem thus generalisable for the domain we are looking at (Section 9.3.1). However, for some transition strategies the perception of the agent is judged significantly differently among the two scenarios. For example, strategy 6, a statement of the fact that the agent has seen some other artwork (`object in KB`), has in some contexts a negative influence on the agent's level of *fun* and *competence*, where this is not the case in other contexts. In the two scenarios that were used for this experiment the type of information, the utterance types, and the number of utterances are equal. Therefore, further research will be needed to show what exactly the underlying reason is that the same strategies lead, in a different scenario, to a difference in the perception of the agent.

In summary, we have found that the strategies that score well on all dimensions and in both scenarios are those that ask for the experience of the user, and those that refer to the preferences of the interaction participants. Whether the preference

is the agent's or the user's, and whether or not the new topic is preferred over the current one, transition strategies that integrate any type of preference maintain the coherence of the dialogue while maintaining/establishing a positive perception of the agent. The fact that certain transition strategies can connect topics about very different artworks while maintaining positive perceptions of the dialogue and the agent shows that the agent's topic manager can indeed be allowed to select any topic required to engage the user at any moment in the conversation (Chapter 8).

It should be noted that we did not manipulate the timing of the transition strategies, and we only considered written verbal behaviour. Considering effects of the timing of the topic switch [Clark, 1996] and non-verbal behaviours [Kendon, 1972] might further influence the human's perception of the topic transitions.

9.6 Transition Utterance Generation

In Chapter 8 we have presented a topic selection component of the topic manager that selects *what* topic (and subtopics) the agent should introduce in the interaction. In this section we present the transition generation component that decides *how* the agent introduces the topics in the interaction.

Based on the results of the empirical study described above we obtained a list of validated transition strategies that the topic manager can choose from to make a coherent transition between a pair of topics. Table 9.3 shows this set of strategies.

As shown in the previous Section and as listed in Column 1 (Table 9.3) some of the strategies turned out to be suitable to make a coherent transition between any pair of topics. Strategy 3 (referring to the `agent's associations`) was found to be inappropriate to connect any pair of topics, but gave indications that it could be used to connect similar topics, namely topics whose artworks have at least one characteristic in common.

Because the strategies make transitions between two topics by making reference to elements of the agent's reasoning process, each strategy could be realised by multiple surface forms. Different surface forms could realise the same transition strategy, i.e. imply the same argumentation to introduce the following topic. We believe surface forms could thus also be created to connect topics in other domains than the domain we used in the evaluation study (topics discussing artworks, Section 9.3.2), as long as the surface forms are modelled according to the descriptions (logical forms) of the strategies (adapted to the domain of conversation), and the agent uses (or can pretend to use) the same topic selection reasoning variables. For the

transition generation component of our topic manager, presented below, we consider surface forms that talk about artworks in a museum, the context to which the topic manager is applied.

In the following subsections we respectively define the steps for the topic manager to choose the appropriate transition strategy from the set of validated strategies (Table 9.3), and generate the corresponding surface form (utterance), at any given moment in the interaction.

Table 9.3 : Set of transition strategies in the Topic Manager. KB = Knowledge Base, o_i is the current topic of the interaction and o_j is the one to be introduced, $_a$ = Agent, $_u$ = User, $Char$ = the name of the characteristic (e.g. *"period"*), Obj = the description of the object (e.g. *"The Last Judgement by Michelangelo"*).

No. Table 9.2	Topic pair	Reasoning Variable	Orientation	Transition Strategy	New No.
(1)	All topics	Agent	Agent Preferences	$(Preference_a(o_i) > e) \wedge (Preference_a(o_j) > e)$ E.g. "I also like (another work by/from/with the same $Char$), the Obj.	(1)
(2)	All topics	Agent	Agent Preferences	$Preference_a(o_j) > Preference_a(o_i)$ E.g. "Personally, I prefer (another work by/from/with the same $Char$,) the Obj	(2)
(3)	Similar topics	Agent	Agent Associations	$Association_a(o_i, o_j)$ E.g. "This work reminds me of another work by/from/with the same $Char$, the Obj."	(3)
(4)	All topics	User	User Preferences (Engagement)	$(Preference_u(o_i)) > e) \wedge (Preference_u(o_j) > e)$? E.g. "If you like this work, maybe you also like (another work by/from/with the same $Char$,) the Obj	(4)

(5)	All topics	User	User Preferences (Engagement)	$Preference_u(o_j) > Preference_u(o_i)$? E.g. "Maybe you prefer (another work by/from/with the same $Char$,) the Obj?	(5)
(7)	All topics	User	Objects in KB	$Experience_u(o_i) \wedge Experience_u(o_j)$? E.g."Have you also seen (another work by/from/with the same $Char$,) the Obj?	(6)

9.6.1 Transition Strategy Selection

1. First of all, the agent's choice of transition strategies to introduce a new topic in the ongoing interaction is reduced by looking at the relation between the preceding topic and the topic that needs to be introduced (column 2, Table 9.3). If the underlying artworks of these topics have a characteristic in common, all strategies could make the transition between the topics. If the artworks have no characteristics in common, strategy 3 that refers to the agent's associations is excluded from the candidate strategies.

2. The next step is deciding between an agent or a user oriented strategy (column 3, Table 9.3). This is derived from the extent to which the agent is oriented towards the user (user's engagement) or itself (agent's preferences). This agent orientation is stored in variable $w(t)$ that we already defined in Section 8.5.1 as it also plays a role in the *selection* of the topics and subtopics of the interaction. Ranging from 0 to 1, 0 corresponds to a complete user orientation, and 1 corresponds to a complete agent orientation. Here, $w(t)$ serves as the probability with which an agent oriented transition strategy is chosen. For example, if $w(t) = 0.5$, half of the time the topic manager chooses a user oriented strategy and the other half of the time an agent oriented strategy. In this way the orientation of the agent is reflected in both the way the agent selects a topic and in the way it presents (introduces) the topic.

3. For the remaining three strategies (agent or user oriented) the topic manager checks for which strategy its premises are fulfilled. For example, strategy 2 (Table 9.3) that implies that the agent prefers the subsequent topic can only be selected if the agent indeed prefers the subsequent topic over the preceding

one.

4. A final strategy is chosen by avoiding repetition of strategies as much as possible in order to favour the variety within the interaction.

9.6.2 Transition Strategy Surface Form

When the transition strategy is selected the surface form of the strategy is formed: When the underlying artworks of the two topics have at least one characteristic in common, the surface form of the strategy always includes a reference to this common characteristic to insist even more on the artworks' (i.e. topics') "fellow membership of some category" [Levinson, 1983] (see Section 9.2.3); e.g. *"another artwork from the same artist"*. When the artworks have more than one characteristic in common, the characteristic that is referred to in the utterance depends on

1. The salience of the characteristic;
2. The order in which the characteristics of the former topic have been discussed.

With respect to the first criterion, in the museum context of project 'A1:1', we consider a common *artist* a more salient characteristic than another common characteristic such as a common *period*. This is because the number of artworks belonging to a particular artist is considerably smaller than for example the number of artworks belonging to a particular period, style or genre. The artist of an artwork is thus a more distinctive (salient) category and therefore preferred over other characteristics for the realisation of a transition utterance.

When all common characteristics are equally salient (e.g. a common *style* and *period*) the transition utterance will make reference to that characteristic (i.e. subtopic) that has been discussed most recently in the dialogue history of the preceding topic. In this way the transition utterance not only refers to the commonality between the topics but can also be considered as a continuation of the subtopic that started discussing the characteristic that both objects have in common. In this way the transition utterance is formed to try to favour the smoothness and continuity of the interaction.

With respect to the topic that follows the transition strategy, when a characteristic has been referred to in the transition strategy, this characteristic will not be discussed (again) as a subtopic (mentioned also in Section 8.5.2). This is to avoid redundancy in the interaction.

For the generation of the chosen strategy's surface form we use templates where the topic manager fills the slots with information about the artwork of the following topic. In the end we obtain an agent utterance that can look like *"I also like another artwork with the same style, the Good Samaritan by Vincent van Gogh."*. In Chapter 10 we describe the generation of the agent's turns in more detail.

The first topic of the interaction does not need a transition strategy but its introduction is performed with a strategy that has aspects in common with transition strategies: it also reflects either a user or agent orientation depending on w (Sections 9.6.1 and 8.5.2), and is equally generated by means of a template which allows any topic to be selected as first topic of the interaction. This results for example in *"I have just seen the Good Samaritan by Vincent van Gogh."* or *"So have you seen the Samaritan by Vincent van Gogh?"*

9.7 Conclusion

In this chapter we have looked at how a selected topic of discussion can be introduced by an agent in an ongoing non-task oriented dialogue. In the context we are looking at, each topic consists of the discussion of an artwork from a museum.

Based on social and conversational analytic literature we first constructed a set of candidate transition strategies. We then checked the consequences of each of these transition strategies on the perception of the dialogue and the agent. We have found that the strategies that score well on all dimensions and all tested circumstances are those that ask for the experience of the user, and those that refer to the preferences of the interaction participants. They maintain the coherence of the dialogue while maintaining/establishing a positive perception of the agent, even if the preceding and following topics have nothing in common except from the fact that they are both artworks. This shows that the agent's topic selection component of the topic manager can indeed be allowed to select any topic required to favour engagement at any moment in the conversation (Chapter 8).

We used the outcomes of the empirical study to model a component of the topic planner that automatically selects appropriate transition strategies for the conversation agent whenever the topic selection component initiates a topic switch. This topic transition model takes into account the relation between the preceding and subsequent topic (common object characteristics or not), the agent's orientation (user vs. agent), the variables in the agent's reasoning process to select a topic, and the dialogue history.

In the following chapter we first propose an implementation of the topic manager. In Chapter 11 we perform an evaluation of the topic manager that includes this topic transition component.

10 Topic Manager: Implementation

10.1 Introduction

In human-agent interactions the choice of the interaction topics can influence the user's engagement during the interaction (Chapter 7). We have therefore developed a topic selection model that takes into account the user's engagement, the agent's preferences and the agent's associations to decide when the agent should talk about what topic (Chapter 8). In Chapter 9 we proposed a way of introducing the selected topics on an utterance level. In this Chapter we describe the implementation of these two components, forming a complete topic manager, in a virtual agent platform. The architecture we propose leads to behaviours of an ECA (Embodied Conversational Agent) allowing agent initiated (agent topic initiation) human-agent interaction.

In the sections below we first describe the agent platform that we use for integration (Section 10.2), and a task network that serves as the dialogue manager within this platform (Section 10.3.2). In Section 10.4 we then describe how the topic manager relates to these components on an architectural level.

10.2 VIB Platform

To use the topic manager in a conversational virtual agent we integrate the model in the java-encoded VIB (Virtual Interactive Behaviour) platform [Pecune et al., 2014]. VIB enables the creation of ECA's by handling all the steps required to transform an agent's intention to the rendering of the agent's multimodal behaviour.

10.2.1 SAIBA

For this, VIB is based on the SAIBA architecture [Pecune et al., 2014]. As shown in Figure 10.1, SAIBA [Kopp et al., 2006] describes three stages of generating behaviour by two mediating languages FML (Function Markup Language) [Heylen

et al., 2008] and BML (Behaviour Markup Language) [Vilhjálmsson et al., 2007]: The `intention planner` generates *agent intentions* that are encoded in FML-files [Heylen et al., 2008]. FMLs form the input of the `behaviour planner` that generates *agent behaviours* that are encoded in BML-files [Kopp et al., 2006]. BML-files are sent to the player (`behaviour realisation`), which lead to rendered agent animations.

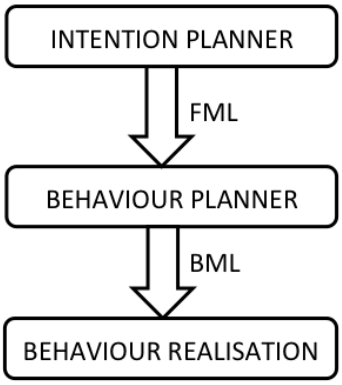


Figure 10.1 : SAIBA architecture

Looking in more detail, the FML-files (see Figure 10.2 for an example) encode the agent's speech, accompanied by tags that serve to determine the speech-accompanying non-verbal behaviours. The latter are based on the Affective Presentation Markup Language [De Carolis et al., 2004] and correspond to 4 communicative functions [Poggi, 2001].

10

— Information about the speaker's beliefs: Behaviours that provide information about for example the speaker's degree of certainty with respect to what he/she is talking about.

— Information about the speaker's intentions: Information about the speaker's goal through for example a performative or a focus of the speaker's utterance.

— Information about the speaker's affective state: The speaker can show his/her emotional state through particular facial expressions.

— Metacognitive information about the speaker's mental state: The speaker can try to remember or recall information.

A communicative function can be realised by different signals. For example, an emphasis of a word may be accompanied by a raised eyebrow, a head nod, or a combination of both signals. On the other hand, the same signal may be used to convey different meanings. For example, a raised eyebrow may be a sign of surprise, emphasis, or a suggestion.

The behaviour planner parses the FML to instantiate the communicative functions into the appropriate signals, written in BML (see Figure 10.3). BML tags indicate the modality of the agent's behaviour (head, torso, face, gaze, body, legs, gesture, speech or lips) and encode the begin and end of the signals. Further em-

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<fml-apml>
    <bml>
        <speech id="s1" start="0.0" language="english" voice="openmary" type="SAPI4" text="">
            <description level="1" type="gretabml">
                <reference>tmp/from-fml-apml.pho</reference>
            </description>
            <tm id="tm1"/>
                Hello.
            <tm id="tm2"/>
                My name is Leonard.
            <tm id="tm3"/>
        </speech>
    </bml>
    <fml>
        <deictic id="p1" type="selftouch" start="s1:tm2" end="s1:tm3" importance="1.0"/>
    </fml>
</fml-apml>
```

Figure 10.2 : FML example

```
<?xml version="1.0" encoding="UTF-8"?>
<bml xmlns="http://www.mindmakers.org/projects/BML" character="Greta"
    composition="blend" id="bml1">
    <gesture end="2.251" id="p1_0" lexeme="self_Pos_R" start="-0.041">
        <description priority="1" type="gretabml">
            <reference>deictic=self_Pos_R</reference>
            <intensity>1.000</intensity>
            <SPC.value>0.000</SPC.value>
            <TMP.value>0.050</TMP.value>
            <FLD.value>-0.050</FLD.value>
            <PWR.value>0.100</PWR.value>
            <REP.value>-0.100</REP.value>
            <OPN.value>0.000</OPN.value>
            <TEN.value>0.000</TEN.value>
        </description>
    </gesture>
    <speech xmlns="" id="s1" language="english" start="0.0" text=""
        type="SAPI4" voice="openmary">
        <description level="1" type="gretabml">
            <reference>tmp/from-fml-apml.pho</reference>
        </description>
        <tm id="tm1"/>
            Hello.
        <tm
        id="tm2"/>
            My name is Leonard.
        <tm id="tm3"/>
    </speech>
</bml>
```

Figure 10.3 : BML example

bedded tags further specify the non-verbal behaviour within each modality. The behaviour realiser generates the animations that are described in the BML message.

10.2.2 Modular

VIB has a modular architecture, where each module represents a functionality of the ECA. A particular ECA is thus defined by a particular set of modules [Pecune et al., 2014]. The output of one module is the input for another module and processes the flow of information that represents an event. Figure 10.4 shows a basic, SAIBA compliant, example of a module architecture in the VIB platform [Pecune et al., 2014].

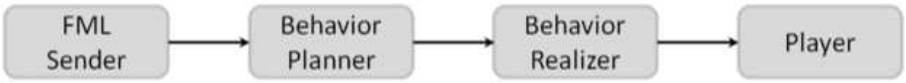


Figure 10.4 : An example of a set of modules that model an ECA in the VIB platform [Pecune et al., 2014].

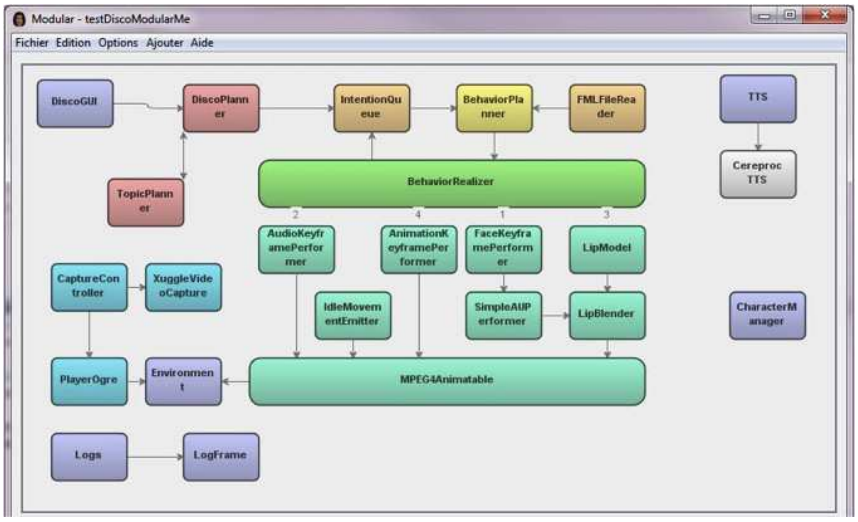


Figure 10.5 : VIB's modular

A graphical user interface (GUI) called the VIB's 'modular' allows to easily assemble the modules of choice to create an ECA [Pecune et al., 2014]. Figure 10.5 shows a picture of this *modular* containing an example set of modules.

VIB's modular architecture allows constant additions to the agent's functionalities. Thus, to endow an agent with the functionalities of the topic manager (described in the previous two chapters) we have created a module in the VIB platform that is responsible for the management of topics in an interaction with a human user.

This module consists of algorithms that describe *when* the agent should initiate and end *what* topic (Chapter 8), and *how* to introduce it in the interaction (Chapter 9).

To execute the topics, this topic management module is connected to a dialogue manager module. The latter module manages the agent's conversational behaviour in VIB-based human-agent interactions by means of the hierarchical task network Disco [Rich and Sidner, 2012]. In Section 10.4 we describe how we make both modules work together to send the desired agent intentions to the behaviour planner. In the following Section we first introduce the dialogue manager module without the connected topic manager.

10.3 Dialogue Manager Module

The dialogue manager module handles the alternation of user and agent contributions in the interaction. It receives input in the form of user answers, and outputs the agent's intentions (FML) that result in agent behaviours (Section 10.2.1). To decide what input leads to what output the module uses Disco, a hierarchical task network [Rich and Sidner, 2012].

10.3.1 Input and Output

The dialogue manager module receives as input information about the user's turns: It receives the user's choice for an answer that is selected among a set of potential user answer types that are predefined (anticipated to) for that moment in the interaction (further detailed in Section 10.3.2). The choice for an answer type can be obtained (in external modules) by, for example, speech recognition, or manual selection (keyboard or mouse clicks) of one of the predefined answer types. In the case of speech recognition, for example keywords of the user's turn can be spotted which then map to one of the predefined user response options (Figure 10.6).

Based on the input (answer) from the user, the Disco component selects what agent turn should be produced in response, forming the output of the module. For this, in the dialogue manager module Disco selects an FML-file from an intention library, which specifies the communicative intent of the agent's behaviour including the agent's speech [Heylen et al., 2008] (Figure 10.6). As explained above and shown in Figure 10.6, by means of the SAIBA framework the by Disco selected FML-file results in agent behaviour.

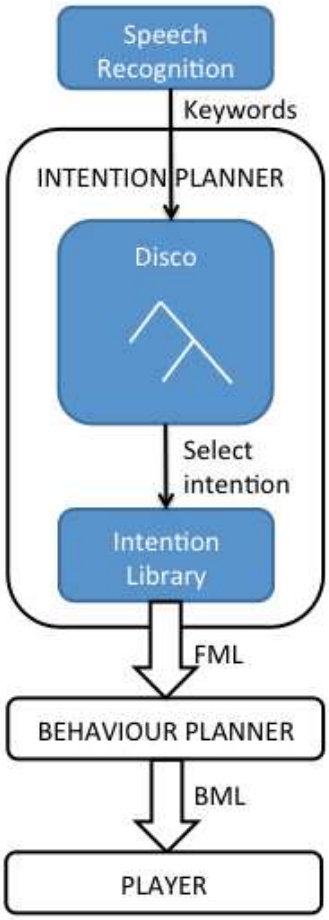


Figure 10.6 : SAIBA compliant architecture in VIB incorporating DISCO

10.3.2 Disco

Within VIB's dialogue manager module, the hierarchical task network Disco [Rich and Sidner, 2012] determines what user turn leads to what agent turn. Disco handles namely, in a hierarchical order, predefined tasks that are executed by alternating agent and user contributions. Disco thereby enables the agent to manage predefined and strictly ordered interactions with a user.

The tasks of the interaction handled by Disco are listed in an XML-file. Disco reads the scripted hierarchically ordered steps of the interaction as a tree of tasks where one task can be decomposed in multiple sub-tasks (i.e. children tasks). Figure 10.7 shows an example task tree (dialogue tree). Each node represents a task. The circled nodes are tasks that are further decomposed into sub-tasks (children

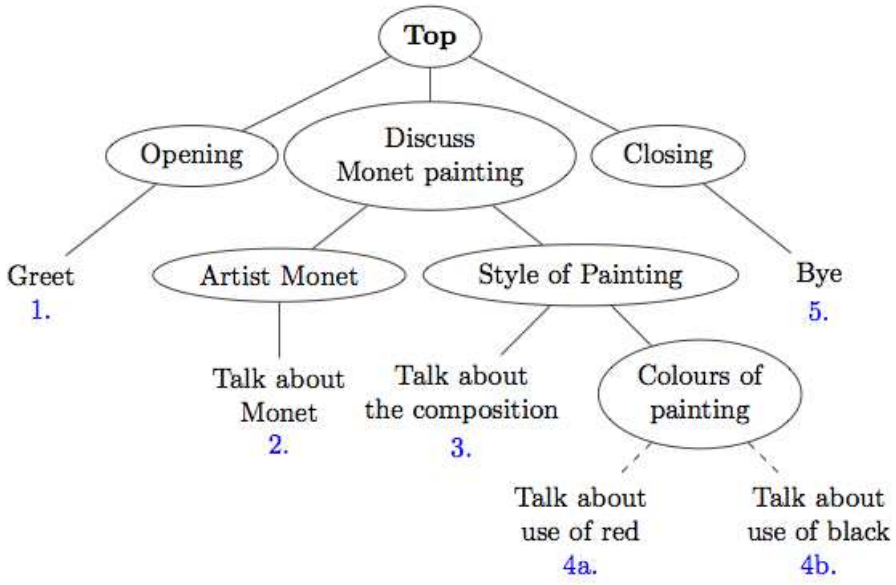


Figure 10.7 : Example of a Disco task network (dialogue tree in our application). Circled nodes represent tasks that are further decomposed, non-circled nodes represent primitive tasks that are realised by agent intentions resulting in agent behaviours, and user reactions. Dashed edges represent alternative tasks. The numbers indicate the order of task execution.

nodes). The tasks that are presented at the leafs of the tree are tasks that are not further decomposed and are called primitive tasks (non circled nodes in Figure 10.7). Primitive tasks consist of alternating agent turns and possible user turns (further explained below).

For any task, its parent task forms the goal of the child task. Executing all the tasks (children tasks) of a goal (parent task) completes the goal (parent task). As illustrated by the numbers in Figure 10.7, the primitive tasks of the task tree (dialogue tree) are executed in a linear order from left to right to complete their respective goals (parent tasks, circled nodes) from bottom to top. Dashed edges indicate alternative tasks: only one of the alternative tasks needs to be executed to complete their goal.

In Figure 10.7, the `top` task of the tree that covers the overall interaction is decomposed into three children tasks: `opening` the interaction, `discussing` a painting, and `closing` the interaction. These task form the goals of their children tasks. In this example the `opening` goal (task) can be completed by performing its primitive child task that consists of user and agent utterances by which the agent and user

`greet` each other (task 1). Then, the tasks of the second goal (`discussing` the painting, tasks 2, 3 and 4), are executed from left to right, thereby first completing the goal `Artist Monet` by executing task 2. For `Style of Painting` task 3 is executed followed by task 4a *or* 4b. Performing task 4a *or* 4b completes respectively the goal (parent task) of talking about the `colours of the painting`, the `style of painting` and `discussion of the painting`. When finally the task of the `closing` phase, namely saying `bye` (task 5), is completed as well, the `top` goal of the interaction is achieved.

