



HAL
open science

A touching agent : modelling the social touch modality for immersive human-agent interactions

Fabien Boucaud

► **To cite this version:**

Fabien Boucaud. A touching agent : modelling the social touch modality for immersive human-agent interactions. Human-Computer Interaction [cs.HC]. Université de Technologie de Compiègne, 2021. English. NNT : 2021COMP2666 . tel-03723514

HAL Id: tel-03723514

<https://theses.hal.science/tel-03723514v1>

Submitted on 15 Jul 2022

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

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

Par **Fabien BOUCAUD**

*Un agent touchant : modélisation du toucher social
dans les interactions humain-agent
en environnement immersif*

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC



Soutenue le 3 décembre 2021

Spécialité : Informatique : Unité de recherche Heudyasic (UMR-7253)

D2666

ED 71 - Sciences de l'ingénieur

**THÈSE DE DOCTORAT DE
L'UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE**

Spécialité

Informatique

École doctorale Sciences de l'ingénieur (Compiègne)

Fabien Boucaud

Un Agent Touchant :

**Modélisation du toucher social dans les interactions
humain-agent en environnement immersif**

Directrices de thèse : **Indira Thouvenin & Catherine Pelachaud**

Soutenue le 3 décembre 2021 devant le jury composé de :

M. Pierre DE LOOR, Professeur, LabSTICC, ENIB	Rapporteur
M. Daniel MESTRE, DR CNRS, ISM, Université Aix-Marseille	Rapporteur
Mme Françoise DÉTIENNE, DR CNRS, Telecom Paris	Examineur
M. Anatole LECUYER, DR INRIA, INRIA Rennes, IRISA	Examineur
M. Charles LENAY, Professeur, UTC, COSTECH	Examineur
Mme Indira THOUVENIN, Enseignant chercheur HDR, UTC, HEUDIASYC	Directrice de Thèse
Mme Catherine PELACHAUD, DR CNRS, ISIR, Sorbonne Université	Directrice de Thèse

**A touching agent : modelling the social touch
modality for immersive human-agent
interactions**

Fabien Boucaud

December 3, 2021

Résumé Long

Si la récente pandémie de COVID-19 et les confinements successifs qui l'ont accompagnée ont montré la puissance de nos technologies de communication pour maintenir du contact social dans des situations extrêmes, nous avons aussi pu réaliser à quel point elles restent aujourd'hui insuffisantes pour combler les sentiments de solitude et d'isolement. Le contact physique interpersonnel, le toucher social, est essentiel au bien-être humain (Montagu (1986), Cascio et al. (2019)), nous permet de développer et raffermir nos liens relationnels (Chatel-Goldman et al. (2014)) et est un puissant canal de communication des émotions (Hertenstein et al. (2009)). L'absence de la modalité haptique dans nos technologies de communication actuelles apparaît de fait comme un défi à surmonter dans les années à venir.

A ce titre, la recherche dans le domaine du haptique s'est considérablement développée dans les deux dernières décennies. Ce domaine s'articule en deux grands types de technologies: les technologies de détection du toucher humain (avec des interfaces tangibles et textiles connectés comme ceux développés par Silvera-Tawil et al. (2014) ou Cang et al. (2015)) et les technologies de génération haptique qui permettent de produire une sensation de toucher pour un humain (avec de nombreux types de stimuli parmi retour de force (Bailenson et al. (2007)), vibrations (Huisman et al. (2013)), alliage à mémoire de forme (Suhonen et al. (2012)), interfaces tangibles (Teyssier et al. (2018)), etc.). A l'aide de ces deux types de technologies il devient possible d'échanger du toucher à distance via une médiation technologique: on peut utiliser le toucher pour spécifier le type de toucher que l'on veut envoyer et recevoir une vraie sensation haptique.

Au-delà des interactions sociales entre humains cependant, de nouveaux types d'entités sociales artificielles commencent à apparaître comme des robots et des personnages virtuels. D'abord principalement utilisés dans le domaine du divertissement, notamment des jeux vidéos, ces agents artificiels sont maintenant déjà équipés de nombreuses modalités d'interaction (paroles, gestes, expressions faciales, etc.), et il n'est plus si rare d'entendre quelqu'un s'adresser à un assistant vocal ou être engagé émotionnellement dans des interactions avec des personnages de jeux vidéo. En particulier munir les Agents Conversationnels Animés (ACA), des agents virtuels munis d'un corps visible, de la modalité tactile pourrait donc améliorer encore leurs capacités émotionnelles et relationnelles.

A ce titre le but du présent travail de thèse est donc de déterminer sous quelles conditions des interactions sociales basées sur le toucher peuvent être mises en place entre un humain et un agent conversationnel animé. Nous nous intéressons plus particulièrement à la question de recherche suivante: comment peut-on déterminer quand et comment toucher un humain de manière utile, autonome et respectueuse. A cet effet, nous proposons une structure théorique et une implémentation fonctionnelle d'un agent touchant capable d'interagir dynamiquement avec un humain en temps réel, au sein d'un environnement immersif. En particulier, nous décrivons nos développements d'un module de perception des comportements de l'humain, nos parti pris théoriques et nos développements d'un modèle de décision qui prenne en compte la cohérence et l'acceptabilité du

toucher en contexte, et nos développements d'une interface haptique SOFTLY qui puisse être utilisée en immersion dans un environnement virtuel.

Les évaluations de ces développements sont décrites et utilisées pour mettre en lumière les capacités prometteuses de ce framework, ainsi que ses limites actuelles et les perspectives d'amélioration pour des travaux futurs.

Questions de Recherche

Cette thèse prend place dans le contexte du projet de recherche ANR Social Touch. Ce projet se positionne au croisement de l'interaction humain-machine (IHM) et de l'informatique affective. Son objectif est d'étudier la manière dont le sens du toucher peut être intégré à des systèmes interactifs pour permettre une communication sociale et émotionnelle à la fois entre humains via la machine et entre humains et machines. Cette thèse se positionne plus précisément dans le champ de l'IHM et sa motivation principale est d'étudier comment on peut munir des ACA de capacités de toucher social pour améliorer leurs capacités émotionnelles et relationnelles. Pour entreprendre cette étude, nous proposons une preuve de concept d'un ACA capable d'échanger dynamiquement des touches sociaux avec un humain et d'adapter son comportement au fil de l'interaction. L'agent est ainsi capable de percevoir les touches de l'humain et d'y réagir, mais aussi de toucher l'humain en retour de manière appropriée. Nous pensons qu'élaborer une telle boucle interactive sera une contribution importante à la fois pour le domaine des agents virtuels et pour les domaines du toucher social et de ses technologies.

Nous proposons de construire cet agent avec des capacités de toucher dans un contexte d'environnement immersif, de sorte à permettre l'élaboration d'environnements et scénarios entièrement contrôlés mais interactifs. Nous espérons que l'utilisation d'environnements immersifs avec un agent qui peut toucher et être touché nous permettra d'étudier le toucher social d'une manière plus écologique que ce qu'il est traditionnellement possible via des études observationnelles ou de laboratoire.

Notre question de recherche principale est donc la suivante : de quoi un agent virtuel a-t-il besoin pour être capable de faire partie d'une telle boucle interactive basée sur le toucher ? Comment peut-on donner à un humain la sensation d'être touché par un agent virtuel, malgré leur caractère intangible ? Traditionnellement, un agent autonome a besoin de trois capacités: perception, décision et action. Adapter ces capacités pour le toucher nécessite de répondre aux questions suivantes:

1. Comment peut-on donner un sens du toucher virtuel à un agent conversationnel animé, de sorte à ce qu'il puisse percevoir les touches réalisés par l'humain dans l'environnement immersif ?
2. Quand et comment toucher un humain d'une manière à la fois appropriée au contexte et acceptable ?
3. Et comment instancier une décision de toucher d'une manière qui sera perceptible par l'humain ?

Contributions

Pour répondre à ces questions, la présente thèse met en avant les contributions suivantes:

-
- Une revue de la littérature sur le toucher social, les technologies du toucher social et les agents conversationnels animés (Chapitre 2).
 - Une preuve de concept d'un framework mettant en place une boucle interactive intégrant le toucher social comme modalité d'interaction entre un humain et un agent virtuel au sein d'un environnement immersif. (Chapitre 3) Cette structure comprend plus spécifiquement:
 - Un module de perception pour un sens du toucher virtuel (Chapitre 4) avec:
 - * Détection et enregistrement des collisions avec le corps de l'agent et de leurs caractéristiques (vitesse, durée, localisation, etc.).
 - * Estimation de la pression exercée sur le corps virtuel de l'agent.
 - * Pré-interprétation et algorithme de reconnaissance des gestes de toucher simple, basé sur les ensembles flous.
 - * Estimation des distances proxémiques et de la direction de regard entre l'humain et l'agent.
 - Un modèle de décision (Chapitre 5) adapté à la prise en compte des interactions de toucher conçu avec:
 - * Un modèle computationnel d'émotions pour prendre la composante émotionnelle du toucher en compte.
 - * Un modèle de connaissances pour l'agent qui estime dynamiquement le rapport entre l'humain et l'agent en fonction du niveau d'attention de l'humain pour l'agent et de l'état émotionnel de l'humain.
 - * Un choix de l'intention de communication à transmettre à l'humain en fonction de l'état actuel de l'interaction.
 - * Des décisions qui prennent la réceptivité au toucher, le rapport et l'intention de communication en compte pour déterminer quand un toucher est acceptable et approprié.
 - Un dispositif haptique (SOFTLY) capable de générer une variété de retours haptiques pour des interactions de toucher social humain-agent (Chapitre 6):
 - * Un manchon tactile basé sur des technologies vibrantes (*voice-coil*).
 - * Facilement portable et sans fil.
 - * Capable de générer une impression de mouvement continu.
 - * Des signaux haptiques élaborés pour simuler différents types de toucher (tapotement, caresse, coup, toucher maintenu).
 - Une structure logicielle adaptable et facilement reproductible pour connecter le module de perception, le modèle de décision et le dispositif haptique.
 - Des évaluations des capacités du framework, ses composants et ses choix techniques (Chapitre 7):
 - Une évaluation de la crédibilité des touchers d'un agent en environnement immersif.
 - Une évaluation des propriétés de base du manchon tactile SOFTLY.
 - Un cas d'utilisation simulé d'évaluation du modèle de décision et en particulier de la cohérence du processus de décision.
-

-
- Une étude observationnelle à la troisième personne consacrée à l'étude de la perception d'observateurs humains quant aux décisions de toucher du modèle de décision concernant leur cohérence, leur acceptabilité et leurs fonctions de toucher.
 - Des propositions de lignes directrices pour de futurs travaux sur des agents touchants et l'utilisation de toucher social simulé.

Revue de la littérature du toucher social

Pour déterminer ce dont un ACA a besoin pour être capable d'avoir des interactions basées sur le toucher intéressantes avec un humain, nous proposons dans le chapitre 2 une revue détaillée de la littérature du toucher social entre humains et des technologies du toucher social. Dans ce chapitre, nous mettons d'abord en lumière les propriétés physiques et anatomiques avant de montrer l'importance cruciale du toucher pour le bien-être et le développement socio-psychologique des humains (Montagu (1986), Cascio et al. (2019)). Nous montrons également que le toucher social est particulièrement adapté à l'expression des émotions (Hertenstein et al. (2009)) et au processus de construction des relations (Montagu (1986), Chatel-Goldman et al. (2014)). En ce sens, si la littérature montre bien les nombreux effets positifs du toucher, elle met également en lumière le fait que le toucher reste pour autant toujours un canal de communication qui relève de l'intime et que, de ce fait, tous les touchers ne sont pas aussi acceptables les uns que les autres pour la personne qui les reçoit. Il y a d'abord une composante culturelle, les normes sociales, qui encadre l'acceptabilité d'un toucher en fonction du genre, de la situation, des rapports de pouvoir, de la relation, etc. (Dibiase and Gunnoe (2004), Remland et al. (1995), Meyer et al. (2017)) Au-delà encore de ces effets culturels, chaque personne a sa propre réceptivité au toucher (Webb and Peck (2015), Martin and Anderson (1993), Andersen and Leibowitz (1978)) singulière, personnelle, qu'il convient de mesurer et prendre en compte pour comprendre l'influence qu'un toucher aura sur une personne.

Du côté du domaine des technologies du toucher social, nous montrons que la littérature y a connu un rapide développement dans les dernières années, avec l'apparition de nombreux nouveaux types de technologies et d'applications. Le domaine s'articule plus particulièrement en deux grands champs d'études consacrés aux deux fonctions du toucher humain: la détection du toucher (et la reconnaissance du type de toucher qui y est souvent associée) et la génération de la sensation de toucher. Du côté de la détection du toucher, il y a de très bons résultats avec l'association de textiles intelligents et d'algorithmes de machine learning pour détecter et interpréter les touchers humains (Silvera-Tawil et al. (2014), Cang et al. (2015)). La détection de toucher en environnement immersif est beaucoup moins développée en revanche, avec seulement deux occurrences d'études proposant un sens du toucher virtuel (Nguyen and Wachsmuth (2009), Sykownik and Masuch (2020)). Du côté de la génération de toucher en revanche, il n'y a pas de façon parfaite de simuler des sensations de toucher "naturelles", dans le sens où le sens du toucher humain est un système complexe qui inclut tout à la fois la température, la douleur, la proprioception, la détection des formes et des textures, etc. Il y a toutefois de très nombreux exemples de dispositifs haptiques qui proposent d'adresser ces différentes fonctions de la peau humaine : dispositifs à retour de force pour donner une impression de tenir un objet, dispositifs thermiques, dispositifs vibrotactiles,... Construire un dispositif à la fois portable et capable d'utiliser ces différentes technologies en même temps reste cependant aujourd'hui un challenge.

En ce qui concerne l'application de ces domaines de recherche à l'interaction humain-agent, nous identifions un point aveugle dans la littérature : là où il commence à y avoir de nombreuses études sur les effets d'un toucher venant d'un robot ou d'un agent sur un humain, ainsi que sur la façon dont les humains perçoivent et interprètent ces types de toucher simulés, il y a très peu voire aucun travaux qui implémentent une boucle interactive complète avec un agent capable de s'adapter à un toucher reçu et toucher en retour dynamiquement et de façon autonome. Nous choisissons de nous positionner dans cet angle mort actuel de la littérature en proposant un modèle de décision qui permettra de compléter la boucle interactive et ainsi étudier la manière dont des capacités de toucher peuvent améliorer les capacités socio-relationnelles d'un agent.

Architecture et choix techniques pour un agent touchant

Pour rendre possible une boucle interactive intégrant du toucher social en environnement immersif, nous proposons dans le chapitre 3 d'adopter l'architecture classique des agents autonomes qui requiert des capacités de perception, de décision et d'action, et d'étudier la façon d'on peut y intégrer le toucher. Nous choisissons également d'utiliser un dispositif immersif de type de salle immersive pour ses puissantes capacités d'immersion et son avantage (par rapport à des casques de réalité virtuelle classiques) de laisser l'utilisateur en capacité de voir son propre corps. De cette façon, le problème de l'*embodiment*, c'est-à-dire le sentiment d'incarnation dans l'avatar virtuel de l'utilisateur, ne se pose pas pour ce premier travail exploratoire. Cette décision d'utiliser un environnement immersif de réalité virtuelle est particulièrement guidé par l'argument tenu par [Huisman et al. \(2014a\)](#) que la RV pourra profiter à l'étude du toucher social en nous permettant de soigneusement construire des scénarios d'interaction engageants et crédibles qui pourront mener à des réactions plus naturelles de la part des sujets que ce qu'il est actuellement possible de réaliser en interaction directe en laboratoire ou avec des études observationnelles.

Pour intégrer le toucher à une architecture d'agent autonome et un environnement immersif, nous procédons au développement d'un module de perception pour donner un sens du toucher virtuel à notre agent, et ainsi le rendre sensible aux touchers humains et à leurs propriétés physiques. Pour la capacité de décision, nous avons eu le sentiment que le lien fort entre émotion, relation et toucher était crucial et devait donc être pris en compte. Le toucher est en effet une modalité puissante et intime qui demande d'être extrêmement prudent dans la manière dont on va l'utiliser avec les autres. Nous avons de ce fait proposé d'utiliser un modèle computationnel d'émotion pour rendre notre agent touchant au plein sens du terme: capable de toucher à la fois physiquement et émotionnellement. Enfin, pour rendre perceptible les touchers de l'agent auprès de l'humain (la capacité d'action), nous utilisons à la fois la plateforme d'animation et de contrôle d'agents virtuels GRETA pour l'aspect visuel des comportements de l'agent, et nous développons SOFTLY, un manchon vibrotactile, pour l'aspect haptique du retour.

Cette architecture est implémentée dans une structure logicielle élaborée pour être facilement reproductible et intégrable avec d'autres équipements immersifs (comme des casques) et d'autres dispositifs haptiques. Nous pensons que c'est en effet un point particulièrement important pour l'avancée de ce domaine de recherche quand on travaille avec des équipements onéreux comme peut l'être une salle immersive.

Un module de perception pour un sens du toucher virtuel

Dans le chapitre 4 nous montrons comment nous avons implémenté un sens du toucher virtuel pour des interactions immersives en améliorant des travaux pré-existants de la littérature (Sykownik and Masuch (2020), Nguyen and Wachsmuth (2009)) qui utilisent une grille de détecteurs de collisions virtuels (des "colliders") pour former une peau virtuelle pour l'agent. Avec cette approche on peut détecter quand et où un toucher a lieu ainsi que la vitesse, la durée et la présence ou l'absence de mouvement,...

L'évaluation présentée dans la section 7.1 a cependant montré les limites de notre première implémentation, la résolution de la grille étant un facteur hautement limitant qui rendait les données difficiles à interpréter. De plus, cette approche ne permettait pas d'avoir une estimation de la pression exercée, du fait de l'intangibilité de l'agent, ce qui est une propriété particulièrement importante pour le toucher. Nous avons donc proposé une deuxième implémentation de notre module de perception utilisant un système de coordonnées locales aux colliders et en utilisant une méthode de god-object avec un ressort pour avoir une estimation de la pression exercée sur l'agent. Nous avons de plus mis en place une détection des détections de regard et des estimations des distances interpersonnelles.

Enfin, nous avons implémenté un algorithme simple utilisant des ensembles flous pré-définis pour avoir une pré-interprétation basique des valeurs brutes enregistrées par le module de perception et ainsi faciliter le processus décisionnel.

Un modèle de décision pour un agent touchant

Dans le chapitre 5 nous décrivons le processus de décision qui prendra en compte les entrées enregistrées par le module de perception pour modéliser les connaissances de l'agent. Notre modèle de connaissance pour l'agent est élaboré pour permettre de déterminer si le toucher est cohérent avec le contexte d'interaction et acceptable étant donné la réceptivité au toucher de l'humain et le niveau de la relation entre lui et l'agent. En ce qui concerne l'acceptabilité, nous avons notamment établi un modèle d'estimation du niveau de *rappor*t (terme anglais qui désigne l'entente en quelque sorte) entre l'humain et l'agent en fonction du niveau d'attention exprimé par l'humain vis-à-vis de l'agent et de l'état émotionnel de l'humain.

Nous détaillons également la manière dont nous avons modélisé les processus d'interprétation et d'évaluation cognitive des entrées du modèle pour générer les connaissances et l'état émotionnel de l'agent en fonction des événements dans l'interaction et, en particulier, des actions de l'humains. En combinant la réceptivité au toucher estimée pour l'humain à ces connaissances, et en particulier au niveau estimé de rapport entre l'humain et l'agent, nous pensons que notre agent sera plus à même de déterminer quand toucher sera acceptable pour l'humain. Cette estimation de l'acceptabilité et de la cohérence d'un toucher est réalisée au moment de la sélection et de l'activation de règles de décisions élaborées à partir de la littérature sur le toucher social en fonction de l'état de l'interaction et des différentes connaissances de l'agent. Ce modèle a été implémenté dans un modèle computationnel d'émotion déjà existant et testé dans la littérature, FATiMA (Dias et al. (2014)), que nous avons modifié pour répondre à nos besoins particulier quand nécessaire.

Le processus de décision a lieu en deux étapes : on détermine d'abord l'intention de communication la plus adaptée à la situation actuelle, puis on décide quelle modalité (toucher, gestes, parole) est appropriée à cette intention (selon la littérature, comme l'étude de Jones and Yarbrough (1985)) et on valide le choix en fonction de l'acceptabilité

de la modalité. Les types de toucher spécifiques à utiliser sont déterminés par l'intention de communication selon les résultats de la littérature sur le toucher social.

Le manchon vibrotactile SOFTLY

Le chapitre 6 présente SOFTLY, un manchon vibrotactile facile à porter et composé d'une grille de bobines acoustiques (voice coils) que l'on peut utiliser pour générer un retour vibratoire. La technologie des voice coils nous permet d'utiliser de nombreux types de formes d'ondes et de fréquences différents pour produire des sensations intéressantes. On peut ensuite combiner nos patterns haptiques avec les animations de notre agent pour instancier de manière synchronisée les décisions de toucher du modèle.

Une évaluation présentée dans la section 7.2 nous a permis d'investiguer les propriétés psychophysiques de notre dispositif haptique avec des utilisateurs humains et de calibrer certaines de nos valeurs.

Framework fonctionnel et évaluations

Toutes ces contributions prises ensemble forment un framework fonctionnel à la fois technique et théorique, qui a été développé pour être facile à importer dans des installations similaires d'interactions immersives avec des agents, même si l'équipement spécifique diffère de celui présenté ici.

Si, pour l'instant, la pandémie de COVID a rendu impossible d'évaluer le modèle dans les conditions prévues initialement (en interaction directe en environnement immersif), nous présentons tout de même dans le chapitre 7 les différentes évaluations des composants du framework qui ont montré des résultats intéressants. En particulier, nous présentons une étude observationnelle à la troisième personne réalisée en ligne, dans laquelle le modèle de décision était évalué par des participants humains auxquels nous présentions des interactions enregistrées à l'avance dans lesquelles le modèle de décision prenait des décisions de toucher dynamiquement. Nous avons interrogé les participants sur leurs impressions concernant la cohérence et l'acceptabilité des décisions de l'agent pour le toucher et nous avons également évalué leurs estimations des niveaux émotionnels et d'attention de l'humain simulé présenté dans les enregistrements. Les résultats ont montré que nos choix initiaux pour le modèle sont dans l'ensemble corrects, mais nécessitent quelques ajustements. En revanche, seule une évaluation dans le contexte immersif complet pourra nous permettre de conclure avec plus de certitude concernant les capacités de notre modèle.

Avec cette thèse, nous espérons tout de même pouvoir démontrer l'intérêt et l'étendue de ce domaine de recherche et la possibilité de construire un tel ACA capable de prendre part à une boucle interactive intégrant le toucher social, bien qu'il reste de nombreux défis à relever à la fois pour les capacités de perception, décision et de génération des comportements.

Remerciements

Je remercie tout d'abord mes directrices de thèse, Indira Thouvenin et Catherine Pelachaud, pour m'avoir offert l'opportunité de travailler sur cette thèse, pour m'avoir guidé tout au long de ce travail de recherche et pour m'avoir soutenu dans mes propositions et idées.

Je remercie également Quentin Duchemin, pour toute son aide et toute la qualité de son travail technique au cours des différents moments où il a pu collaborer à mon travail de recherche, et sans lesquelles les travaux présentés ici n'auraient sans doute pas pu voir le jour, mais également pour nos discussions et moments d'échanges toujours chaleureux, agréables, intéressants voire musicaux.

A ce titre, mes remerciements vont aussi à Quentin Tafiani, Valentin Cecchin et Maxime Grandidier pour leurs différentes contributions de grande qualité aux travaux présentés ici.

Si je ne suis pas sûr d'à quel point ma famille a jamais compris la teneur exacte de mon travail et de mon sujet de recherche, je les remercie néanmoins pour leur soutien et leur présence dans ma vie, sans lesquels je me sentirais bien souvent perdu et dépassé par les événements.

De même, j'adresse de profonds sentiments de gratitude à tous les amis qui m'ont accompagné tout au long de ces trois années de travail, qui m'ont aidé à me changer les idées quand nécessaire et à me faire rire, mais qui savent aussi m'écouter et me faire part de leurs avis éclairés à l'occasion.

Enfin, je ne sais pas si j'aurais pu arriver au bout de ce présent travail de thèse, en particulier étant donné les circonstances mondiales compliquées, sans la présence et le soutien de Jingjing, même si celui-ci passait parfois par me rappeler aussi que j'étais toujours libre de laisser tomber !

Merci du fond du coeur.