```
1  <agent  id="MondrianComposition" text="mondrian_style_composition.xml" >
2      <user text="squares">
3          <agent text="mondrian_style_explain_squares.xml">
4              <user text="Indeed"/>
5              <user text="Do not know"/>
6          </agent>
7      </user>
8      <user text="abstract">
9          <agent text="mondrian_style_like_abstract.xml">
10             <user text="Me too"/>
11             <user text="Do not like"/>
12             <user text="Do not know"/>
13         </agent>
14     </user>
15 </agent>
```

Figure 10.8 : Primitive task consisting of agent turns and user answer types.

As mentioned before, a primitive task consists of alternating agent and user contributions to the interaction. Different user contributions (turns) can lead to different agent responses. To illustrate this, Figure 10.3.2 gives an example of what the (non-compiled) code of a primitive task looks like, corresponding to primitive task 3 from the example tree in Figure 10.7. The first line shows that the task starts off with an agent turn by pointing to an FML-file (.xml) (here talking about the composition of the painting). Then, for the user there are 2 types of responses anticipated: Either the user talks about the "squares" in the painting (line 2), or the "abstract" style of the painting (line 8). In Section 10.3.1 we explained how one of these user answers types can be selected. When the answer type "squares" is selected the agent responds with a turn that elaborates on this subject (line 3, pointing to an FML). When the answer type "abstract" is selected the agent responds differently (another FML, line 9). A selection of a user answer in reaction to one of these agent

contributions then, (resp. one out of lines 4-5, or one out of lines 10-12) completes the primitive task.

10.4 Topic Manager Module

As we have shown in the previous section and Figure 10.7, Disco's tree structure offers a strict task-oriented approach, where the tasks of the interaction are performed in a rigid predefined order. While Disco can select a task from a fixed set of alternative tasks at a specific predefined moment in the interaction (dashed lines in Figure 10.7) it does not allow the possibility of adjusting (breaking-off, starting, changing) the ongoing tasks of the interaction on the fly.

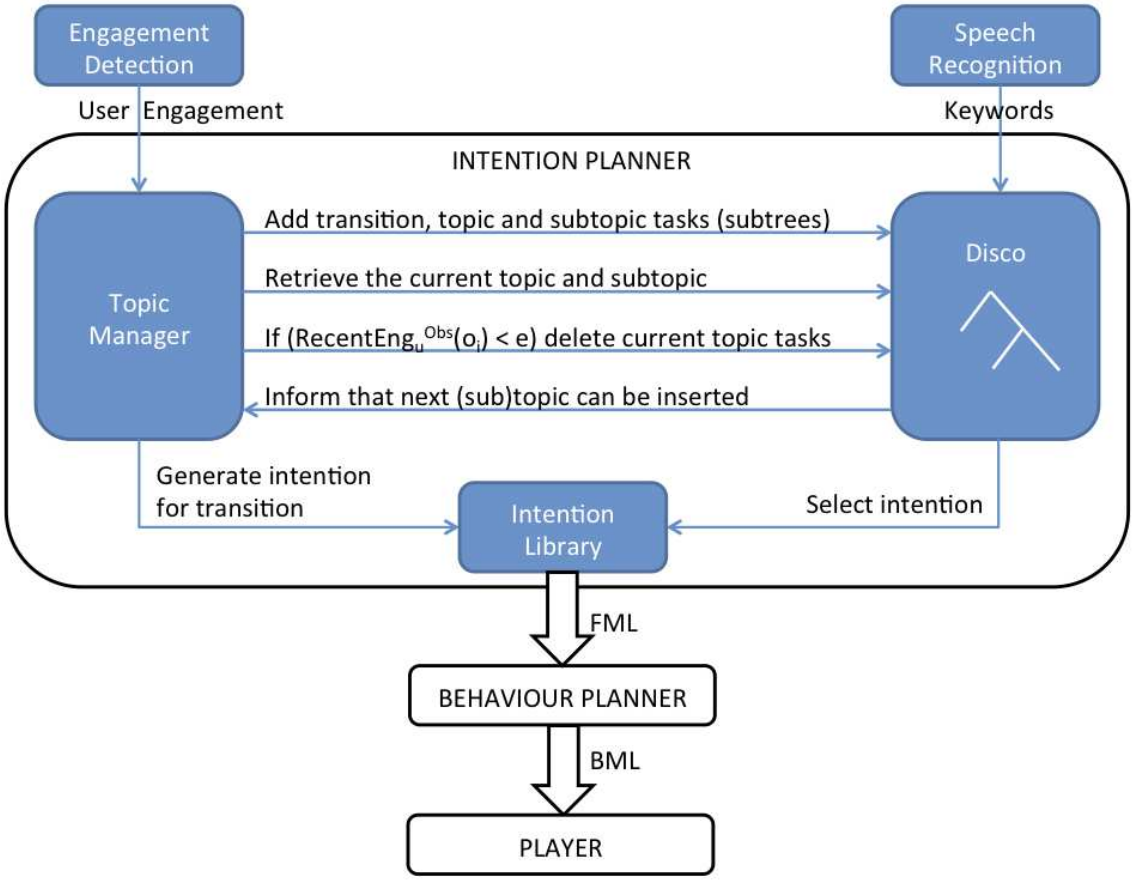


Figure 10.9 : Architecture

Therefore, to make the agent able to manage interactions whose topics are continuously adapted to the ongoing interaction we integrated a topic manager module that is connected to the dialogue management (Disco) module: The topic man-

ager module adds flexibility to Disco's fixed order of task execution by continuously exploiting and modifying the dialogue tree that Disco processes (hierarchical task network, XML-file) during the interaction.

Figure 10.9 gives an overview of all the actions between the topic manager and Disco that together ensure that Disco's dialogue tree is modified on the fly. In this way we continue using Disco's predefined order of task execution where possible, but add flexibility where necessary.

As shown in Figure 10.9, the topic manager receives input from two sources: information about the structure of the ongoing interaction from Disco, and the engagement level of the user sent by an engagement detection module. Below we respectively explain the actions of the topic manager (Figure 10.9) that follow from these two inputs: topic management driven by the dialogue (Section 10.4.1) and topic management driven by detected user engagement (Section 10.4.2).

10.4.1 Dialogue driven Topic Management

As mentioned above, in order to obtain flexible dialogue we use a strictly ordered Disco dialogue tree (task tree) and have the topic manager adapt this tree on the fly. The first step is thus creating the initial tree: we define all the parts of the dialogue that have a fixed structure and encode these as multiple non-connected subtrees (in the XML-file). Figure 10.10 shows an example of such subtrees (look only at the black lines). It shows one subtree for the `top` of the interaction, one for a `topic transition`, one for each topic corresponding to an artwork discussion (`TopicArtwork`), and one for each `subtopic` corresponding to the discussion of an artwork's characteristic.

The topic manager then uses each of these subtrees as building blocks for the overall interaction tree [1], as shown by the coloured lines in Figure 10.10. For this, we first indicate that the interaction is executed by following the steps of the top tree, the one that is headed by the `top` goal and not by one of the other subtrees. The topic manager then selects during the interaction which blocks (subtrees), representing topics, subtopics, or topic transitions, are inserted within this `top` tree fragment.

To know when a block (subtree) can be inserted given the structure of the ongoing interaction, the task in the top tree that precedes the place where a potential (sub)topic could be inserted, informs the topic manager (see Figure 10.9) that a

1. features added for us by Charles Rich, the developer of Disco, to make this technically possible

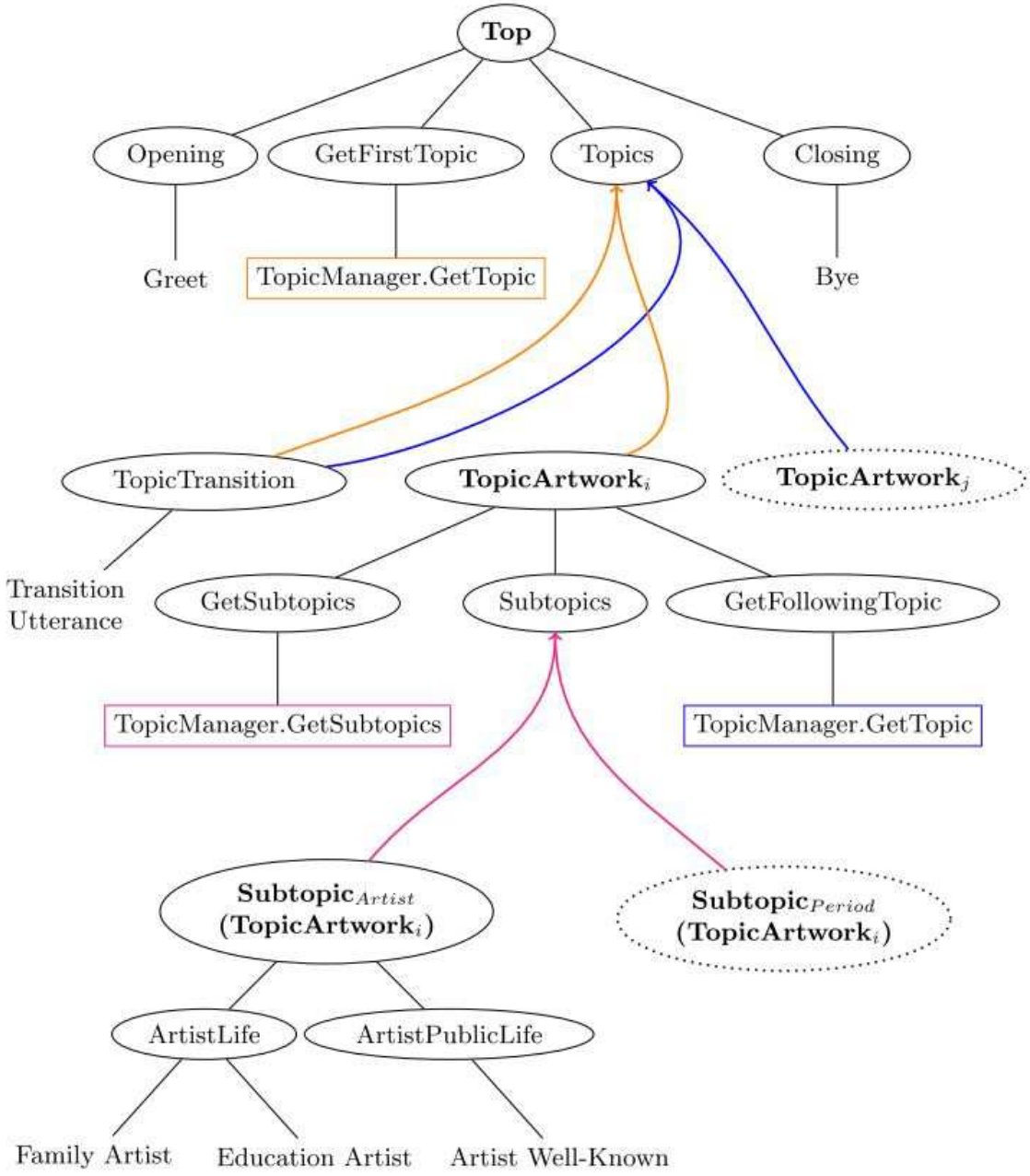


Figure 10.10 : Dialogue tree assembled from subtrees. Circled nodes are non-primitive tasks. Dotted circled nodes are non-primitive tasks that are further decomposed into tasks that are not shown in the figure. Non-squared leafs represent primitive tasks that are realised by agent and user turns (as in Figure 10.3.2). Squared leafs are primitive tasks that execute a script.

potential place for subtree insertion arrives (i.e. follows the current task). This task of informing is performed by a script in the dialogue tree that calls the topic manager, represented by the coloured rectangles in Figure 10.10. Calling the topic manager at this moment in the interaction allows the topic manager to select and subsequently insert a (sub)topic at the last minute, taking into account the most recent estimations of the user's engagement (Chapter 8 Section 8.5.4 describes how).

When the topic manager then decides to add a transition, topic and/or subtopic to the ongoing interaction, it inserts the corresponding transition, topic or subtopic subtree into the top tree. In this way the interaction is assembled in real-time, adapted to the ongoing interaction. The insertion procedures are shown by the coloured lines in Figure 10.10 where each colour corresponds to the script (squared leaf) that has triggered the insertion. The topic manager finds the appropriate position (see Figure 10.10) to insert the subtrees in the top tree by parsing the top tree top-down. Because after the insertion, the transition, topic and/or subtopic subtrees form part of the top tree, Disco can then handle the execution of their corresponding tasks.

As shown in Figure 10.3.2, a primitive task that consists of human-agent turns (non-squared leafs in Figure 10.10) points to an FML-file to specify an agent intention which results eventually in the agent's behaviour (i.e. turn, see Section 10.3.2). The agent's intention (FML-file) is selected from a fixed set of FML-files (i.e. intention library). However, the topic manager can also automatically generate FML-files during the interaction and place them in the library, where they can be selected by the Disco module (shown in Figure 10.4). More specifically, when the topic manager decides that an introduction or transition utterance is needed in the interaction (see Figure 10.9), just before the topic manager inserts the transition subtree into the top-tree, the topic manager automatically generates the FML-file for the transition. This file describes the custom-made transition utterance for that situation: for a particular set of topics, taking into account a particular dialogue history (previous transition utterances, see Section 9.6). Disco then subsequently executes the agent turn of the transition task by pointing to this FML-file.

At the moment when the topic manager decides that no further topics need to be addressed in the interaction, it does not insert any more subtrees to the top-tree and the interaction passes to the following child task of the top tree (e.g. `closing` in Figure 10.10), thereby completing the interaction.

10.4.2 Engagement driven Topic Management

Besides the dialogue structure, there is another reason to modify the top tree, this time triggered by the topic planner's second input. This is when the topic manager receives information about the user's engagement and calculates that the user's engagement is too low to carry on with the current topic (described in Chapter 8 Section 8.5.3).

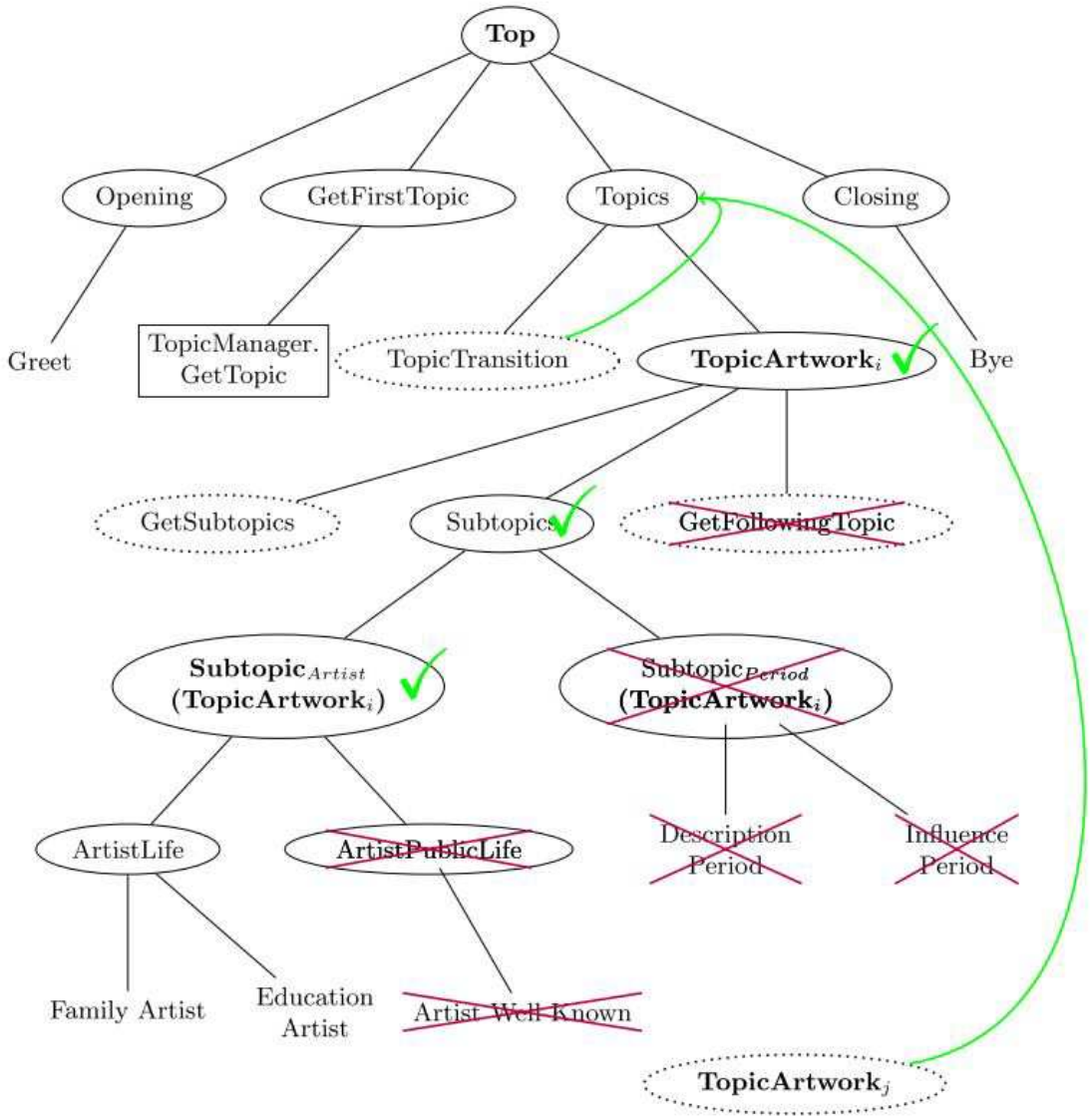


Figure 10.11 : Dialogue tree where the topic manager breaks-off the current topic and inserts a subsequent topic when the Focus of the interaction is on `Education Artist`.

Whenever the topic manager receives a measurement of the user's engagement, it retrieves the current topic and subtopic of the interaction from Disco's dialogue tree (see Figure 10.9), and calculates if the user's engagement during the discussion of the current topic is high enough to carry on with the topic (described in Sections 8.5.3 and 8.5.4). If not, the topic manager needs to stop the current topic and optionally introduce a different topic (see Figure 10.9).

The topic manager module stops the current topic by deleting all the tasks of the current topic that have not yet been executed from the top tree and marking the goals that have been partially completed as completely completed. It then calculates if a new topic should be introduced to the interaction and if so, inserts the corresponding subtrees in the same way as described above. This way of modifying the dialogue tree is shown schematically in Figure 10.11.

To perform the breaking-off of a topic, we developed an algorithm for the topic manager that parses and modifies the tree bottom-up. This is necessary because Disco methods only allow access to the dialogue tree (task network) from its top, or from its bottom by the task that is currently in progress (`focus`); and they only allow marking of (deleting/marking as completed) primitive nodes (i.e. tasks that have no child tasks that need to be executed). Algorithm 1 shows the pseudocode of the algorithm, that also allows breaking-off the topic when the `focus` of the interaction (task in progress) is still on the transition strategy that introduces the topic.

Where the previous subsection described how the topic manager manages the topics of the interaction whenever the dialogue structure requires this, in this subsection we showed how the topic manager also manages the topics of the interaction in real-time whenever the user's engagement level requires this. The topic manager continuously controls the current topic of the interaction and adapts it where necessary, thereby creating flexible user adapted interaction.

10.5 Conclusion

In this chapter we have proposed an implementation of the topic manager whose models were described in the previous two chapters 8 (topic selection) and 9 (topic transitions). More precisely, we integrated a topic manager module in the VIB platform and developed an architecture that allows this topic manager to manage the agent's *timing*, *selection*, *initiation* and *breaking-off* of topics in human-agent interaction.

To manage the agent's overall conversational behaviour, we let the topic manager