Contents

1	Context of the Thesis	1
1.1	Introduction	1
1.2	Research Questions & Contributions	3
1.3	Publications & Dissemination	5
2	Theoretical Background	7
2.1	The Sense of Touch and Social Touch Among Humans	8
2.1.1	The Sense of Touch: biological properties	8
2.1.2	Social touch: the importance of touching and being touched by others	9
2.2	Social Touch Technologies	14
2.2.1	Social Touch Generation	15
2.2.2	Social Touch Perception	18
2.3	Simulated Social Touch: Human-Agent Interactions	19
2.3.1	Embodied Conversational Agents and their social abilities	19
2.3.2	Human-Agent Tactile Interactions	21
2.3.3	Position of our work in the simulated social touch literature	26
2.4	Computational Models of Emotion	27
2.4.1	Theoretical Models of Emotion	27
2.4.2	Computational implementation of models of emotion	29
3	Designing Human-Agent social touch interactions in VR	31
3.1	Requirements for a Touching Agent	31
3.2	Architecture of a Virtual Touching Agent	33
3.3	Choosing an adequate VR setup	35
4	Perception Module for a Virtual Touching Agent	37
4.1	A Virtual ECA's Sense of Touch	38
4.1.1	Basic Principles of a Virtual Sense of Touch	38
4.1.2	First version of the Touch Perception Module	38
4.1.3	Improving the Touch Perception Module	39
4.1.4	Interpretation of the Touch Properties	42
4.2	Perception of additional human inputs	45
5	A Touching Agent's Decision Model	47
5.1	Decision Process of a Touching Agent	48
5.1.1	Overall design choices for the decision process	48
5.1.2	Overall implementation framework	49

CONTENTS

5.2	Internal States of the Agent	51
5.3	Appraisal Rules	53
5.4	Beliefs of the Agent	57
5.4.1	World Model & Interpretation rules	58
5.4.2	Determining the information needed to model the acceptance of a touch	58
5.4.3	The dimensions of rapport	59
5.4.4	Estimation of the human's emotional state	59
5.4.5	Estimation of the human's attentiveness towards the agent	61
5.4.6	Evaluating rapport & static touch avoidance	61
5.4.7	Implementation: Interpretation rules and Rapport calculation	61
5.5	When and How to Touch a Human	63
5.6	Decision Rules	65
6	Haptic Device and Agent Animation for Simulated Social Touch	73
6.1	Audiovisual Animation of the Agent	73
6.2	Designing a Haptic Device for Simulated Social Touch	75
7	Evaluations of the Touching Agent	81
7.1	Preliminary study	82
7.1.1	Experimental Protocol	82
7.1.2	Setup and Questionnaire	84
7.1.3	Participants	85
7.1.4	Results	85
7.2	Validation of the Haptic Sleeve Softly	89
7.2.1	Experimental Protocol: Qualification of haptic properties	89
7.2.2	Results	91
7.2.3	Discussion	93
7.3	Use Case Simulation of the decision model	94
7.4	Coherence and Acceptability: Online evaluation of the model	98
7.4.1	Design of the study	99
7.4.2	Research Questions & Measures	101
7.4.3	Procedure	103
7.4.4	Results	103
7.4.5	Discussion	114
8	Conclusion and Perspectives	119
8.1	Summary of Contributions	120
8.1.1	Contribution 1: Review of the Social Touch literature	120
8.1.2	Contribution 2: Architecture and Design Choices for a Touching Agent	121
8.1.3	Contribution 3: A Perception Module for a Virtual Sense of Touch	121
8.1.4	Contribution 4: A Decision Model for a Touching Agent	122
8.1.5	Contribution 5: the Vibrotactile Sleeve SOFTLY	122
8.1.6	Contribution 6: Functional Framework and Evaluations	123
8.2	Current limitations	123
8.3	Perspectives for future works	125

CONTENTS

Bibliography

146

List of Tables

2.1	Emotions and the touch types most commonly associated with them.	12
4.1	Linear partition table of a touch sequence's duration.	43
4.2	Linear partition table of velocity at the time of the touch collision.	43
4.3	Linear partition table of a touch sequence's mean force exerted.	43
4.4	Linear partition table of a touch sequence's mean movement velocity.	44
4.5	First interpretation table	44
7.1	Results of the first task	92
7.2	Results of the second task	92
7.3	Results of the third task	93
7.4	Results of the fourth task	93
7.5	Distribution of the participants' choices regarding the function of the agent's touches they observed for each subvideo of each condition.	109
7.6	Distribution of the participants' answers about whether they would also have used touch in this particular situation had they been in the shoes of Camille, the agent.	113

List of Figures

2.1	Examples of tangible haptic devices (Mobilimb on the left, and a robotic arm on the right, both by Teyssier et al.)	16
2.2	The SenseGlove Nova, a vibrotactile sleeve (TaSST, by Huisman et al.) and the Dialtact system (Lenay et al.)	17
2.3	The PARO robotic seal (left), and the Haptic Creature (right, Yohanan et al.)	23
3.1	Schematic overview of the components proposed for our touching agent . .	34
3.2	Augmented reality (left), virtual reality (middle) and immersive room (right) setups	35
4.1	Virtual skin receptors for virtual agents: Nguyen and Wachsmuth implementation (left) compared to ours (right)	39
4.2	Ethereal body of our agent (left) and second version of our tactile cells (right)	40
4.3	Side and top-down views of the axis used in the definition of the tactile cells relative coordinates	41
4.4	Structure of the code implementation of the touch perception module	42
4.5	Second interpretation table to take into account the velocity of the gesture at the time of the initial collision	45
4.6	Proxemic spaces and their respective dimensions	46
5.1	Schematic overview of the overall process of the decision model in an interactive loop with a human	49
5.2	Software architecture for social touch interactions with a virtual embodied conversational agent	50
5.3	Updating process of the internal states of the agent upon perception of an event	51
5.4	Screenshot of the FAtiMA interface with a few appraisal rules for touch events.	56
5.5	The affect derivation processed of the OCC model, as implemented in FAtiMA (available at https://fatima-toolkit.eu/5-emotional-appraisal/)	57
5.6	The circumplex model of affect and its emotional octants (by Knez and Hygge (2001))	60
6.1	The SAIBA model of behaviour animation planning and realization	74
6.2	First version of the SOFTLY tactile sleeve	76
6.3	Second version of the SOFTLY tactile sleeve	77
6.4	Complete software architecture for a touching agent in immersive environment	78
6.5	Process and framework of a human-agent social touch-based interactive loop in immersive environment	79

LIST OF FIGURES

7.1	The immersive environment of the study, with the agent	83
7.2	Credibility of the virtual agent's behaviour according to participants.	86
7.3	Recognition rate of the emotion transmitted by the agent.	86
7.4	Shots extracted from the videos of the online study of the model's decisions, with the agent (Camille) in a blue shirt and the simulated participant (Laura) in a teal one.	101
7.5	Evolution of the ratings of Laura's mood across conditions and subvideos.	105
7.6	Evolution of the ratings of Laura's attentiveness across conditions and subvideos.	106
7.7	Evolution of the ratings of the level of rapport between Laura and Camille across conditions and subvideos.	107
7.8	Evolution of the ratings of the degree of "correctness" of Camille's touch decision across conditions and subvideos.	110
7.9	Mean correctness ratings for each identified touch function in the Introduction subvideos	112

Context of the Thesis

Le corps est l'expérience de toucher indéfiniment à l'intouchable, mais au sens où l'intouchable n'est rien qui soit derrière, ni un intérieur ou un dedans, ni une masse, ni un Dieu. L'intouchable, c'est que ça touche. On peut aussi employer un autre mot pour dire cela : ce qui touche, ce par quoi on est touché, c'est de l'ordre de l'émotion. Émotion est un mot très affaibli pour nous, mais émotion, cela veut dire : mis en mouvement, mis en branle, ébranlé, affecté, entamé..

Jean-Luc Nancy

Contents

1.1 Introduction	1
1.2 Research Questions & Contributions	3
1.3 Publications & Dissemination	5

THIS Chapter introduces the context where this Thesis is placed, in particular the research questions we address and the approach followed to answer these questions. The motivation of this work and the main steps we realised to accomplish it are introduced. A list of the contributions of this Thesis and the publications that resulted from it is also given.

1.1 Introduction

In our digital societies, we are perpetually connected to each others. We can call our friends on the phone or send them a text message, we can connect with people from the other side of the world online and exchange ideas at any hour of the day or the night. Beyond human-human communication, we can even talk to our vocal assistants about the weather or our tastes in music. We can emotionally bond with virtual characters in video

games. Even in the most dire situations, with the forced lockdowns provoked around the world by the COVID pandemic, our digital communication technologies have allowed them to keep on working, studying or at least staying updated on the situation. And despite all that, even in the middle of the digital crowds, we can still feel the sharp sting of loneliness and isolation.

As human beings, and maybe even simply as living beings, we crave the contact of others, their physical presences and their touch. Touch is our most fundamental way to apprehend the world, even when all our other senses fail us, we will never lose the ability to touch, to feel the world around us, to explore its textures and shapes, to experience the outside and its resistance. It is also an incredibly powerful social sense, thus often reserved to our close relationship, that we can use to express emotions (Hertenstein et al. (2009)) and to bond (Chatel-Goldman et al. (2014)) with others, to fight or to help, to feel and to share. Even in the field of virtual reality, which may seem like the domain the furthest away from touch considering its intangible nature, haptics are an increasingly important field of research, further improving the feeling of presence and immersion in virtual environments Garcia-Valle et al. (2018). And yet while it probably has never been as easy to interact with others than today, be it with humans, pets, robots, chatbots, plush toys, video game characters, vocal assistants, or even the simplest of our technological tools, touch, despite its critical importance in our well-being and social development (Montagu (1986), Cascio et al. (2019)), remains crucially (and often cruelly) absent from most of our communication technologies. We believe that integrating touch, and especially social touch, to the panel of our digital interaction modalities will be both one of our biggest challenges and one of the most essential task to accomplish in order to feel content in our increasingly digital world.

And while this will necessarily entail the development of touch mediation technologies, so that we may touch our fellow humans over the distance, it should also become an important way to interact with the new kinds of social entities that are beginning to appear. In recent years, many different types of artificial companions have been created to take up roles that we would have thought reserved to human beings. They can be machines, robots, or virtual agents: entirely fabricated characters who nonetheless can provide us services and make us feel like we really are connecting and interacting, as they are endowed with more and more social abilities (speech, gestures, social strategies of collaboration and bonding, etc.). As we have a very tactile relationship with our technical tools (we hold our pens, our smartphones, we tap on our keyboards, we wear our clothes), so it would make sense that we would also start to interact tactually with those new artificial companions. Such touching companions could even help us in the domains of healthcare (both physical and mental), entertainment, technical training, collaboration, etc.

The goal of the present Thesis is to explore the conditions that would make meaningful and respectful social touch interactions with embodied conversational agents possible,

and investigate how this would improve their relational and emotional abilities. Neither the production of haptic stimuli nor the perception of human touch are particularly new topics of interest, but even in social interactions among humans, the question of determining when and how to touch another without hurting them doesn't have a clear answer and is at the heart of current societal issues (MeToo, BlackLivesMatter,...). By investigating how we can enable social touch interactions between humans and agents, we thus hope to contribute to both the field of research on autonomous embodied conversational agents social abilities and that of social touch interactions at large, both in practical and theoretical ways.

1.2 Research Questions & Contributions

The main motivations of this Thesis, which takes place in the context of the ANR Social Touch project, is to investigate how to endow embodied conversational agents with social touch abilities to enhance their emotional and relational abilities. While there are already quite a few research works that have investigated the effects of agent-initiated social touch on human, we should note that those works are mostly built around very controlled protocols, with unidirectional touches coming from an almost entirely scripted agent. Works on truly interactive agents that use touch are, on the contrary, very hard to come by at the present time. We propose a different, more interactive approach: building a proof-of-concept of an embodied conversational agent that would be able to seamlessly exchange social touches with a human by adapting its behaviour during the interaction. This means that the agent would be able to receive touches from the human and react to it, as well as being able to touch the human too. We believe that building such an interactive loop would prove beneficial both to the field of virtual agents and to the field of social touch and social touch technologies.

Furthermore, we propose to build this agent with touch abilities in an immersive environment context, thus allowing us to build controlled environments and scenarios that would still give a lot of freedom of action to the human. The hope is that such use of immersive environment with an agent that can touch will enable us to study social touch in a more ecological way as opposed to the difficulty of studying social touch in a laboratory or with observational studies. Enabling such an interactive loop based on social touch in immersive environment represents a challenge and research topic in itself however, which leads us to our second research question: what does a virtual agent need to be able to perform in such an interactive loop? How can we make it so that the human will feel touched by an agent characterized by its virtuality and a priori intangibility?

This question can only lead to more questions, because a truly interactive agent traditionally needs three abilities: perception, decision and action. In the present Thesis, we propose to address those questions that we will formulate in those terms:

1. How can we build a virtual sense of touch for an embodied conversational agent in an immersive environment?
2. When and how to touch a human so that the touch would be both meaningful and acceptable?
3. And how to instantiate a touch decision in a way that a human could feel and, again, accept?

To answer those questions we put forward the following contributions:

- A review of the literature on social touch, social touch technologies and embodied conversational agents (Chapter 2).
- A functional framework allowing an interactive loop integrating social touch as a modality between a human and an agent in an immersive environment (Chapter 3), which includes:
 - A perception module for a virtual sense of touch (Chapter 4) with:
 - * Collision and touch characteristics detection and recording (speed, duration, location of the touch,...).
 - * Estimation of the force exerted on the agent's virtual body.
 - * Simple pre-interpretation and recognition of touch gestures algorithm based on fuzzy sets.
 - * Estimation of proxemics and gaze direction.
 - A decision model (Chapter 5) adapted to the consideration of touch interactions designed as:
 - * A computational model of emotions to take the emotional component of touch into account.
 - * A model of beliefs that estimates dynamically the Rapport between the human and the agent based on attentiveness and emotional state.
 - * Choice of communicative intention to be conveyed based on the state of the interaction.
 - * Decisions that take touch avoidance, rapport and communicative intention into account to determine when a touch is acceptable and appropriate.
 - A haptic device (SOFTLY) able to generate a variety of haptic feedback for human-agent social touch interactions (Chapter 6):
 - * Voice-coil based tactile sleeve.
 - * Wearable and wireless.
 - * Generation of continuous movement impressions.

- * Design of haptic patterns that simulate different types of touch (tap, caress, hit, sustained touch).
- An adaptable and easily reproducible software structure to connect the perception module, decision model and haptic device together.
- Evaluations of the abilities of this framework, its components and its design choices (Chapter 7):
 - Evaluation of the perception module recordings and the credibility of an agent’s touch in an immersive environment.
 - Evaluation of the basic properties of the SOFTLY tactile sleeve.
 - Use case simulation of the decision model to check the coherence of the overall decision process.
 - Third-person observational study of human observers’ perception of the touching decisions of the decision model regarding coherence, acceptability and functions of the touch.
- Guidelines with propositions for future works regarding touching agents and the use of simulated social touch.

1.3 Publications & Dissemination

The publications in International Conferences and Workshops include:

- Fabien Boucaud, Quentin Tafiani, Catherine Pelachaud, Indira Thouvenin. Social Touch in Human-agent Interactions in an Immersive Virtual Environment. *3rd International Conference on Human Computer Interaction Theory and Applications (HUCAPP 2019)*, Feb 2019, Prague, Czech Republic. pp.129-136
- Fabien Boucaud, Catherine Pelachaud, and Indira Thouvenin. 2021. Decision Model for a Virtual Agent that can Touch and be Touched. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 232–241.

The publications in National Workshops include:

- Fabien Boucaud, Quentin Tafiani, Catherine Pelachaud, Indira Thouvenin. Vers une prise en compte du toucher social dans les interactions humain-agent en environnement virtuel immersif. *Journées de la Réalité Virtuelle (j•RV 2018)*, Oct 2018, Evry, France.

- Fabien Boucaud, Indira Thouvenin, Catherine Pelachaud. Quand et Comment Toucher un Humain ? Un Modèle de Décision pour un Agent Touchant. *Workshop sur les Affects, Compagnons artificiels et Interactions (WACAI 2020)*, CNRS, Université Toulouse Jean Jaurès, Université de Bordeaux, Jun 2020, Saint Pierre d’Oléron, France.

The work presented in this Thesis has also been presented as posters:

- Fabien Boucaud, Indira Thouvenin, Catherine Pelachaud. Touch-aware agent: challenges of social touch in virtual reality, *ConVRgence (VRIC) Virtual Reality International Conference Proceedings*, Laval, France, avril 2020
- Fabien Boucaud, Indira Thouvenin, Catherine Pelachaud. A Touching Agent : Integrating Touch in Social Interactions between Human and Embodied Conversational Agent in an Immersive Environment. *3rd International Association for the Study of Affective Touch Congress (IASAT 2019)*, Sep 2019, Linköping, Sweden.

Theoretical Background

Contents

2.1	The Sense of Touch and Social Touch Among Humans	8
2.1.1	The Sense of Touch: biological properties	8
2.1.2	Social touch: the importance of touching and being touched by others	9
2.2	Social Touch Technologies	14
2.2.1	Social Touch Generation	15
2.2.2	Social Touch Perception	18
2.3	Simulated Social Touch: Human-Agent Interactions	19
2.3.1	Embodied Conversational Agents and their social abilities	19
2.3.2	Human-Agent Tactile Interactions	21
2.3.3	Position of our work in the simulated social touch literature	26
2.4	Computational Models of Emotion	27
2.4.1	Theoretical Models of Emotion	27
2.4.2	Computational implementation of models of emotion	29

This Chapter introduces the subject of touch and social touch. We first investigate the literature regarding the fundamental properties of touch as one of the five human senses, before going more specifically in the direction of touch as a social communicative channel and exploring how the forms it can take in traditional human-human interactions. Departing from what we might be accustomed to in our everyday social experiences, we then look at the research on social touch technologies and how it can both allow humans to exchange touches through the distance and enable touch interactions with artificial companions, either virtual or robotic. Along the way, we

discuss the position of our work and the contributions it can make in this increasingly popular research domain, notably regarding the complete interactive loop and decision models for human-agent social touch.

2.1 The Sense of Touch and Social Touch Among Humans

2.1.1 The Sense of Touch: biological properties

What we call in our everyday lives the “sense of touch”, one of the five human senses alongside sound, sight, smell and taste, designates in fact a complex system of perception : the somatosensory system, which encompasses all the organs (skin, muscles, joints,...) concerned by the perception of tactual sensations. Functionally, we can differentiate two types of tactual sensations that constitute the sense of touch : the tactile or cutaneous perception and the kinesthetic perception (Loomis and Lederman (1986)).

- Cutaneous (or tactile) perception refers to all the sensations perceived at the surface of the skin, such as pain, temperature, texture, following a pattern from the tip of the finger, etc. In anatomic terms, those perceptions are the product of the multiple types of cutaneous receptors present in our skins. Thermoreceptors perceive warmth, nociceptors perceive pain and further types of mechanoreceptors all contribute to the perception of textures and discriminatory properties of our touch, allowing us to determine the shapes of our surrounding environment.
- Kinesthetic perception, also often called proprioception, refers to the awareness of body spatial position and movement : feeling where our head, limbs and other parts of our body are positioned relative to each other, allowing us to move with coherence and stability as well as enabling us to estimate the weight and/or length of objects based on dynamic perceptions. Kinesthetic perception is the product of the wide variety of proprioceptors found in our muscles, tendons and joints.

Although they are extremely different in terms of functions, there are very few cases where kinesthetic and cutaneous perceptions are separated, and what we usually experience in our everyday lives when it comes to touch is **haptic** perception : the combination of tactile and kinesthetic sensations.

Haptic perception thus fulfills multiple critical functions for human beings by allowing us to perceive the boundaries of the world and all the objects that inhabit it, as well as the properties of those objects: weight, texture, temperature, etc. This makes us equally aware of the outside’s potential dangers (thanks to the perception of pain for example) and its potential affordances (what we can do to interact with the world).

In fact, touch is the first sense to develop itself in the fetus before birth, and it is also the primary sense through which the newborn infant will progressively learn to differentiate

2.1. THE SENSE OF TOUCH AND SOCIAL TOUCH AMONG HUMANS

his bodily self from the rest of the world (Montagu (1986)). On the scale of the species, research has shown how the technical development of humans through the manufacturing of tools seemed to correlate with the development of their brains' size (Leroi-Gourhan (1964)). And finally, even in our contemporary “digital” work (“digit” etymologically coming directly from the Latin word for finger), we still interact mainly with our technical tools through touch, be it for manipulating a hammer or to send a message to a friend via our tactile smartphone screens.

Despite this considerable importance of touch in the development of human beings and in our everyday lives, research regarding the senses, their properties and uses, has remained scarce for the longest time, and is still lacking. This is explained by the traditions of Western thinking, which, influenced by religions, considered the body as a place of vice and lowly desires that should be ignored, as opposed to the virtuous and transcendent mind (Cranny-Francis (2011)). Since the middle of the 20th century however, the deconstruction of the old idea of the mind/body dichotomy and the multiplication of works and essays regarding the embodied experience of living, often produced in the fields of feminist and social studies, has led to the appearance of more and more research regarding the senses and, eventually, touch.

There is still much to learn regarding the sense of touch, both in terms of physiology and of its meanings in our embodied experiences of the world and of society. Nevertheless, the importance of touch in our lives goes beyond the ability to perceive the boundaries and textures of objects, it is also one of our most essential modalities when interacting with other people.

2.1.2 Social touch: the importance of touching and being touched by others

As social individuals, humans can use multiple modalities to express themselves and communicate with other people: we can speak, modulate our speech with prosody, use gestures, gaze, write, and we can touch. Beyond its discriminative and proprioceptive functions, touch can take a wide range of meanings when used in a social setting, and it is essential to our social lives as it is to our proper functioning as material bodies (Cascio et al. (2019)).

Here, we will call “social touch” all the uses of touch in a social context. It can be a handshake to greet somebody, a tap of encouragement on the arm, or any type of interpersonal touch, even accidental, as even the most fortuitous touch can have a sizable impact on the perception we have of another person or on the feelings we experience.

As is the case with most of humans' communicative modalities, social touch is not easy to formally study. Creating natural-enough circumstances to elicit natural touch interactions in the laboratory is difficult, while observational and self-report studies have their own inherent biases and challenges to overcome. Despite this, since the 1950s there has been abundant research on the meanings of touch in social interactions, from a wide range

of different research domains. When reviewing this dense literature, it appears that the “meanings” of a touch are not that easy to define.

2.1.2.1 The inherent ambiguity of social touch

In their review of the research regarding the communicative meanings of touch, [Hertenstein et al. \(2006b\)](#) present the existing evidence regarding the importance of touch for emotional communication, attachment, bonding, compliance, power, intimacy, hedonics, and liking. At a more detailed level, an observational study by [Jones and Yarbrough \(1985\)](#) attempted to classify occurrences of touch based on their communicative functions. They identified twelve “relatively unambiguous” meanings : support, appreciation, inclusion, sexual interest or intent, affection, playful affection, playful aggression, compliance, attention-getting, announcing a response, greetings, and departure. Beyond those, they also observed four additional ambiguous meanings: reference to appearance, instrumental ancillary, instrumental intrinsic, and accidental. It doesn’t stop there either as they further noticed that a single touch could very well have multiple different meanings, such as greetings/affection. While this study dates back to 1985, it is still, to our knowledge, one of the only works to have seriously tried to observe and to classify occurrences of touch by their communicative functions. The presence of hybrid and ambiguous categories questions whether the argued twelve unambiguous meanings really are without ambiguity, but it also serves to show how many varieties there can be in the meanings of touch and how difficult it is to actually formalize them. In their study regarding affective interpersonal touch [Fisher et al. \(1976\)](#) had already written on the matter that “while the message of touch has an inherent ambiguity, touching usually has a fairly clear intentionality. In effect, touch implies that a communication is intended, but the content of the communication may not be clear”.

Therefore in the same way that in terms of physical properties we can distinguish the temperature, the texture, the shape or the weight of an object that comes into contact with our skin, a social touch can also be interpreted in multiple ways at the same time in terms of affective and functional/communicative meanings.

2.1.2.2 The Meanings of Social Touch: social development, well-being and love

For mammal newborn babies, touch from their parents at birth is essential to activate the bodily functions that are not yet fully operational such as breathing or defecation ([Montagu \(1986\)](#), [Cascio et al. \(2019\)](#)). A lack of social touch in the early stages of life can have very serious consequences on the health of the baby, and can even, in extreme cases, lead to death ([Crandall \(1897\)](#)). Even well into adulthood, there is evidence that social touch has positive physiological effects by helping lower stress, blood pressure, perception of pain and heart rate ([Ditzen et al. \(2007\)](#), [Field \(2010\)](#), [Shaltout et al. \(2012\)](#)). On the other hand, from the psychological point of view, social touch has also been evidenced

2.1. THE SENSE OF TOUCH AND SOCIAL TOUCH AMONG HUMANS

as essential to the socio-psychological development of individuals. Touch deprivation in childhood has been correlated with stress, depression and a lack of social skills later in life (Floyd (2014), Carlson and Earls (1997)).

In a famous although ethically questionable study by Harlow (1958) on rhesus monkeys, babies having been separated from their mothers at birth were then presented with different surrogate “mothers”: one made of wire that also provided feeding, and one made of soft cloth that did not provide feeding. It was then observed how the babies were consistently reaching for the surrogate mother involved in the affective moments when given the choice or faced with stressful situations, instead of the feeding mother. The extent to which this happened led the researcher to even argue having identified the nature of love, in that comforting contact was eclipsing even nursing behaviours in terms of preference. This work was profoundly influential in the field of psychology where almost at the same time the notion of transitional objects was founded by Winnicott (2016). This concept involves the transfer of the mother-child bond into a surrogate object such as a toy or plush which helps the baby to learn to distinguish the boundaries between himself (his body, his desires and needs) and the outside world which does not always obey his command. Both of those works show how connected social touch is to the notions of affect, relational bonding, social development, and, ultimately, love.

2.1.2.3 The Affective Meanings of Social Touch

It is therefore no surprise that there is now abundant research in the field of social touch on the affective meanings of touch. In languages with Latin origins, touch is often semantically associated with emotion. We can say that we feel touched when we feel moved by an emotion and we can describe a work of art as touching. This linguistic intuition that touch is related to emotion has been more and more supported by the recent findings regarding the C-tactile afferents (or C-fibers). It was indeed recently discovered that some specific mechanoreceptors of the skin which are mostly found in the hairy parts of the body, seem to be responsible at the neurobiological level for the identification of social touches and their affective valence (McGlone et al. (2014), Löken et al. (2009)). Those CT-afferents react with the most strength to stroking gestures of a speed between 1-10 cm/s, which is also the range of speeds that human subjects have rated as the most pleasant. This suggests that we are biologically “wired”, so to say, for enjoying social touch.

Beyond the pleasure we can experience from such affective stroking, touch is also able to convey specific emotions to others. In an experiment where a participant was made unable to see nor hear while another participant tried to convey to them specific emotions via touch, it was observed that the recognition rates of the emotions were reaching above chance level even when the two participants didn't know each other (Hertenstein et al. (2009), Hertenstein et al. (2006a)). Further studies have confirmed those findings and given evidence that a closer relationship between the two people helped them to better

convey the same emotions as well as convey a greater variety of emotions (McIntyre et al. (2019), Thompson and Hampton (2011)).

Two other things were learned from those studies. First, it appears that specific types of touch gestures are more frequently used to convey specific types of emotions (App et al. (2011), Hertenstein et al. (2009), Bianchi-Berthouze and Tajadura-Jiménez (2014), Hauser et al. (2019)). The touch types/emotions conveyed associations evidenced in those works are summarized in Table 2.1. However it also becomes apparent, and that is the second thing we learned from those studies, that the inherent ambiguity regarding the actual message of a touch is at play here too. The principles of equipotentiality and equifinality are particularly visible in those associations of touch types with the emotion they are used to convey: one specific touch type can be used to express different emotions, and at the same time, a single emotion can be expressed through different touch types.

Emotion	Touch Types
Anger	Hit, Tap, Push, Shake, Squeeze
Fear	Contact, Lift, Press, Shake
Sympathy	Contact, Hug, Tap, Rub, Stroke
Sadness	Contact, Hug, Lift, Nuzzle, Rub, Stroke
Disgust	Contact, Kick, Push, Slap, Toss

Table 2.1 – Emotions and the touch types most commonly associated with them.

2.1.2.4 Interpretation, Acceptability and Effects of Social Touch

So then, the question becomes: how do we decisively interpret a touch in our daily interactions? Just as with other communicative modalities, multiple factors come into play when interpreting a touch. Touch meanings have a definite cultural component (Remland et al. (1995), Dibiasse and Gunnoe (2004)), they are influenced by the context they take place in and are also negotiated during the actual interaction (Meyer et al. (2017)).