```
ParentTask = parent of FocusTask;
forall Levels of the dialogue tree do
    if ParentTask is a part of the TopicTask then
        forall DescendentTask of ParentTask (= all tasks below ParentTask)
         do
            if DescendentTask ≠ FocusTask and DescendentTask ≠ done and
             DescendentTask == primitive then
                Mark DescendentTask;
            end
        end
    end
    if ParentTask == TopicTransitionTask then
        ParentTask = TopicTask (= Child task of ParentTask);
    end
    if ParentTask ≠ TopicTask then
        Parent = parent of ParentTask;
    else
        break;
    end
end
```

Algorithm 1: Topic manager module's procedure to break-off a topic by modifying a Disco dialogue tree (task network). The *levels of the dialogue tree* refers to the number of vertical tasks from the top task to its lowest primitive task. The *TopicTask* refers in Figure 10.11 to the third child node of the top task, and the *TopicTransitionTask* to the first child of this node (*TopicTask*).

add flexibility to a strictly task-oriented dialogue tree (Disco) by continuously exploiting and modifying this dialogue tree. In this way we continue using a predefined order of task execution where possible (e.g. within subtopics), but add flexibility where necessary (e.g. among topics).

We have shown how the topic manager performs modifications to the dialogue tree whenever 1) the structure of the ongoing interaction requires this, and 2) whenever the user's engagement requires a change of topic. In this way the topic manager manages the topics of the interaction in real-time, creating flexible user adapted in-

teraction.

11 Topic Manager: Evaluation

11.1 Introduction

In Chapter 7 we have seen that the choice of the topics in human-agent interaction (precisely the user's preferences for the objects under discussion) is correlated to the user's engagement during the interaction. For so far as possible, personalised topic management is thus not to be neglected in a conversational agent that wants to engage its users. In Chapters 8 until 10 we have therefore proposed a topic manager for conversational agents in the context of information-giving chat; where the agent's goal is to transfer a maximum amount of information to the user while the exact choice of information is of less importance. The freedom of topic choice that information-giving chat provides makes that locally the interaction is not strictly task-driven but can be driven by social variables of the interaction.

The topic manager we proposed allows an agent to adapt the topics of the interaction on the fly to favour the user's engagement. By looking at the structure of the ongoing dialogue, and the perceived (observed/detected) engagement level of the user the topic manager decides *when* to introduce a topic (Section 8.5.3 and Chapter 10). By taking into account the agent's perceived level of user engagement as well as the agent's own preferences and associations it decides *what* topic to introduce in the interaction (Chapter 8). And by taking into account the commonalities of subsequent topics, variables of the agent's topic selection process, and the dialogue history it decides *how* to introduce the topics in the ongoing interaction (Chapter 9).

In this chapter we evaluate the topic manager in an application where a virtual agent plays a visitor in a museum who tries to pass on a maximum amount of cultural information to the user (which is the application described by project A1:1 that forms the application of this thesis). In Chapter 7 we have already seen that the user's engagement is favoured if an agent talks about artworks that the user prefers. In the present evaluation we therefore focus on the question if the topic

manager is indeed perceived as doing what it does, namely adapting the topics (to different extents) to the user's engagement (and thereby to his/her preferences) and the agent's own preferences, by associating the interaction topics. Secondly, we verify if the topic manager has an influence on the human's perception of the agent and its dialogue.

11.2 Methodology

In order to evaluate the effects of the topic manager we consider different human-agent interactions where in each interaction the agent uses a different way of topic management (conditions, see following subsection). To evaluate the effects of the agent's topic management only and control for all the other variables of the interactions we assemble video fragments of computed virtual agent turns and acted human user turns, forming videos of presumed human-agent interactions. These videos form the stimuli of the evaluation study. The stimuli are shown to (third party) human observers who we ask for their perception of these interactions (as in e.g. [Garber-Barron and Si, 2013]).

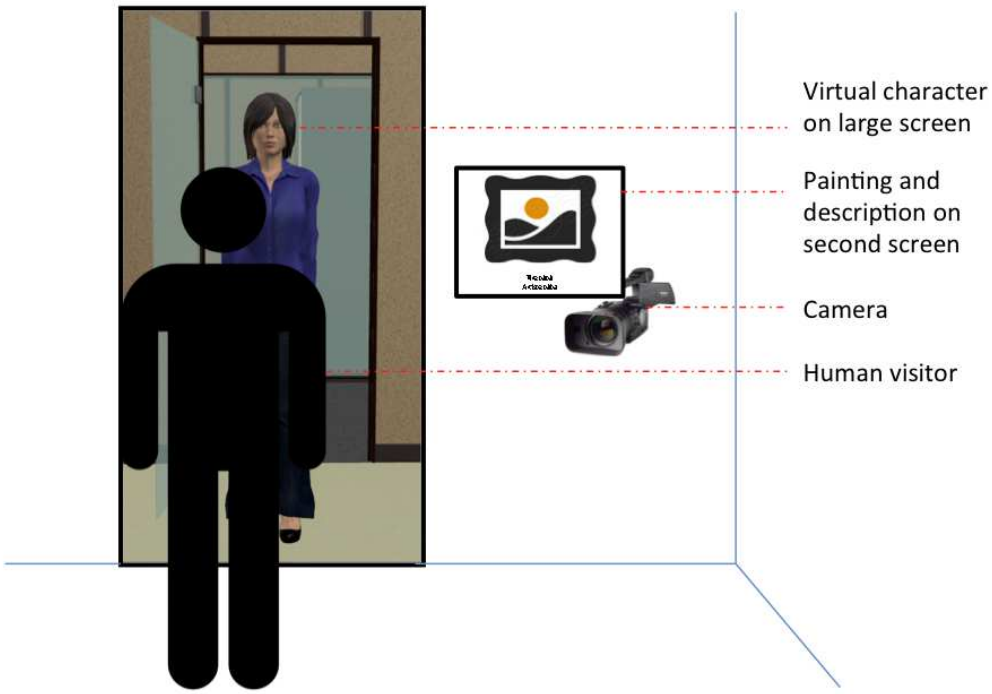


Figure 11.1 : Pretended setting of the interactions.

Both the virtual agent and the human interactor are modelled as visitors of a

museum. The pretended setting of the interactions is shown in Figure 11.1. The virtual agent has a human-size and interacts by means of speech, gestures and facial behaviours. It can show the artworks it talks about on a second screen. The human user talks to the agent as if he would talk to another human, using natural speech.

Creating stimuli videos where the human interactor's behaviour is controlled and having third parties (observers) evaluate the assembled interactions is necessary because there are some practical issues that make free interaction and evaluation by the human interactors themselves (second party evaluation) problematic:

First of all, at the moment, interacting with a virtual agent is still something that is new and exciting for most people. We therefore 'risk' that the human user displays engaged behaviour during the entire interaction. If there is no moment where the human user is clearly less engaged, we cannot sufficiently evaluate the perception of an engagement favouring model.

Secondly, in order to compare the effects of different topic managements in a truly balanced way, we would need the human users to reproduce the same level of engagement for the same topics (object discussions) in different interactions. This is necessary because in order to test the effects of the model, the input of the model, namely the engagement level of the user, should be consistent over the different conditions (interactions). However, in a within-participant design it is likely that every time (for every condition) a user is exposed to the same artwork discussion, he/she is less engaged. In a between-participant design, it is unlikely that users show the same degrees of engagement for the same set of topics, which not only influences the model's output but also biases the users' judgements of the interactions. These problems could be reduced to a small extent by obtaining a very large number of participants in each condition, but this is a costly procedure in situated face-to-face interaction.

The only way to guarantee that a human user is less engaged during some potential artwork discussions than others and guarantee that the user has consistent engagement levels among the different interactions (conditions), is by prescribing the human user's behaviour for each user turn. We record these acted user turns and use them in reaction to the same (computed) agent turns among different videos (interactions/conditions) that serve as the stimuli of the evaluation study.

Below we describe all the steps of this methodology: the conditions we consider, setting the agent system for the preparation of the stimuli, producing the video fragments (stimuli) consisting of acted human user turns and computed agent behaviour, and the experimental setup.

11.2.1 Conditions

In order to evaluate the effects of the topic manager and its configurations we consider 4 different conditions (assembled human-agent interactions). In each condition the agent has a different way of managing topics. In 3 conditions the agent manages the topics with different configurations of the topic manager, and 1 other condition forms a control condition without the topic manager. Table 11.1 gives an overview of the different ways of topic management in each condition.

Condition Name (Independent Variable)	**Condition**	**(Sub)Topic Selection** (What)	**Abandoning topics** (When)	**Transition Strategies** (How)
`TMw=0`	With topic manager, $w = 0$	User oriented	If user engagement too low	Validated, User oriented
`TMw=0.5`	With topic manager, $w = 0.5$	User-Agent oriented	If user engagement too low	Validated, User/Agent oriented
`TMw=1`	With topic manager, $w = 1$	Agent oriented	If user engagement too low	Validated, Agent oriented
`WithoutTM`	Without topic manager	Random topics (controlling different selection than other conditions), Fixed subtopics	No	Non validated, Object oriented

Table 11.1 : The conditions of the evaluation study. The configurations of the conditions with topic manager result in the different ways of topic management described in Rows 1/3, Columns 3,4 and 5. For condition `WithoutTM` we endowed the agent with a different way of managing topics, described in Row 4, Columns 3,4,5.

Table 11.1 lists for each condition (row), its name (Column 1), its description

(Column 2), and the different ways of topic management that result from the conditions: *what* (sub)topics are selected ((sub)topics oriented towards the user, the agent, both, or randomly selected, Column 3), if and *when* topics are abandoned prematurely (Column 4), and *how* the topics are introduced in the interaction (Column 5, orientation of (non)validated transition strategies).

In the first 3 conditions (`TMw=0`, `TMw=0.5`, `TMw=1`) the agent uses the topic manager (TM) that is described in Chapters 8, 9, and 10. These conditions correspond to 3 different configurations of the variable w of the topic manager. w represents the agent's orientation towards the agent itself versus the user (Section 8.5). It plays a role in the agent's selection of the (sub)topics of the interaction (Section 8.5) and its selection of transition strategies to introduce the topics (Section 9.6.2). Condition `TMw=0` ($w = 0$) reflects an agent that is completely oriented towards the user. Condition `TMw=0.5` ($w = 0.5$) reflects an equal user and agent orientation, and `TMw=1` ($w = 1$) reflects an agent that is completely oriented towards itself. These configurations of the agent's topic manager lead to the different ways of topic management that are listed in Columns 3,4 and 5 of Table 11.1.

Condition `WithoutTM` forms the control condition. Here the agent does not use the topic manager from the previous chapters, but uses instead random topic selection; does not abandon topics prematurely even if the user shows low engagement (Section 8.5.3); and uses transition utterances that were not approved for the topic transition model of the topic manager (Chapter 9).

11.2.2 Stimuli Preparation: Agent System Settings

We create a total of 4 stimuli, one for each of the 4 conditions (listed in Table 11.1). A stimuli is formed by assembling video fragments of acted user turns and computed agent turns. For the agent turns we thus have to define the settings of the agent system that computes these turns. In the section above we have already defined the independent variable that we want to evaluate: the way of managing topics (differs among conditions). We therefore keep all other settings of the agent system constant among the conditions:

— Static variables that the agent's topic manager takes into account

— Input of the agent system

— Agent intentions that may be used to realise (sub)topics

Below we respectively define all these constant variables that the system can use to compute the agent turns.

11.2.2.1 Topic Manager: Knowledge Base

3 out of the 4 conditions (stimuli) use the topic manager. The topic manager employs several static variables (in its knowledge base, see Section 8.5) that we keep constant among all conditions (interactions), even if the condition `WithoutTM` may not act upon all these variables:

Artworks : In the application that we consider for the evaluation study (museum setting), and just as in the previous chapters, each topic corresponds to the discussion of an artwork of the museum and a subtopic corresponds to the discussion of an artwork's characteristic. We therefore define the artworks and their characteristics that form the potential (sub)topics that the agent can select for discussion.

We keep the number of artworks (potential topics) and characteristics (potential subtopics) very limited to facilitate a human evaluator in judging the agent's reasoning about the (sub)topics: 7 artworks with each 4 characteristics (*period*, *artist*, *genre* and *style*), shown in Appendix E.

For the same reason of simplification, 6 of the artworks clearly belong to one out of two categories: On the one hand we have art-nouveau and post-impressionist portraits (Figures E.1a until E.1.c, Appendix E), and on the other hand baroque and renaissance styled religious paintings (Figures E.1e until E.1g). The 7th artwork (Figure E.1d) falls in between both categories: it is a post-impressionist religious painting.

To limit the length of the interaction videos (stimuli) to more or less 5 minutes, we set the maximum number of topics that can be addressed in the interactions to 4 for the conditions `TMw=0`, `TMw=0.5` and `TMw=1`, and to 3 topics in the condition `WithoutTM` (because in the latter condition the topic fragments are not stopped before they are discussed completely).

Agent's Preferences : The second fixed variable used by the topic manager that we keep stable among all conditions (though not acted upon in `WithoutTM`) are the agent's preferences. With respect to the agent's preferences we suspect that not every museum would appreciate an agent that has negative preferences for some of its artworks. We therefore decided to model an agent that has a slight preference for some artworks over others, but has no negative preferences. Such agent preferences ensure that the agent can consider any potential topic that could engage the user.

We make the preferences of the agent for the artworks accumulative, meaning that they are composed from the agent's preferences for each of the artwork characteristics separately (see Section 8.5.1). The agent has high preferences for the

religious genre and the baroque and renaissance styles. For all other characteristics the agent has a neutral positive preference (0.5). These preferences lead to 3 slightly preferred paintings ($0.65 \leq \texttt{PREF}_a(o_e), \texttt{PREF}_a(o_f), \texttt{PREF}_a(o_g) < 0.68$), where o_e corresponds to the artwork from Figure E.1e, o_f to Figure E.1f, and o_g to E.1g ; and 3 paintings with neutral agent preferences ($\texttt{PREF}_a(o_a), \texttt{PREF}_a(o_b), \texttt{PREF}_a(o_c) = 0.5$), corresponding to artworks from Figures E.1a until E.1.c. For the post-impressionist religious painting (Figure E.1d) the agent has a preference that falls in between both levels ($\texttt{PREF}_a(o_d) = 0.6$).

Agent's Associations : The agent's similarity scores (Section 8.5.1) between the artworks and their characteristics are also constant among the conditions (not used in `WithoutTM`). The similarity scores are responsible for the agent's associations and reflect the two groups of artworks that we have mentioned above: on the one side art-nouveau and post-impressionist portraits and on the other hand baroque and renaissance styled religious paintings. The similarity scores between the objects are derived from the similarity scores between the individual characteristics (Section 8.5.1):

Characteristics have a similarity of 1 towards themselves. Similarities between `period` pairs are relative to their chronological distance. With respect to the `styles`, only the pairs *renaissance–baroque* and *art-nouveau–post-impressionism* have high similarity scores (0.7). The only `genre` pair *portrait–religious painting* has a similarity of 0. `Artist` pair *Da Vinci–Michelangelo* has similarity of 0.5, while all other `artist` pairs have similarity 0.

11.2.2.2 Topic Manager: Input

We also keep the input to the agent system constant among all conditions, namely the user's engagement during the different (sub)topics (Section 8.5.4). The topic manager uses this variable in its topic management (Chapters 8 and 9). In this evaluation the user's speech does not influence the agent's topic management, nor any other aspects of the agent's behaviour, making the user's engagement the only input for the agent system (see also Section 11.2.2.3). In this way the evaluation is focused on the agent's behaviour in reaction to the user's engagement only.

The user is an acting human interactor, which allows us to prescribe (roughly) the level of engagement that the human interactor conveys during every user turn. We manually enter these user engagement levels as input to the agent system after each user turn.

To determine what engagement level the user should convey during each user turn (belonging to a (sub)topic), we first define the user's preferences for all the artworks and characteristics. Given that the user's engagement is positively related to the user's preferences (Chapter 7), the user is prescribed to convey engagement levels that correspond to the user's preferences (high or low, Section 11.2.3.1 explains how).

We assign the human interactor (user) preferences that oppose the preferences of the agent because this causes a clear difference between the interactions that use an agent oriented configuration (condition `TMw=1`) and those that use a more user oriented configuration (`TMw=0`).

Besides this, as mentioned before, in order to properly test an engagement driven topic model the user needs to display lower engagement during some topics than others. We therefore do assign negative preferences to the user, while the agent's lowest preferences are neutral.

The above two considerations lead us to define high user preferences for the 3 art nouveau and post-impressionist portraits (Figures E.1a until E.1.c from Appendix E), low user preferences for the 3 baroque and renaissance religious paintings (Figure E.1e, until Figure E.1g), and an average preference for the post-impressionist religious painting (Figure E.1d). These preferences are consistent with (because based upon) the user's preferences for the objects' characteristics: Objects with high/low user preferences have characteristics with high/low user preferences. Only the artwork for which the user has an average preference that is neutral, the user's preferences for the artwork's characteristics vary: He has a higher preference for the characteristics `artist` *Van Gogh* and `style` *post-impressionist*, low preferences for the `genre` *religious*, and an averagely preference for the `period` *19th century*.

Precise values of user preferences are not defined though, since they only serve to give an indication of the level of engagement that the human interactor is prescribed to convey during the (sub)topics, which forms the input of the agent system: When the agent speaks about an artwork's characteristic for which the user has a high preference, the human interactor shows high engagement, and when the agent speaks about a characteristic for which the human interactor has a low preference he conveys a low level of engagement.

To keep the evaluation transparent (i.e. avoid complexity), after each user turn where the user conveys a higher level of engagement we enter a user engagement level of: $\overset{meas}{\texttt{Eng}}_u\ (t) = 0.8$ to the agent system. Similarly, after each user turn with a lower level of user engagement we enter $\overset{meas}{\texttt{Eng}}_u\ (t) = 0.2$, and after user turns with

average engagement we enter $\overset{meas}{\texttt{Eng}}_u\ (t) = 0.5$. If the user's utterance contains an explicit preference, we assign a certainty of $Cert(\overset{meas}{\texttt{Eng}}_u\ (t)) = 0.8$ to the engagement estimation. If not, we fill in a certainty of $Cert(\overset{meas}{\texttt{Eng}}_u\ (t)) = 0.3$. A measurement certainty of 0.3 makes that the agent in most of the cases has to detect at least twice an engagement level below average (< 0.5) before it changes the topic.

11.2.2.3 Agent Intentions

The last setting of the agent system that we keep constant among all conditions is the set of potential agent intentions that the agent can select intentions from to realise the (sub)topics that it addresses in the interaction.

We describe each potential agent intention in an FML file (see Section 10.2). When the agent's intention (FML) is selected during the interaction, it results in multimodal agent behaviour consisting of synthesised speech (French), gestures and facial expressions (explained in Chapter 10). We define the agent's intentions by considering the following criteria:

— Because we are only interested in the effects of the choice, timing and introduction of the topics in the interaction (see Table 11.1), to avoid biases we keep the agent's intentions to realise different topics as equal as possible without repeating information.

— Each of the (sub)topic agent intentions is designed to provide some cultural information about the artwork and/or its characteristic under discussion. The transmission of cultural information is, after all, the final goal of the interaction (Section 11.1).

— The agent does not ask the user questions in order to avoid that the turn with a question is selected in some conditions and not in others. Showing unbalanced occasional interest in the user could bias the human perception of the agent's behaviour between the conditions [Peters et al., 2005b].

— The agent does not show an ability to react to the content of the user's utterances (e.g. response to a question) to represent possible limitations of a natural language understanding component and to focus the evaluation on the agent's behaviour in reaction to the user's engagement only.

— Since the agent has no negative preferences (see Section 11.2.2.1) and tries to engage the user, the agent shows engaged behaviour for every turn: besides

providing information about the artworks on its own initiative, its gestures and facial expressions show interest in the topics [Peters et al., 2005a].

There are several forms of non-verbal behaviour that can (along with the agent's speech) realise the same agent intention. This makes that among the different conditions (interactions) that display the same agent intentions, the agent's non-verbal behaviour may vary slightly while the agent's speech is the same.

As we mentioned in Section 11.2.1, the only potential agent intentions that are not the same among the conditions are those that form the agent's transition utterances. This is because the content of these utterances is automatically generated by the different ways of topic management (conditions). For the conditions with topic manager, the set of templates that is used to generate the validated transition utterances is listed in Chapter 9 Table 9.3. For condition `WithoutTM`, less coherent (Section 9.4), non user or agent oriented utterances are generated to introduce the topics, e.g. *"Another artwork is the Self Portrait by Van Gogh"*. To avoid repetition the forms of the non-validated transition strategies vary, changing the word order and choice (e.g. *"artwork"* vs. *"painting"*).

11.2.3 Stimuli Production: Video Fragments

In the previous Section 11.2.2 we have prepared the creation of the stimuli by setting the variables that the agent's topic manager takes into account (Section 11.2.2.1), the user's engagement that forms the input of the agent system (Section 11.2.2.2, and the potential agent intentions to realise to agent's turns (Section 11.2.2.3). These preparations allow us in this section to produce video fragments of acted human user turns and computed agent turns. The acted human user turns convey the user's engagement levels, which serve as the input of the agent system (Section 11.2.2.2). The computed agent turns are the output of the agent system. We assemble the fragments of recorded user and agent turns into videos that form the stimuli of the evaluation study.