Gender differences regarding touch uses and touch interpretation have been widely studied (Derlega et al. (1989), Dibiasse and Gunnoe (2004), Stier and Hall (1984), Martin and Anderson (1993), Sanderson and Jorgensen (1997), Major (1981)) and it appears that men are usually more likely to initiate touch towards women, especially at the beginning of a relationship, but are very averse to touch from other men. This men-men touch aversion is theorised as being related to a latent homophobia in some cultures (Derlega et al. (1989)), and it should be noted that on the other hand, women were found to have a similar acceptance of being touched by women and by men. However, another difference that was found is that for men, a touch coming from a woman was always positively received, while for women it depended on the level of the relationship they had with the person touching them (Bradac et al. (1984)). A study by Martin and Anderson (1993) showed however that those gender differences were significantly correlated with the psychological gender of the person, and not so much with the biological gender. This means

2.1. THE SENSE OF TOUCH AND SOCIAL TOUCH AMONG HUMANS

that all those differences are likely highly cultural and the fact that some of those studies date back to 20 or 30 years ago now should be taken into consideration: cultural biases may very well have changed in the elapsed time. A recent study by [Suvilehto et al. \(2015\)](#) on the topography of acceptable touches depending on the level of the relationship between the two interactors showed however that most of those observations still hold true: women are shown to be more open to being touched in terms of areas and relationships for both men and women touches, while men are averse to men-initiated touch. In all cases though, the interpretation of a touch, and here its acceptability for the recipient, depends on the type of relationship shared by the toucher and the touchee and the state of this relationship. A touch coming from a stranger will always be less acceptable than the same touch coming from a close one.

Regarding the way relationships influence the interpretation of touch, it should be noted that here relationship has a large meaning and can also be part of the specific context of interaction. A touch coming from a colleague at work is going to be interpreted entirely differently as it would be if it had come from a concurrent in a competition, from the boss of the place we work at, from a healthcare professional, or from a romantic partner.

Personality traits also play an important part in how we perceive and accept social touch. It was for example shown that certain attachment styles led to decreased perceived pleasantness when experiencing CT-optimal stroking ([Krahé et al. \(2018\)](#)) and, similarly, persons with autism have different perceptions towards social touch. But even beyond our cultural and physiological backgrounds we each have a singular relationship with touch. As there are very tactile persons who cannot refrain from touching people they interact with, there are also people that generally do not enjoy touch that much, or even experience it as unpleasant and annoying. To model these personal preferences regarding touch, the notion of “touch avoidance” or “touch receptivity” has been put forward, and multiple scales and questionnaires have been created to try and measure these individual preferences ([Webb and Peck \(2015\)](#), [Martin and Anderson \(1993\)](#), [Andersen and Leibowitz \(1978\)](#)).

Concurrently to the affective meanings of touch, we have seen how touch also can have more functional meanings, such as attention-getting, turn-taking or task-related meanings. Again, a single occurrence of a touch can both have an affective component and a functional meaning. This has made social touch important and interesting in some professions such as in the domain of healthcare. In those situations, touch can serve both as a tool of examination of the body, and as a way of conveying reassurance, comfort and/or securing compliance.

There is a wide literature regarding the “Midas touch” effect, i.e. the fact that social interpersonal touch significantly improves the perception we have of a person and increases the odds of the recipient of the touch complying with a request, even if they did not take explicit notice of the touch and don’t remember it ([Crusco and Wetzel \(1984\)](#), [Guéguen](#)

(2004), Kleinke (1977)). Not only that, but the recipient of the touch has also been shown to be more perseverant when the demand was to accomplish a task.

In summary, touch is a very important sense to human beings. With its discriminative properties, humans are able to perceive textures and learn to differentiate the outside world from themselves, and ultimately touch allows us to manipulate this outside world. Furthermore, between humans, social touch also plays an essential role in the development of our social skills and our well-being. Although the meaning of social touch can be modulated by culture, relationships and context, it is also constructed and negotiated directly during the interaction by the people involved, and it ultimately helps us bond with each other and convey our emotions as well as express many different communicative intentions.

However, in an increasingly digital and technological world where countless hours are spent on computers and smartphones, and where at times direct social contact is further hindered, such as it happened with the recent pandemic situation, social touch appears to be getting harder to come by. With the pandemic, we had first-hand experience of what this can mean for both our social lives and our health, as isolation often led to depression. What place is there for social touch in a world where we spend more time texting or having video conferences than being physically present with other people? How can we maintain an actual contact with other people even when distance or circumstances sets us apart? And in the future, what other new forms could social touch take to help us in our everyday lives?

2.2 Social Touch Technologies

As we have seen, maintaining social contact is a critical need for humans' well-being and social touch helps us connect, bond, and express emotions with others. It thus makes sense that we would want to implement social touch abilities in our communication tools, so that a phone call could be accompanied by social touch interactions for example.

Enabling users to exchange touch over the distance via a technological tool is called mediated social touch. To allow such exchanges to happen, mediated social touch technologies obviously need to both allow users to input their desired touch patterns and to be able to generate the haptic stimulus for the user. The challenge is thus two-fold: how can we perceive a human touch? And how can we reproduce touch with technology? And ultimately, once the technological challenge are overcome, will mediated touch have the same effects and uses as actual social touch?

From our perceptual point of view, touching is always also being touched. Upon putting our hand on the arm of someone we perceive the skin of the person, its texture and temperature, but we also produce, at the same time, a similar sensation for them. While we uniformly describe this as the sensation of touch, there are in fact two different functions performed simultaneously: we both feel the thing that we touch (its texture,

temperature, shape, weight, etc.) and deliver the haptic sensation of our touch to that thing (the texture, temperature, shape, weight, etc. of our own hand for example). While this might appear as extremely obvious and natural to human readers, those two functions will not be addressed by the same kinds of technologies.

It is therefore no wonder that the field of haptics research is often polarized in two categories: the perception of touch technologies (how to sense and interpret a touch) and the haptic feedback generation (producing a touching sensation on a human via technology).

2.2.1 Social Touch Generation

While there is more and more research in the field of haptic generation, there is still no perfect way to imitate human touch with technology. In its most basic use, touch allows us to feel when we come into contact with something outside of us that prevents us from continuing our movement. In essence, touch is the sense that allows us to perceive the things that resist us. However, as we have seen in Section 2.1.1, we can then further discriminate between the perception of textures, vibrations, weight, proprioception, temperature, etc. Our sense of touch is the combination of multiple extremely complex perceptive systems and therefore simulating a complete haptic sensation via technology is a huge challenge.

Maybe the most obvious way to simulate a touch is thus to use actual tangible objects. [Eichhorn et al. \(2008\)](#) showcase a small arm covered with textile activated by a servomotor that can be commanded over the distance. Similarly [Teyssier et al. \(2018\)](#) propose small mechatronic fingers that can be plugged on a smartphone to deliver stroking gestures on the wrist of the user based on the inputs of the other user. In the works by [Bonanni et al. \(2006\)](#) an air bladder is inflated to generate a squeezing sensation and in a research work by [Suhonen et al. \(2012\)](#) two shape memory alloy (SMA) MigaOne actuators are worn on the arm and are able to tighten themselves around it. Another popular research avenue is that of robotic devices, such as the robotic arm used by [Teyssier et al. \(2020\)](#), which can simulate human touching gestures. While those kinds of devices manage to actually deliver a natural sense of touch, they are often limited by their materiality: you have to be holding or wearing them already to be able to use them, they might take time to activate (air bladder) and are somewhat limited both in the ways you can equip them on a person (a robotic arm is quite massive) and in the kinds of touch they can simulate (SMAs can only squeeze).

A somewhat similar type of technology that can be used to simulate a sensation of resistance is force-feedback, which essentially consists in opposing some amount of force against the user's body in order to make it seem like there is an obstacle. This is described in research papers as early as in 1996 with [Hashimoto and Manoratkul \(1996\)](#) in the work of which two users could exchange handshake through the distance thanks to a force-feedback device. Other studies such as the ones by [Bailenson et al. \(2007\)](#) and [Brave et al.](#)



Figure 2.1 – Examples of tangible haptic devices (Mobilimb on the left, and a robotic arm on the right, both by Teyssier et al.)

(2001) make use of force-feedback joysticks and have shown how users can indeed convey understandable emotions using those kind of haptic signals. More widely, force-feedback devices can take the form of mechanical arms, be used in video game controllers or even take a wearable form with the recent SenseGlove¹ or HaptX². With the glove form, force-feedback is applied to the fingers, preventing them from closing the hand. This produces the sensation of holding something with potential variations of texture (soft or hard) and size. While this technology can lead to an impression of contact and of textures and weight, its downside is also that it usually has to be held or at least put in a stable position and is difficult to make into a wearable device (the gloves are big in size, often wired (in their current state) and relatively heavy).

When it comes to thermal stimulation, the main technology used is the Peltier component, which is able to produce both heat and cold. This is notably used in the works by Bonanni et al. (2006) and Suhonen et al. (2012). Those studies showed how thermal stimuli can indeed be used by users to convey emotions and meaning, with cold being more attributed to negative emotions and heat to positive ones. However, it was also shown that temperature is difficult to use, as it takes time to cool down or heat up (making it unsuitable for instances that would require fast communication) and it needs to be calibrated at each use to take the current body temperature of the person using it into account. Changes in temperature are indeed only perceptible with a difference of 3-4 degrees Celsius compared to the current body temperature baseline, which is different person-to-person and even for an individual is subject to change during the day. Furthermore, too high differences in temperature cannot be used as it could present risks for the user and it can induce sweating which is not always appreciated.

Currently however, the haptic technology most represented in the literature is vibrotactile technology. This might seem strange because we all know vibration motors for their presence in our smartphones, and it doesn't actually feel like a natural touch stimulation.

¹<https://www.senseglove.com/>

²<https://haptx.com/>

2.2. SOCIAL TOUCH TECHNOLOGIES

Despite this unnatural sensation of vibrations, vibrotactile technology has a lot of advantages that make it a very convenient way to allow people to exchange touch. It can be used to send “touch emojis” (Wilson and Brewster (2017)), although combined to other types of haptic feedback ; it is made as a vibrotactile sleeve "TaSST" in the work by Huisman et al. (2013) and perceptual illusions of continuous movement are used to reproduce diverse types of touch (hit, stroking, pat, etc) and their intensities ; in another example of a sleeve, Rantala et al. (2013) showed that emotions could still be transmitted with vibrotactile stimulation; multiple studies (Haans et al. (2008), Haans et al. (2014)) have even shown how effects such as the “Midas Touch” effect can also be experienced through mediated touch communication with vibrotactile stimulation. The fact that vibrotactile stimulation is easy to set up (small components that can easily be put together and prototyped even with low initial knowledge of electronics), easy to be put in a wearable form, cheap, and can be used and understood similarly to natural social touch have made it a popular and somewhat reliable option for mediated social touch. Its biggest drawback is still that it doesn't feel like a natural touch however, and that can be off putting to people.

An interesting take on haptic stimuli generation is that of the Tactos, Intertact and Dialtact systems developed by Lenay (2019) and colleagues, which make use of braille cells to enable minimalist but meaningful mediated touch interactions.



Figure 2.2 – The SenseGlove Nova, a vibrotactile sleeve (TaSST, by Huisman et al.) and the Dialtact system (Lenay et al.)

An interesting take on haptics in virtual environments is that of pseudo-haptics, such as it is defined by Pusch and Lécuyer (2011). No specific devices are needed with pseudo-haptics, as no actual haptic stimulus is produced. Instead, by leveraging specific multi-modal cues (mainly audio and visual) we can generate an illusion of haptic feedback in the user. An exemple would be that of having the impression that there is a resistance when our hand is visually slowed when passing through a specific substance in VR. Researches in the field of pseudo-haptics are becoming more and more popular (Ustek et al. (2018), Gomez Jauregui et al. (2014), Haruna et al. (2020)), with very promising results. There are however very little research on social touch applications of pseudo-haptics at the moment, as those works are more focused on object manipulation and environment exploration.

There are other ways to produce haptic stimuli, for example by using pneumatic or air jet technologies, but in the end the ones discussed here are, to our current knowledge,

the most popular and effective ones. None of those technologies produce a sensation that could be mistaken for an actual human touch, and combining them is for now quite difficult considering their respective drawbacks. Nevertheless, it was shown by [van Erp and Toet \(2015\)](#) that we could still expect people to be able to understand mediated social touch and to use it to produce affective meanings in the same way that they do in human-human direct interaction.

Having seen what the options are regarding haptic sensations production, the question then becomes how to give meaningful input interfaces to the users. Very often, to allow people to use their touching habits, this will take the form of a sensing device aimed at perceiving human touch.

2.2.2 Social Touch Perception

To enable the perception of touch, the most popular strategy is to use some kind of tangible textiles that will act as an artificial skin [Silvera-Tawil et al. \(2014\)](#). There are many ways to build those kinds of artificial skins. A very simple example is presented in the work by [Eichhorn et al. \(2008\)](#) where a potentiometer is directly covered by textile and used as an input device to record a touching gesture. On the other hand, a much more sophisticated design is presented by [Stiehl and Breazeal \(2005\)](#) with a sensitive skin prototype able to determine four modalities of haptic information: pain, temperature, contact and kinesthetic information (pressure). To perceive that information, electric field sensors and Quantum Tunneling Composite sensors are used to determine when contact happens and with which strength at the impact. Temperature is measured with thermistors and the kinesthetic information with potentiometers. Pain is modeled as any of those recorded signals being higher than a certain threshold. In the work by [Cang et al. \(2015\)](#) pressure-location sensors are used to build a low-cost multitouch fabric and touch recognition is achieved with good rates of recognition of touch types. Furthermore, it is shown that machine-learning algorithms are able to reach even better rates of recognition when trained on single individuals, which highlights the possibility of identifying the toucher based on its touch stimuli. Similarly, [Silvera-Tawil et al. \(2014\)](#) showcase another recognition of touch types algorithm with good results using a stretchable sensitive skin with electrical impedance tomography to record the touch. In their work on the vibrotactile sleeve TaSST, [Huisman et al. \(2013\)](#) also proposed an input layer to their device, allowing users to prepare touch types to send to another person equipped with a TaSST. The input layer was composed of a grid of 12 Lycra pads, filled with conductive wool so that when compressed, the resistance of the wool would change, making it an affordable and flexible sort of force-sensitive reflector.

As an alternative to tangible sensing interfaces, we can note the works of [Nguyen and Wachsmuth \(2009\)](#) and of [Sykownik and Masuch \(2020\)](#) which present virtual avatars equipped with virtual skin receptors that can determine when a contact happens. While

this approach allows to detect a touch and some of its properties (duration, velocity, etc.), the drawback is that perception of pressure is difficult to achieve. Nevertheless, thanks to this system users can then use mediated touch in shared virtual environment.

2.3 Simulated Social Touch: Human-Agent Interactions

Touch is thus an essential modality to social communication, and efforts are made towards the enhancement of our communication technologies with touch perception and generation abilities, thus enabling mediated social touch.

Nowadays though, new kinds of social entities are beginning to appear that we can interact with. Those are the artificial agents: entities made able to act in some capacity, often thanks to a program or algorithm of varying complexity. While their most basic forms may only take the form of an algorithm producing a single automated response to some specific event, there are now more and more kinds of “intelligent” or autonomous agents, equipped with artificial intelligence algorithms that make them able to dynamically interact (in specific domains) with humans and/or other agents. With their varying levels of autonomy and abilities those artificial agents can take many forms and be found in many different kinds of context : it ranges from the most basic chatbot that automatically refers you to the frequently asked questions of a website to the most advanced robots or virtual characters.

2.3.1 Embodied Conversational Agents and their social abilities

Embodied conversational agents (ECAs) are among the most suited kind of artificial agents for social interactions. According to their name, ECAs are:

- *Embodied*, meaning that they have some form of a visible body, either complete (as a character in a video game) or partial (an animated head for example). This distinguishes them from agents that only perform verbal behaviour (vocal assistants for example);
- *Conversational*, thus social and able to engage in conversations with one another or with humans, by the mean of language or multi-modal attitude, thanks to their embodied nature (Natural Language Processing, non-verbal behaviours, etc...);
- *Agent*, meaning that they are equipped with rationality, proactivity, autonomy functionalities. This distinguishes them from avatars, which are virtual bodies remotely controlled by a real human.

Thus, while traditional conversational agents are already able to socially interact with other agents or humans through conversation (think about a chatbot or the virtual assistant Samantha in the movie “Her”, only represented by her voice), be it written or spoken,

ECAs have the extra ability to make use of non-verbal communication: gestures, facial expressions, gaze, body movement in general (Lee and Marsella (2006)). It should be noted that while this term is usually reserved for virtual agents that we encounter in the digital worlds, their definition does not strictly exclude robots, whose fundamental physicality make embodied by nature. As the differences between social robots and ECAs will be interesting to discuss later, we will only use the term of ECAs to talk about virtual agents.

In affective computing studies (Picard (1997)) on human-agent interactions, multiple behaviour models have been designed to allow such artificial social agents to gain the human's trust, to bond with them and be able to build some level of relationship even in short-term interactions. Those kinds of models are usually based on the psychological and sociological theories of human-human social interactions. Thus while simple conversational agents are able to use verbal strategies of bonding or emotional communication, embodied conversational agents have the ability to use multimodality to complement those relational strategies with adequate non-verbal behaviours (Pecune (2013) Bickmore and Cassell (2001)).

That being said, the very idea of forming and developing an actual relationship with an artificial agent might seem a bit presumptuous considering the fact that human-agent interactions are usually studied on a short-term basis (one-time interactions). Another social construct that has more meaning for a short interaction with a virtual agent is: rapport. According to Tickle-Degnen and Rosenthal (1990) rapport is the sensation of "mutual attentiveness, positivity and coordination" in an interaction between two people, with attentiveness and positivity being especially important in the first interactions, and coordination becoming more important when a bond has already been created.

There is a rich literature on the non-verbal correlates of rapport and how non-verbal communication helps to build it, not only in the field of human-human social interactions (Tickle-Degnen and Rosenthal (1990) LaFrance (1979)), but also in the field of human-agent interactions (Gratch et al. (2007a) Gratch et al. (2007b) Gratch et al. (2013) Gratch et al. (2006) Cassell and Thorisson (1999) Cassell et al. (2007)). Those studies show how rapport can be built based on appropriate mimicry of body movements and expression of attention through posture and gestures, which evokes behavioural alignment as described by Garrod and Pickering (2009) for example. But rapport also involves verbal behaviour and strategies (Bickmore and Cassell (2001)) among which small talk is an important component (Bickmore and Cassell (1999)). Rapport thus involves elements from situational and inter-cognitive alignment as well, and computational models of rapport have explored all of those dimensions of rapport: non-verbal mimicry and its timing (Madaio et al. (2017), Zhao et al. (2016)), backchannels and feedback gestures (Cassell and Thorisson (1999), Gratch et al. (2007b), Gratch et al. (2007a)) or conversational strategies (Bickmore and Cassell (2001), Bickmore and Cassell (1999), Kang and Gratch (2011), Lucas et al. (2017)).

With this ability of building rapport and the rest of the social capabilities of those kinds of agents, it was shown that humans actually respond naturally to them, perceive them more positively, and even adopt self-disclosure behaviours during the conversation [Lucas et al. \(2017\)](#) [Gratch et al. \(2007b\)](#) [Gratch et al. \(2007a\)](#). This gives them a lot of possibilities in terms of practical applications.

In the field of healthcare first [Lucas et al. \(2017\)](#) showed that virtual agents allow humans to report on their potential mental health problem more freely than they do with human listeners (less fear of stigmatisation). Similarly, [Breazeal \(2011\)](#) pointed out how the emotional and social abilities of those kinds of social robots make them particularly well suited for interacting with patients in a compassionate way, which is essential for improving their well-being.

When it comes to task-oriented cooperation as well, ECAs with rapport-building abilities have been shown by [Cassell et al. \(2020\)](#) to be able to successfully help children in literacy learning and to overall perform well in pedagogical situations. And while virtual assistants such as Siri and Alexa are already more and more present in households, they still lack this social intelligence. In a work by [Pecune et al. \(2019\)](#) regarding such a virtual assistant providing movie recommendations, design guidelines were discussed regarding how such social abilities would make those agents better at explaining their choices. It should especially be noted that granting social and emotional abilities to agents could also help regarding the problem of the explainability of algorithms decisions, which is an essential question with the increasing popularity of machine-learning decision making in all the spheres of society.

Despite all those social abilities of ECAs though, one modality today appears to still be widely lacking from their skillsets: touch. As reviewed in Section 2.1 touch is a fundamental component of non-verbal communication in human interaction and critical to human well-being. We argue that bringing the tactile modality to human-agents social interactions should therefore be a priority and would help them in bonding and building rapport, and in expressing and understanding emotions.

2.3.2 Human-Agent Tactile Interactions

As it was discussed in Section 2.2, the field of haptics is technologically divided between technologies aimed at the perception of human touch and technologies aimed at the production of haptic stimuli. Obviously, those two types of technologies, which can be used for mediated social touch, can also be used for what we would call simulated social touch: the use of social touch by automated systems. Furthermore, both perception and generation technologies are needed to enable human-agent bidirectional touch in a complete interactive loop. In order to autonomously interact with a human, an agent indeed traditionally needs perception abilities (a way to detect the environment and the human's actions, including touch here), reasoning abilities (to interpret its perceptions and to de-

cide how to react accordingly) and display abilities (a way to realize its decided behaviour in a form that the human can also perceive). At the moment however, the field of simulated social touch features very few research works showcasing complete interactive loops with bidirectional touch. Most simulated social touch works focus on either the perception part of the interaction or the touch generation part of it.

2.3.2.1 Human-Robot Tactile Interactions

Interestingly, the most represented kind of agents in simulated social touch is not the virtual agents but the robots. When it comes to the perception of human touch indeed, we saw in Section 2.2.2 that the most popular options are tangible artificial skins. This type of technology fits robots particularly well as they have physical bodies that can be covered with those types of devices. It is especially well suited to animal-inspired robots where textile-based artificial skin can be used in fur-like designs. Pet-inspired robots have the advantage of not being expected to have complex social behaviours (as pets are not able to talk anyway, for example). Those kinds of robots are notably developed to be used in health-care domains such as hospitals or nursing homes, where it was shown that robotic pets such as the PARO seal have real therapeutic effects in helping alleviate the feeling of isolation and providing social and affective touch [Petersen et al. \(2016\)](#). While in their most basic form, those robotic pets may look like a child toy plushie and not much more, efforts are made to endow them with more and more social and haptic abilities. The “Haptic creature” developed by [Yohanan and MacLean \(2012\)](#) and [Sefidgar et al. \(2016\)](#) for example, is a robot covered with fur designed to evoke a cat-sized animal and equipped with a system allowing it to simulate “breathing” by producing up-and-down movements of the body (like a chest going up and down when breathing). This was shown to have a significant calming effect on humans, even more so than the same robot not equipped with the breathing simulation ([Sefidgar et al. \(2016\)](#)). In a study regarding how users would like to communicate with the Haptic creature and how they would expect it to react to their touching behaviours, the researchers elaborated from the participants’ responses five categories of intention/emotional meanings: protective, comforting, restful, affectionate, and playful touches ([Yohanan and MacLean \(2012\)](#)). They further identified that the participants expressed that they would expect and like the Haptic creature to reciprocate the emotion communicated by their touch. Another example of such robotic pet is the Cuddlebot ([Allen et al. \(2015\)](#)) which is also cat-sized and able to “breath” but further equipped with full-body sensitive fabric and made able to move its head, arch its back and purr. The sensitive fabric is used with classification algorithms to enable the recognition of the communicative intent based on the way the robot is touched, which would make the Cuddlebot able to adopt an adequate behaviour in response. The specifics of the behavioural decision model that would make the robot autonomous and enable an interactive loop (the robot is touched, evaluates the touch, and



Figure 2.3 – The PARO robotic seal (left), and the Haptic Creature (right, Yohanan et al.)

answers back, and repeat) is however not presented nor discussed in the paper. Similarly [Stiehl and Breazeal \(2005\)](#) and [Knight et al. \(2009\)](#) works feature plush bears robots equipped with sensitive fur and neural network algorithms able to classify the touches in classes of intentions and affective meaning (e.g.: punishment light, tease pleasant, etc.). Although among those the Huggable is described as able to nuzzle, to orient itself towards humans for touch, and to perform hug movements, no decision algorithm regarding when to perform those actions is described.

Outside of those pet-inspired robots, we can also find research aimed at social robots having more anthropomorphic features. A robotic arm is used by [Teyssier et al. \(2020\)](#) and was shown to be able to convey emotional meanings to human participants. In the works by [Block and Kuchenbecker \(2019\)](#) and [Block and Kuchenbecker \(2018\)](#) a robot with a humanoid shape is used to investigate how users react to simulated hugs. It was shown that participants could appreciate the hugs coming from the robot, but that they prefer soft and warm hugs from squeezing robots equipped with textile and thermal output, as opposed to hugs coming from a non-covered metallic robot. They also reported preferring hugs that end as soon as they present a withdrawal movement, illustrating the importance of the robot being aware of the human's touch and able to interpret and react to it. This last point also raises ethical questions regarding consent and acceptability in human-agent social touch interactions, which we will come back to later on. Finally, in a study by [Fukuda et al. \(2012\)](#), participants played an ultimatum game with a robot that made them unfair offers. It appeared that when the robot touched the participant upon making its offer, the offer was perceived as less unfair, which is consistent with the observations regarding the "midas touch" effect among humans.

2.3.2.2 Human-ECA Tactile Interactions

As we have seen in Section 2.3.1 however, robots are not the only kinds of agents that we can interact with. The virtual nature of ECAs however leads to less works focusing on the question of the perception of touch, and much more focusing on the generation of touch. In the work by [Bickmore et al. \(2010\)](#) a hybrid ECA was made of a graphical representation animated on a computer screen attached to a robotic-looking body. It was further equipped with an air-bladder haptic device able to squeeze the hand of a user. The hybrid ECA was employed in a context of empathic communication during which the participants who identified as comfortable with touch reported better ratings of their estimation of the relationship level between them and the agent when it touched them. This result was however not observed with participants having low scores on comfortability with touch. This again shows that people seem to react to interactions with social agents as they do with humans in everyday life. It also showcases that the personal touch avoidance of people should be taken into account by agents with social touch abilities.

Regarding the affective ratings of simulated haptic stimuli, [Huisman et al. \(2016\)](#) showed that the pleasantness of simulated stroking with a vibrotactile sleeve is rated consistently with research on actual stroking pleasantness with the CT-optimal speed being the most pleasant. Going further in the direction of immersive interactions, [Huisman et al. \(2014b\)](#) and [Huisman et al. \(2014c\)](#) used Augmented Reality (AR) for touching ECAs where the agents were seen through a tablet. The haptic sensation was generated with a vibrotactile sleeve whose activation was synchronized with the movements of the agents. In those works the focus is on the effects of simulated haptic stimuli and the touch sensing and recognition is not discussed, but again, none of those works feature a complete interactive loop: the agent cannot dynamically and autonomously react to the variety of the human's behaviours. Instead, the participants played an AR video game with or against an ECA that could or could not touch them, based on controlled conditions. In the cooperative setting, the agent was perceived more positively when it touched the participant, which is consistent with the "midas touch" effect observed in human-human social touch. Unexpectedly there were no significant differences in how the cooperative and the competitive agents were appreciated, but again, in all cases, the agents that touched participants were perceived as warmer.

With applications for the training of medical students, a "mixed reality human" concept that involved social touch interactions was studied by [Kotranza and Lok \(2008\)](#), [Kotranza et al. \(2009b\)](#) and [Kotranza et al. \(2009a\)](#). In those studies, a dummy is used as a tangible interface and equipped with a mechanical arm while the student practices his medical examination with a HMD that allows him to see the animated virtual human. The student can practice his medical touches for the examination and the virtual human is programmed to use the mechanical arm in a way that evokes fear about the examination. The students were observed using social touch in the same way they would with actual human

patients and with the same communicative intentions. The conditions of the experiment involved both pseudo-haptics (only the visual cue of the virtual human arm moving) and active-haptics sensations, and both were able to convey understandable affective meaning to participants, although active haptics appeared a little bit more effective. Finally, the active-haptics touch helped the participants feel like the virtual human was an actual patient and they rated the experience as realistic, believable and pedagogically beneficial. Another study by [Hoppe et al. \(2020\)](#) showed that participants rated ECAs that touched them in virtual reality (VR) as more human-like. Finally, [Zhou et al. \(2020\)](#) showcased a VR adventure in which an ECA was taking the form of a forearm-mounted robot with an expressive face. This virtual companion was synchronized with the movement of an actual forearm-mounted haptic robotic device consisting of four robotic actuators driven by servo motors, which rendered specific tactile patterns to communicate primary emotions (fear with squeezing and trembling, happiness with rapid poking, etc.). While participants were easily able to recognize those emotions and felt that it improved their experience, especially regarding the virtual companion, all the interactions were scripted and, again, no actual interactive loop was happening. Furthermore, the visual and audio cues synchronized with the haptic feedback most certainly helped disambiguate the meanings of the touches.

Regarding the few works on perception of a human touch involving ECAs, we can note the work by [Nguyen et al. \(2007\)](#) on endowing an ECA with a virtual sense of touch in an immersive room. To do so, they used a principle extremely similar to the one described by [Sykownik and Masuch \(2020\)](#) for virtual avatars: they covered the agent's body with shapes, acting as virtual receptors, able to detect when there is an overlap between the coordinates of an object and their own. Following up on this work, [Nguyen and Wachsmuth \(2011\)](#) showed how such body-conscious ECAs could be automatically trained to learn their boundaries and those of all things and agents present in their environment. This enabled them to take into account proxemics and understand who can reach what in the environment and how to coordinate with others in terms of spatial cooperation. While knowing how to reach another is a prerequisite to dynamic touch interactions, further work in this direction was not discussed by the authors. In the work by [Bailenson and Yee \(2008\)](#) on the other hand, participants were asked to touch different kinds of objects and characters in VR with HMDs and it was observed that social norms seemed to be respected: they touched the virtual characters with less force than they did non-human objects and they used even less force when touching their faces.

How we can see, virtual reality is a field where more and more works regarding social touch with ECAs are appearing. It should be noted however that while haptics are an increasingly popular research domain in the field of VR, most of those research works are focusing on the question of the manipulation and interaction with the immersive environment and the objects we can find inside it, and research on haptic interaction with social ECAs is still very scarce at the moment. This is evidenced by the recently published sur-

veys by Bouzbib et al. (2021) and Dangxiao et al. (2019) on the uses of haptics in virtual reality, which barely mention social interactions with ECAs. Despite this, what ECAs lack in physicality compared to robots, they make up for by being easier to control and animate in conjunction with their immersive environments, which can also be entirely controlled. Their downsides however is that they will often require more substantial haptic equipment to make them tangible to the user. This can end up being a challenge, because while some heavy pieces of equipment can be set up for haptic interactions dedicated to static parts of the environment, it is far more difficult to set them up for dynamic interaction with a character that can move around in the immersive environment and is human- or animal-sized. For example, for now it is still challenging to use tangible interfaces to represent moving characters (or even parts of moving characters) in VR outside of the controllers used for the navigation and interaction with the world. Heavy or static equipment like robotic arms or wired devices are thus particularly difficult to use in VR, as the user is probably also going to move around a lot. In this regard, the pseudo-haptics researches mentioned in Section 2.2.1 represent a great hope for the future of social interactions with ECAs in immersive environment. Notwithstanding, bringing wearable haptic devices in immersive environments is still a perfectly valid option that makes ECAs overall quite even in terms of touching abilities compared to their robotic counterparts.

2.3.3 Position of our work in the simulated social touch literature

What did we learn from this review of the existing literature? Overall, studies involving social touch interactions with ECAs and social robots almost always focus on one of the directions of touch: either the sensing of touch or the generation of touch (and its effects). Bi-directional touch is, for now, almost unrepresented. It was however shown that endowing ECAs with social touch abilities seem to indeed lead to them being perceived more positively by the humans they interact with and allow them to have positive effects on their well-being, as seen with the robotic pets in nursing homes and hospitals. Emotions can be conveyed to humans by such agents using social touch and, at the same time, humans seem to interact with those ECAs with touching abilities in a way consistent with the way they interact in human-human social touch interactions. Applications of simulated social touch technologies could thus positively impact domains such as training and task-related cooperation, healthcare, research or entertainment. These findings are consistent with those of other reviews of the literature on human-computer social touch interactions by van Erp and Toet (2015) and Erp and Toet (2013).

While both sensing and haptic stimulation technologies are still young and far from being able to produce sensations of natural touch or perfect interpretations of touch types and intents, we feel like the next step in the research on social touch with ECAs is to build an actual interactive loop with touch. As it is underlined by Huisman et al. (2014a), the classic structure of an autonomous agent involves three main components: an ability to

perceive the world and the actions happening inside it; an ability to reason about what was perceived, interpret it, and then decide how to behave; an ability to manifest the chosen behaviour in a way that is perceptible by the other agents inside the world. Those three abilities are what makes a complete interactive loop, in which an agent and a human (or another agent) can autonomously interact with each other by adapting to their respective behaviours during a certain amount of time, possible. We believe that allowing an agent to both sense touch and produce touch, and most importantly to autonomously reason about when and how to touch, would be an essential and interesting contribution to the literature. This should further improve the quality of the interactions we can have with those robotic and virtual companions and it should help them achieve further emotional communication and rapport-building.

While we have seen here how the sensing and the production of haptic stimuli can be achieved technologically, we still need to figure out what kind of decision model would be best suited for such a touching ECA.

2.4 Computational Models of Emotion

As it was discussed in Section 2.1.2, there is a strong connection between touch and emotion. Touch is a good channel of emotional communication, and the latin languages have showcased it directly in the way they are spoken. In English we can talk about a “touching experience” or a “touching movie”, expressing the way we felt emotionally moved by something. We argue that building an ECA with touching abilities should make it able to be truly that: touching both in the physical sense, and in the emotional sense of the word. A touching agent would thus be able to take affects into account when interacting with others, and we have seen how this plays a big part in relational bonding. But what kind of decision model would be suited to emotional reasoning? How can we integrate this importance of the emotional side of an interaction directly in the reasoning process of the agent and have it influence both how it perceives events and how it acts?

2.4.1 Theoretical Models of Emotion

If at first computer science mostly developed systems able to reason about world states, ontologies and objective processes, it didn't take that long before the field of affective computing was formed by Picard (1997). In affective computing, the focus is put on the importance of emotions regarding decision making processes. And indeed, even for human beings there are strong arguments in the fields of philosophy, cognitive science and psychology that affects play an important role in our cognition. Emotions allow us to be more efficient in some situations by rapidly shifting the priority of certain goals over other ones and can thus influence our decisions. Some argue for example that the triggering of a strong emotion such as fear or anger allows the body to rapidly modify its physiological

state and enter a state of readiness for specific actions: the adrenaline rush that will allow us to suddenly sprint to escape danger or make use of our strength to defend ourselves (Scherer (2000), Izard (2010)).

Implementing a decision process able to take emotions into account therefore necessarily requires some basic understanding of what human emotions are and how they function. But the study of emotions is difficult, and although they have been discussed for centuries now, their precise definition and the question of how emotions come to be and interact with cognition are still uncertain (Izard (2010)). This section will summarize the most important theoretical paradigms regarding definition and generation of human emotion, by studying psychology and cognitive science literature.

Psychological theories regarding emotions are many (Scherer (2000), Izard (2010)) but can generally be classified in three categories: theories that look at emotions as discrete entities, dimensional theories of emotions and appraisal theories. What differentiates those paradigms is how they define and consequently generate emotions.

For discrete theories of emotions, emotions are approached through specific labels such as “anger”, “joy”, “fear”, etc., which are considered to reference specific, identifiable phenomena on the physiological and expressive levels (facial expressions, state of the body,...) (Scherer (2000)). Rules are then observed and deducted to determine how the emotion is triggered. The main critics regarding those discrete theories is that they hide the complexity of emotion as well as the influence of personality and culture on the experience of emotions. As of today, there is still no definite proof that what we usually call “anger” is a specific phenomenon, innate and uniform. Emotion is still a blurry notion that seems composed of many different layers (what are mood, sentiment or affect compared to emotion?).

Dimensional theories of emotions instead consider that emotions can be represented in a multi-dimensional space depending on their properties. Maybe the most common representation is that of the Pleasure-Arousal-Dominance space (Mehrabian and Russell (1974)) which builds a map of emotions based on their coordinates on the scales of pleasure (or valence, how positive it is), arousal (or activity, agitation) and dominance. This dimensional model takes into account the complexity of emotions better than the discrete theories and works well for a number of emotions, but ultimately the model relies on the dimensions that were chosen to be considered (here, pleasure-arousal-dominance). For example, dominance is not necessarily as useful to represent some types of emotions, which will make them harder to represent as convincingly in the model. Furthermore, cultural or personal differences are still difficult to take into account with such dimensional models.

This is the reason why, at the moment, the main paradigm appears to be that of appraisal theories of emotion. According to appraisal theories, emotions are generated in response to the appraisal (or cognitive evaluation) of an event based on the culture, the goals and the personality of the person. The individual is going to evaluate how this event

might influence his goals (will it help the fulfillment of his goals?) and how this event scores on certain variables (for example, valence, or desirability for others). As an example, upon coming face-to-face with a bear while hiking in the mountains, we are likely to appraise this event as not desirable, as going against our goal of staying alive, and as quite unexpected. With low scores on those three appraisal variables, emotions (probably fear here) are going to be produced accordingly, with varying intensities depending on the exact scores. The number and nature of the appraisal variables that are being used in such appraisal models can be different depending on the authors (although valence and goal conduciveness are almost always present), but the fundamental principle of appraisal is always the same. Those theories are able to model a lot of different complex emotions and depict the link between cognition and emotions better than other theories, with emotions being produced through a quasi-cognitive process. They however also consequently tend to approach the human mind in a computational way, which might be somewhat biased by our current society and technological achievements. This is nonetheless what makes the strength of those appraisal theories, especially when it comes to building a computational model of emotion.

2.4.2 Computational implementation of models of emotion

Taking inspiration from the theoretical models of emotion, affective computing has been putting efforts into building computational models of emotions (CMEs) that would allow artificial agents to have an emotional state and be able to take emotions into account. As there are many psychological theories of emotions, there are also many different kinds of CMEs (Kowalczyk and Czubenko (2016) , Ong et al. (2019)). As expressed by McDuff and Czerwinski (2018), there is great hope that such emotionally sentient systems will be able to reach even higher levels of performance in complex tasks, to make better decisions regarding human-machine interactions and to generally offer better services to humans. For embodied conversational agents specifically, those CMEs would be the missing piece that will allow them to autonomously select their expressive abilities in a given situation. With the CMEs that have already been built and tested, it was shown that those emotional abilities help them both develop their relational abilities with humans (Paiva et al. (2004)) and behave in a human-like manner (Gratch and Marsella (2005)). It was even shown that CMEs could have beneficial effects on the use of VR in psychotherapy (Hudlicka (2005)). We thus further believe that such CMEs would be a critical addition for a truly touching agent.

What are the currently existing computational models of emotions and what are the differences between them? First, we can note that, to our knowledge, there is no CME based on the theories of discrete emotions, and that the appraisal theories are at the basis of most CMEs. A notable exception is ALMA (Gebhard (2005)) an open-source dimensional model of emotion which allows the creation of rules mapping an event to the

dimensions of the emotion. Furthermore, ALMA's emotional abilities have been evaluated as plausible by human participants (Gebhard and Kipp (2006)). The number of emotions that ALMA can represent is limited however, and the software is built in a way that does not allow further modules/functionalities to be developed and integrated easily, if required.

On the side of appraisal theories, many different models exist (see Castellanos et al. (2018) for a recent review) with different design choices regarding the implementation of the appraisal process (use of fuzzy sets or not, choice of a specific psychological theory over another,...). Two examples that seem particularly interesting are EMA (Marsella and Gratch (2009)) and FATiMA (Dias et al. (2014)). EMA is a very complete and powerful CME, based on the appraisal theory of Lazarus (Lazarus (1991)), and able to represent a lot of different emotions and complex emotional scenarios. It has also been extensively tested and evaluated (Gratch and Marsella (2005)). Its complexity is however also one of its drawbacks, as it requires a lot of work to model a given scenario. It is also not openly available at the moment. FATiMA on the other hand is based on a different appraisal theory (the OCC model of emotion by Ortony et al. (1988)) and much less powerful in terms of the number of emotions it can represent. It is however also much less complex to use, and it was specifically designed to be modular and openly accessible and modifiable as a toolkit, even for people with no prior knowledge of emotional theories. It still features all of the basic functions of such a model, with an autobiographical memory for the agents and the ability to define both the appraisal and decision rules. It was also already used and tested in terms of believability in multiple kinds of scenarios (Guimarães et al. (2019)).

Recently, new kinds of CMEs have also started to appear which are not as based on theoretical models. Those are machine-learning based CMEs, which either use neural networks (Hieida et al. (2018)) or reinforcement learning (Moerland et al. (2018)) and which have the usual advantages over rules-based models of being more adaptable as long as they have more data available. On the other hand, they lack the explainability that rules-based models have, in that they act as black boxes of sorts regarding their decision process.

CMEs thus appear as powerful tools to build emotional agents able to reason about their own emotional states as well as more sensitive to their interlocutors' emotional state. This may help them emotionally connect with humans and thus further their relational abilities and make them easier to interact with. The multiplicity of implementations of those CMEs however, while being an indicator of the health of the field, also somewhat hinders efforts towards the improvement of those models. As it is discussed in (Reisenzein et al. (2013)), efforts should be made to make CMEs openly available for other researchers to help their improvements, instead of always starting back from square one. We argue that a touching agent will highly benefit from such emotional abilities and that working with an already existing CME would both accelerate the development of the agent and participate in the overall improvement of the CMEs' field.

Designing Human-Agent social touch interactions in VR

Contents

3.1 Requirements for a Touching Agent	31
3.2 Architecture of a Virtual Touching Agent	33
3.3 Choosing an adequate VR setup	35

IN this Chapter we discuss our design choices regarding the development of a Touching Agent, both able to use haptic feedback in communication and to touch humans in the emotional meaning of the word. Chapter 2 evidenced how bi-directional interactive loops with social touch are still extremely rare in the human-agent social interactions literature. We will thus present our efforts towards overcoming the challenges of autonomous social touch interactions with a virtual agent and the framework developed in that regard.

3.1 Requirements for a Touching Agent

To build a complete interactive loop with an autonomous artificial agent, the agent needs the abilities to perceive, decide and act. With its perception ability, the agent can indeed sense and record what happens in the world around it and what the other human(s) (or agents) are doing. With its decision ability, it is then able to reason about what was perceived, about the current state of the situation, and determine what course of action to take in response. With the ability of action finally, the agent can act on its decision and make its behaviour perceivable by the one(s) they are interacting with. Once a full

iteration of this process has happened, the human or the other interacting agent can act again, which will be perceived by the agent and prompt it to react again with the same process, thus explaining why we talk about an interactive *loop*.

In human-human communication, touch is one modality among many others (Knapp and Hall (2010)). It shares with those other modalities an ability to convey specific meanings via symbolic/ritualized uses but also shares a common ambiguity when looked at on its own. We thus believe that there is *a priori* no need to alter the classic structure of artificial autonomous agents to integrate touch to it. The question then becomes: what specific systems will be needed for the implementation of touch? More precisely, how can we implement and connect together a touch perception module, a decision model able to consider touch and its implications, and a device able to generate a sensation of touch for a human?

Chapter 2 evidenced how both the fields of robotics and virtual ECAs have started exploring the possibilities of haptics and social touch interactions. However, we have also seen how for now their focus is usually on only one of the directions of the interaction, either the sensing of touch or the generation of haptic stimuli. Furthermore, there currently are almost no works that focus specifically on the reasoning and decision making process for an agent that can touch.

In the domain of virtual reality, haptics have, for now, mostly been developed and researched for interacting with the environment and the manipulation of objects. We argue that virtual reality and immersive environments are nevertheless very well suited to the study of touch interactions (as also discussed by Huisman et al. (2014a), who highlight all the benefits to be gained by investigating social touch in immersive environments), as they allow for minute control of every part of an ECA's social behaviour and of the immersive environment. 3D modelling and animation further facilitate the use of somewhat credible human-looking agent, where humanoid appearances can be more challenging in the field of robotics. We thus position ourselves towards the development of a virtual ECA for immersive interactions.

We have seen in Sections 2.2.2 and 2.2.1 respectively the options we can choose from and increment upon to make a virtual agent able to physically perceive a human touch on one hand and generate haptic stimuli with affective properties on the other hand. Sections 3.2 and 3.3 will thus respectively detail a proposition of architecture for a touching agent and our technological choices regarding the immersive setup. To enable an actual human-agent interactive loop with bi-directional touch however, we emphasize the importance of thinking about how to connect all those haptics-related elements with an adequate decision model, so as to build a complete functional framework. While we do not aim to reach the top results of the state-of-the-art in the sensing nor in the haptic generation, we argue that such a proof-of-concept of a complete system allowing an interactive loop based on social touch with a virtual agent is an interesting and important contribution to the literature.

3.2 Architecture of a Virtual Touching Agent

How should we design the interaction process of our interactive loop then? The question here is basically what kind of inputs do we want to feed to the decision model so that it may make appropriate decisions regarding the use of touch? How to model the decision making of the agent for social touch decisions? And regarding the result of the decision, what kind of behaviour do we want to simulate?

As was discussed before, we argue that a computational model of emotion would be ideal for a touching agent in order to take the affective component of social touch into account. That being said, social touch is also closely related to relationships and bonding. Section 2.1.2 showed how the relationship between toucher and touchee affects whether the touch is going to be accepted and how it is going to be interpreted. We also saw how the history of the study of the senses was heavily influenced by cultural considerations, notably religious backgrounds, and how it only really began with the apparition of feminist, anti-discrimination, sociological and philosophical works that were interested in our embodied existences, pushing aside the old body/mind dichotomy. Those works are still fairly recent, and the study of social touch is still lacking in many aspects, with a lot to be investigated and discovered still. Recent events like the MeToo and BlackLivesMatter movements have clearly shown how our societies are still far from achieving equality among individuals, and how our bodies are still the focal point of discrimination and attacks. And just as individuals and the societies they form need to question their prejudices and everyday behaviours, the field of research and development also need to question the social impact of the new technologies that we produce and how we design them. Despite what might be commonly thought, there is no such thing as a neutral technical object or technology. We believe that this is especially true when working on the development of social touch technologies, and maybe even more so when working on a system that can autonomously touch human users. In the context of mediated interactions in VR, violations of personal space by others, including simulated touches without haptic feedback, have already been felt by users as actual harassment [Blackwell et al. \(2019\)](#). We believe that a virtual touching agent should always be designed so as to respect social norms and personal preferences and avoid being felt as intrusive (or worse) by the user it interacts with.

A touching agent should therefore be able to consider the level of the relationship that has formed between it and the human user. As discussed in Section 2.3.1 the notion of rapport, defined as “mutual attentiveness, positivity and coordination” by [Tickle-Degnen and Rosenthal \(1990\)](#), is better suited to short-term interactions. While we do not intend to work on a very precise way to dynamically measure rapport, we argue that a touching agent should be able to adopt rapport-building strategies and to have at least a basic estimation of the level of rapport that has been built with the human user. This will in turn allow the agent to better determine when and which types of touch are acceptable at

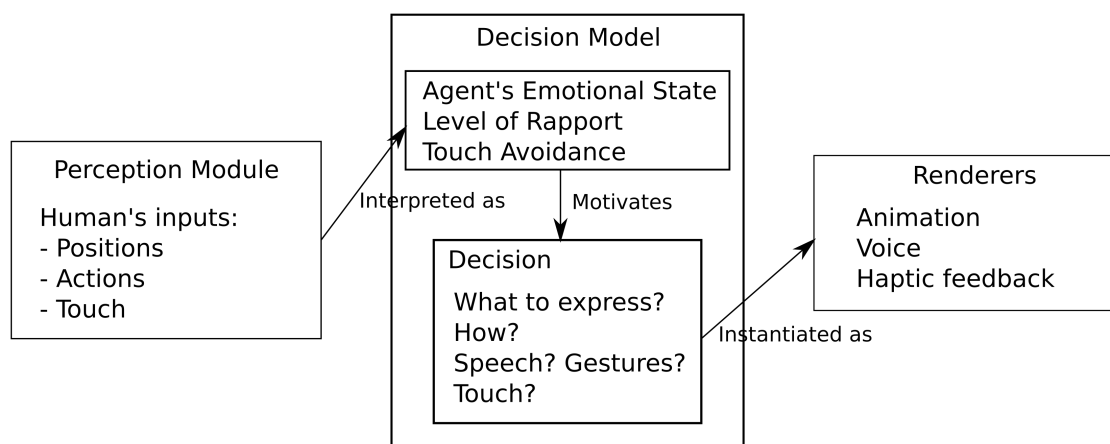


Figure 3.1 – Schematic overview of the components proposed for our touching agent

each point of the interaction. To enable this measuring of rapport, we propose to follow the definition cited above and more specifically focus on looking at the inputs that would allow us to evaluate the level of attentiveness of the human towards the agent, as well as their emotional state (for the positivity). Based on these pieces of information, on their respective positions in the environment and on its own emotional state, the agent will then be able to decide if touch would be appropriate to the situation. Furthermore, we believe that the ECA should also be able to talk, so as to use multimodality and give context to the interaction. This will thus require both a way to animate the agent's body and produce its voice, and a way to produce haptic stimuli.

We thus propose the basic components and process shown in Figure 3.1, with the agent being composed of a perception module, an appraisal-based computational model of emotions and appropriate renderers for haptic generation and animation of the virtual agent. Inputs sensed via the perception module of the human will then be sent to the decision model which will compute them in order to generate the emotional state and the beliefs of the agent. Based on those elements a decision will then be made to determine a communicative intention and an adequate modality to express it. Finally, the selected behaviour is going to be rendered through animation of the agent and haptic feedback (if necessary).

The perception module will be discussed in Chapter 4, whereas the question of how the decision model will actually compute those inputs and evaluate the level of rapport will be addressed in details in Chapter 5. Finally, our choices regarding the haptic feedback and animation of the agent will be described in Chapter 6. Before that however, the next section will address our technological choices regarding the immersive environment.

3.3 Choosing an adequate VR setup

Virtuality devices, in the context of immersive experiences at least, are often uniformly referred to with the term of virtual reality by the average individual. In actuality, those technologies can be organized on a spectrum depending on their degree of virtuality. At one end of this spectrum we find the absence of virtuality, the real world that we experience everyday, and at the other end of the spectrum we find complete virtuality, which would today correspond to head-mounted displays (HMDs) of virtual reality such as the HTC Vive ¹ or the Oculus Quest ², where not only the world but even our bodies are replaced by digital avatars. In-between those two ends however, we can find many different kinds of display devices with varying levels of virtuality such as augmented reality (AR) glasses, which integrate virtual elements in the real world, or mixed reality CAVETM-like immersive rooms (Cruz-Neira et al. (1993)), which display an entirely virtual world but still allow users to see themselves.



Figure 3.2 – Augmented reality (left), virtual reality (middle) and immersive room (right) setups

As we saw in Sections 2.2 and 2.3 some works have already tackled the question of social touch in most of those levels of virtuality. The AR setup of Huisman et al. (2014b) is interesting but the agent is only visible when the tablet or phone is oriented towards it, which limits the immersiveness of the interaction. We argue that having the human feel immersed and in control of their body in a 1:1 scale environment is critical when working on social touch interactions. In this regard, AR and VR HMDs would thus appear as better alternatives. The AR devices have the advantage of letting the human see their own body, whereas it has to be recreated with an animated avatar in a VR HMD setup. This requires specific efforts towards making sure the human is going to have a sense of embodiment with this virtual avatar, which for now still constitutes a research topic on its own, as evidenced by the works of Fribourg et al. (2020) and Sykownik and Masuch (2020) for example. On the other hand, VR HMDs have been more popularized and are the generally

¹<https://www.vive.com/>

²<https://www.oculus.com/>

cheaper commercially available option at the moment. Furthermore, AR setups obviously don't allow an easy monitoring of the actual environment the interaction is going to take place in, as it could potentially happen anywhere, whereas VR setups give full control over the context of the interaction. As we have seen, the context is extremely important for the interpretation of social touch and being able to control it in the interaction is a crucial feature for our work. Finally, the most expensive and least commercially available option is that of the immersive rooms, which present the best of both worlds in the sense that the user is both immersed in a controlled immersive environment and still able to see their own body.

Again, our goal with the present work is to build a proof-of-concept of an architecture allowing a human-agent interactive loop with social touch, with a focus on the decision model which appears to be the least discussed in the literature for now. Considering that we were already equipped with an immersive room and how adequate this setting is to the study of social touch interactions, with the users being able to always see their bodies while being immersed in the interaction, we estimated that this was the most ideal setup for our work and decided to use an immersive room for our developments. That being said, and being aware that VR HMDs are by far the most commonplace option at the moment, we also decided to make sure to develop a generic software framework that would not rely on a specific VR/AR or haptic setup. This should guarantee the reproducibility of our work for further research and interested third-parties, although transferring our work to another degree of virtuality would require studying the implications in terms of sense of embodiment and level of immersion.