11.2.3.1 Acted User Turns

As explained in Section 11.2 the user's behaviour needs to be consistent among the interactions (conditions). For each potential agent turn we thus create a video fragment of a user turn (acting human interactor) that forms a reaction to this agent turn. For each interaction video (stimuli) the fragments of the user turns are then

taken from this same set of potential user turns. In this way, among the conditions, the human interactor makes exactly the same turn in reaction to an equal agent turn.

To obtain the set of potential user turns we prescribe the user's utterances and record a human interactor who acts out these utterances. The utterances are developed by following two criteria:

— The utterance should be a reaction to the preceding agent's utterance;

— The utterance, together with the user's non-verbal behaviour, should give an indication of the user's engagement;

With respect to the second criteria, as mentioned in Section 11.2.2.2 the user's engagement varies according to the user's preferences for the artworks and their

Topic - Subtopic	**User Engagement**	**Speaker**	**Utterance**
The Music - Period	High	Agent:	At the beginning of the 20th century Vienna was a lively cultural center.
		User:	Indeed, you also had Freud and Egon Schiele.
		Agent:	Even architecture was renewing, take for example the Secession building.
		User:	Yes, I like its dome!
The Last Supper - Genre	Low	Agent:	There are multiple symbolic details in this work. For example, Jesus and Juda reach for the bread at the same moment.
		User:	I see.
		Agent:	In this way Jesus announces that one of his disciples will betray him.
		User:	Yes.

Table 11.2 : Agent-user utterances (translated) for two example subtopics. The utterances are enriched with non-verbal and prosodic behaviours. The user's levels of engagement do not need to be inferred from these utterances as the agent's perceived engagement of the (acting) user is provided to the evaluation participants (Section 11.2.3.3).

characteristics. Each user turn should thus contribute to conveying different levels of engagement: We develop the user's highly engaged utterances by extending its lengths (thereby extending the interaction length [Bickmore et al., 2013]), demonstrating interest in the (sub)topics [Peters et al., 2005a], expressions of emotions [Peters et al., 2005a], and expressions of self-disclosure [Higashinaka et al., 2008]. Low engaged utterances consist of minimum contributions, just enough to keep the interaction going. In Chapter 5 we have already seen that such strategies are able to successfully create a more and less engaged interaction participant.

Table 11.2 shows examples of user (and agent) utterances for two subtopics. While the user's verbal behaviour is prescribed, for the user's non-verbal behaviour we asked the actor to behave according to whether or not the user likes the artwork (user preferences).

11.2.3.2 Computed Agent Turns

To create the video fragments of computed agent turns we use an ECA that is based entirely within the VIB platform, from agent intentions all the way to rendered agent animations (see Chapter 10). Figure 11.2 shows what the agent looks like.

Figure 11.2 : The ECA used for the stimuli.

The agent turns in each condition are obtained by running the agent system that incorporates in each condition a different way of topic management (Table 11.1). In each interaction the agent starts the interaction, and after each produced agent turn we enter input in the form of the user's engagement level (see Section 11.2.2.2). The resulting agent turns are recorded to serve in the stimuli videos (see Section 11.2.3.3).

In each interaction the computed agent turns lead to different (orders of) topics, (orders of) subtopics, timings of topic switches, and transition strategies. The detailed courses of the resulting interactions are shown in Appendix E Table E.1. Table 11.3 shows a summary of Table E.1, listing only the most apparent features of the 4 interactions: the choice and order of the topics (Column 2), if they are abandoned prematurely (Column 3), the orientation of the transition strategies (Column

4), and the overall 'detected' (entered) engagement levels of the user during the topics (Column 5).

Condition (Independent Variable)	**Topic** (What)	**Topic fragment stopped/ completed ✓** (When)	**Transition Orientation** (How)	**Average user engagement** ($\overset{meas}{\texttt{Eng}}_u(t)$)
`TMw=0`	- The Last Supper	stopped	User	Low
	- Portrait of Adele Bloch-Bauer I	✓	User	High
	- The Music	✓	User	High
	- Self Portrait	✓	User	High
`TMw=0.5`	- The Last Supper	stopped	User	Low
	- The Samaritan	✓	Agent	Neutral
	- Self Portrait	✓	User	High
	- The Music	✓	Agent	High
`TMw=1`	- The Last Supper	stopped	Agent	Low
	- The Last Judgement	stopped	Agent	Low
	- The Storm	stopped	Agent	Low
	- The Samaritan	✓	Agent	Neutral
`WithoutTM`	- The Last Supper	✓	-	Low
	- Self Portrait	✓	-	High
	- The Storm	✓	-	Low

Table 11.3 : Course of the 4 interactions. The 4th column shows the overall (average) engagement of the user during the topic. Measurements of the user's engagement are sent after each user turn ($\overset{meas}{\texttt{Eng}}_u(t)$, Section 11.2.2.1).

Some conditions can lead to varying agent intentions at specific moment in the interaction every time the system is run. In these cases we control this output to obtain stimuli that best reflect our conditions. For example, condition `TMw=0.5` leads to a $50-50\%$ chance of an agent or user oriented transition strategy every time a

transition turn is created (see Section 9.6). For the evaluation study we made sure that in the resulting interaction this same percentage is represented by 2 user and 2 agent oriented strategies. Also, in the condition `WithoutTM`, random topic selection could result in the same topics as those that are selected in other conditions. To make sure that the interaction shows that the agent makes different (random) topic choices, we let the agent start with the same topic as in the other interactions (a topic with low user engagement), but then have it address respectively 2 topics that are not selected at these positions in the other conditions: one topic with high user engagement, concluded by a topic with low user engagement. For `WithoutTM` the fixed order of subtopics within each topic is set to *period* - *type* - *artist* - *style*, but any other fixed order would work as well.

11.2.3.3 Assembled Videos

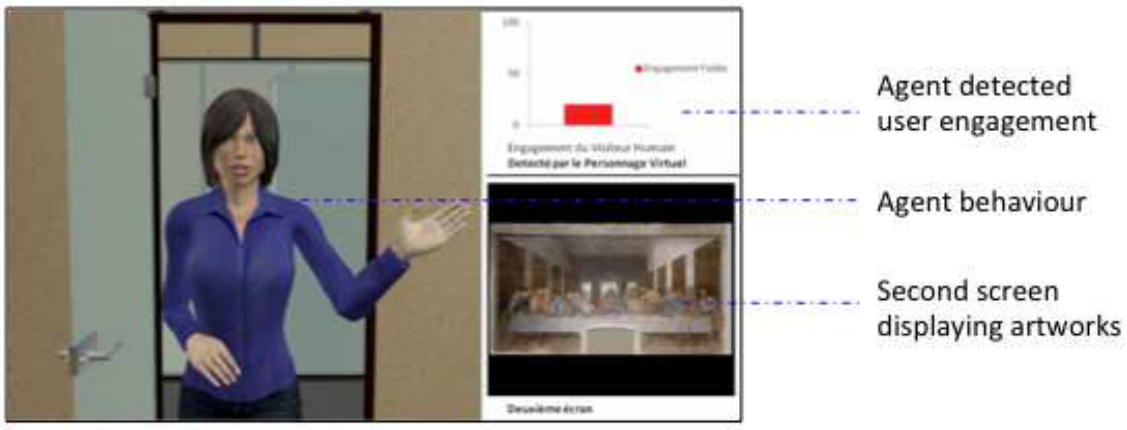
The videos of the human-agent interactions, which form the stimuli of the evaluation study, are now assembled by pasting the video fragments of acted human interactor turns after (in reaction to) the corresponding fragments of computed agent turns. In this way, video fragments of the user and the agent alternate. Figure 11.3 shows two video frames during respectively an agent turn and a user turn.

Next to footage of the interacting agent and user, the videos display the presumed (i.e. pretended) detection of the user's engagement by the agent. This allows the evaluation participants to judge the agent's behaviour while knowing that the agent detects these levels of user engagement. In this way the evaluation is focused solely on the agent's behaviour in reaction to the user, and not on the user's acting skills or the accuracy of the agent's detection of the user's engagement.

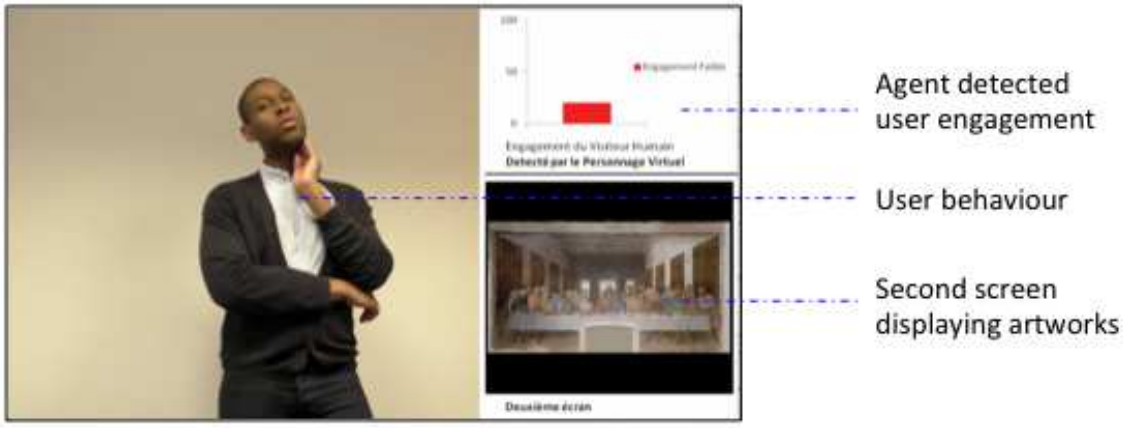
A third element of the videos shows a second screen (besides the screen that displays the agent) that the agent uses during the interaction to display the artworks that it is talking about. As shown in Figure 11.1 this second screen is part of the pretended setting of the interactions.

11.2.4 Experimental Setup

To evaluate the topic manager we expose participants to the stimuli (videos) by means of an online website. A questionnaire serves to retrieve the participants' perception of the interactions.



(a) Video frame of when the agent has the turn.



(b) Video frame of when the user has the turn.

Figure 11.3 : Frames of a video that forms a stimulus.

11.2.4.1 Website

To recruit a maximum number of participants we developed an online evaluation protocol. By means of a website each evaluation participant is exposed to only one interaction video (stimuli) to avoid biases on the perception of the agent from previous interactions (between-participant design). Another reason for this is that it would be difficult for a participant to compare the interactions, because a difference in topics between the conditions implies variations in the entire range of multimodal behaviours of the user and the agent. Conditions are assigned randomly while distributing the participants equally over the conditions.

The first page of the website is a consent page. Following this page, we guide the participants through the 3 steps described below (in French). Printscreens of

these steps are shown in Appendix E Figures E.2 and E.3.

1. We explain that the interaction that the participants are going to evaluate takes place in a museum, between a human visitor and a virtual visitor. The evaluation participants can visit the museum themselves by scrolling through a window that displays the artworks and their descriptions. The artworks correspond to the 7 artworks that form the potential topics of the interaction (but we do not say this). The artwork descriptions consist of listings of the artworks' characteristics that form potential subtopics, as in Figure E.1 of Appendix E.
2. We show Figure 11.1 and explain the supposed setting of the human-agent interaction: "*After the human visitor has seen all the paintings of the museum, he enters a room where there is a human sized virtual visitor. The human visitor and the virtual visitor start to talk. We recorded this interaction. The figure (11.1) shows the setting of the interaction: both visitors, a camera and a second screen where the agent can show artworks.*"
3. We show one of the videos (stimuli) and ask the participants to answer questions about the agent, its behaviour and its utterances (see below for the questionnaire). The participant is told that he/she can replay the video and "revisit" the museum as many times as he/she desires.

After the questionnaire is completed, on a next page the participants are posed some demographic questions and are provided a field to optionally comment on the evaluation.

11.2.4.2 Questionnaire

As reminder, we have already seen previously (Chapter 7) that the user's engagement is favoured if an agent talks about artworks that the user prefers. In this study we therefore controlled the user's behaviour, including his engagement, to focus on other issues: if the topic manager's way of managing topics is indeed perceived by human observers, namely if the it adapts the topics to different extents to the user's engagement and the agent's own preferences by taking into account its associations; and to find out if the topic manager has an influence on the human observers' perception of the agent and its dialogue.

To evaluate these issues we developed a questionnaire that consists of 4 blocks of questions. A first block verifies if the topic manager is indeed perceived as managing

the topics of the interactions in the ways that underlie each condition. We ask if the participant finds that *the virtual character:*

Q1.1 *has its own preferences for the artworks;*

Q1.2 *associates artworks with each other;*

Q1.3 *adapts the topics of the interaction to its own preferences;*

Q1.4 *adapts the topics of the interaction to the user's engagement;*

Q1.5 *adapts the topics of the interaction to the user's preferences;*

Q1.6 *selects the topics of the interaction randomly.*

The questions only talk about "topics" in general. It is up to the participant to decide what he/she considers a topic. We do not distinguish topics from subtopics in the questions as it is may be difficult for participants to track and reason about subtopics, given that subtopics are hard to distinguish in the interaction. For example, a subtopic that talks about the style of an artwork may also reveal something about the artist or period of the artwork.

All questions are answered on a 9-point scale ranging from *not at all* to *completely.*

To verify if the topic manager has an influence on the way the agent's discourse is perceived we check to what extent the discourse is considered *human-like* [Bergmann et al., 2010], *competent* [Bickmore and Cassell, 2005], *social intelligent* [Kaukiainen et al., 1995] and *coherent* [Macias-Galindo et al., 2012]. Respectively each of the concepts is estimated with a balanced pair of questions [Gratch and Kang, 2015], one positive (a) and one negative (b). We ask if the participant finds that *the virtual character's discourse:*

Q2.1 (a) *is a copy of human discourse;*

(b) *is artificial;*

Q2.2 (a) *implies competence;* [Bickmore and Cassell, 2005, Kervyn et al., 2009]

(b) *is disorganised;* [Kervyn et al., 2009]

Q2.3 (a) *emphasizes social intelligence;*

(b) *emphasizes social ignorance;*

Q2.4 (a) *is coherent;* [Macias-Galindo et al., 2012]

(b) *is incoherent.* [Macias-Galindo et al., 2012]

With other balanced question pairs we verify how the agent is perceived with respect to its *dominance* [Wiggins, 1995] (opposed to the context specific "looking for compromises"), *friendliness* [Bickmore and Cassell, 2005], and *warmth* [Fiske et al., 2007]. We do not ask if the agent is engaging, because this answer could be influenced by the (prescribed) way the user reacts to the agent. Instead, we verify two related aspects of engagement [Peters et al., 2009], namely the agent's *motivation* [Peters et al., 2009, Nadig et al., 2010] and *distance* [Gratch and Kang, 2015]. Participants are asked if they find that *the virtual character:*

Q3.1 (a) *is dominant;* [Wiggins, 1995]

(b) *is compromising;*

Q3.2 (a) *is friendly;* [Bickmore and Cassell, 2005]

(b) *is unpleasant;* [Bickmore and Cassell, 2005, Kervyn et al., 2009]

Q3.3 (a) *is warm;*[Bickmore and Cassell, 2005, Fiske et al., 2007, Kervyn et al., 2009]

(b) *is cold;* [Fiske et al., 2007, Kervyn et al., 2009]

Q3.4 *is motivated;* [Peters et al., 2009]

Q3.5 *is distant.* [Gratch and Kang, 2015]

A last question we pose is:

Q4. *Would you like to interact with the virtual character yourself?*

For each participant the questions are listed in 4 blocks but in a different, random order, within and between each of the 4 blocks of questions.

11.3 Hypotheses

We divide the hypotheses for this evaluation study in two types: one type looks at the difference between agents with the topic manager (conditions `TMw=0`, `TMw=0.5`, `TMw=1`, together indicated as `TM`) and the one without (condition `WithoutTM`), and the second type looks at the effects of the different configurations of the topic manager with topic manager (among conditions `TMw=0`, `TMw=0.5`, `TMw=1`). Some of the questions mentioned above serve to verify one type of hypothesis, others verify both. Table 11.4 gives a schematic overview of our hypotheses, split in the two types (resp. Column 3 and 4). The coloured information displays the results (confirmation or rejection of the hypotheses, see Section 11.4).

Q. No.	Dependent variable	Hypo. & Results WithoutTM → TM	Hypo. & Results TMw=0 → TMw=0.5 → TMw=1
	Topic Model variables:		
1.1.	Presence agent preferences	↗̸	↗
1.2.	Presence agent associations	↗	–
1.3.	Adapting topics to agent preferences	↗̸	↗̸
1.4.	Adapting topics to user engagement	↗	↘̸
1.5.	Adapting topics to user preferences	↗	↘
1.6.	Random topic selection	↘	–
	Agent's discourse:		
2.1.	Is human-like	↗̸	–
2.2.	Implies competence	↗̸	–
2.3.	Emphasizes social intelligence	↗̸	–
2.4.	Is coherent	↗̸	–
	Agent's:		
3.1.	Dominance	↘̸	↗̸
3.2.	Friendliness	↗̸	↘̸
3.3.	Warmth	↗̸	↘̸
3.4.	Motivation	↗̸	–
3.5.	Distance	↘̸	–
4.	**Desire to interact**	highest for ~~TMw=0.5~~ TMw=1	

Table 11.4 : Hypotheses for higher (↗) or lower (↘) scores for TM in comparison with WithoutTM (column 3), and for higher (↗) or lower (↘) scores as the agent is more oriented towards itself instead of to the user (column 4). The conditions in the 4th column correspond respectively to *complete user orientation (w = 0)* → *agent/user orientation (w = 0.5)* → *complete agent orientation (w = 1)*. The colours indicate whether an hypothesis is confirmed (↗ ↘) or rejected (↗̸ ↘̸) by the results.

Related to the first block of questions, we expect that with the topic manager the agent is perceived as *having its own preferences for the artworks*, *associating*

artworks and *adapting the topics of the interaction to the agent's preferences, user's engagement* and *preferences*, where this has to be perceived less without topic manager. This is because only the conditions `TM` use these variables for their topic management. Without the topic manager we expect that it is perceived that the agent *selects its topics randomly* where this is less the case for an agent with topic manager. The more the agent's topic manager is configured as oriented towards the agent, the higher we expect it to score on *having its own preferences* and *adapting the topics to its own preferences*, and the lower we expect it to score on *adapting the topics to the user's engagement and preferences*.

With respect to the second block of questions, since the agents with topic manager (`TM`) reason about the topics, time them by taking into account the user, and introduce them in a coherent way (see Tables 11.1 and 11.3), while the agent `WithoutTM` does not, we expect that the agent's discourse is considered more *human-like, competent, socially intelligent,* and *coherent* with topic manager than without.

Because the topic planner takes into account (the engagement level of) the user we further hypothesise that the agent is perceived as less *dominant* and *distant* and more *friendly, warm* and *motivated* [Nadig et al., 2010] with topic manager than without (third block Table 11.4). In addition we expect that the more the agent is oriented towards the agent itself instead of to the user, the more it will be considered as *dominant*. When the agent is more oriented towards the user we expect the agent to be perceived more *friendly* and *warm*.

With respect to the participants' desire to talk with the virtual agent themselves (block 4), we form a different type of hypothesis: Among all conditions we hypothesise that people prefer to talk with an agent where the agent adapts the topics of the interaction to the user's preferences but also manifests its own preferences, as represented by `TMw=0.5`.

11.4 Results

We collected results online from 80 participants, 20 per condition (interaction). 35% of the participants are male and 65% female, aged 18-66. 88% of the participants have French as their mother tongue. 81% of the participants identified themselves with French culture while the remaining participants were divided over 10 other cultural identities. 19% of the participants visits a museum less then once a year. The majority of the participants (48%) visits a museum 1-3 times a year, 29% of the participants tends to visit a museum between 3 and 10 times, and 5% of the

participants visits a museum more then 10 times a year.

The red and green in Table 11.4 give an overview of the obtained results in comparison with the hypotheses, which we discuss here in more detail.

First, we compared (with Kruskal-Wallis tests) the results of the three interaction videos that display agents with topic planner (`TM`) with the results of the video that displays an agent without topic planner (`WithoutTM`) and found, contrary to our hypothesis, that there is no significant difference regarding if the agent has its own preferences for artworks (Q1.1) and if it adapts the topics to its preferences (Q1.3). In line with our hypothesis however, `TM` agents score significantly higher on associating artworks with each other (Q1.2, $p < 0.001$) and adapting the topics to the user's engagement (Q1.4, $p < 0.001$) and preferences (Q1.5, $p < 0.01$) than `WithoutTM` agents. As expected `TM` agents score significantly lower on selecting topics randomly than an agent `WithoutTM` (Q1.6, $p < 0.05$).

When we compare the results of the three agents with different configurations of the topic planner (different agent/user orientations) we find that the three differ significantly with respect to if the agent has preferences for the artworks (Kruskal-Wallis test, Q1.1, $p < 0.05$) and if the agent adapts the topics of the interaction to the user's preferences (Q1.5, $p < 0.05$). As hypothesised the scoring for the agent's preferences increase from $w = 0$ to $w = 0.5$ to $w = 1$, while the adaptation to the user's preferences decrease. Kruskal-Wallis multiple comparison post-hoc analyses show that on both dimensions the differences between $w = 0$ and $w = 1$ are significant ($p < 0.05$). With respect to the adaptation of the topics to the agent's own preferences (Q1.3) and the user's engagement (Q1.4) no significant differences have been found between the different configurations.

In contrast to our hypotheses, the perception of the agent's discourse (Q2.1: human-likeness, Q2.2: display of competence, Q2.3 social intelligence and Q2.4: coherence), and the agent's perception regarding its dominance (Q3.1), friendliness (Q3.2), warmth (Q3.3), motivation (Q3.4), and distance (Q3.5) is not perceived as significantly different between `TM` and agent `WithoutTM` (Kruskal-Wallis tests). Similarly, against expectations, the perception of the agent's dominance, friendliness and warmth does not differ among the different agent orientations (`TMw=0`, `TMw=0.5`, `TMw=1`, Kruskal-Wallis tests).

Lastly, we compared all conditions regarding the participants' desire to interact with the agent themselves. There is a tendency that the conditions score differently among each other (Kruskal-Wallis, Q4, $p = 0.06$). When we compare only the conditions with topic planner the difference becomes significant ($p < 0.05$). Contrary

to expected, the agent with $w = 1$ scores highest but pairwise comparisons are not significant (Post-hoc Kruskal-Wallis multiple comparison).

11.5 Discussion

11.5.1 Perception of Topic Management

The results of comparing agents with topic manager (`TM`) and without (`WithoutTM`) show that the actions of the topic planner are indeed perceived in the interactions: the topic manager ensures that the agent is perceived as associating topics with each other, adapting the topics of the interaction to the user's engagement and preferences; and not selecting topics randomly. Different configurations of the topic manager (`TMw=0`, `TMw=0.5`, `TMw=1`) result indeed in different perceptions of agent preference manifestation and adaptation of the topics to the user's preferences.

No differences are found with respect to if an agent has its own preferences for the artworks, and if it adapts the topics accordingly between the conditions `TM` and `WithoutTM`. A possible explanation for this, is that because in `WithoutTM` the agent does not take into account the engagement or preferences of the user at all (with respect to timing, selection and introduction of the topics), the impression may rise that the agent acts upon other priorities thereby following some underlying agent preferences. In reality, in this condition no agent preferences are exploited at all.

The fact that among the different topic planner configurations (`TMw=0`, `TMw=0.5`, `TMw=1`) the presence of agent preferences is perceived significantly different but no effect is noticed on the adaptation of the topics according to these preferences, may be explained by the interpretation of "adapting the topics of the interaction". In fact, the agent preferences are reflected to different extents in the *selection* and *introduction* (transition strategies) of the topics, which may explain the significant effect on presence of agent preferences among conditions; but in none of the conditions the agent preferences play a role in adapting the topics in the sense of *abandoning* an ongoing topic prematurely to start a different one (Table 11.3). This is because the agent's preferences have already been taken into account in the *selection* of the ongoing topic.