Perception Module for a Virtual Touching Agent

Contents

4.1	A Virtual ECA's Sense of Touch	38
4.1.1	Basic Principles of a Virtual Sense of Touch	38
4.1.2	First version of the Touch Perception Module	38
4.1.3	Improving the Touch Perception Module	39
4.1.4	Interpretation of the Touch Properties	42
4.2	Perception of additional human inputs	45

IN this Chapter we present the perception module we developed for our agent. While our goal is not to build state-of-the-art algorithms for the recognition of emotional behaviours and social touch gestures and their meanings, we still need to have a basic perception module that allows the agent to have some amount of meaningful and useful inputs for its decision process. This entails building a sense of touch for the agent, but also identifying other types of inputs that will allow us to make a basic estimation of the level of rapport between the human user and the agent.

4.1 A Virtual ECA's Sense of Touch

4.1.1 Basic Principles of a Virtual Sense of Touch

Many different parameters can be used to characterize a touch. We can talk about the texture that we felt, about how much pressure was exerted, about where and for how long we were touched, about the speed or strength of the touch. Those parameters and how we can use them to recognize social touch gestures and their meaning has been discussed at length in the literature of social touch and social touch recognition (Hauser et al. (2019), Elbani (2019)). To make our virtual ECA sensible to touch, we must give it the ability to perceive at least some of those parameters, so that those inputs can be reasoned upon and used to determine how to best react to it.

To that end, we identified six touch parameters: **intensity**, **body location**, **presence of movement**, **velocity of movement**, **duration** and **pressure**. We define intensity as the kinematic velocity of the touching gesture at the moment of the first contact, simply put it is what comes immediately to mind when we compare a simple contact with a hit. Body location is almost self-explanatory: the body part that is currently being touched. A touching gesture may not always stop and be static on our skin, in the case of stroking motions for example we need to be able to determine whether the gesture is dynamic (meaning that there is movement along the agent's body) and if so, with which velocity. Finally, duration is for how long the agent is being touched and pressure corresponds to the amount of force applied after the initial impact, as this can be used to detect squeezing touches.

Now how can we make a virtual ECA able to perceive those parameters of a human touch in an immersive room? Since we have a similar setup, we propose to follow the work of Nguyen et al. (2007), where an agent was endowed with the ability to perceive touch in an immersive room. To achieve this, they first created an ethereal body with invisible capsules around the body parts of the agent, allowing it to detect incoming touches, and then covered the model of the agent with virtual shapes acting as collision detectors, which we will call colliders. Those colliders are able to determine when the coordinates of another object overlap with their own and thus when a contact is occurring. When we make sure that those colliders are not rendered visually, they make for a virtual skin that can determine when, where and for how long a touch is happening. As mentioned in Section 2.2, Sykownik and Masuch (2020) proposed a similar approach (although simplified in terms of what can be perceived) with VR HMDs, which proves that the solution is fit for other VR setups as well with only minor adaptations to be made.

4.1.2 First version of the Touch Perception Module

We reproduced this work in the Unity3D engine by first creating an ethereal body around our agent's body which detects when the human's hand comes close. This allows us to

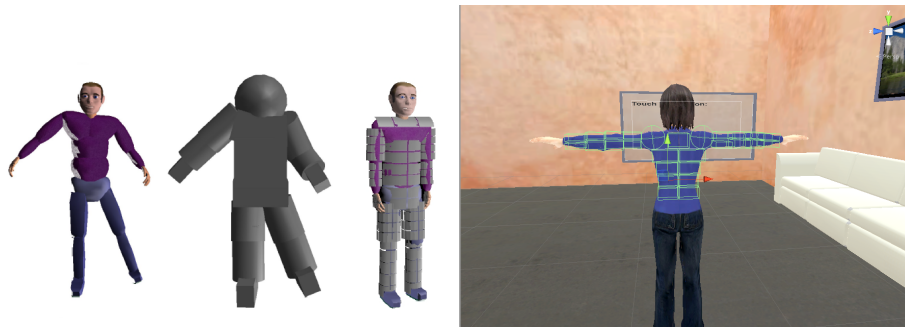


Figure 4.1 – Virtual skin receptors for virtual agents: Nguyen and Wachsmuth implementation (left) compared to ours (right)

give the information of potential incoming touches to the decision model, to determine if it is necessary to react (move back or accept the touch with a smile for example).

Next, we covered our agent with multiple box colliders that would act as tactile cells, as seen in Figure 4.1. The hand of the human is co-located in the virtual world thanks to the motion tracking cameras of the immersive room and the equipment of the user (seen in Figure 3.2). In turn, when the coordinates of the user's hand overlap with those of the agent's tactile cells, a touch is detected and can be recorded. In this implementation however, what each tactile cell records is only whether it was activated or not, and thus the precision of the detection of touches is limited by the resolution of the grid of tactile cells: the more cells we put on the agent the more precisely we can know where the touch is happening. This means that for detecting a movement on the agent's body, we look at the successive activations of tactile cells, and it is not possible to detect movement inside a single cell. This can be misleading on occasion, for example in the case of the user first touching with their fingers and then putting their entire hand on the agent, which could be interpreted as a movement. Having the tactile cells themselves be in charge of the recording of the touches also proves challenging when it comes to centralizing the information regarding a single touch event. Those problems and more were highlighted in the course of the preliminary study presented in Section 7.1. While they could be solved by spending time on testing different grid densities and developing further thresholds regarding when to consider successive activations as movement and when not to, our initial tests led us to consider another solution.

4.1.3 Improving the Touch Perception Module

With the help of Valentin Cecchin, a second version of the touch perception module was developed. On the technical level we first decided to deport the recording of touch information to the objects that come into contact with the agent's tactile cells. Furthermore, instead of having tactile cells only able to give us a binary recording of touches (activated or inactivated), we use larger tactile cells but record the local coordinates of the collisions,

so that we can know where the hand of the user is inside a tactile cell and still be able to know when it moves.

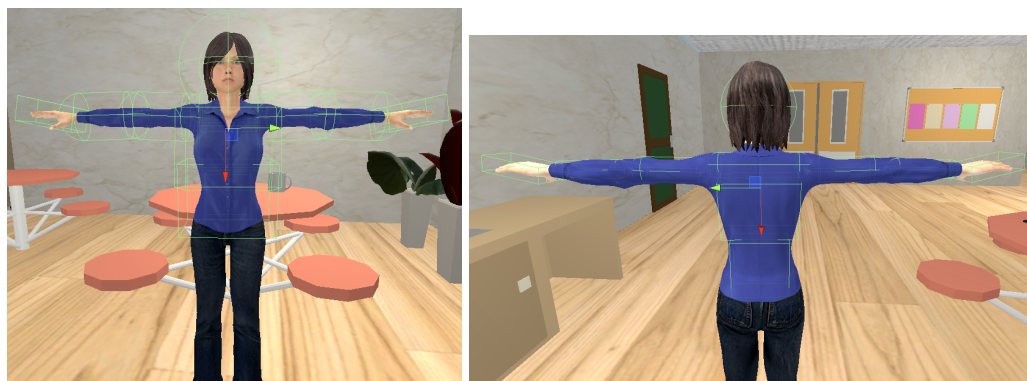


Figure 4.2 – Ethereal body of our agent (left) and second version of our tactile cells (right)

In more technical terms, we built a hierarchy of tactile cells based on joints of the agent’s animated skeleton. What happens is that when the user’s hand collides with the tactile cell, two positions are calculated relative to the center of the local coordinate system of the cell. Those positions are expressed as scalars taking a value between 0 and 1 and indicate how far the hand is from the hierarchically superior joint. The first position scalar indicates whether the hand is near the origin of the tactile cell it is in (0 meaning it is at the base of the current tactile cell) or near the next joint (1 meaning it is on the verge of entering the next tactile cell). The second position scalar determines the rotation angle around the cell. In the case of an arm for example, we identify 0 as the hand being on the “inside” part of the arm (the part of the arm closest to the torso) and 1 as the hand being on the “outside” (the part of the arm oriented away from the body). See Figure 4.3 for a graphical depiction of those axis. We then normalize those measures thanks to the anatomic parameters (i.e. the metric dimensions) of the agent’s 3D model, which makes it easier to interpret them. Based on the social touch literature, we identified which were the body parts more likely to be touched during the interactions and dedicated a tactile cell for each of them: the lower back, upper back, hands, forearms, upper arms and head of the agent.

With this implementation, we can:

- Calculate the intensity of a touch by calculating the speed of the hand at the time of the collision with the tactile cell;
- Precisely know which body part is being touched;
- Determine the presence of movement even inside a body part’s tactile cell;
- Calculate the velocity of such movement thanks to the coordinates of the user’s hand;
- Know when the touch first happened and calculate for how long it lasted.

4.1. A VIRTUAL ECA'S SENSE OF TOUCH

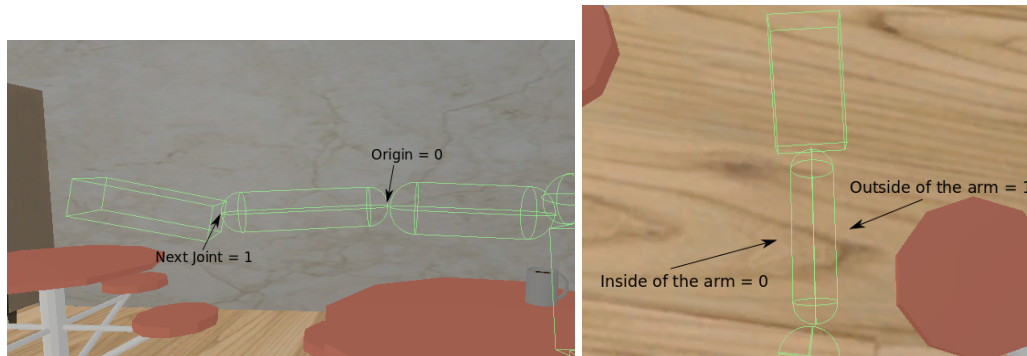


Figure 4.3 – Side and top-down views of the axis used in the definition of the tactile cells relative coordinates

Among the six parameters that we had initially identified as fit for the characterization of a touch, only pressure can not be estimated via this system. Even though we can detect a collision, the agent is still virtual and the user's hand goes through the agent's model. With no physical resistance to the user's hand, we are a priori unable to measure any kind of pressure, thus making certain touch types, such as squeezing, impossible to perform or detect. We however decided to try and have some form of estimation of how much force the user is applying on the agent's model by using a god-object method [Zilles and Salisbury \(1995\)](#).

Even though the human user would see their hand go through the agent's body, we felt that it might still happen that when expressing a particularly powerful emotion they would use strong gestures and simulate shaking or squeezing touches. In order to have an estimation of the force used by the human on the agent, we created a virtual spring joint attached to the user's hand coordinates. This spring is invisible to the user but closely follows their hand, being "pulled" by this god-object. Now, when the human user touches the agent, it is the spring that will activate the tactile cells of the agent. However, where the human's hand goes through the agent's 3D model, the spring joint will not and will stay blocked against the virtual skin of the agent, as would happen with a real world touch. This difference in the coordinates between the tracked real hand of the human and the spring joint will be used to estimate how much force is applied. This is calculated via a scalar proportional to the distance between the spring and the real hand. We use a pseudo Hooke's law to determine the force based on a rigidity coefficient and a movement. To simplify the use of this force estimation, we use a rigidity coefficient whose inverse equals the maximal authorized distance between the real hand and the spring. With this method we do obtain a magnitude of the force exerted, but it should be noted that this remains extremely imprecise.

In the end, the structure of the tactile perception is the one shown in Figure 4.4 The hand of the user is created in the virtual world as a HandManager object, to which is attached a spring joint. Both of those are managed by a HandTouchManager that com-

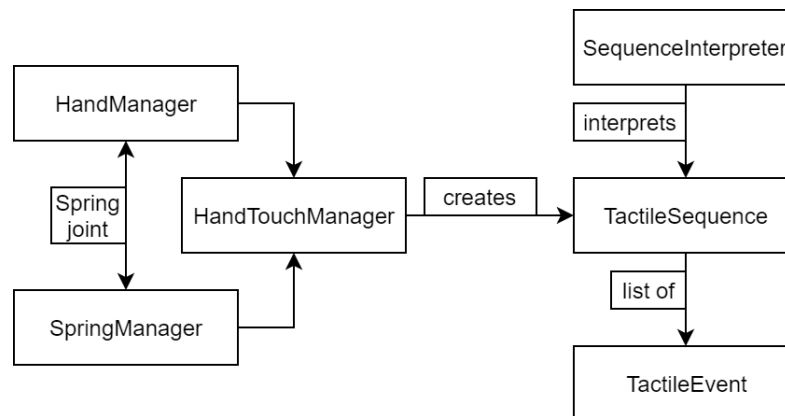


Figure 4.4 – Structure of the code implementation of the touch perception module

municates with the tactile cells of the agent in order to record touches. When a collision happens with a tactile cell, a `TactileEvent` is created that records all the data pertaining to this instance (force, velocity, body part, timeframe, etc.) and it is stored in a new `TactileSequence`. As long as the tactile cells are not all deactivated, we consider that the `TactileSequence` is still active and we record the next `TactileEvent` inside it. A new `TactileEvent` is generated as soon as there is a change in force, velocity or location of the touch that is significant enough (based on arbitrary thresholds set up to limit the number of recordings). When all the tactile cells indicate that there is no touch anymore, we complete the `TactileSequence` with the total duration of the touch. With this, we have all the parameters that we can possibly record to characterize a touch. All that is left to do is interpret those values.

4.1.4 Interpretation of the Touch Properties

Interpretation of a touch then happens in two steps: first we want to have a preliminary interpretation of the values in terms of touch gesture, and then we want to send all of this information to the decision model for interpretation of emotional meaning. This latter part will be described in Chapter 5. Regarding the former, we use a `SequenceInterpreter` structure that will look at the raw data of the detected collision and try to determine which touch type it corresponds to. More specifically, here are the values that we are going to take into account:

- Duration of the `TactileSequence` in seconds (`totalDuration`);
- Velocity of the hand at the time of the initial collision (`initialImpactVelocity`);
- Mean normalized contact force as estimated with the god-object method and the spring joint (`meanForce`);
- If movement: velocity of movement along the agent's body (`meanVelocity`).

4.1. A VIRTUAL ECA'S SENSE OF TOUCH

The touch types that we will consider here are the following: tap, hit, caress, rub, scratch, brush, press and a default value of unknown/not recognized. Again, our work is not focused on the recognition of touch and we do not aim to reach state-of-the-art performances. We however argue that having supplementary indicative information in that regard is interesting and will allow us to already set up a prototype for this information in the decision model, in anticipation of future iterations of the system where a better recognition algorithm could be integrated. To that end we propose a system of interpretation based on fuzzy sets, where we partition each touch parameter in sets of qualitative classes. Tables 4.1 to 4.4 present those partitions.

First, totalDuration is partitioned in three classes: short, medium and long. For our initial tests, we considered values up to two-second durations, but future developments will most likely consider longer durations.

	Min	v1	v2	v3	v4	Max
	0,00	0,10	0,60	0,80	1,60	2,00
Short	1	1	0	0	0	0
Medium	0	0	1	1	0	0
Long	0	0	0	0	1	1

Table 4.1 – Linear partition table of a touch sequence's duration.

Then, initialImpactVelocity is separated into: weak, medium and strong impact velocities. The values are expressed in meters by second. Again those values are arbitrary choices that would require further tests and calibrating.

	Min	v1	v2	v3	v4	Max
	0,00	0,20	0,70	0,80	3,50	4,00
Weak	1	1	0	0	0	0
Medium	0	0	1	1	0	0
Strong	0	0	0	0	1	1

Table 4.2 – Linear partition table of velocity at the time of the touch collision.

Similarly, meanForce is expressed as low, medium or high and the values calculated with the pseudo Hooke's law are comprised between 0 and 1.

	Min	v1	v2	v3	v4	Max
	0,00	0,10	0,50	0,50	0,90	1,00
Low	1	1	0	0	0	0
Medium	0	0	1	1	0	0
High	0	0	0	0	1	1

Table 4.3 – Linear partition table of a touch sequence's mean force exerted.

Finally, meanVelocity is separated into: static, CT-Optimal and fast. Values are again expressed in meters by second and the CT-Optimal one are based on the literature on social touch (3-10cm/s, see Section 2.1.2 for more details on this range of velocities).

	Min	v1	v2	v3	v4	Max
	0,00	0,02	0,03	0,1	0,4	1,00
Static	1	1	0	0	0	0
CT-Optimal	0	0	1	1	0	0
Fast	0	0	0	0	1	1

Table 4.4 – Linear partition table of a touch sequence’s mean movement velocity.

The interpretation is then realized through two interpretation tables. The first is a three-dimensional table which we split in three sub tables here for the sake of visual clarity. The second, simpler, interpretation table takes into account the initialImpactVelocity. A union is realized to combine the results of both tables.

totalDuration		Short		
		meanForce		
meanVelocity		Low	Medium	High
Static		Brush	Tap	Press
Slow		Brush	Tap	Hit
Fast		Tap	Hit	Hit

totalDuration		Medium		
		meanForce		
meanVelocity		Low	Medium	High
Static		Brush	Tap	Press
CT-Optimal		Caress	Rub	Rub
Fast		Rub	Scratch	Scratch

totalDuration		Long		
		meanForce		
meanVelocity		Low	Medium	High
Static		Brush	Press	Press
Slow		Caress	Caress	Rub
Fast		Rub	Scratch	Scratch

Table 4.5 – First interpretation table

This interpreter can easily be modified if necessary, either to change the touch types or by changing the values in the tables of the parameters. All the classes are intersected so as to attribute a score for each type of touch. Each type’s score is determined via the union of the score of all the corresponding activated cells of the table. This way we

4.2. PERCEPTION OF ADDITIONAL HUMAN INPUTS

totalDuration	initialImpactVelocity	Weak	Medium	Strong
Short		Brush	Tap	Hit

Figure 4.5 – Second interpretation table to take into account the velocity of the gesture at the time of the initial collision

authorize the algorithm to test a maximum amount of cases, without excluding touch types preemptively.

In summary, this touch perception module gives us information about possible incoming touches via the ethereal body, a complete `TactileSequence` with the physical properties of the touch and a pre-interpretation of the touch type that we can send to the decision model. The module remains to be properly evaluated in context in the future.

4.2 Perception of additional human inputs

In order to help the decision model evaluate the level of rapport between the agent and the human we however need to perceive more different types of inputs than just touch. The ones that we consider as the most interesting are the proxemics and the gaze direction of the human.

The proxemics values observed in human-human interaction (intimate, personal, public and social spaces) by [Hall \(1990\)](#) have indeed also been observed to be respected by human users in virtual environments [Cafaro et al. \(2016\)](#). Since we track the position of the human in the immersive environment thanks to their stereoscopic 3D glasses equipped with motion tracking markers, we can directly use the positions of the human and the agent to calculate the distance between them. We then determine which space the human is in relative to the agent by using the traditional threshold values of proxemic spaces (see [Figure 4.6](#)).

For the gaze direction, we are somewhat limited by the fact that we don't have access to eye tracking to precisely determine where the human is looking at, but we can still use the forward vector of the user's glasses to have an approximation of this direction. We simply use dot products between this vector and the forward vector corresponding to the head of our agent to obtain a certainty coefficient between 0 (the user is not looking at the agent at all) and 1 (the user is looking straight at the agent). We also added supplementary tests to determine if the user is rather looking towards the body of the agent or towards its face, but those would only yield significantly different results from our basic estimation with more precise eye tracking data.

Finally, a note about speech inputs. While we realize how important speech is in human communication and how the field of speech recognition has come far, we chose not to implement it in our work for now, as the scope of our research is already quite

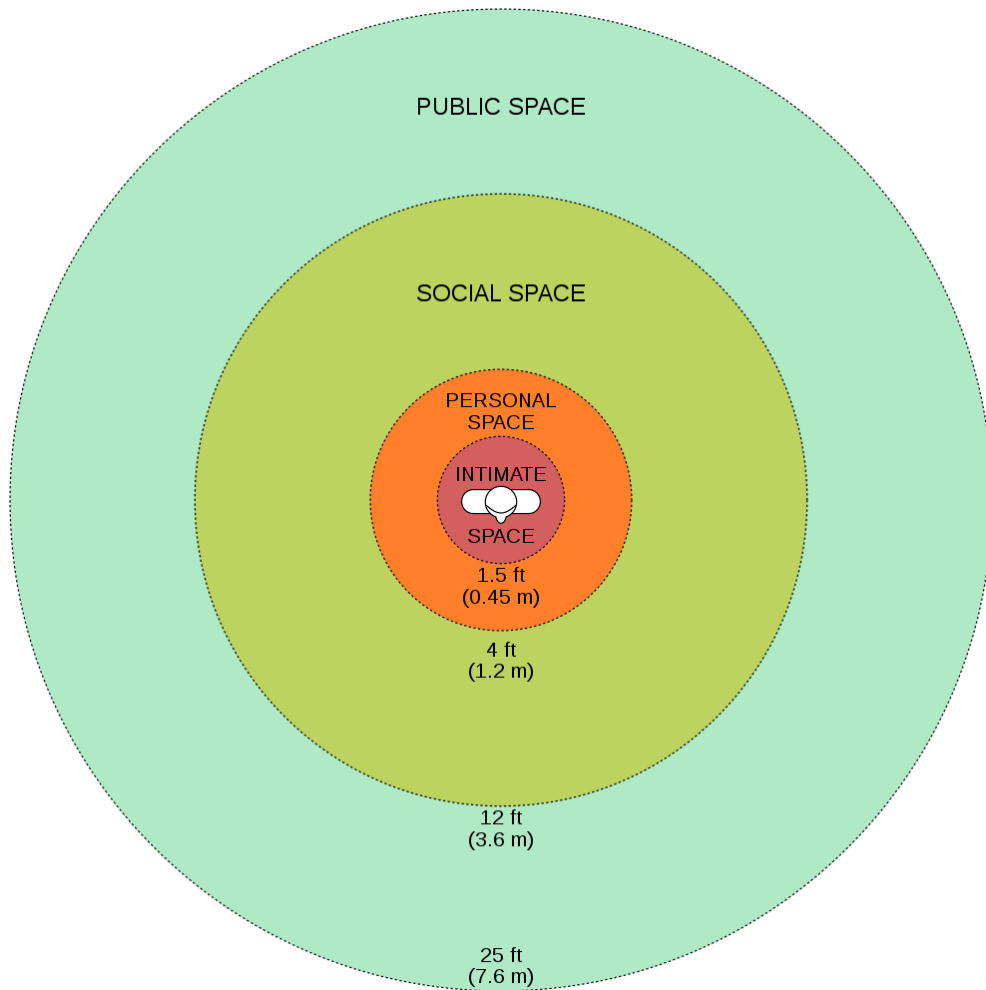


Figure 4.6 – Proxemic spaces and their respective dimensions

vast as it is. We argue however that adding such a module of speech recognition in later iteration of the system should prove relatively easy, as it would only bring supplementary inputs to consider in the definition of the decision model and would not force us to change the overall architecture of the system.

In summary, in this chapter we have shown how we implemented a sense of touch for our virtual agent via colliders acting as tactile cells. This collision information is then pre-interpreted in terms of touch types and sent to the decision model along with proxemics and gaze direction information.

Chapter 5

A Touching Agent’s Decision Model

Contents

5.1	Decision Process of a Touching Agent	48
5.1.1	Overall design choices for the decision process	48
5.1.2	Overall implementation framework	49
5.2	Internal States of the Agent	51
5.3	Appraisal Rules	53
5.4	Beliefs of the Agent	57
5.4.1	World Model & Interpretation rules	58
5.4.2	Determining the information needed to model the acceptance of a touch	58
5.4.3	The dimensions of rapport	59
5.4.4	Estimation of the human’s emotional state	59
5.4.5	Estimation of the human’s attentiveness towards the agent	61
5.4.6	Evaluating rapport & static touch avoidance	61
5.4.7	Implementation: Interpretation rules and Rapport calculation	61
5.5	When and How to Touch a Human	63
5.6	Decision Rules	65

IN Chapter 2 we saw how most works regarding simulated social touch focus on only one of the directions of touch, either the perception of a human touch or the production of haptic feedback and its effects on the recipient. With our work we propose an architecture for an agent able to participate in an interactive loop based on social touch with a human, thus enabling bi-directional social touch and behaviour

adaptation based on the situation. Chapter 3 described our technological choices for the immersive experience, and Chapter 4 developed the agent's perception module. What will connect the perception module to the adapted behaviour that the agent will display to the human is the decision model, where the inputs are interpreted and used to select the agent's actions, thus enabling dynamic interaction. It will be the task of this chapter to describe what is maybe our most important contribution: the design of the decision model, and especially our choices regarding how to determine when to use the touch modality, as well as the overall implementation of the model.

5.1 Decision Process of a Touching Agent

5.1.1 Overall design choices for the decision process

One of the first questions that we can ask ourselves when considering the design of a decision model is: what kind of model do we want to build? At the moment, machine-learning algorithms (in their many forms) are the most popular option in the field of artificial intelligence, thanks to their very strong performances and ability to produce variability as long as there is data to train the model. Reciprocally, one of their downsides is that the less data is available for the training, the harder the model will be to train to achieve high performance. Another commonly noted drawback of those models is that their decision process has very low explainability, as it relies on the computation of many variables that we, as human beings, cannot interpret as is. They thus produce correct results and classifications, but we don't know exactly how they reached their conclusions. On the other hand, the alternative to machine-learning models is that of rule-based decision models in which the decision process is performed based on pre-defined logical rules regarding the current knowledge (or beliefs) of the model. Those models are thus very easily explainable and understandable for the users, but require a lot of definition work for the rules. Those models are therefore always focused on a specific field or situation and rely on the expertise of the human mind that produces the rules. Furthermore, rule-based models cannot learn from their experiences and if a situation arises that is not present in the rules, the model simply won't be able to interpret nor react to it.

Considering that touch is an intimate modality that may lead to feelings of uneasiness or even harassment, we argue that special focus should be put on making the model as understandable and transparent for the human as possible. In addition, and to our knowledge, no databases featuring documented human-human touch interactions currently exist. Furthermore, systematic study of when specific types of touches are occurring and for which meanings is still very scarce. We argue that building a rule-based model would thus fulfill our needs regarding its explainability and also be an interesting contribution to the literature on social touch as it will lead to the construction of a formal set of rules

5.1. DECISION PROCESS OF A TOUCHING AGENT

regarding when and how to touch depending on the situation. The proposed architecture and decision process for such a model are presented in Figure 5.1.