An interpretation of "adapting topics" as *abandoning* topics prematurely may also explain the lack of significant differences among `TMw=0`, `TMw=0.5`, and `TMw=1`, with respect to adapting the topics according to the user's engagement. The element

in the topic manager that decides to continue or break-off a topic is indeed the same among all conditions with topic planner: the topic is continued except when the agent's recent measurements of the user's engagement are too low (Section 8.5.3 and Table 11.1).

In contrast to the perception of the agent adapting the topics to the agent's preferences and the user's engagement, participants did perceive a difference between `TMw=0`, `TMw=0.5`, and `TMw=1` in the agent adapting the topics to the user's preferences. In these conditions the topics are indeed abandoned due to the user's preferences (inferred from the user's engagement) in order to be replaced by topics that correspond to different extents to the user's preferences.

11.5.2 Perception of the Agent's Discourse

We further hypothesised differences in the way the agent's discourse is perceived between `TM` agents and the agent `WithoutTM`. The results have not confirmed such differences. We assume that this is due to the fact that the agent's discourse differs only on a (sub)topic level: The topics differ among the interactions but the type of agent turns are similar among the conditions and for each topic, with respect to both form and content (see Section 11.2.2.3).

The only thing that clearly differs among the conditions on an utterance level, are the agent's transition strategies. `TM` conditions communicate a reason of why the agent introduces the following topic (see Section 9.3.2) while `WithoutTM` does not (Section 11.2.2.3). In the evaluation study from Chapter 9 we have seen that in a small dialogue fragment, the latter transition strategy scores lower on (amongst others) naturalness, coherence and smoothness than the ones that are selected for the topic planner (Section 9.4). Here however, due the low frequency of transition utterances in comparison to the rest of the interaction ($\pm$ 3 out of 30 agent turns), the transition strategies may have had too little impact on the perception of the entire interaction to cause any effects on the perception of the overall *dialogue of the agent*.

11.5.3 Perception of the Agent

The same explanation as in the subsection above (Section 11.5.2) can be given for the absence of effects on the perception of the agents between `TM` and `WithoutTM` with respect to aspects like warmth and competence. While the transition utter-

ances used in WithoutTM have, in comparison to the strategies used in TM, lead to a less friendly and warm agent in an isolated small interaction fragment (Chapter 9, Section 9.4), this effect is non-apparent in these larger interactions. The same holds for dominance, motivation and distance between agents with the topic manager (TM) and without (WithoutTM).

Besides, while the participants do find differences among TM and WithoutTM in adapting the topics to the user (the user's engagement and preferences), these differences are not sufficient to lead to significant different perceptions of the agent. This observation suggests that what matters for the perception of the agents is not on a topic level, but on a lower, utterance level (as in Chapter 9). The transition strategies alone have shown not to be enough to cause such a significant effect on the utterance level. In the same way, among the agents TMw=0, TMw=0.5, and TMw=1 the transition strategies that show orientations towards the agent or user, turn out not to be sufficient to express significant differences in the perception of the agent's dominance, friendliness and warmth.

11.5.4 Desire to interact with the Agent

The scores of the observers' (i.e. evaluation participants') desire to interact with the agent themselves differs significantly among TMw=0, TMw=0.5, and TMw=1, and almost significantly among all conditions. Against expectation it is the completely agent-oriented configuration TMw=1 that scores highest among all conditions, though not significantly different in pairwise comparisons. While we expected that TMw=0.5 would score highest, we can think of several explanations for TMw=1's popularity. For example:

- The evaluation participants may find it more appealing that an agent has clearly its own preferences (TMw=1) than that it follows the user's preferences. Even if we have seen in Chapter 7 that for a user, speaking about his/her preferred objects can lead to more user engagement than when the agent addresses less preferred objects, the agent could be considered a more interesting and engaging conversation partner if it manifests its own preferences instead.
- The evaluation participants (observers) can have preferences that correspond to the agent's preferences, which can (un)consciously be appreciated by the participants [Reeves and Nass, 1996];
- The evaluation participants may like the fact that the agent takes into account the user by showing it abandons a topic every time the user's engagement is

low. This happens more frequently in `TMw=1` than in the other conditions because here the agent selects the topics based on only the agent's own preferences.

- The evaluation participants may just like that the interaction of `TMw=1` is shorter than the others.

The effect of the topic manager and its configurations on the observers' desire to interact with the agent themselves is however limited. This outcome may be related to an absence of significant differences with respect to the way in which the agent is perceived on dimensions like warmth and competence (Section 11.5.3). We expect again that variations on the utterance level, instead of only on the topic level, could cause further effects on the observers' desire to interact with the agent themselves.

11.5.5 Participants' Remarks

Lastly, some evaluation participants submitted comments on the evaluation study. Most comments relate to technical limitations of the agent's interaction capabilities: Unnatural synthesised speech (Text-To-Speech), incapacity to react to the content of the user's turns (Natural Language Understanding), and the agent's appearance.

It should be noted that one person remarks that he/she finds the agent's detection of the user's engagement arbitrary. This participant was exposed to condition `TMw=0.5`, the only condition where the agent detects varying levels of user engagement during one topic (differs per subtopic). Another evaluation participant however, who was assigned to the same condition `TMw=0.5`, seems to the contrary to appreciate and believe the agent's detection of the user's engagement: "[...]*Congratulations on this virtual character who is very interesting and succeeds in progressively capturing the attention of the human visitor (at the beginning a bit sceptic), next to all its knowledge!*" [1]

Some other remarks illustrate the outcomes of the evaluation study that agents with topic manager seem to adapt their verbal behaviour more to the user than an agent without the topic manager: A participant that evaluated `WithoutTM` wrote: "*It's annoying to see it elaborate topics that do not interest the listener, I would definitely leave but that must be uncomfortable...*"; while a participant that looked at `TMw=0` writes that the virtual character "[...] *adapts to the artistic preferences of the hearer* [...]".

1. all citations are translated

11.6 Conclusion

In the previous chapters we have developed and implemented a topic planner for a conversational agent that tries to engage the user while providing him/her with information. The topic planner is developed to adapt the topics of the interaction on the fly by deciding what (sub)topics to talk about, when, and how to introduce the topics in the ongoing interaction.

In this chapter we evaluated the topic manager in an agent that plays a visitor in a museum interacting with a human visitor whose behaviour is controlled. We considered 4 conditions: a control condition without topic planner (random topic selection, no breaking off topic fragments before they are completed, and non-validated topic transitions); and three agents with topic manager, representing different configurations, corresponding to a complete user orientation, an agent-user orientation, and a complete agent orientation. In order to avoid biases among the conditions we kept all other variables in the interactions consistent among the conditions. We asked human observers (third parties) for their perception of the interactions.

The outcomes of the evaluation study show that the *topic manager's actions* are indeed perceived in the interactions: the topic manager makes sure that the agent is perceived as associating topics with each other, as not selecting topics randomly, and as adapting the topics of the interaction to the user's engagement and preferences. Different orientations (agent and/or user orientation) of the agent (topic manager) indeed ensure different agent preference manifestations and different extents of adapting the topics to the user's preferences.

The topic manager and its configurations do not significantly influence the observers' perception of the *agent's dialogue* and the *perception of the agent* itself regarding aspects such as warmth and competence. For these dimensions, adaptations on other levels than (only) the topic level are likely to have more impact (e.g. adaptations on an utterance level, or regarding forms of non-verbal behaviour).

PART VI : CONCLUSION AND PERSPECTIVES

TELECOM ParisTech

12 Conclusion and Perspectives

In this thesis we have aimed at contributing to a better understanding of engagement by modelling verbal strategies to enhance the user's engagement in non task-oriented human-agent interaction. Such knowledge of how to use language in order to contribute to a social goal of the interaction serves at making agents more human-like and effective [Bickmore and Cassell, 2001].

In the three sections below we respectively summarise our findings and contributions, and give directions for future research.

12.1 Summary

— We have started this document with an overview and analysis of the definitions of engagement. We have identified their implications, commonalities, and differences, and noted that studies usually focus on a particular aspect of engagement without being able to cover the entire range of interpretations. For our research we selected one definition that suites the context and requirements that we consider, and used this definition throughout all the studies we performed in this thesis. This is [Poggi, 2001]'s definition where engagement is "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction".

— Following the various definitions, we have given an account of how engagement is manifested in an interaction by giving an overview of the concepts and behaviours that have been associated with engagement in previous research.

— To position our work we have addressed previous efforts that are related to ours, by discussing embodied conversational agents (ECA's) in museum contexts, and ECA's that consider verbal engagement strategies.

— We have performed several studies to explore potentially engagement-favouring verbal behaviours. We have looked respectively at aspects of verbal behaviour that consider the *form*, *timing*, and *content* of utterances:

- We have carried out an empirical study to verify the existence of a link between the speaker's politeness and his/her perceived engagement level of the hearer. We found that in specific contexts, our hypothesis is confirmed that if the speaker's perceived engagement level of the hearer is lower, the speaker perceives his/her face-threatening act as weightier. This indicates that in these contexts, an agent that wants to maintain the user's engagement, needs to speak more politely to a user who is less engaged than to a user who is more engaged.
- We performed a systematic empirical analysis of interruptions that considers both the amount of overlapping speech and the content of the interruptions in order to reveal their effects on the perception of the interrupter and the interruptee. We found out that by using a cooperative interruption strategy, as opposed to a disruptive one, the interrupter is perceived as more engaged and more involved in the interaction, and the more an interruption overlaps the more dominant and less friendly the interrupter and the interruptee are perceived.
- With a perceptive study we showed the existence of a positive correlation between the user's preferences for an artwork (a physical object) and the user's engagement during the discussion of this object with an ECA.

— The latter observation motivates the development of a topic manager to personalise the topic of conversation in information-giving chat in order to engage the user. The topic manager decides what to talk about, when, and how to introduce a new topic in the ongoing interaction:

- We developed the topic selection component of the topic manager that takes into account the agent's dynamically updated perception of the user's engagement, the agent's own preferences and its associations. The agent selects the topic that maximises the agent and user's combined engagement where the agent's orientation towards itself versus the user is adjustable. An ongoing topic is abandoned when the observed engagement of the user is too low.
- We have checked how the selected topic can be introduced on the dialogue level without loosing the coherence of the interaction. Based on the reasoning of the agent's topic selection we have constructed a set of potential topic transition strategies, checked the consequences of these strategies on the perception of the dialogue and the agent, and proposed

a computational topic transition model for the topic manager.

- We have proposed an integration of the topic manager in a virtual agent platform. The architecture allows the agent to manage the topics' timing, selection, initiation and breaking-off.
- We have conducted an evaluation study of the topic manager that shows that the actions of the topic manager are indeed perceived in the interaction. The agent is perceived as not selecting topics randomly but associating topics and adapting them to the user. Depending on the configurations of the topic manager the agent can give to different extents the impression that it has its own preferences. In sum, the topic manager makes that the agent is endowed with human-like features, and contributes to the interaction from its own point of view.

— We have contributed to project A1:1, realising a human-sized multi-modal ECA in a museum.

12.2 Contributions

12.2.1 Engagement

We have augmented insight into how engagement can be manifested and influenced in interactions by means of verbal behaviours of the interaction participants. Previous research on engagement is often oriented to non-verbal behaviours, or considers a range of verbal behaviours at once (Chapters 3 and 4). We have provided a more precise analysis by considering individual verbal behaviour aspects.

We have provided insight into how engagement is related to a verbal behaviour that refers to the *form* of utterances (politeness behaviour), to a behaviour that refers to the *timing* of utterances (interruption behaviour), and a behaviour that refers to the *content* of utterances (topic management).

12.2.2 Verbal behaviours

On the other hand, we have augmented knowledge about several verbal behaviours by showing if and how these behaviours impact and/or are dependent on the engagement of the interaction participants. For example, politeness behaviours, interruptions, and ways to introduce topics in an interaction have been considered in previous work primarily from conversational analytic or pragmatic points of view.

We however showed what these behaviours implicate with respect to a social goal of the interaction, namely the engagement of interaction participants.

Besides approaching verbal behaviours from an engagement point of view, we also provided insight into how some verbal behaviours could be exploited by conversational agents from a generation point of view. For example, where to our knowledge topic transition strategies in non-task oriented dialogue had been studied exclusively from an observational point of view (how humans tend to change topic), we have proposed a way of generating transition strategies (in human-agent interaction). The other way around, we have contributed to the domain of natural language generation by considering the generation of verbal interaction behaviours (i.e. dialogue). Dialogue (especially non-task oriented dialogue) forms an application that still receives relatively little attention in the domain of natural language generation.

12.2.3 Topic Management: Dialogue and Cognition

Based on the indications that we obtained with respect to how verbal behaviours could favour engagement we developed a computational model to implement engagement favouring strategies in human-agent interaction: We have proposed a computational model for a conversational agent to manage the topics of a non-task oriented interaction. The topic manager decides what to talk about, when, and how to introduce a new topic in the ongoing interaction. It does this by taking into account the agent's perception of the user's engagement and the agent's own mental state including its preferences and associations, without loosing eye for the dialogue history and coherence of the ongoing interaction. The resulting human-agent interaction is thereby driven by a social variable (engagement), by considering the agent's cognitive variables and the structure of the ongoing dialogue. In this way we contributed to a multidisciplinary view where an agent's dialogue system and cognitive system are intertwined.

12.3 Perspectives

The perspectives of the work that is presented in this thesis start where this work ends. Limitations and perspectives can therefore be found within the studies we have conducted and the models we developed, but also as in the form of studies that build further on our findings. On a short term we consider the limitations

within the studies we performed and models we have developed, on the long term we consider future work that *continues* or builds further our work:

12.3.1 Short Term

12.3.1.1 Stimuli

First of all, with respect to the perceptive studies that we performed to explore the relation between verbal behaviours and engagement, the stimuli that we consider in each study are limited. This is an intrinsic limitation of empirical studies due to limited length of experiments and limited number of potential participants. This limits both the number and the nature of the stimuli, leaving elaborations of stimuli for future work.

For example, in Chapter 5 the dialogue fragments and utterances that incorporate politeness strategies are only instances of all the possible dialogue fragments and utterances that could be used to test the perceived weight of the 3 types of face-threatening acts that we consider. Besides, we look at only 3 face-threatening acts while results on more face-threatening acts would provide us further information.

As a second example, the interrupting utterances that we used for evaluation in Chapter 6 are not the only ways to formulate their respective interrupting functions and strategies, and the amounts of overlap that we modelled in the stimuli are not always the only amount possible to represent a particular interruption type.

Similar limitations of stimuli hold for the studies described in Chapter 7 on user preferences, Chapter 9 on transition strategies, and Chapter 11 where the topic manager could also be evaluated with other topic fragments for example.

The limitations of stimuli imply that we should be careful in generalising the obtained results and leaves the opportunity to confirm or reject the results with other stimuli in future work.

12.3.1.2 Conditions: Independent Variables

Another intrinsic limitation of evaluation studies besides the stimuli, are the conditions that are considered. When we evaluate an independent variable, we keep the other variables stable among the conditions to avoid biases. Some of these variables form however interesting directions for future research.

For example, we kept the onset of the interruptions (Section 6.4.1.1) constant across conditions to focus on the effects of the interruption types and strategies.

Though, manipulating the moment when the interruption occurs during the interruptee's turn might also influence the perception of both agents.

Similarly, we would like to explore the effects of the timing of the topic switch on the perception of topic transition strategies [Clark, 1996] where in Chapter 9 we concentrated on comparing transitions strategies that all occur at comparable moments during the interaction.

In human-human interaction personal characteristics such as the gender, age and status of the interactants can play a role in the perception of participants' interrupting behaviour [Zimmermann and West, 1996, Beattie, 1981], variables that we did not manipulate in the evaluation studies but could provide further insight in future research.

12.3.1.3 Topic Manager Extensions

With respect to the topic manager that we proposed (Part V) there are several extensions we would like to consider in future work. A first extension is coming back to a previous topic, thus being able to continue the discussion of an artwork later on in the interaction when this artwork has already been addressed earlier on in the interaction.

A second extension is considering agent preferences that may change during the interaction. While integrating dynamic agent preferences would not change much the computational model of topic management, further research would be needed to decide what makes that an agent changes its preferences during the ongoing interaction.

Another extension we would like to implement is modelling if and how the agent should refer to its own preferences and/or the user's preferences in the interaction outside of the topic transition strategies.

12.3.1.4 Applications

We would like to exploit our findings in the modelling of ECA's. We can incorporate the verbal behaviours that we have identified as having an effect on the engagement of the interaction participants.

As an example, we would like to model an ECA that is able to manage interruptions [Cafaro et al., 2016]. We could implement for example different reactions of the agent in response to a user's interruption, or have the agent proactively interrupt

the user by deploying a specific interruption strategy.

12.3.2 Long Term

12.3.2.1 Automatic Agent Behaviour Generation

The topic manager handles the topics and subtopics of the interaction. At the moment, all the agent turns to realise the topics and subtopics are predefined. Only the transition utterances are automatically generated on the fly. In the future we would like to work towards generating the (sub)topic agent turns automatically and in real time as well. We could automate the generation process to a certain extent by exploiting for example information about the artworks and variables of the agent's mental state such as its preferences. The more the agent is able to learn about the user, the more we could also exploit information about the user to generate the agent's utterances on the fly.

Another future step in the real-time generation of agent behaviour is to automatically assign an agent intention to an (generated) agent utterance. The agent's intention is already automatically resulting in agent multi-modal behaviour.

12.3.2.2 More Verbal Engagement Strategies

We have modelled several verbal engagement strategies. There may be much more however, as more verbal behaviour aspects could potentially influence the value that the user contributes to "wanting to be with" the agent and "continuing the interaction". In the future we would like to model more ways of favouring the user's engagement by means of verbal behaviour.