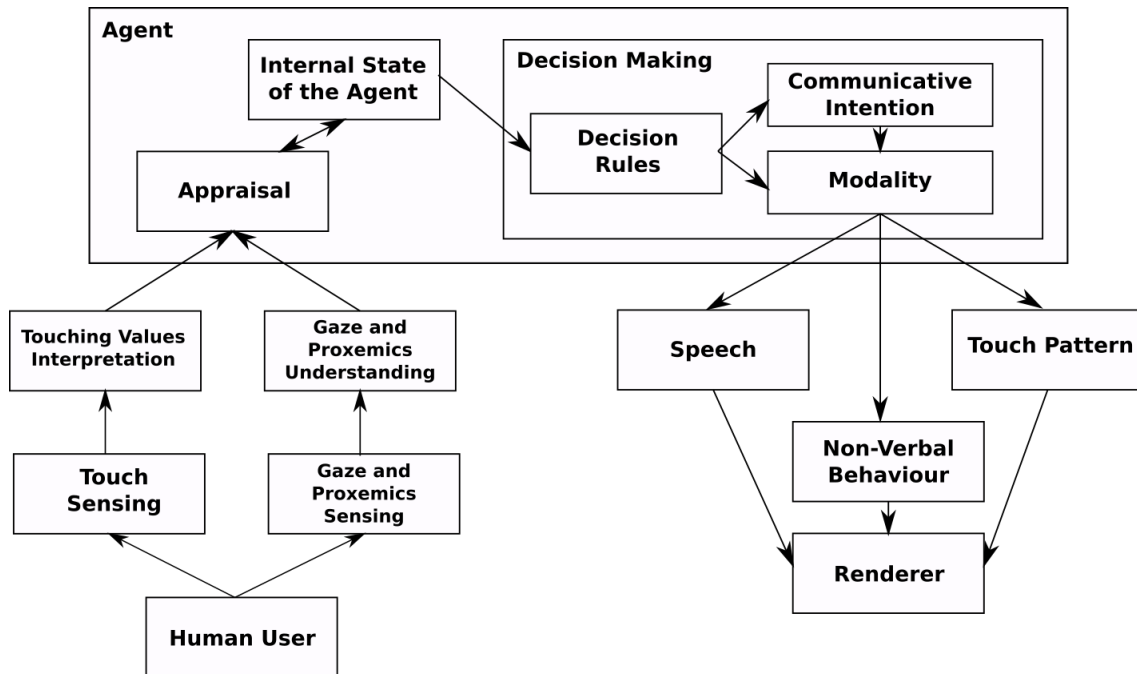


Figure 5.1 – Schematic overview of the overall process of the decision model in an interactive loop with a human

In order to enable our agent to be truly “touching” in all the meanings of the word, i.e. able to touch physically at appropriate times and to have a believable emotional behaviour that may “touch” the human it interacts with (emotionally), we propose to design the model as a computational model of emotions. Upon receiving inputs from the perception module (in the form of proxemics, gaze direction, touch or scenario-related information), the model is going to evaluate how those events influence the agent’s emotional state and beliefs about the situation and the human’s mental and emotional states. More particularly, we want to estimate the current emotional state of the human and the current level of rapport between them and the agent. Once the internal states of the agent have been updated, they will be used to determine which decision rules to activate. This process first produces a communicative intention (what the agent aims to express), which is then used to determine an adequate modality (among speech, non-verbal behaviour and/or touch). Finally, the chosen behaviour is sent to the renderers for animation and haptic feedback.

5.1.2 Overall implementation framework

To implement such a model, we went back to our review of the literature about computational models of emotions (Section 2.4) and identified that an appraisal-based model would most likely be the most appropriate. Since the development of a CME is a research subject on its own and valid options are already available, we want to use an already

existing CME that would allow us to implement our own rules and modify some of its functionalities if necessary. This way our work can also help the development of this field of research. The CMEs that appeared as the most suitable for our work were EMA and FAtiMA, discussed in Section 2.4, with EMA being a very powerful model able to model and compute a lot of different emotions and situations, but overall more oriented towards simulation, and FAtiMA being a less sophisticated model but open-source, slightly more oriented towards interactive scenarios and with a modular architecture making it easy to adapt to most needs. We argue that for this proof-of-concept work we do not require a very sophisticated process of emotion generation and we thus felt like the FAtiMA toolkit ¹ was the most adequate option at this point in time and decided to use it for the realization of our decision model.

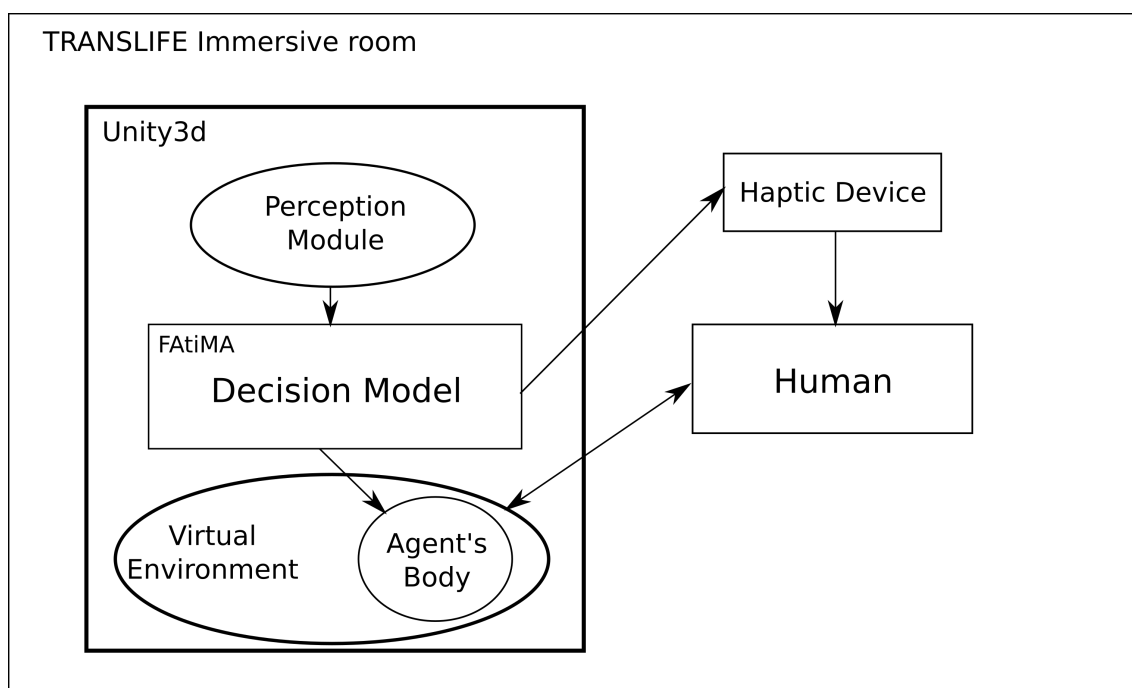


Figure 5.2 – Software architecture for social touch interactions with a virtual embodied conversational agent

With this technical choice settled, we can now propose the software architecture shown in Figure 5.2. In this setup, the interactive application and perception modules are developed in the Unity3D game engine. FAtiMA is easily integrated into Unity via DLLs and only requires defining the properties of the model via configuration and scenario files that can be stored in the Unity project folders. This way we only need to properly format the events recognized by the perception module before calling the appropriate FAtiMA's methods to update the internal states of the agent and trigger the decision process. Once a decision has been reached, Unity will only need to send the decided behaviour to the animation and haptic feedback renderers, which will be discussed in Chapter 6. With this architec-

¹<https://fatima-toolkit.eu/>

ture, it is easy to adapt our work to other VR modalities or haptic devices: as Unity3D can easily be used with any type of VR device, only little modifications to the existing code would be necessary ; FAtiMA can receive inputs from literally any source as long as those inputs are formatted in a way that FAtiMA can understand and the rules are adapted ; similarly, with any type of off-the-shelf touch device, it would only be required to translate the model's decisions into valid signals for the specific device.

5.2 Internal States of the Agent

What should the “internal states of the agent” be? They represent what the agent is currently aware of and can be separated in four different categories: the agent's emotional state, the agent's current knowledge base, the agent's goals and the agent's memory. The agent's emotional state consists in either one emotion or a combination of emotions experienced by the agent and an overall mood, which is influenced by said experienced emotions. The agent's current knowledge base (or KB) will store the current beliefs of the agent regarding the state of the situation: this is where we will find the beliefs regarding the perceived human's emotional state, the current level of rapport, the positions of objects in the environment, etc. The agent's goals will define what the agent wants to achieve during the interaction and also store for each goal how far it currently is from being achieved and how important it is (compared to other goals). Finally the agent's memory will be a simple record of all events perceived by the agent and having taken place during the interaction. While we focus here on short-term interactions, we believe that touching agents will prove even more interesting and useful when adapted to long-term interaction with users and having an accessible memory will thus become increasingly important in future works.

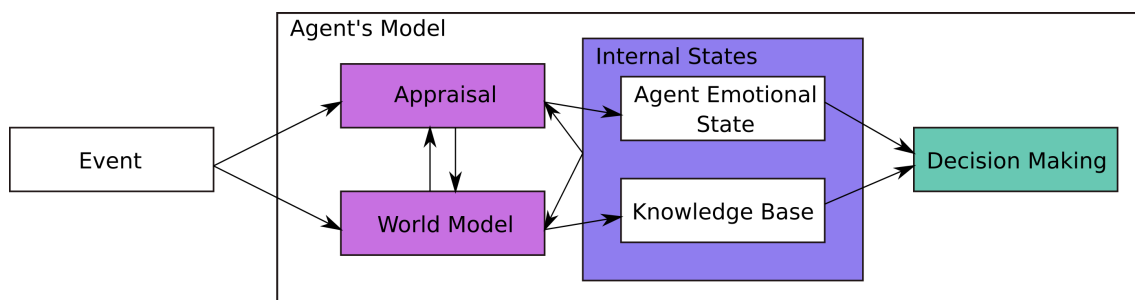


Figure 5.3 – Updating process of the internal states of the agent upon perception of an event

The process by which those internal states are updated during the interaction is shown in Figure 5.3, although goals and memory were left out for the sake of clarity. The emotional state is derived from the application of the appraisal rules when an event is transmitted by the perception module to the decision model. Those rules define how an event

should be evaluated in emotional terms based on the goals of the agent and how this event scores for each appraisal variable. While the goals of the agent influence how the appraisal of an event is made, the appraisal process may also update the current progression of the goals. For example, if an event is evaluated as inconvenient for the realization of a goal, it will be interpreted negatively and may push back the current progression of the goal. This appraisal process and its implementation in FAtiMA will be further detailed in Section 5.3. Concurrently to the appraisal process, a World Model (the collection of all interpretation rules) will derive from the perceived event how to update the knowledge base with new or modified beliefs. If an object has been moved from a certain position to a new one and the agent observed it, for example, the new position is going to be updated in the knowledge base by the corresponding interpretation rule of the world model. Finally, all the events are directly stored in the memory as they are received from the perception module.

Our model uses proxemics, gaze direction, touch and scenario-related information as inputs. Most of those inputs have been discussed in Chapter 4, where we described how the perception module detects and records this input data. As directly receiving the raw values would make it difficult to define appraisal and interpretation rules, we instead use the pre-interpretations produced by the perception module to classify the values and format everything in a single event with multiple parameters to be sent to the decision model. The following tuples are examples of event structures:

- DistanceChanged([c])
 - Where [c] is the interpretation class of distance, among Intimate, Personal, Social and Public.
- LookDirectionChanged([b], [h], [e])
 - Where [b] is True or False and indicates if the human is looking at Camille's body
 - Where [h] is True or False and indicates if the human is looking at Camille's head
 - Where [e] is True or False and indicates if the human is looking at Camille's eyes
- TouchEvent([st], [en], [part], [type], [durint], [invint], [meanfint], [meanvint])
 - Where [st] is the start time of the touch (since the start of the simulation). It acts as an identifier for all the touch sequence.
 - Where [en] is the end time in seconds since the start of the simulation.
 - Where [part] is the body part currently touched, among Member, Torso, Head, Arm, Forearm and Hand.

- Where [type] is the interpretation class of the touch, among Unknown, Tap, Hit, Caress, Rub, Scratch, Brush and Press.
 - Where [durint] is the interpretation class of the current duration, among Short, Medium and Long.
 - Where [invint] is the interpretation class of the impact velocity, among Weak, Medium and Strong.
 - Where [meanfint] is the interpretation class of the mean force, among Low, Medium and High.
 - Where [meanvint] is the interpretation class of the mean velocity, among Static, CTOptimal and Fast.
- EtherealBodyEvent([st], [dur])
 - Where [st] is the start time since the start of the simulation.
 - Where [dur] is the current duration of the stay into the ethereal body, in seconds.

During an interaction with a human user, the agent's beliefs (e.g. about the perceived level of human attentiveness) and the emotional state of the agent are constantly updated during the interaction, based on what events are perceived and sent to the decision model. Based on those prototypes we can thus tailor the appraisal rules to define how the agent is going to emotionally react, and the interpretation rules to determine what the agent should understand from a specific event.

5.3 Appraisal Rules

To generate the agent's emotional state, our CME uses an appraisal process. Appraisal rules are defined to determine how each type of occurring event scores in terms of appraisal variables. As mentioned in Section 2.4 appraisal variables often include valence or goal conduciveness but each different appraisal theory uses its own specific set of appraisal variables. The theoretical model of emotion generation used in FATiMA is the OCC ([Ortony et al. \(1988\)](#)) model and its implementation in FATiMA uses five appraisal variables: desirability, desirability for others, praiseworthiness, goal conduciveness and liking. Anything perceived by the agent's perception module and transmitted to the decision model is considered as an event. For each kind of event, such as the human moving towards the agent or starting to look at something else in the environment, specific rules need to be designed to determine how to attribute the appraisal variables scores based on the parameters of the event and the current internal states of the agent. The generic prototype of an appraisal rule is thus the following:

```
Event: Event([parameters])
Conditions (prototypic examples):
- [parameters] = [specific value]
- Belief([par]) = [value]
- ...
Appraisal Variables Effects:
- Desirability = [value]
- Desirability for others = [value]
- Praiseworthiness = [value]
- Goal conduciveness = [value]
- Like = [value]
```

Where the event section specifies the event concerned by the rule, the conditions section enumerates which conditions need to be specified for the rule to be activated and the appraisal variables effects section specifies the appraisal variables scores for this event. For example, if we model our agent so that one of its goals is to keep the human user's attention, the event 'human looking elsewhere in the environment' would be appraised as scoring negatively in desirability and goal conduciveness because it is interpreted as a sign of inattention. However, it will be different if this event takes place in a task-oriented scenario which features an object related to the task. In that situation if the object is being discussed and an event of 'the human is looking at the object' is triggered, it should instead be considered as the human being attentive (joint attention) and thus desirable and positive for the agent's goals.

If we now take a more specific example for a TouchEvent([st], [en], [part], [type], [durint], [invint], [meanfint], [meanvint]) tuple as defined in Section 5.2, it probably becomes obvious that defining rules for every possible combination of parameters would be almost impossible considering the sheer number of parameters. What we can do however is create rules for each specific parameter and its potential values, and then concatenate the results of each activated rule. For now, in the appraising of a touch event we chose to consider two variables: desirability and liking. A touch event detected by the agent will thus be considered as desirable when occurring on an acceptable body part (arms, hands) and undesirable when occurring on a less accepted body part (torso, head for example). However, as relationships impact the acceptability of a touch, a touch on the shoulders and the back can become desirable if the level of estimated *rapport* is considered as high. Similarly, if we look at the mean velocity of a stroking movement, we will evaluate a CT-optimal speed as likable and an estimated touch type of hit should be evaluated as unlikable and undesirable. Here are a few formatted examples:

```
Event: TouchEvent([st], [en], [part], [type], [durint], [invint],
[meanfint], [meanvint])
Conditions:
```

5.3. APPRAISAL RULES

- [part] = Head

Appraisal Variables Effects:

- Desirability = -6

- Like = -6

Event: TouchEvent([st], [en], [part], [type], [durint], [invint],
[meanfint], [meanvint])

Conditions:

- [meanvint] = CTOptimal

Appraisal Variables Effects:

- Desirability = 3

- Like = 6

In terms of implementation, those examples are already almost perfectly formatted as in FAtiMA all the different types of rules rely on logic programming, with each rule being determined by a set of conditions expressed as logical and/or mathematical statements. A system of substitution of the variables (what we put between '[' and ']') with values will find which conditions are validated and thus activate the corresponding rules. Examples of actual implemented appraisal rules for touch events are shown in Figure 5.4, with prototypes of events tuples where we specify only the parameters that interest us. By default, all the activated rules results are combined and applied. Based on those results the emotion generation algorithm is used to compute the new emotional state of the agent (emotions, their intensities and the updated mood of the agent). In the OCC model implemented by FAtiMA, the computation of emotions is realized in the way described in Figure 5.5. Here we won't address the calculations made by the model as they are already implemented and functional inside of FAtiMA and we simply make use of them.

Through the definition of the goals and appraisal rules for an agent, it is thus possible to define different personalities for different agents. As an example consider that we build two different agents and that one has this appraisal rule:

Event: TouchEvent([st], [en], [part], [type], [durint], [invint],
[meanfint], [meanvint])

Conditions:

- [type] = Tap

Appraisal Variables Effects:

- Like = 8

While the other agent has this appraisal rule:

Event: TouchEvent([st], [en], [part], [type], [durint], [invint],
[meanfint], [meanvint])

Conditions:

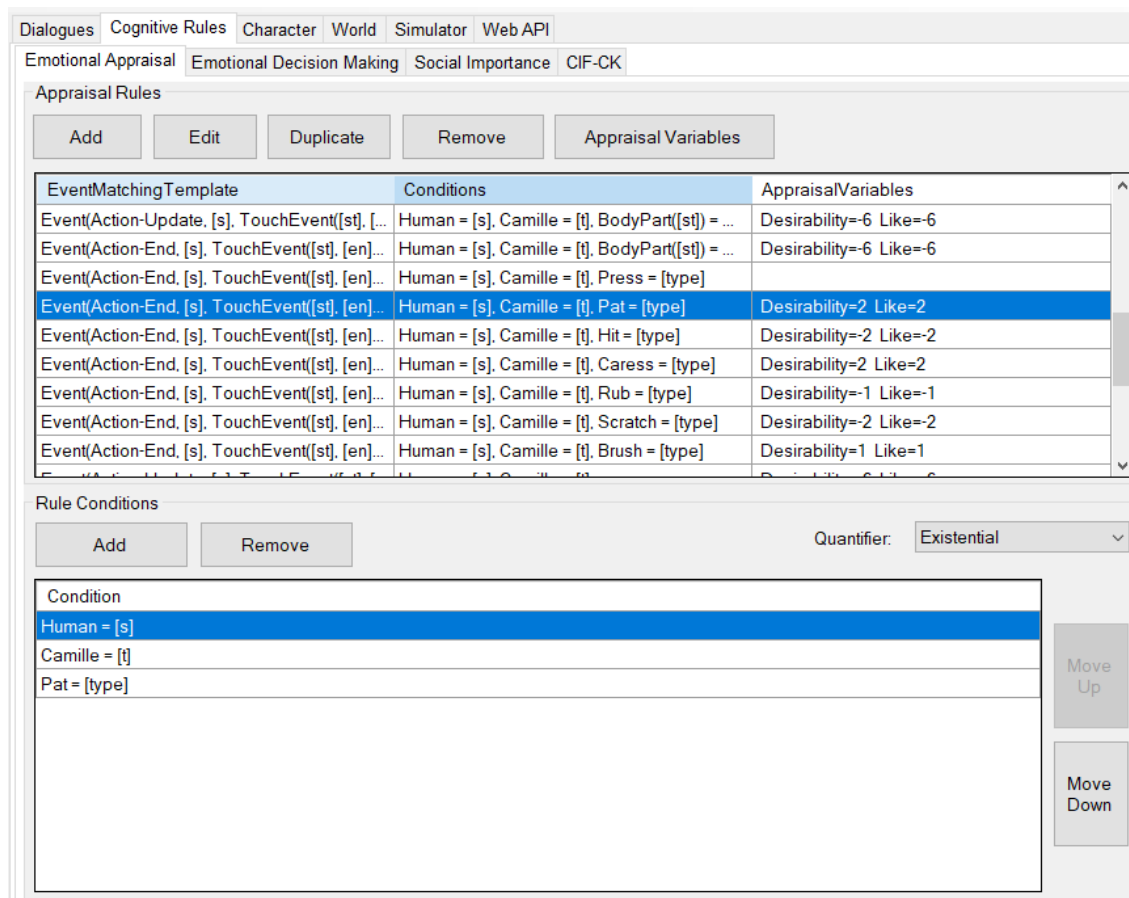


Figure 5.4 – Screenshot of the FAtiMA interface with a few appraisal rules for touch events.

- [type] = Tap

Appraisal Variables Effects:

- Like = -8

- Goal conduciveness = -3

Although both of those agents react to the same touch event, this event is appraised entirely differently and one is going to experience a negative emotion while the other is going to experience a positive emotion. We essentially defined character traits. By deliberately making it hard (or easy) for an agent to experience some specific emotions, we can make them feel like being overall cool- or hot-headed, for example. Similarly, their personalities can be further defined by setting up a specific emotional state for the agent at the beginning of the interaction (for example, it will start the interaction by being extremely joyous), or by defining different goals. A goal is defined as follows:

GoalName

Importance : [value]

Goal Progression : [value]

5.4. BELIEFS OF THE AGENT

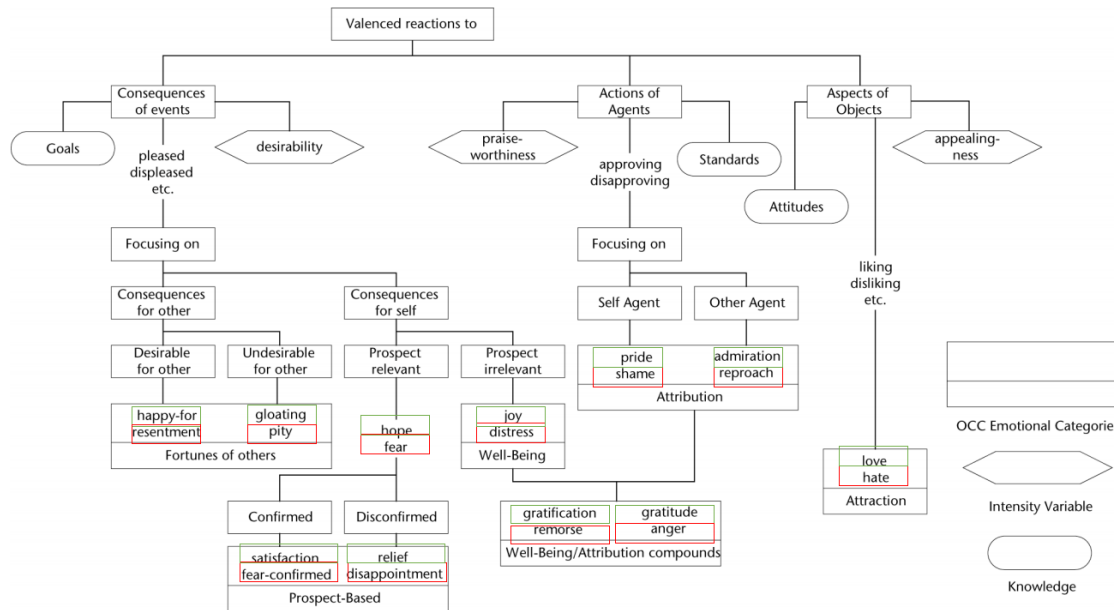


Figure 5.5 – The affect derivation processed of the OCC model, as implemented in FAtiMA (available at <https://fatima-toolkit.eu/5-emotional-appraisal/>)

The name will be used to refer to the goal in the expression of conditions, the importance define the relative importance of this goal compared to the other goals of the agent ("Survival" would be the highest importance for most living beings for example) and goal progression is a value that indicates how far the agent is from having achieved this goal (if it makes sense for this particular goal). As a side note, in FAtiMA, this particular property of a goal is called likelihood. In the available implementation of the model however, it was not possible to retrieve the previous value of the likelihood to add or subtract values to it, thus forcing us to reset the value every time. We went and modified the code to make a quick fix by creating a function that could return the current likelihood of the goal and add or subtract from it.

While we believe that using goals and differing appraisal rules for the definition of multiple types of agents' personalities would make their behaviours more believable and would be interesting to study in relation to touching behaviours, we choose not to investigate that topic in our present work, instead leaving it up to future works that could specifically focus on that matter.

5.4 Beliefs of the Agent

In our model, the agent's knowledge is represented via beliefs (stored in a knowledge base) and updated by interpretation rules (stored in a world model) when an event is perceived by the perception module. A belief can be a very straightforward value, such as the current proxemic space that the human is in, relative to the agent, or the current state

of a task. On the other hand it can also be a value calculated based on other beliefs or on the interpretation of specific events. In the end, the purpose of both the beliefs of the knowledge base and the emotional state of the agent are to be used by the decision rules to evaluate which conditions are fulfilled and, therefore, which action to undertake.

5.4.1 World Model & Interpretation rules

In Section 5.3, we saw that appraisal rules are in charge of generating and updating the emotional state of the agent. When it comes to the knowledge base, their counterparts are the interpretation rules, which are stored in a World Model. Those interpretation rules define all the possible actions and events that the agent can perceive and interpret to generate beliefs. If we have an event that can be perceived by the perception module, we therefore also need to define interpretation rules in the world model to derive beliefs from this event. In short, the interpretation rules are the ones that will look at a `DistanceChanged(Intimate)` event, for example, and determine that this should update the current beliefs of the agent regarding the position of the human. As for appraisal rules, we can define multiple interpretation rules for a same event, as long as we also specify the conditions of their activation. The prototypic interpretation rule would thus look like this:

```
Event : Event([parameters])
Conditions:
- [parameters] = [value]
- Mood(Agent) = [value]
- ...
Effects:
- Belief([var]) = [newvalue]
- ...
```

5.4.2 Determining the information needed to model the acceptance of a touch

As expressed previously, we want to enable touch-based interactions that will not make the human uncomfortable or feel invaded. Before taking the decision of touching the human, the agent should therefore have an idea of the human's acceptance of touch. In Section 2.1 we saw that the acceptance of a touch is based on three elements: the interactive context (in other words: is the touch justified by an emotion expressed or a functional need?), the level of the relationship between toucher and touchee, and the personal preferences of the individual (which we will address as their baseline or *static touch avoidance*). Recording the contextual information is the role of the perception module but interpreting it in a meaningful way is the role of the world model's interpretation rules: based on the events perceived, what is the agent going to understand of the situation? In other words, which

beliefs should be created or updated in reaction to an event? In our case, what we are most interested in are beliefs that can help the model decide when and how to touch. Since acceptance of a touch depends on the interactive context, the level of relationship and the static touch avoidance of the human, those are the things that we want to try to estimate and record in the beliefs of the agent. Information relating to the context will often be scenario-dependent, so we won't detail that here, but it also includes the human's emotional state. If the human expresses sadness, this could indeed be an appropriate time to touch. How can we derive from our perception module's events the level of relationship between the human and the agent and the static touch avoidance and emotional state of the human?

5.4.3 The dimensions of rapport

In our present work, we are mostly interested in developing a proof-of-concept of a social touch-based interactive loop between a human and a virtual agent. For now, we will therefore focus on short-term interactions. As discussed before in Section 2.3, when working on short-term interactions with agents, the notion used to model the level of the relationship between the human and the agent is that of rapport. According to [Tickle-Degnen and Rosenthal \(1990\)](#), rapport is composed of three main elements: positivity, attentiveness and coordination, with coordination being less important at the beginning of an interaction. Considering that, again, it is unlikely that our agent will interact with humans for more than a few minutes, we choose to focus on estimating the rapport via the evaluation of the positivity and attentiveness expressed by the human towards the agent. While attentiveness should not appear too troublesome, positivity is an ambiguous term that requires clarification. In our understanding of the word, it directly relates to the notion of valence (or pleasure) and thus to the emotional state of the human. Since this is something that we were already interested in modelling in our agent's beliefs (see Section 5.4.2, let us start from there.

5.4.4 Estimation of the human's emotional state

To deduce a potential emotional state of the human, we propose to use the touch events and other emotional cues expressed by the human and perceived by the agent. To map those events to an emotional representation however, and to update it over time, we will again need a theoretical model of emotions. In this particular case, we believe that a dimensional model would be well suited to our needs, as we can directly derive values from emotional cues and map those values in a dimensional space. While the most well-known dimensional model is probably the Pleasure-Arousal-Dominance model (PAD), the relationship between touch and dominance is still unclear and studies report different and sometimes contradictory results ([Summerhayes and Suchner \(1978\)](#), [Dibiase and Gunnoe \(2004\)](#)). We instead decide to use the circumplex model of emotion that classifies emo-

tions on a scale of Valence-Arousal (VA) (as described by Posner et al. (2005)). With this model, we can define interpretation rules in the world model that will derive scores of valence and/or arousal from the events perceived by the perception module. Those scores can then be used to update totals (*MoodValence* and *MoodArousal* for example) stored in the beliefs of the agent. The theoretical model of the circumplex however, despite being geometrically represented, is inherited from psychological research where there are no actual values nor calculations. The repartition of emotions on the VA space is thus not rigorously defined in computational terms. To overcome this difficulty, we choose to represent the actual emotional state of the human with the emotional octants of the circumplex model which are shown (with their respective corresponding emotions) in Figure 5.6 (with the A axis for arousal and the P axis for pleasure (valence), thus: HA = high activation, AP = activated pleasant, P = pleasant, UAP = unactivated pleasant, LA = low activation, etc.). To map our values of arousal and valence to an emotional octant, we can then define fuzzy sets and partition tables (similar to the ones used for the classification of distance or touch parameters in the perception module, see Section 4.1) to design a simple classifying algorithm for the current emotional octant of the human.

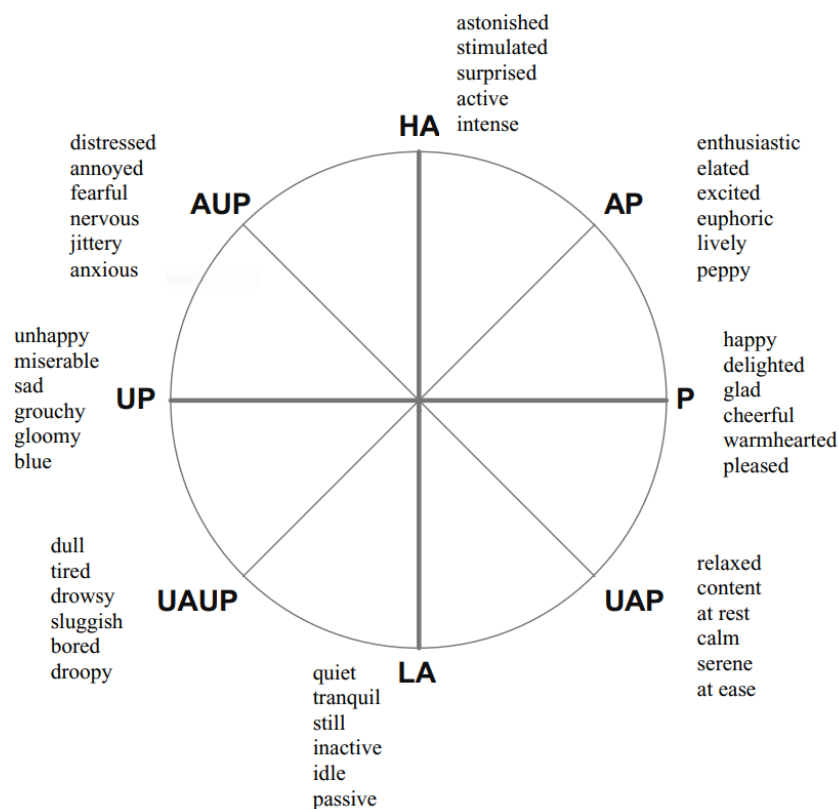


Figure 5.6 – The circumplex model of affect and its emotional octants (by Knez and Hygge (2001))

5.4.5 Estimation of the human's attentiveness towards the agent

As for the estimation of the attentiveness of the human towards the agent, we propose to use proxemics and gaze direction values to compute it, as those are known to relate to attention and engagement Sidner et al. (2003). Similarly to how we proceed for the estimation of the valence and arousal values of the human's emotional state, we thus define interpretation rules for gaze direction and proxemics events that will derive scores to add or subtract to the total attentiveness stored in the beliefs of the agent. An example of such a rule would be: if the human is directly looking at the agent, this should be considered as a high score of attentiveness. Another would then be: if the human stops looking at the agent, we subtract a score, unless we are in a situation of joint attention and the agent is looking at the discussed object.

5.4.6 Evaluating rapport & static touch avoidance

Going back to our estimation of rapport, we now have beliefs representing an estimation of the emotional valence and the emotional arousal of the human, which further gives us an emotional octant. We also have an estimation of the level of attentiveness of the human towards the agent. For the evaluation of rapport, we needed an estimation of the positivity and the attentiveness of the human. Considering the emotional valence of the human as this positivity, we can now use both of the current scores of the beliefs regarding the human's valence and attentiveness to dynamically compute the level of rapport of the human. The actual calculation is further detailed in Section 5.4.7.

Interestingly, the last information that we are interested in modelling in the beliefs of the agent is that of the static touch avoidance of the human. Where rapport, emotional state and attentiveness are all dynamic measures derived from the events happening in the interaction, we feel like static touch avoidance, being a baseline, should indeed remain static. The challenge then is to determine the value that it should take. While it could potentially be possible to assess the static touch avoidance of someone in the course of a long interaction with them, we believe that this would not be possible at the current time, with short interactions. We therefore propose to either have the agent directly ask the human about it upon starting the interaction, or to assess the static touch avoidance of the human by making them fill a touch avoidance questionnaire just before the interaction and directly inputting their score in the agent's beliefs.

5.4.7 Implementation: Interpretation rules and Rapport calculation

In terms of implementation now, the FATiMA model functions by computing the effects of the appraisal and world model's interpretation rules at the same time. Initially, FATiMA's world model only allowed us to define ONE interpretation rule for each type of event, so that we couldn't have multiple concurrent rules for the same event. It also didn't

allow the definition of conditions to regulate when to activate the rules. Considering that touch events can have a lot of different interpretations despite similar parameters, this functioning didn't suit our needs at all. We thus modified the model so that multiple interpretation rules could exist for one event and to implement conditions similar to the ones used for appraisal and decision rules. This required adding a field for each condition in order to specify the character whose knowledge base is concerned by the condition, as multiple characters can observe the same event in the world model, as opposed to what happens in the appraisal and decision process. In the end, the prototypic interpretation rule in our modified version of FAtiMA looks like this:

```
Event: Event([parameters])
Conditions:
- [parameters] = [value] | [Character(s)' KB]
- Mood([Character]) = [value] | [Character(s)' KB]
- ...
Effects:
- Belief([var]) = [newvalue] | [ObserverCharacter]
- ...
```

Where 'character' refers to either an agent or a human, and where each effect is applied on the beliefs of the observers.

We have implemented five main beliefs in the beliefs of our touching agent: human's estimated *attentiveness* and *emotional state* (declined in valence, arousal and emotional octant), *estimated level of rapport*, *current state of the scenario*, and *static touch avoidance of the human*. *Attentiveness of the human* is currently implemented as a discrete scale ranging from 1 to 7, where 1 stands for low and 7 for high: $Attentiveness(Human) = [1, 7]$. The human's estimated *emotional state* is divided into two beliefs implemented as similar discrete scales for valence and arousal: $MoodValence(Human) = [1, 7]$, $MoodArousal(Human) = [1, 7]$. The emotional octant belief is calculated based on the MoodValence and MoodArousal scores and can take the following values : $EmotionalOctant(Human) = HA|AP|P|AUP|LA|UAUP|UP|AUP$ (explained by Figure 5.6)

We want to be able to adapt the measurement of rapport based on what we value more in our specific situation between attentiveness or valence. Certain scenarios may require emphasis on the human's attentiveness, while others require it on the level of the human's experienced valence. It is thus calculated as a function of the values taken by the human's estimated *attentiveness* and the estimated *emotional state's* valence over time. We also want our agent to avoid using touch in an invasive manner and thus we need to emphasize the distinction between low and high values of rapport, while keeping a continuous spectrum of values so that the model can make varied decisions even within low and high values. The value of rapport thus ranges between 1 (lowest) and 100 (highest), with

threshold values that will be added when *attentiveness* and *valence* reach higher levels.

$$\begin{aligned}
 Rapport(Human) = & Attentiveness(Human) * x & (5.1) \\
 & + MoodValence(Human) * y \\
 & + ProxemicSpaceValue + ThresholdValues
 \end{aligned}$$

In this equation, x and y are weights that represent, respectively, the amount of *attentiveness* or *valence* that influence *rapport*. Those weights can be set differently in order to alter the calculation depending on the situation. In our case, we want *attentiveness* to have the same overall importance as *valence*. *ProxemicSpaceValue* is either 0, 5 (personal space) or 10 (intimate space). This is a deliberate choice that makes the agent positively receptive to a human coming close to it, to facilitate touch interactions. Modifying those values would allow us to modify the personality of the agent regarding social touch. The actual values are to be empirically adjusted based on further evaluations.

This equation means that *rapport* is low (resp. high) if either *attentiveness* of the human is low (resp. high) or the *human's estimated emotional state* is expressing a negative (resp. positive) mood. Finally, the *static touch avoidance* of the human is a discrete measure which can take the values of low (tactile person), medium (average touch acceptability) and high (overall averse to touch). The *static touch avoidance* is determined before the interaction starts via a questionnaire and is inputted as a fixed value which will modulate the level of *rapport* required for a touching action to be chosen.

5.5 When and How to Touch a Human

As discussed in Chapter 2, the principles of equipotentiality and equifinality are at play in social touch: a specific touch type can be used to express multiple different meanings whereas a specific meaning can be expressed by multiple different touch types. Furthermore, those meanings of social touch can be both emotional and functional. Table 2.1 showed the mapping of touch types/emotions expressed observed in the literature of human-human social touch. There is however little descriptive research on the instances in which the other functions of touch take place and of such mappings for those functions. To our knowledge, the only works having investigated this question are those by Jones and Yarbrough (1985), already discussed in 2.1, and Héron et al. (2022), who partially based their investigations on Jones and Yarbrough (1985)'s study and identified additional functions of touch during a study of human-human mediated touch.

Considering both their sheer number and the lack of data, it would be a far too tall order to try and model decision rules for all those types of functions at once. We therefore decided to focus our work on a selection of those communicative intentions: attention-getting, turn management, emotional emphasis and supporting touch. Attention-getting refers to behaviours displayed in order to try and get the attention of the interlocutor,

turn management refers to behaviours aimed at taking or giving the floor to the interlocutor in spoken interaction, emotional emphasis refers to non-verbal behaviours aimed at highlighting the emotional component of a spoken statement, and supporting touch refers to behaviours aimed at displaying support towards the interlocutor. The notion of supporting touch is somewhat ambiguous however, and we thus decided to clarify it by separating it into the notions of comforting touch [Cekaite and Kvist Holm \(2017\)](#); [Shaltout et al. \(2012\)](#), encouraging touch [Guéguen \(2004\)](#), calming touch, and a general notion of touching for maintaining rapport. This means that when the decision model selects one of those communicative intentions as the next behaviour to be performed, the touch modality is going to be considered as a valid option to express this intention.

Similarly, we won't model a multitude of touch types and we will instead select the following four: a hit, a tap, a caress and a "neutral" sustained touch. Based on [2.1](#), we believe that they should prove sufficient to express a majority of our selected communicative intentions and emotions.

Those touch types being overall appropriate communicative channels to express those functions does not mean, however, that touch itself, as a modality, should always be used to express them. Furthermore, depending on the situation, other modalities may often be sufficient to express any of those meanings, as evidenced by all the works related to endowing conversational agents with gestural and speaking abilities. As discussed earlier, touch can be a rather invasive modality and the agent should be able to take that into account when deciding upon which modality to use to convey a specific intention. We thus propose to make the decision making process happen in two steps.

First, the decision process will take as input the current state of the interactive scenario to determine all the actions objectively available to the agent at this point. Based on the agent's beliefs regarding its interlocutor and on the agent's own emotional state (as calculated by the appraisal process described in [Section 5.3](#)), decision rules select among those available options a specific communicative intention to convey. For example, if the scenario is not progressing properly and the internal beliefs of the agent indicate an apparent lack of attentiveness from the human, the agent will decide to try to get the human's attention. On the other hand if agitation (high arousal) is perceived in the emotional state of the human, the agent could instead try to adopt a calming behaviour.

Once this communicative intention has been selected, the second step of the decision process is going to determine how to instantiate this intention into a multimodal behaviour. The agent's available communicative channels are speech, gaze direction, gestures, facial expression, and/or touch. It can also move around in the environment. Again, decision rules will specify which of those options are suited for each possible communicative intention, and what are the specific conditions under which they can be used. Among those conditions, the most important ones to consider for the use of the touch modality are the level of rapport and the value of the human's static touch avoidance. It is based on those values that we first determine if touch can happen in an acceptable and non-harmful

way for the interlocutor. Only then are the physical conditions that make a touch possible examined (for example, is the agent in range to perform the touch?). In summary, after those two decision steps, the decision model outputs a communicative intention and the modality to use for its expression. This is then unified with the current step of the scenario to determine the actual utterance to produce (if necessary) and all of this information is sent to the concerned renderers (which will be described in 6) for the instantiation of the multimodal behaviour.

While this answers the question of determining when our agent is going to touch the human, we have yet to explain how a specific touch type is selected. In fact, we will externalize this last step of the decision process directly in the renderers, as this is a micro-decision that FATiMA is not as well suited to manage. The selection of the touch type is thus described in Section 6.1.

In summary, we propose a decision model able to look at the current internal beliefs of the agent regarding the human's perceived actions and emotions (as well as the agent's own ones) to determine which communicative intention to convey next and when touch is an adequate and acceptable communicative channel for it. While we discussed which communicative intentions we focus on here and which types of touch we are going to use, we still need to delve deeper into the actual processing of the inputs and internal states of the model to understand how they are computed to produce a decision.

5.6 Decision Rules

Now that we have determined the general process by which we want our decision making to happen for touch, we need to figure out how to translate this into actual decision rules. In FATiMA, decision rules have a prototypic structure extremely similar to those of the appraisal or interpretation rules:

```
Decision: Action([parameters])
Priority: [value]
Conditions:
- Belief([par]) = [value]
```

The only difference of note is that of the "priority" field. This value is used when two different decision rules have their conditions activated at the same time. In technical terms, the decision method of FATiMA returns all the activated rules by default, but since most of the time we only want to select one behaviour at a time, we can specify the priority of the rules. When two rules are activated at the same time, FATiMA will compare their priorities and the one with the highest priority will be returned first. If even their priorities are equal, then the rule to return will be selected at random. If need be however, we can always access all the rules activated at a decision step directly from the code. For the sake

of clarity, we won't include the priority field in our future examples, as it is overall not that interesting of an information.

Now that we have seen what a decision rule looks like, let us go back to our topic of interest. As explained in Section 5.5, we consider the following communicative intentions: attention-getting, turn management, emotional emphasis, comfort, calm and maintain rapport. To structure the interaction however, we need to distinguish the situation of locutor and that of listener: a same communicative intention should not be expressed in the same way whether someone has the speaker-turn or is listening.

At the moment, we manage this by considering that the agent and the human exchange speaking-turn smoothly. We do not consider interruptions nor barge-in. Only when the agent has the floor can it use the speech modality, but all other non-verbal abilities are available when it speaks or listens. Since we do not focus on the speech of the human, we can keep the turn management that simple for now. Nevertheless, this makes our first condition to every decision rule be a check of whether the agent has the floor or not.

In the case where the agent doesn't have the floor, a last communicative intention is possible: backchannel. A backchannel is traditionally a type of non-verbal behaviour used to manifest to our interlocutor that we are paying attention to what they are saying. It can feature speech ('uh', 'hum', etc.), head nods, facial expressions, arm gestures or even touch. Those modalities can also be combined. A backchannel can however also have further meanings, like an emotional component. This is especially true for touch, as in the case where listening to someone expressing sadness may prompt us to use touch both to manifest that we are listening and for comforting them at the same time. We thus propose four types of decision rules for the 'backchannel' action:

Backchannel(Nod)

Conditions:

1. Has(Floor) != Agent
2. isSpeaking(Human) = True

The nod backchannel is the default one and is always triggered as long as the human has the speaking turn and is currently identified as speaking.

Backchannel(Smile)

Conditions:

1. Has(Floor) != Agent
2. isSpeaking(Human) = True
3. EmotionalOctant(Human) = AP || UAP || P

Backchannel(Sad)

Conditions:

1. Has(Floor) != Agent

5.6. DECISION RULES

2. `isSpeaking(Human) = True`
3. `EmotionalOctant(Human) = UAUP || AUP || UP`

The 'smile' and 'sad' backchannel will trigger the smiling and sad facial expressions (respectively) and have a supplementary condition which checks for the current human's estimated emotional state. Our assumption here is that if the human expresses negative or positive emotions in their speech, it is going to influence the estimation of their emotional state. We thus check for the current emotional octant to determine the polarity of the current emotional state.

As for the touch, we also want to look at all the possible emotional octants but we also need to look at the baseline value of the touch avoidance of the human and which proxemic space they are in. It is based on those three values that we determine the amount of rapport that is required for a touch to be selected. A higher touch avoidance implies general unease with touch and thus will require a higher level of rapport to consider touch as a valid modality. Similarly, in general terms, the further away the human is from the agent the higher the rapport will need to be to consider making the agent cross the distance in order to touch.

`Backchannel(Touch)`

Conditions:

1. `Has(Floor) != Agent`
2. `isSpeaking(Human) = True`
3. `ProxemicSpace(Human) = Intimate`
4. `StaticTouchAvoidance(Human) = Low`
5. `EmotionalOctant(Human) = UP`
6. `Rapport(Human) >= 40`

`Backchannel(Touch)`

Conditions:

1. `Has(Floor) != Agent`
2. `isSpeaking(Human) = True`
3. `ProxemicSpace(Human) = Public`
4. `StaticTouchAvoidance(Human) = Low`
5. `EmotionalOctant(Human) = UP`
6. `Rapport(Human) >= 55`

`Backchannel(Touch)`

Conditions:

1. `Has(Floor) != Agent`
2. `isSpeaking(Human) = True`
3. `ProxemicSpace(Human) = Public`

4. `StaticTouchAvoidance(Human) = High`
5. `EmotionalOctant(Human) = UP`
6. `Rapport(Human) >= 85`

When the agent has the floor however, we want to make the decision in the two steps presented in Section 5.5. First we will determine the communicative intention and then the modality. Each available communicative intention has its own specific conditions: attention-getting is based on the estimated attentiveness of the human (if it is low, then this intention is triggered), turn management is based on the state of the interaction, emotional emphasis is based on the strength of the agent's emotions and the specific dialog utterance, comfort and calm are based on the human's estimated emotional state, and finally maintain rapport is based on the level of rapport. Here we can mention that the priority field of decision rules can be used to specify ethical values or character traits. For example, we could consider that if the human is inattentive and in a bad mood, it is more important to comfort them than to try to get their attention.

`CommunicativeIntention(Comfort)`

Conditions:

1. `Has(Floor) = Agent`
2. `EmotionalOctant(Human) = UP || UAUP`

`CommunicativeIntention(GetAttention)`

Conditions:

1. `Has(Floor) = Agent`
2. `Attentiveness(Human) <= 3`

`CommunicativeIntention(Calm)`

Conditions:

1. `Has(Floor) = Agent`
2. `EmotionalOctant(Human) = AUP || HA`

Once the communicative intention has been decided, we determine the modality to convey it in a similar way to that of the backchannels, except this time there is an extra condition regarding the communicative intention.

`Modality(Touch)`

Conditions:

1. `Has(Floor) = Agent`
2. `CommunicativeIntention(Agent) = Comfort`
3. `ProxemicSpace(Human) = Intimate`
4. `Rapport(Human) >= 50`

5.6. DECISION RULES

This acts as the default condition: if the rapport is estimated to be at least over 50, then the touch modality should always be a valid option for comforting. However, if comfort is the current intention, it also means that the human's estimated valence is rather low, which impacts the level of rapport negatively (see Equation 5.1). We should thus expect low values of rapport, which does not necessarily mean that touch should not be used, as touch is suited for comforting people. We can therefore further discriminate between the situations in the same way that is done with the backchannel rules:

Modality(Touch)

Conditions:

1. Has(Floor) = Agent
2. CommunicativeIntention(Agent) = Comfort
3. ProxemicSpace(Human) = Intimate
4. StaticTouchAvoidance(Human) = Low
5. Rapport(Human) > 20

Modality(Touch)

Conditions:

1. Has(Floor) = Agent
2. CommunicativeIntention(Agent) = Comfort
3. ProxemicSpace(Human) = Intimate
4. StaticTouchAvoidance(Human) = High
5. Rapport(Human) > 40

In those examples, the static touch avoidance of the human impacts how much rapport is necessary to perform a touch when the agent is already in range of the human. Comfort and attention-getting are communicative intentions that are directly linked to the calculation of rapport, which will lead to lower values of rapport when evaluating the conditions. On the other hand, with communicative intentions such as encouragement, calm or turn giving, we can set conditions with much higher expectations regarding the level of rapport.

Modality(Touch)

Conditions:

1. Has(Floor) = Agent
2. CommunicativeIntention(Agent) = Encouragement
3. ProxemicSpace(Human) = Intimate
4. StaticTouchAvoidance(Human) = High
5. Rapport(Human) > 85

Modality(Touch)

Conditions:

1. Has(Floor) = Agent
2. CommunicativeIntention(Agent) = Encouragement
3. ProxemicSpace(Human) = Intimate
4. StaticTouchAvoidance(Human) = Low
5. Rapport(Human) > 60

Those are only a few examples, but as for any rule-based model, a lot of different decision rules need to be created to take every possible value into account. Once the model will be evaluated, all our rules will be made available as soon as possible.

Again, in FATiMA the implementation of those rules stays very close to the prototypes shown above. An additional detail however is that in FATiMA, every possible action (Backchannel(Touch) for example) has an initiator (the agent or human that performs the action) and a target (the agent(s) or human(s) that the action is oriented towards). When the action is performed, this allows the interpretation rules of the World Model to determine who are the agents/humans that are in a position to observe the action and record it in their knowledge base and memory. This means that a decision rule also specifies its initiator and target(s), but we omitted this detail here as in our case those are always going to be the agent and the human.

Something that we have to detail further however is the way we manage the turn-management and progression of the interactive scenario. In FATiMA, dialog actions are defined inside a sort of state machine acting as a dialog manager: each entry has a current state and next state, which allows taking the sequential nature of the scenario into account. We define a dialog action as a 4-tuple consisting of: *Speak(CurrentState, NextState, CommunicativeIntention, CharacterType)*. While the other parameters are self-explanatory, CharacterType simply refers to whether the initiator is an agent, a human or can be both. For now, in our implementation, those *Speak* actions are taking the place of the CommunicativeIntention action we described before. This might be subject to change to enable more complex behaviours later. As for any other action however, decision rules with conditions regulating when the dialog should be selected can be associated with a dialog entry, so this doesn't change the conditions discussed above. Each *Speak* action entry is associated with a dialog utterance, which we can send to the concerned renderers (see Section 6.1).

Therefore, once a *Speak* action is triggered by a decision rule, the communicative intention selected and the next state of the interaction are stored in the knowledge base of the agent. We then immediately trigger a second decision with the updated knowledge base, which will allow the modality decision rules to look at the currently selected communicative intention and apply the rules described above. In FATiMA, a sequence of a *Speak* action decision rule and a Modality decision rule thus looks like follows:

```
Speak([CurrentState], [NextState], GetAttention, Agent)
Conditions:
```

5.6. DECISION RULES

1. Has(Floor) = Agent
2. DialogState(Human) = [CurrentState]
3. ValidDialogue([CurrentState], [NextState], GetAttention, Agent) = True
4. CommunicativeIntention(Agent) = Unassigned
5. Attentiveness(Human) <= 3

Modality(Touch, [NextState])

Conditions:

1. Has(Floor) = Agent
2. NextDialogState(Agent) = [NextState]
3. CommunicativeIntention(Agent) = GetAttention
4. ProxemicSpace(Human) = Intimate
5. StaticTouchAvoidance(Human) = High
6. Rapport(Human) >= 45

As a reminder, the terms put between brackets in FAtiMA are variables which can be substituted with possible values in accordance with the logic programming logic. In the *Speak* action decision rule conditions 2 and 3 make sure the dialog action is possible at the current step in the scenario (the current dialog state is indeed [CurrentState] and the [NextState] is valid). Condition 4 is a simple check to make sure that we are not already in the process of deciding a modality for an already decided communicative intention. Upon activation of this rule, an interpretation rule is triggered to update the beliefs of the agent with the proper values. Finally, the only notable difference between the prototypes shown before and the modality decision rule here is that we keep the information regarding the next dialog state by interrogating the knowledge base.

Finally, it should be noted that communicative behaviours are very linked to the notion of time. We can start talking and only end our speech minutes later, the same way that we can start a touch and keep touching for an extended period of time. In FAtiMA, such continuous events can be modeled through the Action-Start and Action-End properties of actions: an Action-Start signifies the beginning of an action and Action-End its end. So far, we haven't mentioned that at all, but this is because all our decision rules are to be considered as Action-End for now. In the current official implementation of FAtiMA toolkit, once an Action-Start has been performed, the agent is made unable to take another decision, which is incoherent with the way actual behaviours happen: we are very well able to talk and touch at the same time, sometimes for long periods of time. Similarly, the model only allows us to specify the beginning and end of an action. This is problematic since the parameters of a touch can change in the course of its realization, such as a sustained touch becoming a caress. We thus altered the code to allow an agent to perform any number of continuous or non-continuous actions at the same time and added a third type of action: Action-Update. This way we unify continuous actions with an ID and can update their properties over time to inform the agent of what happens. This combined

with frequent updates from the perception module allows the agent to have a better idea of how the interaction is happening. For now however, we only use those time-related properties for the formatting of events perceived by the perception model and for the interpretation and appraisal rules associated to them. In effect, this currently does not change anything to the multiple rules that were described in this chapter.