We have seen in both previous research (Chapter 3) and our findings that adaptation of the agent's verbal behaviour to the user's behaviour can favour the user's engagement. We have implemented an adaptation to the user on a topic level, but on other levels (e.g. prosody, word choice, syntax) we could implement user adaptations as well. Such strategies are however dependent on what information the agent can collect from the user. By detecting and analysing the user's behaviour the agent could adapt its behaviour accordingly.

12.3.2.3 Multimodality

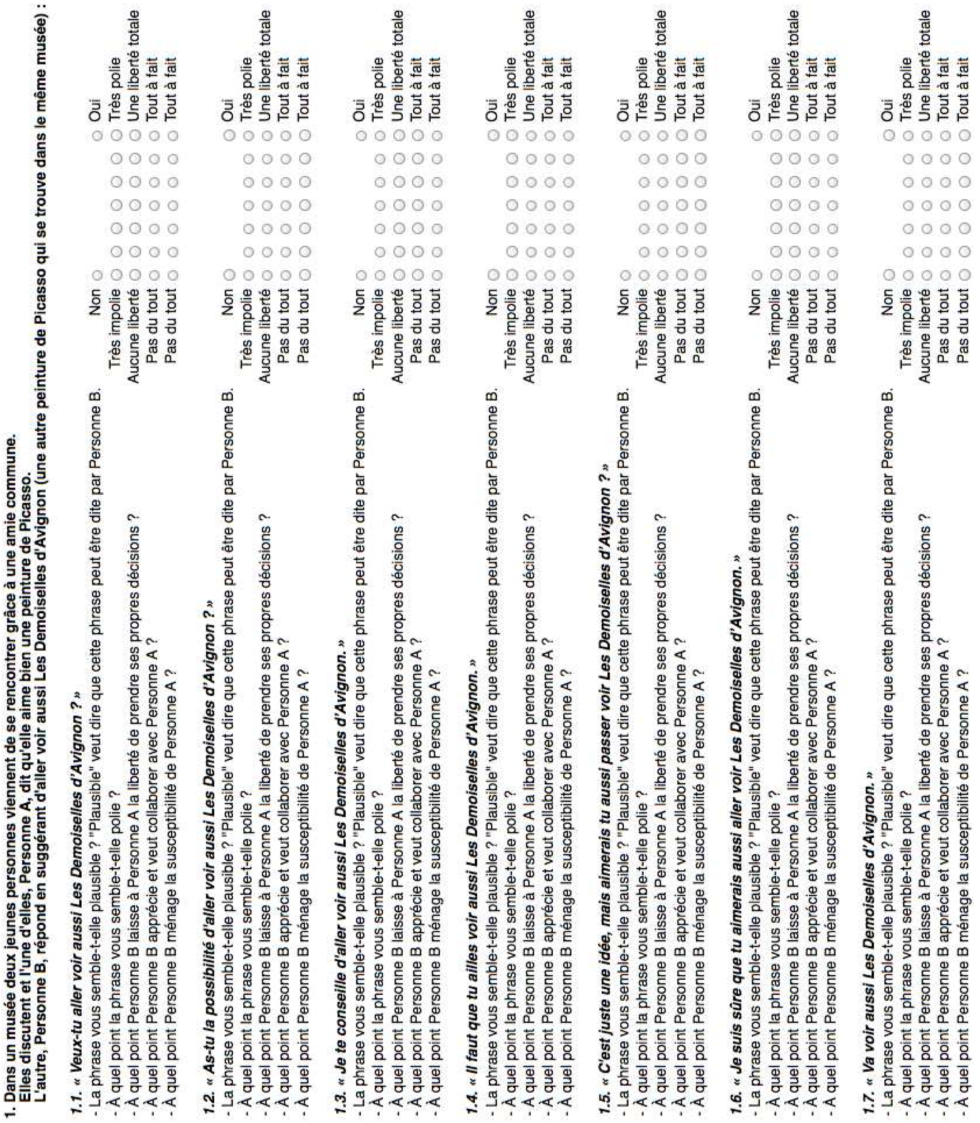
In several of the evaluation studies of this thesis we considered verbal strategies only. In the future we would like to verify the findings in multimodal interactions and enrich the verbal strategies with engaging non-verbal strategies. Non-verbal behaviour such as prosody, gaze and gestures can influence the way in which verbal behaviour is interpreted and reveal a range of information about the person's attitude and perceptions [Sidner et al., 2005b, Peters et al., 2005b]. Non-verbal expressions of feedback and mimicry for example, can play a large role in building and creating rapport [Gratch et al., 2006], and laughter can play a role in topic shifts [Gilmartin et al., 2013].

PART VII : APPENDICES

TELECOM ParisTech

A

Questionnaires Engagement-Politeness



1. Dans un musée deux jeunes personnes viennent de se rencontrer grâce à une amie commune.
Elles discutent et l'une d'elles, Personne A, dit qu'elle aime bien une peinture de Picasso.
L'autre, Personne B, répond en suggérant d'aller voir aussi Les Demoiselles d'Avignon (une autre peinture de Picasso qui se trouve dans le même musée) :

1.1. « Veux-tu aller voir aussi Les Demoiselles d'Avignon ? »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

1.2. « As-tu la possibilité d'aller voir aussi Les Demoiselles d'Avignon ? »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

1.3. « Je te conseille d'aller voir aussi Les Demoiselles d'Avignon. »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

1.4. « Il faut que tu ailles voir aussi Les Demoiselles d'Avignon. »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

1.5. « C'est juste une idée, mais aimerais tu aussi passer voir Les Demoiselles d'Avignon ? »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

1.6. « Je suis sûre que tu aimerais aussi aller voir Les Demoiselles d'Avignon. »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

1.7. « Va voir aussi Les Demoiselles d'Avignon. »
- La phrase vous semble-t-elle plausible ? "Plausible" veut dire que cette phrase peut être dite par Personne B. Non ○ ○ Oui
- À quel point la phrase vous semble-t-elle polie ? Très impolie ○ ○ ○ ○ ○ ○ ○ Très polie
- À quel point Personne B laisse à Personne A la liberté de prendre ses propres décisions ? Aucune liberté ○ ○ ○ ○ ○ ○ ○ Une liberté totale
- À quel point Personne B apprécie et veut collaborer avec Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait
- À quel point Personne B ménage la susceptibilité de Personne A ? Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

Figure A.1 : Printscreen of a fragment of the validation study described in Section 5.5.1.1.

Dans un musée deux jeunes personnes viennent de se rencontrer grâce à une amie commune.
Elles discutent :

« Pauline : - Bonjour. Je suis Pauline. Ravie de te rencontrer !
Charlotte : - Bonjour, Charlotte.
Pauline : - Heureusement vous avez réussi à venir aujourd'hui. Caroline m'a dit que vous avez eu des problèmes de transport.
Charlotte : - Oui.
Pauline : - Qu'est-ce qui s'est passé ?
Charlotte : - Le train avait du retard.
Pauline : - Oh c'est gênant ! Mais, aujourd'hui le musée est ouvert jusqu'à 22 heures donc il n'y a pas de problème.
Charlotte : - Oui.
Pauline : - C'est la première fois que tu viens ?
Charlotte : - La deuxième.
Pauline : - Moi c'est ma première visite. Pour l'instant je n'ai vu qu'une seule partie du musée, mais j'ai beaucoup aimé. Et toi tu as déjà tout vu ?
Charlotte : - Oui.
Pauline : - Ah c'est bien. Et qu'est-ce que tu as aimé le plus ?
Charlotte : - Guernica.
Pauline : - [...] »

Pauline n'aime pas beaucoup Guernica (une peinture de Picasso). Mais elle veut absolument poursuivre la conversation avec Charlotte.

1. Laquelle de ces options conseilleriez-vous à Pauline à ce moment de la conversation ?

- « D'accord. Moi je n'aime pas cette peinture. »
- « D'accord. »
- « D'accord. Je ne suis pas sûre d'aimer cette peinture tant que ça. »
- « D'accord. Le fait est que je n'aime pas cette peinture. »
- « D'accord. Je crois que je n'aime pas cette peinture. »
- « D'accord. J'aime bien et je n'aime pas à la fois. »

2. Quelle valeur pensez-vous que Charlotte attribue au fait d'être avec Pauline ?

Pas de valeur du tout ○ ○ ○ ○ ○ ○ ○ Une valeur maximale

3. Quelle valeur pensez-vous que Charlotte attribue au fait de continuer l'interaction ?

Pas de valeur du tout ○ ○ ○ ○ ○ ○ ○ Une valeur maximale

4. Selon vous, à quel point l'interaction était engageante pour Charlotte ?

Pas engageante du tout ○ ○ ○ ○ ○ ○ ○ Extrêmement engageante

5. Pour vous, Charlotte et Pauline semblent-t-elles vouloir devenir amies ?

Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

6. Selon vous, Charlotte semble-t-elle intéressée par l'interaction ?

Pas du tout ○ ○ ○ ○ ○ ○ ○ Tout à fait

A

Figure A.2 : Printscreen of a fragment of the evaluation study described in Section 5.5.3.

B

Questionnaires Engagement-Interruptions

Please read the following definitions/instructions prior to answering the questions below.

We show below fragments of informal conversations between two interaction participants, A and B.
Please answer the question for each of the 8 fragments.

To help you answer: An utterance can be disruptive or cooperative according to the following definitions:
- **Disruptive**: A disruptive utterance poses threats to the current speaker's territory by disrupting the process and/or content of the ongoing conversation.
- **Cooperative**: A cooperative utterance is intended to help the speaker by coordinating the process and/or content of the ongoing conversation. s

When you are done do not forget to click on the button 'Submit' at the end of this questionnaire.

Participant n°1

Gender: ○Male ○Female
Age: Choose an option:
Full Name:

1.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: Well, that's debatable, it's not my favorite one.

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

2.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: Do you mean the book written by Lewis Carrol?

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

3.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: A busy rabbit and a smiling cat.

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

4.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: When were you in the Wonderland theme park?

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

Figure B.1 : Printscreen of a fragment of the first round of the validation study (manipulation check) described in Section 6.4.1.2.

Please read the following definitions/instructions prior to answering the questions below.

We show below fragments of informal conversations between two interaction participants, A and B.
Please answer the question for each of the 8 fragments.

To help you answer: An utterance can be disruptive or cooperative according to the following definitions:
- **Disruptive**: A disruptive utterance poses threats to the current speaker's territory by disrupting the process and/or content of the ongoing conversation.
- **Cooperative**: A cooperative utterance is intended to help the speaker by coordinating the process and/or content of the ongoing conversation. s

When you are done do not forget to click on the button 'Submit' at the end of this questionnaire.

Participant n°1

Gender: ○Male ○Female
Age: Choose an option:
Full Name:

1.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: No, you didn't read that book.

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

2.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: No, it does not tell an amazing story at all.

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

3.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: No, there is no such book Alice in Wonderland.

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

4.
A: You know, I've read the book Alice in wonderland.
A: It tells an amazing story about a little girl.
B: No, the story is boring.

*How disruptive/cooperative do you find **B's utterance** in the interaction?* ○ Very disruptive ○ Disruptive ○ Neutral ○ Cooperative ○ Very cooperative

Figure B.2 : Printscreen of a fragment of the second round of the validation study (manipulation check) described in Section 6.4.1.2.

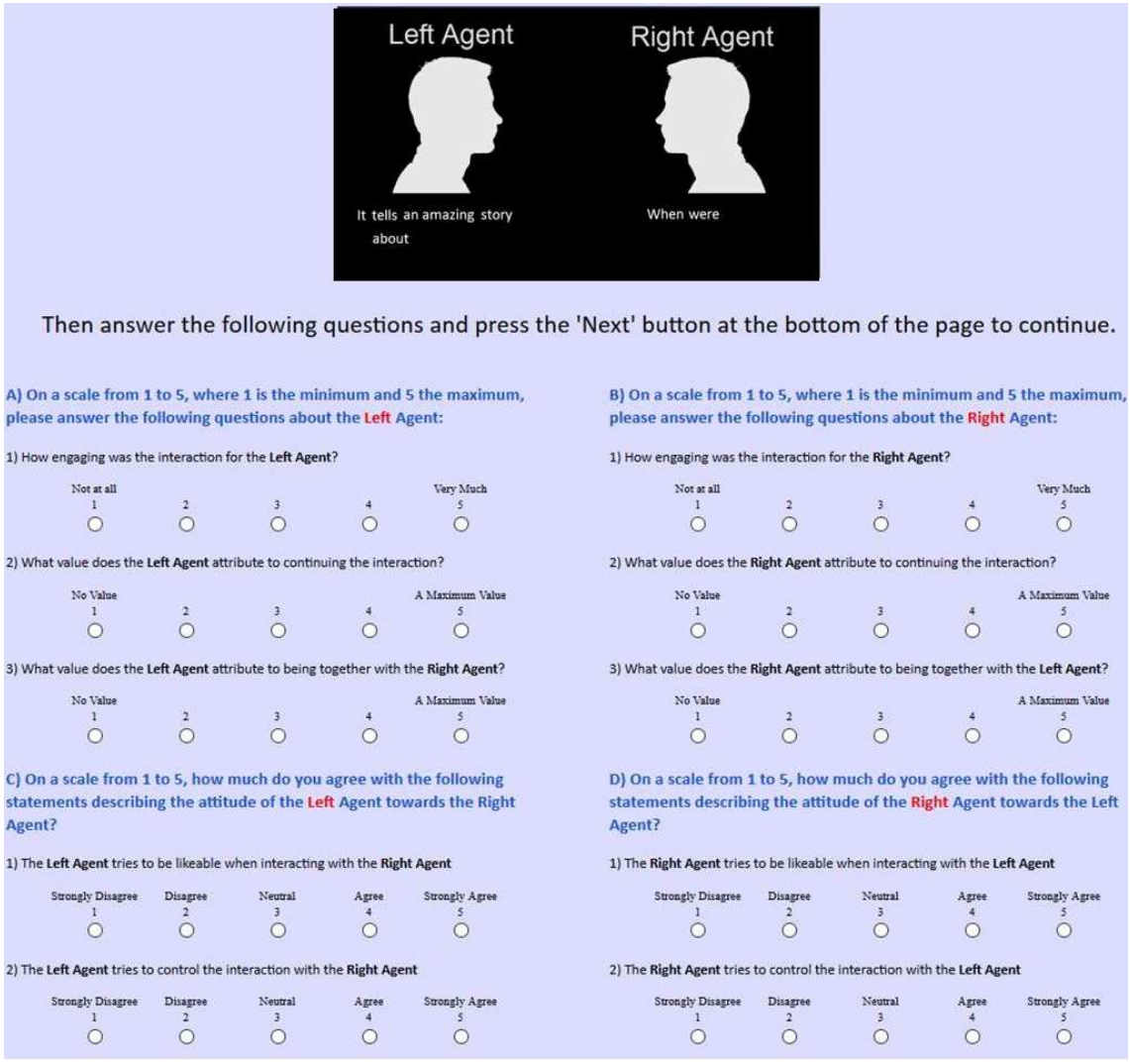


Figure B.3 : Printscreen of a fragment of the evaluation study described in Section 6.4.2.1.

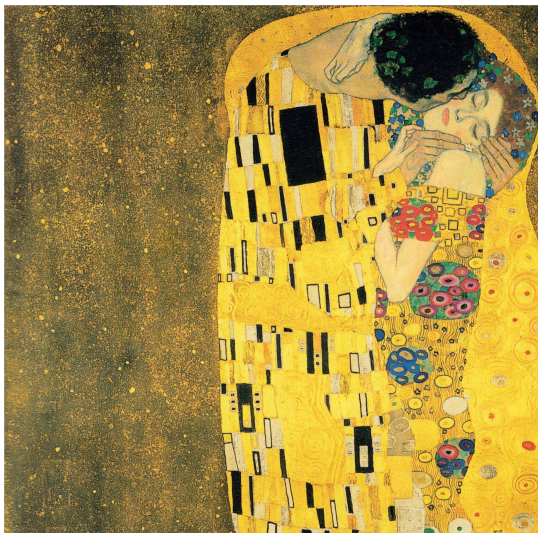
C

Artworks Engagement-Preferences

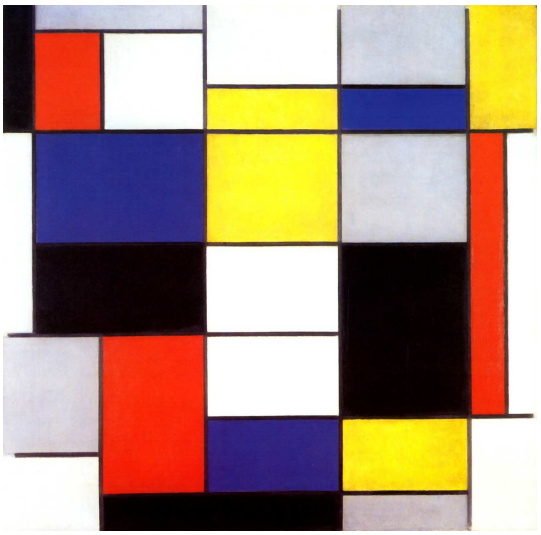
(a) *Soldier drawing his Bow* – Jacques Bousseau

(b) *Balloon Dog* – Jeff Koons

(c) *The Anatomy Lesson of Dr. Frederick Ruysch* – Jan Van Neck



(d) *The Kiss* – Gustav Klimt



(e) *Composition A* – Piet Mondrian

Figure C.1 : The artworks from the improvised museum. Artwork 6.3a was located between the user and the agent. The others were exhibited in another room that was visited before the interaction.

D

Questionnaire Transition Strategies

« Déjeuner sur l'herbe » - Monet

DIALOGUE

La conversation concerne « Déjeuner sur l'herbe », un tableau de Monet.

- [...]

Personnage virtuel : - **Claude Monet était un peintre Français. Il a habité toute sa vie à Giverny, un joli village au nord de Paris.**

Visiteur humain : - **Oui, je sais. J'ai visité Giverny l'année dernière.**

Personnage virtuel : - **Ce tableau a été peint vers 1865.**

Visiteur humain : - **Oui, je l'ai lu.**

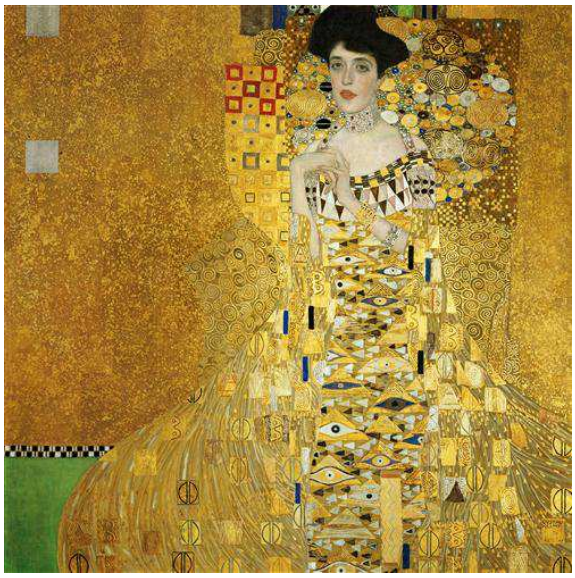
« Chien Gonflable » - Jeff Koons

SUITE DE DIALOGUE	SUITE DE DIALOGUE	SUITE DE DIALOGUE
Personnage virtuel : **- Si vous aimez cette œuvre, peut-être aimeriez-vous aussi le « Chien Gonflable » de Jeff Koons.**	Personnage virtuel : **- Avez-vous aussi vu le « Chien Gonflable » de Jeff Koons ?**	Personnage virtuel : **- Comme autre œuvre il y a le « Chien Gonflable » de Jeff Koons.**
1. Trouvez-vous que le dialogue est naturel ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*1. Trouvez-vous que le dialogue est naturel ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*1. Trouvez-vous que le dialogue est naturel ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
2. Trouvez-vous que le dialogue est cohérent ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*2. Trouvez-vous que le dialogue est cohérent ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*2. Trouvez-vous que le dialogue est cohérent ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
3. Trouvez-vous que le dialogue est fluide ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*3. Trouvez-vous que le dialogue est fluide ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*3. Trouvez-vous que le dialogue est fluide ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
4. Trouvez-vous que le personnage virtuel est amical ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*4. Trouvez-vous que le personnage virtuel est amical ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*4. Trouvez-vous que le personnage virtuel est amical ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
5. Trouvez-vous que le personnage virtuel est chaleureux ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*5. Trouvez-vous que le personnage virtuel est chaleureux ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*5. Trouvez-vous que le personnage virtuel est chaleureux ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
6. Trouvez-vous que le personnage virtuel est stimulant ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*6. Trouvez-vous que le personnage virtuel est stimulant ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*6. Trouvez-vous que le personnage virtuel est stimulant ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
7. Trouvez-vous que le personnage virtuel est compétent ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*7. Trouvez-vous que le personnage virtuel est compétent ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*7. Trouvez-vous que le personnage virtuel est compétent ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait
8. Trouvez-vous que le personnage virtuel est cultivé ? Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*8. Trouvez-vous que le personnage virtuel est cultivé ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait	*8. Trouvez-vous que le personnage virtuel est cultivé ?* Pas du tout ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Tout à fait

Figure D.1 : Printscreen of a fragment of the evaluation study described in Section 9.3.3.

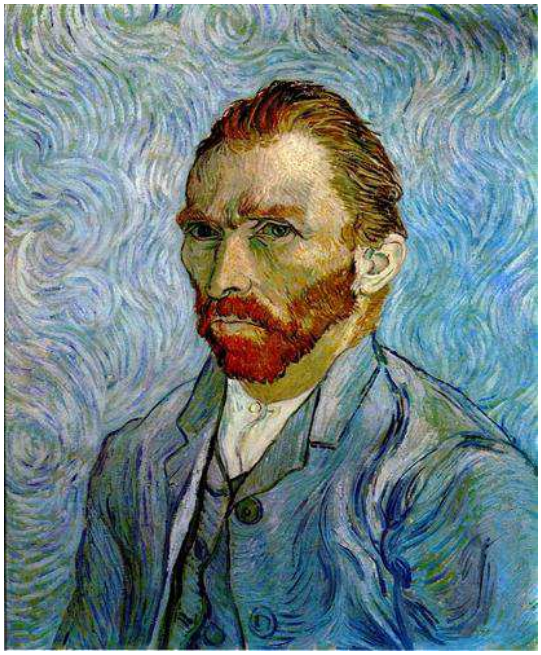
E

Material Topic Manager Evaluation



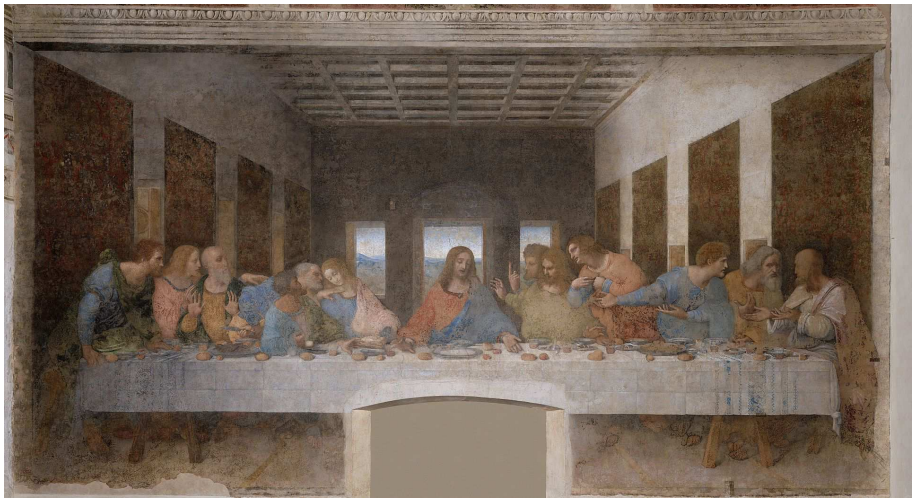
(a) *Portrait of Adele Bloch-Bauer I*
Artist: Gustav Klimt
Style: Art-Nouveau
Genre: Portrait
Period: 20th Century

(b) *The Music*
Artist: Alphonse Mucha
Style: Art-Nouveau
Genre: Portrait
Period: 19th Century

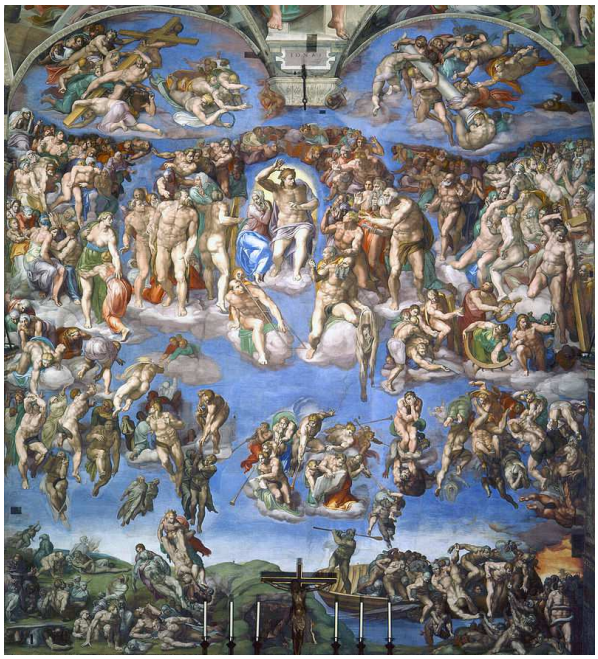


(c) *Self Portrait*
Artist: Vincent van Gogh
Style: Post-Impressionism
Genre: Portrait
Period: 19th Century

(d) *The Good Samaritan*
Artist: Vincent van Gogh
Style: Post-Impressionism
Genre: Religious Painting
Period: 19th Century



(e) *The Last Supper*
Artist: Leonardo da Vinci
Style: Renaissance
Genre: Religious Painting
Period: 15th Century



(f) *The Last Judgement*
Artist: Michelangelo
Style: Renaissance
Genre: Religious Painting
Period: 15th Century

(g) *The Storm*
Artist: Rembrandt
Style: Baroque
Genre: Religious Painting
Period: 17th Century

Figure E.1 : The artworks from the topic manager evaluation study.

Table E.1 : Course of the interactions in the 4 conditions (column 1, TM = topic manager). The subtasks column (4) lists the transitions and the subtopics belonging to each topic (column 2). Each completed subtopic consists of 2 or 3 turn pairs. For each user turn the user's engagement level (column 6) is sent to the agent (resulting in 2 or 3 measurements per subtopic): *low*: $\overset{meas}{\texttt{Eng}}_u\,(t) = 0.2$, *high*: $\overset{meas}{\texttt{Eng}}_u\,(t) = 0.8$, *average*: $\overset{meas}{\texttt{Eng}}_u\,(t) = 0.5$

Condition	**Topic**	**Orientation of Transition**	**Reasoning Variable Transition**	**Subtasks**	$\overset{meas}{\texttt{Eng}}_u\,(t)$ **level**
TMw=0	The Last Supper	User	Object in KB	— Intro. — Artist *stopped*	— Low — Low
	Portrait of Adele Bloch-Bauer I	User	User Preferences (Engagement)	— Transition — Artist — Period — Style — Type	— High — High — High — High — High
	The Music	User	Object in KB	— Transition: Type — Style — Period — Artist	— High — High — High — High
	Self Portrait	User	User Preferences (Engagement)	— Transition: Period — Type — Style — Artist	— High — High — High — High

`TMw=0.5`	The Last Supper	User	Object in KB	— Intro. — Type *stopped*	— Low — Low
	The Samaritan	Agent	Agent Associations	— Transition: Type — Artist — Period — Style	— Low — High — Neutral — High
	Self Portrait	User	User Preferences (Engagement)	— Transition: Artist — Style — Type — Period	— High — High — High — High
	The Music	Agent	Agent Preferences	— Transition: Period — Type — Style — Artist	— High — High — High — High
`TMw=1`	The Last Supper	Agent	Object in KB	— Intro. — Type *stopped*	— Low — Low

	The Last Judgement	Agent	Agent Associations	— Transition: Style — Type *stopped*	— Low — Low
	The Storm	Agent	Agent Preferences	— Transition: Type — Style *stopped*	— Low — Low
	The Samaritan	Agent	Agent Associations	— Transition: Type — Artist — Period — Style	— Low — High — Neutral — High
`WithoutTM`	The Last Supper	-	-	— Intro. — Period — Type — Artist — Style	— Low — Low — Low — Low — Low
	Self Portrait	-	-	— Intro. — Period — Type — Artist — Style	— High — High — High — High — High

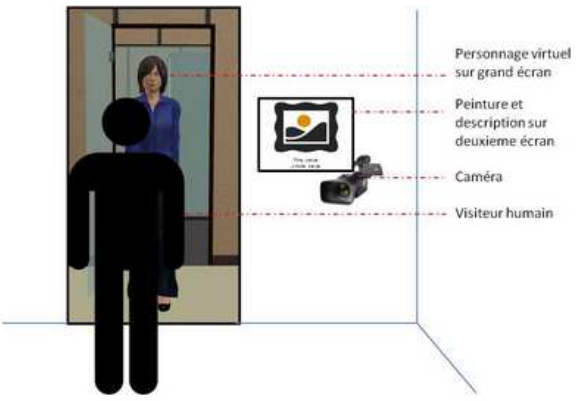
The Storm	-	-	— Intro.	— Low
			— Period	— Low
			— Type	— Low
			— Artist	— Low
			— Style	— Low

1. Visite du musée - Nous allons vous poser des questions suite à une interaction entre deux visiteurs d'un musée: un **visiteur humain** et un autre visiteur joué par un **personnage virtuel**. Avant de regarder l'interaction ci-dessous nous vous invitons à "visiter le musée" en regardant les oeuvres ci-dessous.

<-- Faites glisser pour voir tout les oeuvres. -->

---

2. Disposition de l'interaction - Après que le visiteur humain ait vu les peintures du musée il entre dans la salle où il y a le visiteur virtuel. Ils commencent à discuter. On a enregistré cette interaction. L'image à droite montre la disposition de l'interaction: les deux visiteurs, la caméra, et un deuxième écran où le personnage virtuel affiche les oeuvres du musée.



E

Figure E.2 : Printscreen of a fragment of the evaluation study, described in Section 11.2.4

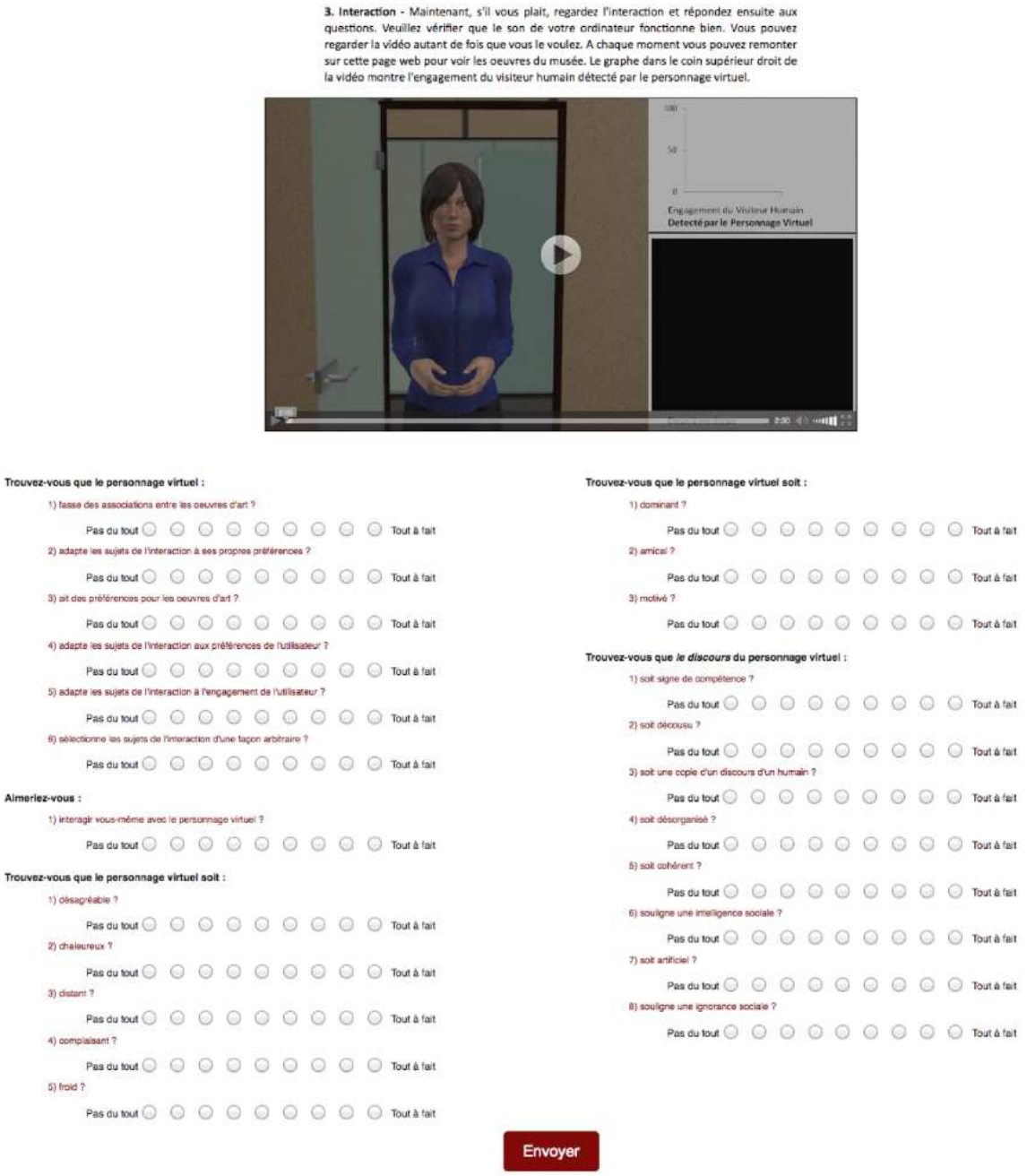


Figure E.3 : Printscreen of a fragment of the evaluation study, described in Section 11.2.4

F

Project A1:1

F.1 Introduction

The work described in this thesis has been performed in the context of the project *A1:1* (http://lifesizeavatar.com). The name indicates "Avatar on a human scale". On the one hand this project motivated the research that is described in this thesis, and on the other hand it required operational work to result in a usable application: a human-sized multi-modal virtual agent that interacts with human visitors in a museum.

Project A1:1 is a French project that lasted 3 years, from September 2012 to September 2015, and is realised by a consortium of companies and universities: Companies Cantoche, Cliris, and Mazedia, and research institutes Télécom ParisTech and Université Pierre et Marie Curie.

The project is performed in collaboration with the museum L'Historial de la Vendée who specified the role of the agent and the setup of the interaction: A human-sized virtual agent plays the role of a museum visitor and talks with a human visitor about the artworks of the museum. The setup of this interaction is illustrated by Figure F.1, showing that one artwork is placed in between the agent and the user.

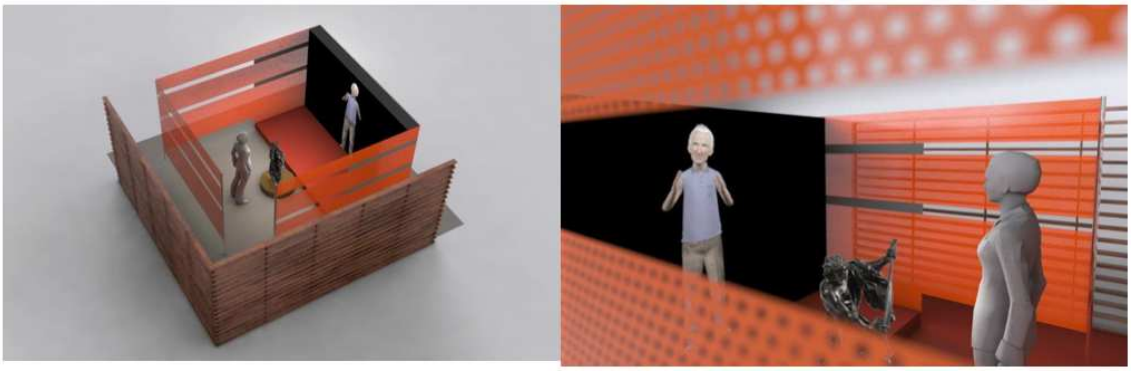


Figure F.1 : Envisioned experimental setup.

F.2 Architecture and Components

Every member of the project's consortium is responsible for some tasks in the realisation of the agent. Figure F.2 gives a high-level overview of the agent's architecture. This architecture indicates the features of the agent, and the repartition of the tasks among the consortium members.

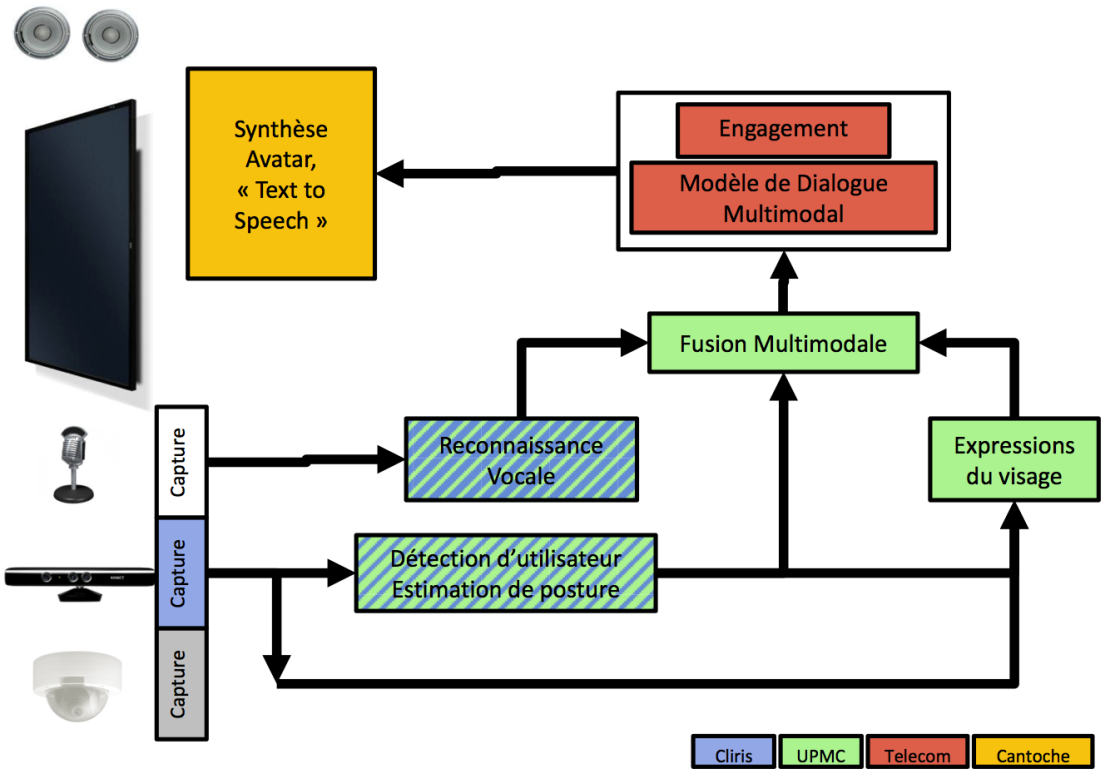


Figure F.2 : Architecture of A1:1's agent system.

The yellow box represents the output of the virtual agent system: Synthesis of the agent's animations displayed on a large screen and synthesis of the agent's speech produced by a text-to-speech tool.

The green and blue boxes are all concerned with the input of the agent system, namely capturing the user's behaviour. Information about the user's speech, postures, and facial expressions is captured with a microphone and a Kinect, combined, and sent to the multi-modal dialogue and engagement component (red boxes).

We (at Télécom ParisTech) are responsible for the management of the agent's multi-modal behaviour (red boxes). Because this is a central part in the agent's architecture, below we further specify the components that handle the output and

input of agent system first, before elaborating on the multi-modal dialogue component that maps an input (user behaviour) to an output (agent behaviour).

F.2.1 Agent Synthesis

The output of the multimodal dialogue component is sent to the component that handles the synthesis of the agent. The synthesis of the project's agent is not realised within the VIB platform (Section 10.2) but is handled by an external tool from Cantoche.

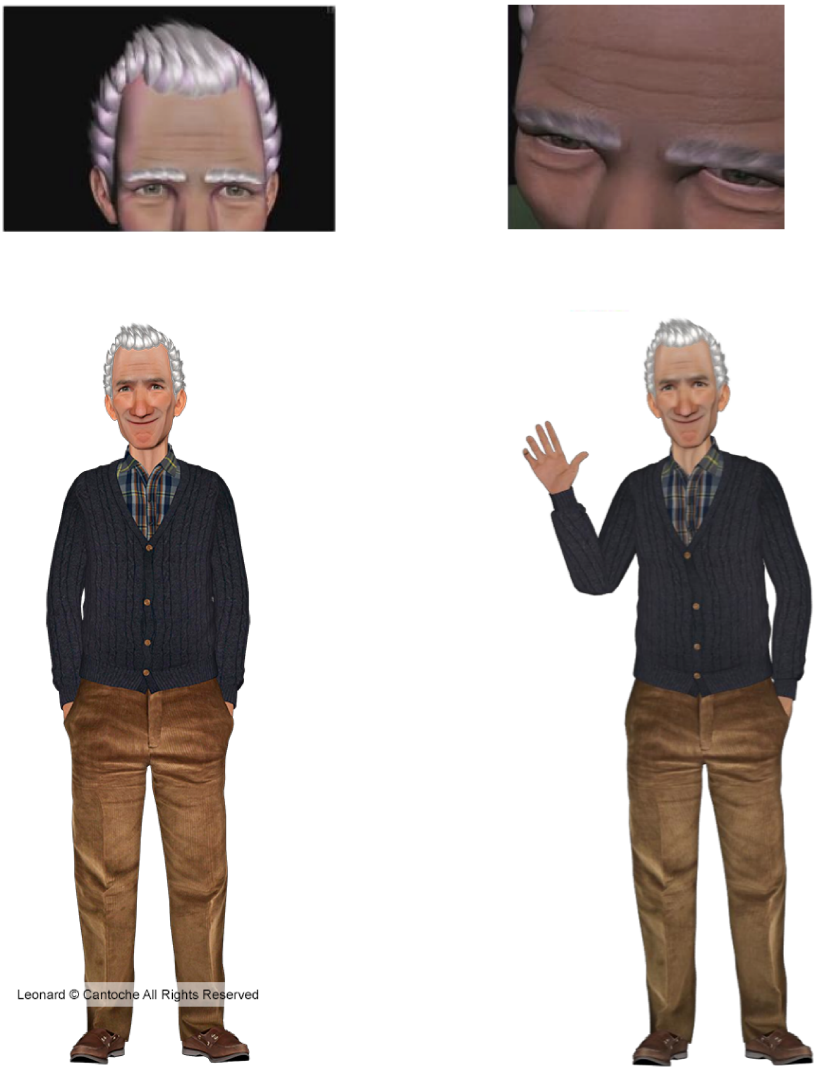


Figure F.3 : The agent Leonard (Copyright Cantoche).

The appearance of the synthesised virtual agent is designed by following the

scientific and technical requirements of the project, and by discussions with the museum. An important requirement in this is that the agent should represent a museum visitor and not a museum guide.

The resulted agent is called Leonard (Copyright Cantoche). It has the appearance of a cartoon-like man between 60 and 70 years old with realistic proportions, as shown in Figure F.3. It is realised in 3d and has a sufficient level of detail with respect to for example materials and hair to be displayed in a human-size. The agent can perform multiple postures and gestures with its skeleton and express several facial expressions. The agent's speech is synthesised by the text-to-speech component Acapela, and the agent has synchronised lip movements.

The agent's speech and animations that are synthesised during the interaction are read from BML-files that are received from the multimodal dialogue component in real-time (Section F.2.3).

F.2.2 User Behaviour Detection

On the other side of the system, the agent detects information about the user's behaviour. It is endowed with speech recognition and non-verbal behaviour recognition.

Figure F.4 : Detection of user's non-verbal behaviour.

With respect to the user's non-verbal behaviour, we use a Kinect to enable the detection of the user's skeleton, action units, and head orientation. Action units identify which parts of the face are activated, thereby recognising for example if the user smiles. The user's skeleton indicates if there is a user in front of the agent and what its posture is.

With respect to the user's verbal behaviour, we use the speech recognition module of the Kinect. It can identify words spoken by the user out of a list of predefined keywords. The more words you predefine, the more words can be recognised, but the harder it becomes for the Kinect to identify (i.e. recognise) the correct word among the set of possibilities.

Information about the user's behaviour is combined into a multimodal fusion component that forwards information about the user's engagement. Besides this, low-level information such as the spotted keywords, is also sent directly to the multimodal dialogue component.

F.2.3 Multi-modal Dialogue Component

For the multi-modal dialogue component we used the VIB platform. We set up an architecture in the platform that receives the user's spotted keywords and information about the user's non-verbal behaviour, decides on the agent's multimodal behaviour, and outputs the BML files that specify this behaviour.

To exchange the information with the other components of the agent (i.e. input and output) we use the Thrift tool[1]. To decide about the agent's behaviour, we use the Disco tool [Rich and Sidner, 2012]. As explained in the implementation chapter (10) the Disco hierarchical task network is integrated into the VIB platform to manage the course of the interactions. Therefore, the scenarios of the interactions that the agent should be able to perform need to be encoded in Disco readable XML files (Section 10.3.2). While we have developed several scenarios (i.e. task networks, dialogue trees) for research purposes, the scenario of the agent that is to be placed in the museum is subject to the cultural message that the museum wants to transfer. Therefore, the consortium member that is responsible for the situated interaction scenario designed together with the museum the interaction scenarios. We made sure that the scenarios follow the technical requirements.

FML files are designed to specify the agent's communicative intentions at each step in the interaction, allowing synchronisation of the agent's verbal and non-verbal behaviour (Section 10.2). Following the encoded scenarios, the Disco module selects the predefined FML file that encodes the agent's intention at that moment in the interaction (Section 10.3). VIB's behaviour planner then translates the agent's intention into a BML file that specifies the agent's behaviour (Section 10.2). The tags of the BML files are adapted to match the lexicon of the facial and gestural

1. https://thrift.apache.org/