Now that we have our decision however, the last question we need to address is that of instantiating this decision in a way that the human is going to be able to perceive.

Haptic Device and Agent Animation for Simulated Social Touch

Contents

6.1	Audiovisual Animation of the Agent	73
6.2	Designing a Haptic Device for Simulated Social Touch	75

THIS Chapter will address the matter of instantiating a decision as a perceivable phenomenon for the human. Given the communicative intention and modality of a chosen behaviour, we indeed still need to determine how it should be manifested in the immersive environment and exactly what kinds of movements our agent should perform. Similarly, in the case of a touch, we need a way to have the human actually feel something. Here, we thus present the animation engine and the haptic device used with our touching agent.

6.1 Audiovisual Animation of the Agent

In our immersive setup the agent is made perceptible to the human user via the rendering and the projection of its graphical 3D model. In order to actually perform actions and behaviours however, this graphical model of the agent needs to be properly animated. In our work, this is done thanks to the GRETA platform [Rosis et al. \(2003\)](#) which computes and generates non-verbal animations for humanoid agents.

The GRETA platform is based on the SAIBA model [Kopp et al. \(2006\)](#) illustrated in [Figure 6.1](#), and allows to calculate and synchronize non-verbal and verbal behaviours. Written dialog is annotated with timemarkers and gestural information is associated with

those timemarkers to synchronize the animation of the agent's body with its speech. The speech and prosody are generated from the written text via the CEREPROC™ text-to-speech software. While the platform and its animation system are getting somewhat old, many useful features are available inside of GRETA: we can directly create custom gestures, use already created and validated ones or import motion capture data. We can then organize those gestures by intentions. For example, we might select a few different gestures that can express a negation and create an intention for it in which those gestures are going to be considered as viable alternatives. It is then possible to annotate a written text with the negation intention and the platform will select one of those pre-selected gestures, either at random or based on selection rules. This enables the automatic generation of variable behaviours even for similar communicative intentions. Furthermore, this will allow us to manage the selection of the specific touch type to use with a given communicative intention: for example, we can indicate that a comforting intention can be instantiated either by a caress or a sustained touch and GRETA will make the decision. For now, we can be satisfied with making this micro decision at random (thus maintaining variability in the behaviours), but further work could develop specific decision rules at this level too.



Figure 6.1 – The SAIBA model of behaviour animation planning and realization

Since all the interaction is managed in the Unity3D engine, we use a Thrift-based communication script to make GRETA and Unity communicate. For each animation frame, GRETA sends animation data in the form of coordinates for the joints of the agent's model. However, this method has the drawback of making it interfere with Unity's own Animator so that both cannot be used at the same time. Since GRETA was originally aimed towards static agents for face-to-face interaction on a screen, it isn't able to produce displacement animations such as walking, which is essential to touch interactions where we need to have the agent be able to position itself close to and towards the human. Our rushed solution to the problem is to circumvent it and have Unity be in control of the legs of the agent and GRETA in charge of the upper body. This is a poor solution implemented because of the lack of time, future works involving GRETA and Unity should address this issue in a more elegant way, with animation blending for example.

For our work, we mostly used gestures already available in GRETA but had to create touch gestures for our hit, tap, caress and "neutral" sustained touch, as there were no already existing ones. GRETA was not conceived for real-time interactions in immersive

environments however and does not allow us to specify a target for our touch gestures at the moment. This means that the touch gestures are performed without awareness of where exactly the arm of the human is. This makes it harder to have completely natural looking touches coming from the agent, although we implemented some failsafe measures for making sure that the agent is facing the user and in close proximity before producing the touch gesture.

In the end, this setup nonetheless allows us to translate the outputs of the decision model into actual graphically rendered verbal and non-verbal behaviours for the agent. This is sufficient for most communication modalities, but not for touch, as this modality requires us to address the question of the haptic feedback.

6.2 Designing a Haptic Device for Simulated Social Touch

We discussed in Chapter 2 all the existing options regarding haptic generation and their respective pros and cons. Considering our use of an immersive room we are limited in the kinds of equipment that can be used inside of it, since physical, large and/or static installations are undesirable in order to preserve the stereoscopic 3D projection. The most ideal options would thus be either some kind of wearable haptic device or a pseudo haptic setup.

When looking at the literature on the matter, pseudo haptic seems like a perfect fit for immersive environments tactile interactions. It is however still a young field of research in which many questions are yet to be answered and where most of the work has been focused on object manipulation. This means that at the moment the use of pseudo haptics to interact socially with an agent needs to be specifically and systematically investigated, especially in an immersive room, to determine the conditions in which pseudo haptics can indeed lead to haptic sensations on the part of the human. Since this is not what we are focusing on with our work, we prefer to use an actual wearable haptic device.

Considering our need for a reasonably sized wearable device that would not tire the user out, we decided to go in the direction of vibrotactile technologies, which have the added benefit of being widely discussed and validated in the literature. While vibrations do not provide a very natural and human-like feeling when it comes to the haptic sensation, it was nonetheless shown to be able to communicate recognizable affective information by [Huisman et al. \(2013\)](#) and [Seifi and MacLean \(2013\)](#). Vibrotactile sleeves are the most discussed form of those vibrotactile technologies and have the advantage of being lightweight and placed on an appropriate body part for interaction with potential strangers, the forearm being rather neutral in terms of intimacy ([Suvilehto et al. \(2015\)](#)).

Furthermore, even with small grids of discrete actuators it is possible to produce a sensation of continuous movement by leveraging the phantom actuators and apparent motion illusions. The phantom actuator illusion is produced when two actuators are placed at specific distances of each other and activated at certain similar intensities. The sensation

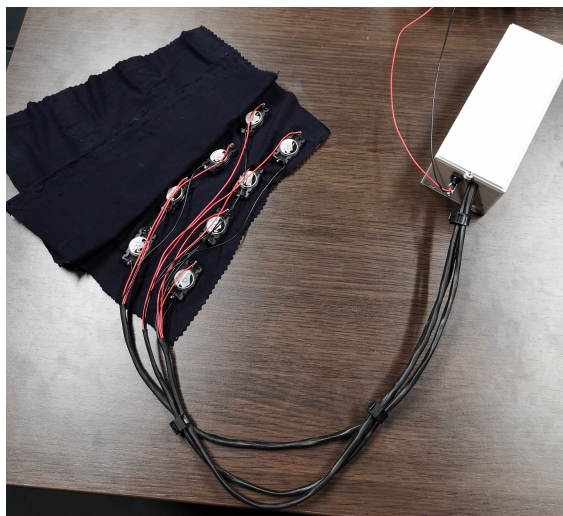


Figure 6.2 – First version of the SOFTLY tactile sleeve

then experienced by the user is not that of two separate vibrations but of a single vibration emanating from somewhere in-between those two real actuators, thus producing the illusion of a “phantom” actuator that doesn’t actually exist. On the other hand, the apparent motion illusion is produced by calculating the optimal successive activation times and intensities of actuators so that one actuator “fades in” while the previous one “fades out”. While the first activated actuator progressively decreases its intensity, the next one starts and slowly increases its intensity so that it will be at the highest intensity when the first actuator has stopped. This way, the user does not feel two discrete activations of two actuators but experiences the illusion of a continuous motion over their skin. By leveraging those two illusions at the same time, we can produce illusions of continuous movements in any direction even with a small grid of actuators. The Tactile Brush algorithm described by [Israr and Poupyrev \(2011\)](#) precisely allows for the automatic calculation of the activation times based on the distance between the actuators and the speed we want to achieve for the movement.

As for the actuators themselves, the most commonly used motors are eccentric rotating mass (ERM) motors but it is difficult to alter their signals in a meaningful way (both in terms of frequency and intensity of the vibration) and their vibrations are thus often quite coarse and unnatural. On the other hand, linear resonant actuators (LRA) allow for a wider range of frequencies to be used and more flexibility in the manipulation of their intensities and activation. After discussing with Professor Vincent Hayward however, we learned of the potential of voice coils (speakers without membrane), in terms of haptic feedback. In a comparison work, [Nunez et al. \(2020\)](#) opposed two vibrotactile devices: one with classic vibrating motors and one with voice coils. In both cases the stimuli were felt like the researchers were expecting it but the device with voice coils was felt as much more pleasant than the other one. In another work on the production of the rendering of

6.2. DESIGNING A HAPTIC DEVICE FOR SIMULATED SOCIAL TOUCH

different textures with a haptic feedback, Romano and Kuchenbecker (2012) voice coils were used on an interactive stylus to make it vibrate to the frequency of a signal recorded during the use of a stylus on a specific surface. This led to the impression by users of being in contact with the original surface even though they were using the stylus on a screen. Voice coils have the ability to reproduce any kind of frequency and signal, thus allowing us to try different options for producing different types of textures that may be felt as less artificial.

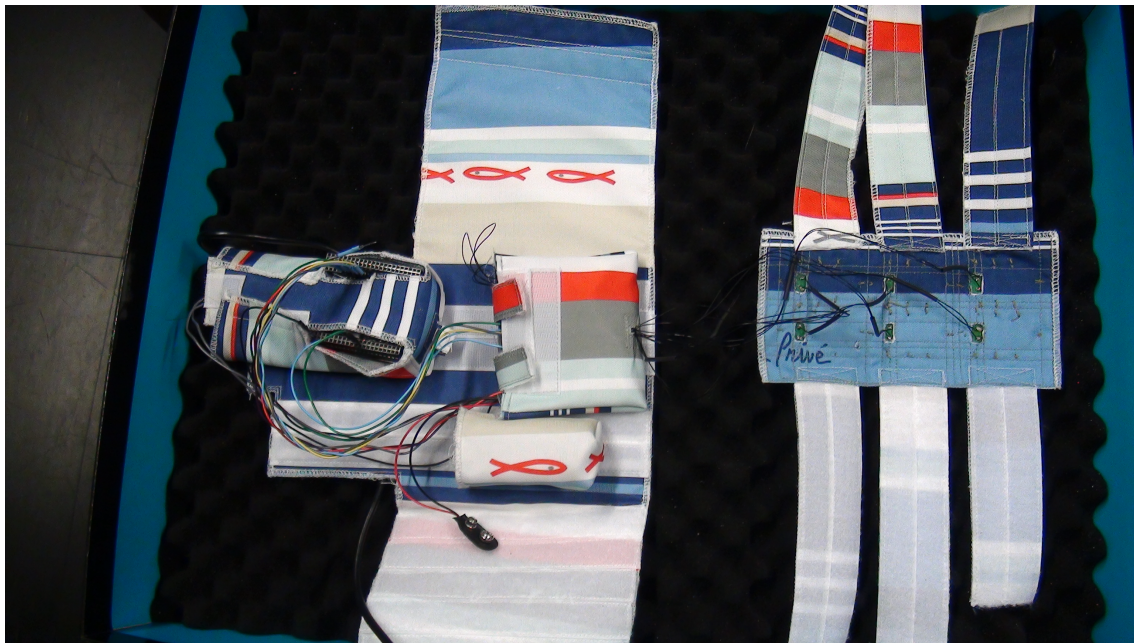


Figure 6.3 – Second version of the SOFTLY tactile sleeve

After a first prototype using classic ERM vibration motors like the ones we can find in smartphones which led to an initial evaluation described in Section 7.1 we decided to build a more robust and interesting vibrotactile sleeve for the forearm, equipped with a grid of 2x3 voice coils and implementing the Tactile Brush algorithm for the production of apparent motion in our stimuli. This new device was called Softly and developed in collaboration with Quentin Duchemin, who dealt with the technical development of the sleeve. This led to a first prototype shown in Figure 6.2 that lacked robustness in terms of the textile that we used, where the electronic components were taking a bit too much space in terms of comfort, and whose voice coils proved to be limited in amplitude. For the second version of the device, we used another more performant type of voice coils¹ and worked with a local sewing workshop for the design and realization of the actual sleeve. We drive the voice coils with amplifiers controlled by a Bela cape mounted on a BeagleBone Black micro-computer. Those class D amplifiers have an I2C interface and a gain control of up to 30dB. The amplifiers are connected to the Bela through a printed

¹HAPCOIL ONE, see <https://www.actronika.com/solutions-for-realistic-haptics#HapCoil-actuators>

circuit board. The circuit board and the Bela are each powered by a cluster of 4 batteries. The voice coils have been sewn into a tarpaulin fabric that has velcro strips, allowing the device to be easily equipped and to fit to any size of forearm. Similarly the circuit board, the Bela and the batteries are put into pockets sewn on another tarpaulin fabric support piece worn on the upper arm. With this design, the device is modular and we can easily make modifications to any individual component without starting from scratch. The final result is shown in Figure 6.3.

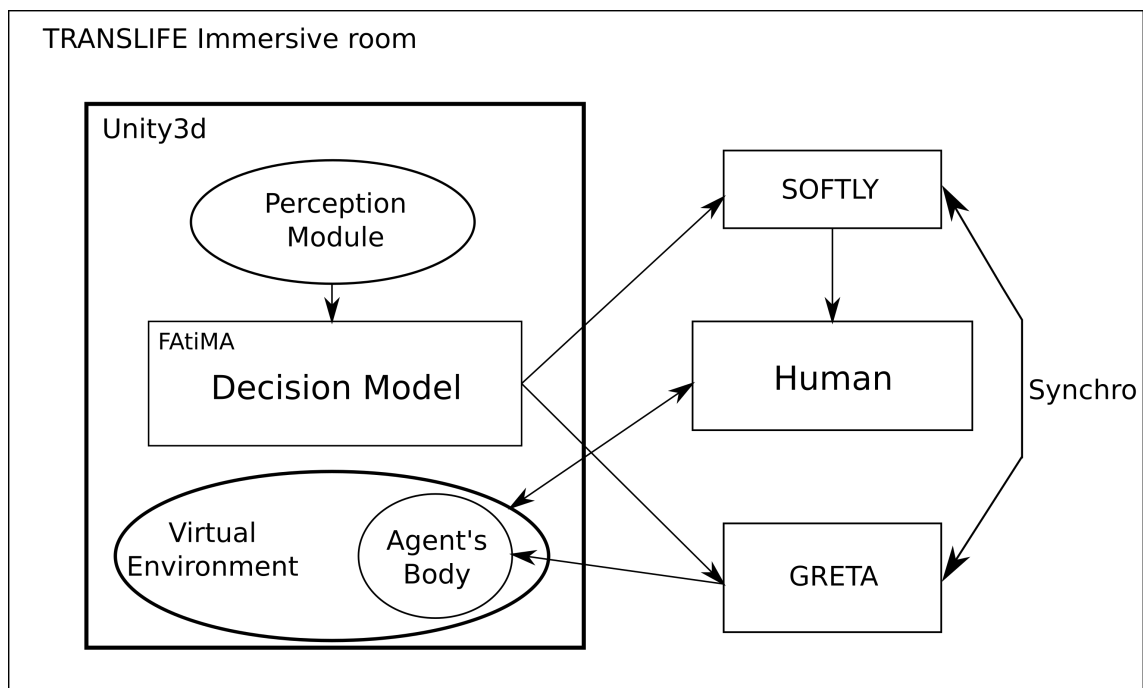


Figure 6.4 – Complete software architecture for a touching agent in immersive environment

To communicate with the device, we use the Open Sound Control protocol. Unity3D then serves as the middleman which receives information from the decision model and sends it to Softly. This information includes the location of the touch, the specific type of touch, the duration, intensity, ramp-up and ramp-down times and modulation parameters. This is computed by the program to produce the signals to send to each of the voice coils. This way, when the model reaches a decision for the agent’s behaviour, that information is simultaneously sent to GRETA for the animation and Softly for the haptic signal, which are then synchronized. As for the stimuli themselves, we produced various signals to try and simulate sensations akin to that of the four types of touch we selected for our agent (see Section 5.5 about this selection): a caress, a hit, a “neutral” sustained touch and a tap. Evaluations of the sleeve and the signals are discussed in Section 7.2.

With that, we have both a way to perceive the human’s actions, to interpret them and choose how to react, and to output the agent’s behaviour with animations and haptic feedback. Figures 6.4 and 6.5 show the complete software framework and interactive process.

6.2. DESIGNING A HAPTIC DEVICE FOR SIMULATED SOCIAL TOUCH

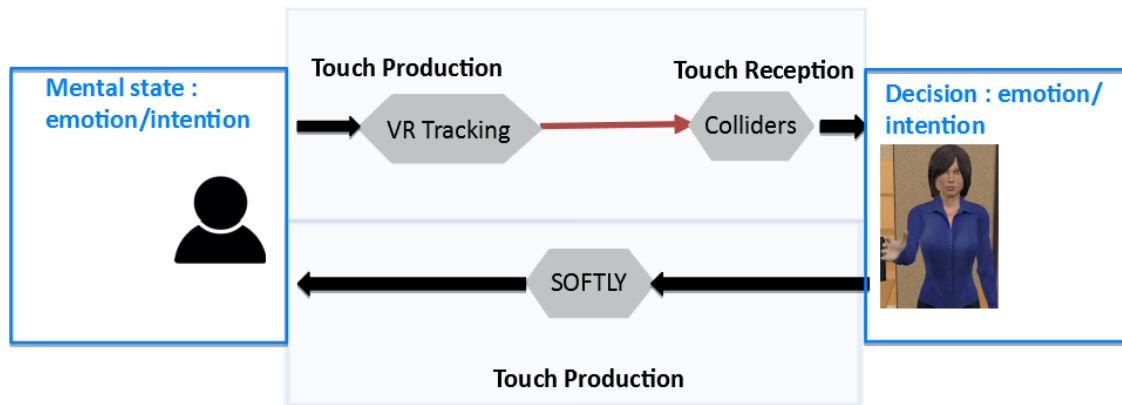


Figure 6.5 – Process and framework of a human-agent social touch-based interactive loop in immersive environment

What is left to do is evaluate the model and the overall framework that we developed, and see what works, what needs to be improved and how our research questions and hypothesis fare in actual context.

Evaluations of the Touching Agent

Contents

7.1	Preliminary study	82
7.1.1	Experimental Protocol	82
7.1.2	Setup and Questionnaire	84
7.1.3	Participants	85
7.1.4	Results	85
7.2	Validation of the Haptic Sleeve Softly	89
7.2.1	Experimental Protocol: Qualification of haptic properties	89
7.2.2	Results	91
7.2.3	Discussion	93
7.3	Use Case Simulation of the decision model	94
7.4	Coherence and Acceptability: Online evaluation of the model	98
7.4.1	Design of the study	99
7.4.2	Research Questions & Measures	101
7.4.3	Procedure	103
7.4.4	Results	103
7.4.5	Discussion	114

THIS Chapter presents all the evaluations realized throughout this thesis work to validate our developments and, most notably, our decision model. Our original research question was to determine whether giving the ability to touch and be touched to an ECA in an immersive context would help it express its emotions to and bond with humans. To answer that however, we first needed to build a framework that would grant such touching abilities to an ECA. This more specifically entailed the

development of a virtual sense of touch to detect when and how the agent is being touched (Chapter 4), the design and implementation of a decision model and appropriate beliefs for the agent to decide when and how to touch the human (Chapter 5), and finally the development and integration of appropriate feedback to manifest the decision with our sleeve SOFTLY and the rendering of the agent’s animations (Chapter 6). To evaluate those contributions, four studies were conducted. In Section 7.1 we present our preliminary work on the basics of bi-directional touch in immersive environment addressing perception of touch and haptic generation (Chapters 4 and 6). In Section 7.2 we present a validation study aimed at determining the current abilities of the latest version of the SOFTLY tactile sleeve (Chapter 6), which was developed based on the observations made during the preliminary study. In Section 7.3 a preliminary evaluation of the decision model (Chapter 5) is made through a use case simulation. Finally, a third-person online evaluation of the decision model (Chapter 5) by human participants is presented in Section 7.4. Because of the pandemic situation and the overall scope of our research, the evaluation of the complete framework in the intended VR context could not be realized in time. We are very aware that without it many of our research questions and the ultimate validity of this work can not be properly answered, but we nonetheless believe that the evaluations presented here will show the value of our work for this field of research.

7.1 Preliminary study

In order to first investigate the basics of bi-directional touch in an immersive context, we had started our work by the development of the first versions of the perception module and the tactile sleeve mentioned in Sections 4.1 and 6.2. To study the credibility of simulated social touch interactions with those basic components, we held a preliminary study in collaboration with Quentin Tafiani. The study was conducted in the immersive room TRANSLIFE of the Université de Technologie de Compiègne with those first versions of the perception module and the tactile sleeve, as well as the animation engine GRETA. The main goal of the study was to determine whether our basic framework would be sufficient for humans to have natural touch interactions with the agent, and whether the agent could reliably communicate emotions with the human participants. The decision model is not used in this study.

7.1.1 Experimental Protocol

The experimental protocol consists in a familiarization phase and two distinct phases of the actual study in which the participant first has to touch and then is touched by a virtual agent. The study has a between subject design in order to prevent fatigue as well as any learning effect. We thus split participants in three groups with each group having

7.1. PRELIMINARY STUDY

a different combination of the emotion participants have to transmit and the emotion transmitted to them by the agent.

Before the actual experiment, we put the participant in the familiarization test environment in order to reduce the novelty effect. There the participant can get familiarized with the virtual environment, the haptic feedback and a virtual agent, although different than the one used in the latter phases.

After this familiarization phase, the actual environment (see Figure 7.1) is displayed and the participant is faced with the virtual agent of the study, who is first oriented with its back towards the human. The participant must thus first get the attention of the agent by placing themselves on the white marking and touching the agent. This will prompt the agent to turn around and to introduce itself as Camille and to explain what the participant will have to do.



Figure 7.1 – The immersive environment of the study, with the agent

Phase 1. The first phase is divided in two sessions of touches by the participant. In each session, the participant is asked to try and express an emotion to the agent via a touch with the vibratory glove. To indicate the emotion that should be conveyed and its intensity (low or high), an emotional scenario is read to the participant. In total, two scenarios are thus read to the participant, one for each session, and both indicate the same emotion but with a different intensity. By using two distinct emotional intensities we hope to determine whether the participant uses different kinds of touch for the different levels of intensity of the emotion. Practically, the participant can do as many as four touches in a session and is told that they are free to use any touch type they feel appropriate, while also being warned that only their hand is tracked and recognized by the agent. The agent reacts to the touch and emotion transmitted with a pre-selected adequate facial expression.

There are three possible emotions, one for each condition (group of participants) of the experiment: sympathy (C1), anger (C2) and sadness (C3). This makes for six emotional scenarios in total. Those emotions were chosen for their very different polarities from each other and for being a priori easy to understand by most people. Emotional scenarios are translated in French based on research works by Scherer et al. (1991). As an example, the following low emotional intensity scenario was used for the "sympathy" emotion (translated from French): "You meet a friend of yours, Camille, that you haven't seen for some time. You express what you are feeling to her." On the other hand, high emotional intensity scenarios involve more emphatic expressions.

Phase 2. In the second phase, it is now the turn of the virtual agent to touch the participant. Similarly to the first phase, there are also two sessions of four touches in this phase, and it is the same emotion that is being expressed in both sessions. This emotion is different from the one expressed in Phase 1 to prevent any kind of learning effect. The difference between the sessions this time is the touch type (gesture and haptic pattern) that is being used. The touch types used by the agent are stroking and tapping for the sympathy; tapping and hitting for anger; and stroking and simple sustained touch for sadness. The touch gestures are made so that they will visually connect with the arm of the participant where the vibratory sleeve is worn. The touch is further accompanied by a facial expression adequate to the emotion being expressed.

As it was discussed in Chapter 2, the physical properties of a touch are not sufficient to unambiguously reach the correct interpretation of social touch. We chose to add other non-verbal cues, facial expressions and gestures to the agent, so that we can evaluate if this setting is already good enough for a minimal setup allowing humans to properly interpret the agent's social touches.

A questionnaire is presented to the participant in-between each session of the experiment and at the end. After having answered the last questionnaire, the participant is debriefed about the experiment.

7.1.2 Setup and Questionnaire

For this preliminary study we use a wizard-of-oz type of procedure where the reactions of the agent are prepared in advance and activated by the person conducting the experiment.

The agent is monitored and animated through the use of the GRETA software platform (see Section 6.1 for more details). As for the questionnaires, they are based on works by Demeure et al. (2011). For the first phase, participants are prompted to describe the types of touch they chose to use, so that their answers can be compared with the information recorded by the perception module. The participants are also asked to evaluate how credible the reaction of the agent to their touch appeared to them. Here we define credibility as the degree to which the participant feels the agent behaved itself in an adequate, human-like way.

In the second phase, participants are this time asked to describe the tactile sensation produced by the vibrotactile sleeve when the agent touched them and to identify it as a touch type. Finally, participants were asked to determine to what degree they felt like the agent was expressing sadness, anger or sympathy, or any other kind of emotion they believed they had felt, and to evaluate to which degree they considered the behaviour of the agent as credible and why.

7.1.3 Participants

The study took an average of one hour by participant and was conducted with twelve participants in total (four by condition), among which there were eight women and four men. Nine of those participants had no prior experience of virtual reality. For this study we had not used a touch avoidance questionnaire and only asked participants to evaluate their receptivity to touch themselves. Ten considered themselves as having a good touch receptivity (they thought they received touch well) and two didn't know. All the participants were between 18 and 39 years old and were of occidental culture. Mean age value was 23,25 and standard deviation was approximately 5,7897.

7.1.4 Results

Subjective data was gathered with 5-items Likert scale items. Since it was a preliminary study and we had very few participants (twelve split in three groups of four), we didn't perform tests of significance and our results should be taken as indicative data.

7.1.4.1 Touching and Being Touched

Unexpectedly, and even though they had been clearly informed that only the glove was tracked and taken into account by the perception module for their touch on the agent, all the participants used a type of touch that we had anticipated as inadequate to virtual reality at least once. In the case of sympathy and sadness that unexpected touch type was the hug, and almost all of the participants tried to use it with the virtual agent.

Seven out of the eight participants that encountered the vibration pattern and animated gesture that corresponded to a stroking recognized it correctly, by name, and more than half of the participants concerned could identify the patterns that simulated both the hit and the tap. However, no participant identified the "neutral touch", which can be explained by the fact that "neutral touch" is not a natural term nor a very specific term.

All the participants that were in the condition where they had to transmit sadness through touch (C3) expressed big difficulties about how to touch the agent for this emotion.