animations of the external agent player (Cantoche).

Together with the scenarios we define the list of keywords for the automatic speech recognition component. This is needed because the words that the user is likely to use in his/her reaction to the agent depend on the preceding agent turns which are dependent on the interaction scenario. The recognised keywords are mapped to a set of anticipated response types of the user, thereby proceeding in the Disco scenario dialogue tree. In this way, as explained in Section 10.3.2, a user's "yes" for example, can lead (i.e. map) to a different agent response than a "no". But also the user's non-verbal behaviour can be mapped to an anticipated user response type. For example, a smile could be interpreted as a confirmation and therefore select the same user response type as the keyword "yes".

With respect to the timing of the agent's multi-modal behaviour, the agent's first turn is triggered (i.e. the first BML is sent) as soon as the multi-modal behaviour component receives from the detection component that a user is detected (present). For the subsequent turns, we made sure that the agent only determines the user's response type when the agent finished its previous turn. Only when the agent's animation and synthesised speech of the previous turn is completed a user response type can be selected that triggers the subsequent agent turn by generating and sending the next BML file to the animation component. This is to avoid overlapping agent and user turns.

F.3 Integrations, Prototypes, Demonstrations

Figure F.5 : Working on prototypes.

Progress during the project has given rise to multiple integration meetings and prototypes. Every year a more advanced working prototype was demonstrated for validation to continue the project (Figure F.5). To conclude the project, the agent was placed in a museum to interact with real museum visitors: During the French

heritage days the system was placed in the museum Carte à Jouer at Issy-les-Moulineaux. As shown in Figure F.6 museum visitors talked into a microphone to interact with the human-sized conversational agent. About 30 visitors between 2 and 80 years old interacted with the system.

Figure F.6 : Museum visitors interact with the final system.

Publications

[Cafaro et al., 2016] Cafaro, A., Glas, N., and Pelachaud, C. (2016). The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *International Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, pages 911–920.

[Campano et al., 2015] Campano, S., Langlet, C., Glas, N., Clavel, C., and Pelachaud, C. (2015). An eca expressing appreciations. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 962–967. IEEE.

[Glas and Pelachaud, 2014] Glas, N. and Pelachaud, C. (2014). Hearer engagement as a variable in the perceived weight of a face-threatening act. *Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, page 179.

[Glas and Pelachaud, 2015a] Glas, N. and Pelachaud, C. (2015a). Definitions of engagement in human-agent interaction. In *ACII International Workshop on Engagement in Human Computer Interaction (ENHANCE)*, pages 944–949.

[Glas and Pelachaud, 2015b] Glas, N. and Pelachaud, C. (2015b). Politeness versus perceived engagement: an experimental study. *Natural Language Processing and Cognitive Science (NLPCS)*, pages 135–148.

[Glas and Pelachaud, 2015c] Glas, N. and Pelachaud, C. (2015c). Topic transition strategies for an information-giving agent. *European Workshop on Natural Language Generation (ENLG)*, page 146.

[Glas and Pelachaud, 2015d] Glas, N. and Pelachaud, C. (2015d). User engagement and preferences in information-giving chat with virtual agents. In *IVA Workshop on Engagement in Social Intelligent Virtual Agents (ESIVA)*, pages 33–40.

[Glas et al., 2015] Glas, N., Prepin, K., and Pelachaud, C. (2015). Engagement driven topic selection for an information-giving agent. *Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, pages 48–57.

[Ochs et al., 2013] Ochs, M., Ding, Y., Fourati, N., Chollet, M., Ravenet, B., Pecune, F., Glas, N., Prepin, K., Clavel, C., and Pelachaud, C. (2013). Vers des agents conversationnels animés socio-affectifs. In *Conference on l'Interaction Homme-Machine*, page 69. ACM.

Bibliography

[Agostaro et al., 2005] Agostaro, F., Augello, A., Pilato, G., Vassallo, G., and Gaglio, S. (2005). A conversational agent based on a conceptual interpretation of a data driven semantic space. In *Advances in Artificial Intelligence*, pages 381–392. Springer.

[Ahmed and Palermo, 2010] Ahmed, S. and Palermo, A. (2010). Community engagement in research: Frameworks for education and peer review. *American Journal of Public Health*, 100:1380–1387.

[Alarcon and Lyons, 2011] Alarcon, G. M. and Lyons, J. B. (2011). The relationship of engagement and job satisfaction in working samples. *The Journal of Psychology*, 145(5):463–480.

[Allwood and Cerrato, 2003] Allwood, J. and Cerrato, L. (2003). A study of gestural feedback expressions. In *First Nordic Symposium on Multimodal Communication*, pages 7–22. Copenhagen.

[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), International Conference on Social Computing (SocialCom)*, pages 918–925. IEEE.

[Amgoud and Parsons, 2002] Amgoud, L. and Parsons, S. (2002). Agent dialogues with conflicting preferences. In *Intelligent Agents VIII*, pages 190–205. Springer.

[Andre et al., 2004] Andre, E., Rehm, M., Minker, W., and Bühler, D. (2004). Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems*, pages 178–187. Springer.

[Andrew and Sofian, 2012] Andrew, O. C. and Sofian, S. (2012). Individual factors and work outcomes of employee engagement. *Procedia-Social and Behavioral Sciences*, 40:498–508.

[Ansarin and Aliabdi, 2011] Ansarin, A. A. and Aliabdi, H. T. (2011). Reader engagement in english and persian applied linguistics articles. *English Language Teaching*, 4(4):p154.

[Argyle, 1988] Argyle, M. (1988). *Bodily communication*. Methuen, New York, 2nd ed. edition.

[Aylett and Pidcock, 2007] Aylett, M. and Pidcock, C. (2007). The cerevoice characterful speech synthesiser sdk. In Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., and Pelé, D., editors, *International Conference on Intelligent Virtual Agents (IVA)*, volume 4722 of *Lecture Notes in Computer Science*, pages 413–414. Springer Berlin Heidelberg.

[Baker et al., 2014] Baker, R. S., Ocumpaugh, J., Gowda, S. M., Kamarainen, A. M., and Metcalf, S. J. (2014). Extending log-based affect detection to a multi-user virtual environment for science. In *User Modeling, Adaptation, and Personalization*, pages 290–300. Springer.

[Bamoallem et al., 2016] Bamoallem, B. S., Wodehouse, A. J., Mair, G. M., and Vasantha, G. A. (2016). The impact of head movements on user involvement in mediated interaction. *Computers in Human Behavior*, 55, Part A:424 – 431.

[Beattie, 1981] Beattie, G. W. (1981). Interruption in conversational interaction, and its relation to the sex and status of the interactants. *Linguistics*, 19(1-2):15–36.

[Bednarik et al., 2012] Bednarik, R., Eivazi, S., and Hradis, M. (2012). Gaze and conversational engagement in multiparty video conversation: an annotation scheme and classification of high and low levels of engagement. In *Workshop on Eye Gaze in Intelligent Human Machine Interaction*, page 10. ACM.

[Bergmann and Kopp, 2012] Bergmann, K. and Kopp, S. (2012). Gestural alignment in natural dialogue. In *Annual Conference of the Cognitive Science Society (CogSci)*.

[Bergmann et al., 2010] Bergmann, K., Kopp, S., and Eyssel, F. (2010). Individualized gesturing outperforms average gesturing–evaluating gesture production in virtual humans. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 104–117. Springer.

[Bickmore and Cassell, 1999] Bickmore, T. and Cassell, J. (1999). Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI Fall Symposium on Narrative Intelligence*, pages 87–92.

[Bickmore and Cassell, 2001] Bickmore, T. and Cassell, J. (2001). Relational agents: a model and implementation of building user trust. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 396–403. ACM.

[Bickmore and Cassell, 2005] Bickmore, T. and Cassell, J. (2005). Social dialongue with embodied conversational agents. In *Advances in Natural Multimodal Dialogue Systems*, pages 23–54. Springer.

[Bickmore et al., 2011] Bickmore, T., Pfeifer, L., and Schulman, D. (2011). Relational agents improve engagement and learning in science museum visitors. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 55–67. Springer.

[Bickmore et al., 2010] Bickmore, T., Schulman, D., and Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6):648–666.

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

[Bickmore et al., 2013] Bickmore, T. W., Vardoulakis, L. M. P., and Schulman, D. (2013). Tinker: a relational agent museum guide. *International Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, 27(2):254–276.

[Bohus and Horvitz, 2009a] Bohus, D. and Horvitz, E. (2009a). Learning to predict engagement with a spoken dialog system in open-world settings. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 244–252. Association for Computational Linguistics.

[Bohus and Horvitz, 2009b] Bohus, D. and Horvitz, E. (2009b). Models for multiparty engagement in open-world dialog. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 225–234. Association for Computational Linguistics.

[Bohus and Horvitz, 2014] Bohus, D. and Horvitz, E. (2014). Managing human-robot engagement with forecasts and... um... hesitations. In *International Conference on Multimodal Interaction*, pages 2–9. ACM.

[Bonin et al., 2012] Bonin, F., Bock, R., and Campbell, N. (2012). How do we react to context? annotation of individual and group engagement in a video corpus. In *Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom)*, pages 899–903. IEEE.

[Branigan et al., 2000] Branigan, H. P., Pickering, M. J., and Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25.

[Branigan et al., 2010] Branigan, H. P., Pickering, M. J., Pearson, J., and McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9):2355–2368.

[Breuing and Wachsmuth, 2012] Breuing, A. and Wachsmuth, I. (2012). Let's talk topically with artificial agents! providing agents with humanlike topic awareness

in everyday dialog situations. In *International Conference on Agents and Artificial Intelligence (ICAART)*, volume 2.

[Brown and Cairns, 2004] Brown, E. and Cairns, P. (2004). A grounded investigation of game immersion. In *Extended abstracts on Human Factors in Computing Systems CHI*, pages 1297–1300. ACM.

[Brown and Yule, 1983] Brown, G. and Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

[Brown and Levinson, 1987] Brown, P. and Levinson, S. C. (1987). *Politeness: Some universals in language usage*, volume 4. Cambridge University Press.

[Burgoon and Hale, 1984] Burgoon, J. K. and Hale, J. L. (1984). The fundamental topoi of relational communication. *Communication Monographs*, 51(3):193–214.

[Burns, 2013] Burns, E. (2013). Marketers mulling arf's 'engagement' definition. `http://www.clickz.com/clickz/news/1707642/marketers-mulling-arfs-engagement-definition`.

[Button and Casey, 1985] Button, G. and Casey, N. (1985). Topic nomination and topic pursuit. *Human Studies*, 8(1):3–55.

[Cafaro et al., 2016] Cafaro, A., Glas, N., and Pelachaud, C. (2016). The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *International Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, pages 911–920.

[Campano et al., 2015a] Campano, S., Clavel, C., and Pelachaud, C. (2015a). I like this painting too: When an eca shares appreciations to engage users. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1649–1650. International Foundation for Autonomous Agents and Multiagent Systems.

[Campano et al., 2014] Campano, S., Glas, N., Langlet, C., Clavel, C., and Pelachaud, C. (2014). Alignement par production d'hétéro-répétitions chez un aca. In *Workshop Affect, Compagnon Artificiel, Interaction*.

[Campano et al., 2015b] Campano, S., Langlet, C., Glas, N., Clavel, C., and Pelachaud, C. (2015b). An eca expressing appreciations. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 962–967. IEEE.

[Carenini and Moore, 2006] Carenini, G. and Moore, J. D. (2006). Generating and evaluating evaluative arguments. *Artificial Intelligence*, 170(11):925–952.

[Carlson and Hunnicutt, 1996] Carlson, R. and Hunnicutt, S. (1996). Generic and domain-specific aspects of the waxholm nlp and dialog modules. In *International Conference on Spoken Language (ICSLP)*, volume 2, pages 677–680. IEEE.

[Casali et al., 2011] Casali, A., Godo, L., and Sierra, C. (2011). A graded bdi agent model to represent and reason about preferences. *Artificial Intelligence*, 175(7):1468–1478.

[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 *SIGCHI Conference on Human Factors in Computing Systems*, pages 520–527. ACM.

[Castellano et al., 2012] Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. W. (2012). Detecting engagement in hri: An exploration of social and task-based context. In *Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom)*, pages 421–428. IEEE.

[Castellano et al., 2014] Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and Mcowan, P. W. (2014). Context-sensitive affect recognition for a robotic game companion. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 4(2):10.

[Castellano et al., 2013] Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F., and Bull, S. (2013). Towards empathic virtual and robotic tutors. In *Artificial Intelligence in Education*, pages 733–736. Springer.

[Castellano et al., 2009] Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). Detecting user engagement with a robot companion using task and social interaction-based features. In *International conference on Multimodal Interfaces*, pages 119–126. ACM.

[Cha et al., 2006] Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M., and Lee, J.-H. (2006). Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. In *Intelligent Tutoring Systems*, pages 513–524. Springer.

[Chakraborty et al., 2007] Chakraborty, S., Bhattacharya, T., Bhowmick, P. K., Basu, A., and Sarkar, S. (2007). Shikshak: An intelligent tutoring system authoring tool for rural education. In *International Conference on Information and Communication Technologies and Development (ICTD)*, pages 1–10. IEEE.

[Chartrand et al., 2005] Chartrand, T. L., Maddux, W. W., and Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. *The New Unconscious*, pages 334–361.

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

[Choi et al., 2012] Choi, A., Melo, C. D., Woo, W., and Gratch, J. (2012). Affective engagement to emotional facial expressions of embodied social agents in a decision-making game. *Computer Animation and Virtual Worlds*, 23(3-4):331–342.

[Clark, 1996] Clark, H. H. (1996). *Using language*, volume 4. Cambridge University Press Cambridge.

[Clemen and Reilly, 2013] Clemen, R. T. and Reilly, T. (2013). *Making hard decisions with DecisionTools*. Cengage Learning.

[Cohen-Mansfield et al., 2009] Cohen-Mansfield, J., Dakheel-Ali, M., and Marx, M. S. (2009). Engagement in persons with dementia: the concept and its measurement. *The American Journal of Geriatric Psychiatry*, 17(4):299.

[Corrigan et al.,] Corrigan, L. J., Peters, C., Küster, D., and Castellano, G. Engagement perception and generation for social robots and virtual agents. In *Toward Robotic Socially Believable Behaving Systems*, volume 1.

[Couture-Beil et al., 2010] Couture-Beil, A., Vaughan, R. T., and Mori, G. (2010). Selecting and commanding individual robots in a vision-based multi-robot system. In *International Conference on Human-Robot Interaction*, pages 355–356. IEEE Press.

[Crook et al., 2012] Crook, N., Field, D., Smith, C., Harding, S., Pulman, S., Cavazza, M., Charlton, D., Moore, R., and Boye, J. (2012). Generating context-sensitive eca responses to user barge-in interruptions. *Journal on Multimodal User Interfaces*, 6(1-2):13–25.

[Crook et al., 2010] Crook, N., Smith, C., Cavazza, M., Pulman, S., Moore, R., and Boye, J. (2010). Handling user interruptions in an embodied conversational agent. In *AAMAS International Workshop on Interacting with ECAs as Virtual Characters*, pages 27–33.

[Damiano et al., 2008] Damiano, R., Gena, C., Lombardo, V., Nunnari, F., and Pizzo, A. (2008). A stroll with carletto: adaptation in drama-based tours with virtual characters. *User Modeling and User-Adapted Interaction*, 18(5):417–453.

[Danescu-Niculescu-Mizil et al., 2013] Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., and Potts, C. (2013). A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.

[De Carolis et al., 2004] De Carolis, B., Pelachaud, C., Poggi, I., and Steedman, M. (2004). Apml, a markup language for believable behavior generation. In *Life-Like Characters*, pages 65–85. Springer.

[De Jong et al., 2008] De Jong, M., Theune, M., and Hofs, D. (2008). Politeness and alignment in dialogues with a virtual guide. In *International Joint Conference on Autonomous Agents and Multiagent systems (AAMAS)*, pages 207–214. International Foundation for Autonomous Agents and Multiagent Systems.

[Decety and Jackson, 2004] Decety, J. and Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2):71–100.

[DeFalco and Baker, 2013] DeFalco, J. A. and Baker, R. S. (2013). Detection and transition analysis of engagement and affect in a simulation-based combat medic training environment. In *AIED Workshops*, volume 7, page 88. Citeseer.

[Delaherche et al., 2012] Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., and Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *Affective Computing*, 3(3):349–365.

[Dellarosa, 1988] Dellarosa, D. (1988). The psychological appeal of connectionism. *Behavioral and Brain Sciences*, 11(01):28–29.

[Diessner et al., 2008] Diessner, R., Solom, R. D., Frost, N. K., Parsons, L., and Davidson, J. (2008). Engagement with beauty: Appreciating natural, artistic, and moral beauty. *The Journal of Psychology*, 142(3):303–332.

[Dietz and Lanzarone, 2004] Dietz, H. and Lanzarone, V. (2004). Measuring engagement of the fetal head: validity and reproducibility of a new ultrasound technique. *Ultrasound in Obstetrics & Gynecology*, 25(2):165–168.

[Downing, 2000] Downing, A. (2000). Talking topically. *CIRCLE of Linguistics Applied to Communication (CLAC)*, 3:31–50.

[Edwards and Barron, 1994] Edwards, W. and Barron, F. H. (1994). Smarts and smarter: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, 60(3):306–325.

[En and Lan, 2012] En, L. Q. and Lan, S. S. (2012). Applying politeness maxims in social robotics polite dialogue. In *International Conference on Human-Robot Interaction (HRI)*, pages 189–190. IEEE.

[Evers et al., 2014] Evers, V., Menezes, N., Merino, L., Gavrila, D., Nabais, F., Pantic, M., and Alvito, P. (2014). The development and real-world application of frog, the fun robotic outdoor guide. In *Conference on Computer Supported Cooperative Work & Social Computing*, pages 281–284. ACM.

[Ferguson, 1975] Ferguson, N. (1975). *Interruptions: speaker-switch nonfluency in spontaneous conversation*. PhD thesis, University of Edinburgh.

[Fiske et al., 2007] Fiske, S. T., Cuddy, A. J., and Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2):77–83.

[Forbes-Riley et al., 2012] Forbes-Riley, K., Litman, D., Friedberg, H., and Drummond, J. (2012). Intrinsic and extrinsic evaluation of an automatic user disengagement detector for an uncertainty-adaptive spoken dialogue system. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 91–102. Association for Computational Linguistics.

[Foster, 2007] Foster, M. E. (2007). Enhancing human-computer interaction with embodied conversational agents. In *Universal Access in Human-Computer Interaction. Ambient Interaction*, pages 828–837. Springer.

[Foster and Oberlander, 2010] Foster, M. E. and Oberlander, J. (2010). User preferences can drive facial expressions: evaluating an embodied conversational agent in a recommender dialogue system. *User Modeling and User-Adapted Interaction*, 20(4):341–381.

[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.

[Gambrell et al., 2004] Gambrell, L. B., Mazzoni, S. A., and Almasi, J. F. (2004). Promoting collaboration, social interaction, and engagement with text. *Treating Suicidal Behavior: An Effective, Time-Limited Approach*, page 119.

[Garber-Barron and Si, 2013] Garber-Barron, M. and Si, M. (2013). Adaptive storytelling through user understanding. In *Artificial Intelligence and Interactive Digital Entertainment Conference*.

[Garrod and Anderson, 1987] Garrod, S. and Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218.

[Gilmartin et al., 2013] Gilmartin, E., Bonin, F., Vogel, C., and Campbell, N. (2013). Laughter and topic transition in multiparty conversation. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 304–308.

[Goffman, 1966] Goffman, E. (1966). *Behavior in public places: Notes on the social organization of gatherings*. Simon and Schuster.

[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.

[Gratch and Kang, 2015] Gratch, J. and Kang, S. (2015). Rapport measures. In *Gratch Lab*, page Unpublished.

[Gratch et al., 2006] Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., and Morency, L.-P. (2006). Virtual rapport. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 14–27. Springer.

[Gratch et al., 2007] Gratch, J., Wang, N., Gerten, J., Fast, E., and Duffy, R. (2007). Creating rapport with virtual agents. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 125–138. Springer.

[Gravano and Hirschberg, 2012] Gravano, A. and Hirschberg, J. (2012). A corpus-based study of interruptions in spoken dialogue. In *InterSpeech*, pages 855–858.

[Hall et al., 2005] Hall, L., Woods, S., Aylett, R., Newall, L., and Paiva, A. (2005). Achieving empathic engagement through affective interaction with synthetic characters. In *Affective Computing and Intelligent Interaction*, pages 731–738. Springer.

[Harper, 2013] Harper, D. (2013). Online etymology dictionary. `http://www.etymonline.com/index.php?allowed_in_frame=0&search=engage&searchmode=none`.

[Harris, 2011] Harris, L. (2011). Secondary teachers' conceptions of student engagement: Engagement in learning or in schooling?. *Teaching and Teacher Education*, 27(2):376–386.

[Harrist and Waugh, 2002] Harrist, A. W. and Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children's development. *Developmental Review*, 22(4):555–592.

[Head, 2007] Head, B. W. (2007). Community engagement: participation on whose terms? *Australian Journal of Political Science*, 42(3):441–454.

[Heinroth et al., 2011] Heinroth, T., Koleva, S., and Minker, W. (2011). Topic switching strategies for spoken dialogue systems. In *InterSpeech*, pages 2077–2080.

[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 *International Conference on Intelligent Virtual Agents (IVA)*, pages 270–280. Springer.

[Higashinaka et al., 2008] Higashinaka, R., Dohsaka, K., and Isozaki, H. (2008). Effects of self-disclosure and empathy in human-computer dialogue. In *Spoken Language Technology Workshop*, pages 109–112. IEEE.

[Hobbs, 1990] Hobbs, J. R. (1990). Topic drift. *Conversational Organization and its Development*, 38:3–22.

[Holroyd et al., 2011] Holroyd, A., Rich, C., Sidner, C. L., and Ponsler, B. (2011). Generating connection events for human-robot collaboration. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 241–246. IEEE.

[Huang et al., 2011] Huang, L., Morency, L.-P., and Gratch, J. (2011). Virtual rapport 2.0. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 68–79. Springer.

[Hyland, 2005] Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2):173–192.

[Jiang et al., 2008] Jiang, B., Pei, J., Lin, X., Cheung, D. W., and Han, J. (2008). Mining preferences from superior and inferior examples. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 390–398. ACM.

[Jokinen and Wilcock, 2012] Jokinen, K. and Wilcock, G. (2012). Constructive interaction for talking about interesting topics. In *Language Resources and Evaluation Conference (LREC)*, pages 404–410.

[Kaukiainen et al., 1995] Kaukiainen, A., Björkqvist, K., Österman, K., Lagerspetz, K., and Forsblom, S. (1995). Peer-estimated social intelligence (pesi). *Turku, Finland: Department of Psychology, University of Turku*.

[Kendon, 1970] Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32:101–125.

[Kendon, 1972] Kendon, A. (1972). Some relationships between body motion and speech. *Studies in Dyadic Communication*, 7:177.

[Kennedy and Camden, 1983] Kennedy, C. W. and Camden, C. T. (1983). A new look at interruptions. *Western Journal of Communication*, 47(1):45–58.

[Kervyn et al., 2009] Kervyn, N., Yzerbyt, V. Y., Judd, C. M., and Nunes, A. (2009). A question of compensation: The social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology*, 96(4):828.

[Kießling, 2002] Kießling, W. (2002). Foundations of preferences in database systems. In *International Conference on Very Large Data Bases*, pages 311–322. VLDB Endowment.

[Klotz et al., 2011] Klotz, D., Wienke, J., Peltason, J., Wrede, B., Wrede, S., Khalidov, V., and Odobez, J.-M. (2011). Engagement-based multi-party dialog with a humanoid robot. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 341–343. Association for Computational Linguistics.

[Kopp et al., 2005] Kopp, S., Gesellensetter, L., Krämer, N. C., and Wachsmuth, I. (2005). A conversational agent as museum guide–design and evaluation of a real-world application. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 329–343. Springer.

[Kopp et al., 2006] Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., Thórisson, K. R., and Vilhjálmsson, H. (2006). Towards a common framework for multimodal generation: The behavior markup language. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 205–217. Springer.

[Landragin, 2013] Landragin, F. (2013). *Man-Machine Dialogue*. ISTE-Wiley.

[Le Maitre and Chetouani, 2013] Le Maitre, J. and Chetouani, M. (2013). Self-talk discrimination in human–robot interaction situations for supporting social awareness. *International Journal of Social Robotics*, pages 1–13.

[Leite et al., 2014] Leite, I., Castellano, G., Pereira, A., Martinho, C., and Paiva, A. (2014). Empathic robots for long-term interaction. *International Journal of Social Robotics*, 6(3):329–341.

[Leite et al., 2013] Leite, I., Martinho, C., and Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5(2):291–308.

[Levinson, 1983] Levinson, S. C. (1983). *Pragmatics (Cambridge textbooks in linguistics)*. Cambridge University Press.

[Li, 2001] Li, H. Z. (2001). Cooperative and intrusive interruptions in inter-and intracultural dyadic discourse. *Journal of Language and Social Psychology*, 20(3):259–284.

[Lombard and Ditton, 1997] Lombard, M. and Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2).

[Lombard et al., 2000] Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J., and Park, S. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *International Workshop on Presence*, volume 240.

[Maat et al., 2010] Maat, M. T., Truong, K. P., and Heylen, D. (2010). How turn-taking strategies influence users' impressions of an agent. In Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., and Safonova, A., editors, *International Conference on Intelligent Virtual Agents (IVA)*, volume 6356 of *Lecture Notes in Computer Science*, pages 441–453. Springer Berlin Heidelberg.

[Macey and Schneider, 2008] Macey, W. H. and Schneider, B. (2008). The meaning of employee engagement. *Industrial and Organizational Psychology*, 1(1):3–30.

[Macias-Galindo et al., 2012] Macias-Galindo, D., Wong, W., Thangarajah, J., and Cavedon, L. (2012). Coherent topic transition in a conversational agent. In *Annual Conference of the International Speech Communication Association (InterSpeech)*.

[Markey et al., 2003] Markey, P. M., Funder, D. C., and Ozer, D. J. (2003). Complementarity of interpersonal behaviors in dyadic interactions. *Personality and Social Psychology Bulletin*, 29(9):1082–1090.

[Martin and White, 2005] Martin, J. R. and White, P. R. (2005). *The language of evaluation*. Palgrave Macmillan Great Britain.

[Mayer et al., 2006] Mayer, R. E., Johnson, W. L., Shaw, E., and Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human-Computer Studies*, 64(1):36–42.

[Maynard and Zimmerman, 1984] Maynard, D. W. and Zimmerman, D. H. (1984). Topical talk, ritual and the social organization of relationships. *Social Psychology Quarterly*, pages 301–316.

[Mazeland, 2003] Mazeland, H. J. (2003). *Inleiding in de conversatieanalyse*. Coutinho.

[McFarlane, 2002] McFarlane, D. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17(1):63–139.

[McGrath and Kuteeva, 2011] McGrath, L. and Kuteeva, M. (2011). Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices. *English for Specific Purposes*.

[McMahan, 2003] McMahan, A. (2003). Immersion, engagement and presence. *The Video Game Theory Reader*, pages 67–86.

[McMahon and Portelli, 2004] McMahon, B. and Portelli, J. P. (2004). Engagement for what? beyond popular discourses of student engagement. *Leadership and Policy in Schools*, 3(1):59–76.

[Michalowski et al., 2006] Michalowski, M. P., Sabanovic, S., and Simmons, R. (2006). A spatial model of engagement for a social robot. In *International Workshop on Advanced Motion Control*, pages 762–767. IEEE.

[Millican, 2013] Millican, J. (2013). Social engagement.

[Milward and Beveridge, 2003] Milward, D. and Beveridge, M. (2003). Ontology-based dialogue systems. In *Workshop on Knowledge and Reasoning in Practical Dialogue Systems (IJCAI03)*, pages 9–18. Citeseer.

[Moore et al., 2004] Moore, J. D., Foster, M. E., Lemon, O., and White, M. (2004). Generating tailored, comparative descriptions in spoken dialogue. In *The Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pages 917–922.

[Mota and Picard, 2003] Mota, S. and Picard, R. W. (2003). Automated posture analysis for detecting learner's interest level. In *Conference on Computer Vision and Pattern Recognition Workshop*, volume 5, pages 49–49. IEEE.

[Mower et al., 2007] Mower, E., Feil-Seifer, D. J., Mataric, M. J., and Narayanan, S. (2007). Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1125–1130. IEEE.

[Murata, 1994] Murata, K. (1994). Intrusive or co-operative? a cross-cultural study of interruption. *Journal of Pragmatics*, 21(4):385–400.

[Murray, 1997] Murray, J. H. (1997). *Hamlet on the holodeck: The future of narrative in cyberspace*. SimonandSchuster. com.

[Nadig et al., 2010] Nadig, A., Lee, I., Singh, L., Bosshart, K., and Ozonoff, S. (2010). How does the topic of conversation affect verbal exchange and eye gaze? a comparison between typical development and high-functioning autism. *Neuropsychologia*, 48(9):2730–2739.

[Nakano and Ishii, 2010] Nakano, Y. I. and Ishii, R. (2010). Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *International Conference on Intelligent User Interfaces (IUI)*, pages 139–148. ACM.

[Ng et al., 1995] Ng, S. H., Brooke, M., and Dunne, M. (1995). Interruption and influence in discussion groups. *Journal of Language and Social Psychology*, 14(4):369–381.

[Niculescu et al., 2010] Niculescu, A., Van Dijk, B., Nijholt, A., Limbu, D. K., See, S. L., and Wong, A. H. Y. (2010). Socializing with olivia, the youngest robot receptionist outside the lab. In *Social Robotics*, pages 50–62. Springer.

[Niewiadomski et al., 2010] Niewiadomski, R., Prepin, K., Bevacqua, E., Ochs, M., and Pelachaud, C. (2010). Towards a smiling eca: studies on mimicry, timing and types of smiles. In *International Workshop on Social Signal Processing*, pages 65–70. ACM.

[Nisbet, 2009a] Nisbet, M. C. (2009a). Communicating climate change: Why frames matter for public engagement. *Environment: Science and Policy for Sustainable Development*, 51(2):12–23.

[Nisbet, 2009b] Nisbet, M. C. (2009b). Framing science: A new paradigm in public engagement. *Understanding Science: New Agendas in Science Communication*, pages 40–67.

[Novielli et al., 2010] Novielli, N., de Rosis, F., and Mazzotta, I. (2010). User attitude towards an embodied conversational agent: Effects of the interaction mode. *Journal of Pragmatics*, 42(9):2385–2397.

[O'Brien and Toms, 2008] O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955.

[O'Donnell et al., 2001] O'Donnell, M., Mellish, C., Oberlander, J., and Knott, A. (2001). Ilex: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(03):225–250.

[Oertel et al., 2011] Oertel, C., De Looze, C., Scherer, S., Windmann, A., Wagner, P., and Campbell, N. (2011). Towards the automatic detection of involvement in conversation. In *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pages 163–170. Springer.

[Oxford-University-Press, 2013] Oxford-University-Press (2013). Oxford dictionaries online. `http://oxforddictionaries.com/definition/english/engagement?q=engagement`.

[Paiva et al., 2004] Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., and Hall, L. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 194–201. IEEE Computer Society.

[Papangelis et al., 2014] Papangelis, A., Zhao, R., and Cassell, J. (2014). Towards a computational architecture of dyadic rapport management for virtual agents. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 320–324. Springer.

[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 *IVA Workshop on Architectures and Standards for IVAs*, pages 16–20. Citeseer.

[Peters, 2005] Peters, C. (2005). Direction of attention perception for conversation initiation in virtual environments. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 215–228. Springer.

[Peters et al., 2009] Peters, C., Castellano, G., and de Freitas, S. (2009). An exploration of user engagement in hci. In *International Workshop on Affective-Aware Virtual Agents and Social Robots*, page 9. ACM.

[Peters et al., 2005a] Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., and Poggi, I. (2005a). A model of attention and interest using gaze behavior. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 229–240. Springer.

[Peters et al., 2005b] Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., Poggi, I., and Tre, U. R. (2005b). Engagement capabilities for ecas. In *AAMAS Workshop Creating Bonds with ACAs*.

[Pickering et al., 2004] Pickering, M. J., Garrod, S., et al. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–189.

[Pilato et al., 2011] Pilato, G., Augello, A., and Gaglio, S. (2011). A modular architecture for adaptive chatbots. In *International Conference on Semantic Computing (ICSC)*, pages 177–180. IEEE.

[Poggi, 2001] Poggi, I. (2001). Mind markers. *The Semantics and Pragmatics of Everyday Gestures*.

[Poggi, 2007] Poggi, I. (2007). *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler.

[Powell and Colin, 2008] Powell, M. C. and Colin, M. (2008). Meaningful citizen engagement in science and technology what would it really take? *Science Communication*, 30(1):126–136.

[Prendinger and Ishizuka, 2005] Prendinger, H. and Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users'affective states. *Applied Artificial Intelligence*, 19(3-4):267–285.

[Prepin et al., 2012] Prepin, K., Ochs, M., and Pelachaud, C. (2012). Mutual stance building in dyad of virtual agents: Smile alignment and synchronisation. In *Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom)*, pages 938–943. IEEE.

[Prohaska et al., 2012] Prohaska, T. R., Anderson, L. A., and Binstock, R. H. (2012). *Public Health for an Aging Society*. Johns Hopkins University Press.

[Ravenet et al., 2015] Ravenet, B., Cafaro, A., Biancardi, B., Ochs, M., and Pelachaud, C. (2015). Conversational behavior reflecting interpersonal attitudes in small group interactions. In Brinkman, W.-P., Broekens, J., and Heylen, D., editors, *International Conference on Intelligent Virtual Agents (IVA)*, volume 9238 of *Lecture Notes in Computer Science*, pages 375–388. Springer International Publishing.

[Reeves and Nass, 1996] Reeves, B. and Nass, C. (1996). The media equation.

[Rehm and Jensen, 2015] Rehm, M. and Jensen, M. L. (2015). Accessing cultural artifacts through digital companions: The effects on children's engagement. In *International Conference on Culture and Computing*, pages 72–79. IEEE.

[Rehm and Krogsager, 2013] Rehm, M. and Krogsager, A. (2013). Negative affect in human robot interaction—impoliteness in unexpected encounters with robots. In *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 45–50. IEEE.

[Rich et al., 2010] Rich, C., Ponsler, B., Holroyd, A., and Sidner, C. L. (2010). Recognizing engagement in human-robot interaction. In *International Conference on Human-Robot Interaction (HRI)*, pages 375–382. IEEE.

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

[Ringeval et al., 2013] Ringeval, F., Sonderegger, A., Sauer, J., and Lalanne, D. (2013). Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–8. IEEE.

[Rose and Spiegel, 2009] Rose, A. K. and Spiegel, M. M. (2009). Noneconomic engagement and international exchange: The case of environmental treaties. *Journal of Money, Credit and Banking*, 41(2-3):337–363.

[Sanghvi et al., 2011] Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *International Conference on Human-Robot Interaction (HRI)*, pages 305–311. IEEE.

[Schaufeli et al., 2002] Schaufeli, W. B., Salanova, M., González-Romá, V., and Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness studies*, 3(1):71–92.

[Schegloff, 2001] Schegloff, E. A. (2001). Accounts of conduct in interaction: Interruption, overlap, and turn-taking. In *Handbook of Sociological Theory*, pages 287–321. Springer.

[Schegloff, 2007] Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis*, volume 1. Cambridge University Press.

[Scherer, 2005] Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729.

[Schmitt and Minker, 2012] Schmitt, A. and Minker, W. (2012). *Towards Adaptive Spoken Dialog Systems*. Springer Science & Business Media.

[Schutz, 1958] Schutz, W. C. (1958). *FIRO: A Three-dimensional Theory of Interpersonal Behaviour*. Rinehart books in the assessment of personality. Holt, Rinehart and Winston.

[Shamay-Tsoory et al., 2009] Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132(3):617–627.

[Shiomi et al., 2006] Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2006). Interactive humanoid robots for a science museum. In *Conference on Human-Robot Ißnteraction*, pages 305–312. ACM.

[Sidner et al., 2013] Sidner, C., Bickmore, T., Rich, C., Barry, B., Ring, L., Behrooz, M., and Shayganfar, M. (2013). An always-on companion for isolated older adults. In *Annual Meeting on Discourse and Dialogue (SIGDIAL)*.

[Sidner and Dzikovska, 2002] Sidner, C. L. and Dzikovska, M. (2002). Human-robot interaction: Engagement between humans and robots for hosting activities. In *International Conference on Multimodal Interfaces*, page 123. IEEE Computer Society.

[Sidner et al., 2004] Sidner, C. L., Kidd, C. D., Lee, C., and Lesh, N. (2004). Where to look: A study of human-robot engagement. In *International Conference on Intelligent User Interfaces*, pages 78–84. ACM.

[Sidner and Lee, 2003] Sidner, C. L. and Lee, C. (2003). An architecture for engagement in collaborative conversations between a robot and humans. *Mitsubishi Electric Research Labs TR2003-13*.

[Sidner et al., 2005a] Sidner, C. L., Lee, C., and Kidd, C. (2005a). Engagement during dialogues with robots. In *AAAI Spring Symposia*.

[Sidner et al., 2005b] Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005b). Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140–164.

[Sidner et al., 2003] Sidner, C. L., Lee, C., and Lesh, N. (2003). Engagement when looking: behaviors for robots when collaborating with people. In *Workshop on the Semantic and Pragmatics of Dialogue (SEMDIAL)*, pages 123–130. University of Saarland.

[Skocpol and Fiorina, 1999] Skocpol, T. and Fiorina, M. P. (1999). Making sense of the civic engagement debate. *Civic Engagement in American Democracy*, pages 1–26.

[Song et al., 2009] Song, X., Maeda, K., Kunimasa, H., Toyota, H., and Han, D. (2009). Topic control in a free conversation system. In *International Conference on Natural Language Processing and Knowledge Engineering*, pages 1–6. IEEE.

[Stede and Schlangen, 2004] Stede, M. and Schlangen, D. (2004). Information-seeking chat: Dialogues driven by topic-structure. In *Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*.

[Strapparava and Valitutti, 2004] Strapparava, C. and Valitutti, A. (2004). Wordnet affect: an affective extension of wordnet. In *Language Resources and Evaluation Conference (LREC)*, volume 4, pages 1083–1086.

[Svennevig, 2000] Svennevig, J. (2000). *Getting acquainted in conversation: a study of initial interactions*, volume 64. John Benjamins Publishing.

[Swales, 1993] Swales, J. M. (1993). Genre and engagement. *Revue Belge de Philologie et D'histoire*, 71(3):687–698.

[Swartout et al., 2010] Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D., et al. (2010). Ada and grace: Toward realistic and engaging virtual museum guides. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 286–300. Springer.

[Tannen, 1981] Tannen, D. (1981). Indirectness in discourse: Ethnicity as conversational style. *Discourse Processes*, 4(3):221–238.

[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 *International Conference on Intelligent Virtual Agents (IVA)*, pages 350–356.

[Tickle-Degnen and Rosenthal, 1990] Tickle-Degnen, L. and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, 1(4):285–293.

[Tindana et al., 2007] Tindana, P. O., Singh, J. A., Tracy, C. S., Upshur, R. E., Daar, A. S., Singer, P. A., Frohlich, J., and Lavery, J. V. (2007). Grand challenges in global health: community engagement in research in developing countries. *PLoS Medicine*, 4(9):e273.

[Traum et al., 2012] Traum, D., Aggarwal, P., Artstein, R., Foutz, S., Gerten, J., Katsamanis, A., Leuski, A., Noren, D., and Swartout, W. (2012). Ada and grace: Direct interaction with museum visitors. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 245–251. Springer.

[Van Baaren et al., 2003] Van Baaren, R. B., Holland, R. W., Steenaert, B., and van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology*, 39(4):393–398.

[Vilhjálmsson et al., 2007] Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., et al. (2007). The behavior markup language: Recent developments and challenges. In *International Workshop on Intelligent Virtual Agents (IVA)*, pages 99–111. Springer.

[Walker et al., 2004] Walker, M. A., Whittaker, S. J., Stent, A., Maloor, P., Moore, J., Johnston, M., and Vasireddy, G. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science*, 28(5):811–840.

[Wallis and Shortliffe, 1984] Wallis, J. W. and Shortliffe, E. H. (1984). Customized explanations using causal knowledge. *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, pages 371–388.

[Welsh and Fearn, 2008] Welsh, J. and Fearn, D. (2008). Engagement: public diplomacy in a globalised world. *London: Foreign and Commonwealth Office*.

[Wiggins, 1995] Wiggins, J. S. (1995). *IAS, Interpersonal Adjective Scales: Professional Manual*. Psychological Assessment Resources.

[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.

[Wiltermuth and Heath, 2009] Wiltermuth, S. S. and Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, 20(1):1–5.

[Wong et al., 2012] Wong, W., Cavedon, L., Thangarajah, J., and Padgham, L. (2012). Flexible conversation management using a bdi agent approach. In *International Conference on Intelligent Virtual Agents (IVA)*, pages 464–470. Springer.

[Yang et al., 2008] Yang, F., Heeman, P. A., and Kun, A. (2008). Switching to real-time tasks in multi-tasking dialogue. In *International Conference on Computational Linguistics*, pages 1025–1032. Association for Computational Linguistics.

[Yannakakis and Togelius, 2011] Yannakakis, G. N. and Togelius, J. (2011). Experience-driven procedural content generation. *Transactions on Affective Computing*, 2(3):147–161.

[Yngve, 1970] Yngve, V. H. (1970). On getting a word in edgewise. In *Chicago Linguistics Society Meeting*, pages 567–578.

[Youngquist, 2009] Youngquist, J. (2009). The effect of interruptions and dyad gender combination on perceptions of interpersonal dominance. *Communication Studies*, 60(2):147–163.

[Yu et al., 2004] Yu, C., Aoki, P. M., and Woodruff, A. (2004). Detecting user engagement in everyday conversations. *arXiv preprint cs/0410027*.

[Zentner and Eerola, 2010] Zentner, M. and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 107(13):5768–5773.

[Zhang et al., 2011] Zhang, S., Jiang, H., and Carroll, J. M. (2011). Integrating online and offline community through facebook. In *International Conference on Collaboration Technologies and Systems (CTS)*, pages 569–578. IEEE.

[Zimmerman et al., 2005] Zimmerman, J., Ayoob, E., Forlizzi, J., and McQuaid, M. (2005). Putting a face on embodied interface agents. *Human-Computer Interaction Institute*.

[Zimmermann and West, 1996] Zimmermann, D. H. and West, C. (1996). Sex roles, interruptions and silences in conversation. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 211–236.

Modélisation des Stratégies Verbales d'Engagement dans les Interactions Humain-Agent

Nadine GLAS

RESUME : Dans une interaction humain-agent, l'engagement de l'utilisateur est un élément essentiel pour atteindre l'objectif de l'interaction. Dans cette thèse, nous étudions comment l'engagement de l'utilisateur pourrait être favorisé par le comportement de l'agent. Nous nous concentrons sur les stratégies de comportement verbal de l'agent qui concernent respectivement la *forme*, le *timing* et le *contenu* de ses énoncés. Nous présentons des études empiriques qui concernent certains aspects du comportement de politesse de l'agent, du comportement d'interruption de l'agent, et les sujets de conversation que l'agent adresse lors de l'interaction.

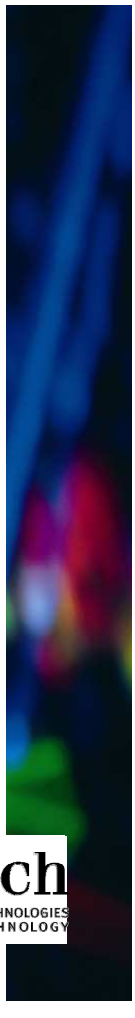
Basé sur les résultats de la dernière étude, nous proposons un Gestionnaire de Sujets axé sur l'engagement (modèle computationnel) qui personnalise les sujets d'une interaction dans des conversations où l'agent donne des informations à un utilisateur humain. Le Modèle de Sélection des Sujets du Gestionnaire de Sujets décide sur *quoi* l'agent devrait parler et *quand*. Pour cela, il prend en compte la perception par l'agent de l'utilisateur, qui est dynamiquement mis à jour, ainsi que l'état mental et les préférences de l'agent. Le Modèle de Transition de Sujets du Gestionnaire de Sujet, basé sur une étude empirique, calcule *comment* l'agent doit présenter les sujets dans l'interaction en cours sans perdre la cohérence de l'interaction. Nous avons implémenté et évalué le Gestionnaire de Sujets dans un agent virtuel conversationnel qui joue le rôle d'un visiteur dans un musée.

MOTS-CLEFS : Interaction humain-agent, agent conversationnel animé, engagement, stratégies de dialogue, préférences, gestionnaire de sujets, transitions de sujets, politesse, interruptions

ABSTRACT : In human-agent interaction the engagement of the user is an essential aspect to complete the goal of the interaction. In this thesis we study how the user's engagement could be favoured by the agent's behaviour. We thereby focus on the agent's verbal behaviour considering strategies that regard respectively the *form*, *timing*, and *content* of utterances : We present empirical studies that regard (aspects of) the agent's politeness behaviour, interruption behaviour, and the topics that the agent addresses in the interaction.

Based on the outcomes of the latter study we propose an engagement-driven Topic Manager (computational model) that personalises the topics of an interaction in human-agent information-giving chat. The Topic Selection component of the Topic Manager decides *what* the agent should talk about and *when*. For this it takes into account the agent's dynamically updated perception of the user as well as the agent's own mental state. The Topic Transition component of the Topic Manager, based upon an empirical study, computes *how* the agent should introduce the topics in the ongoing interaction without loosing the coherence of the interaction. We implemented and evaluated the Topic Manager in a conversational virtual agent that plays the role of a visitor in a museum.

KEY-WORDS : Human-agent interaction, embodied conversational agent, engagement, dialogue strategies, preferences, topic manager, topic transitions, politeness, interruptions



TELECOM ParisTech

EDITE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY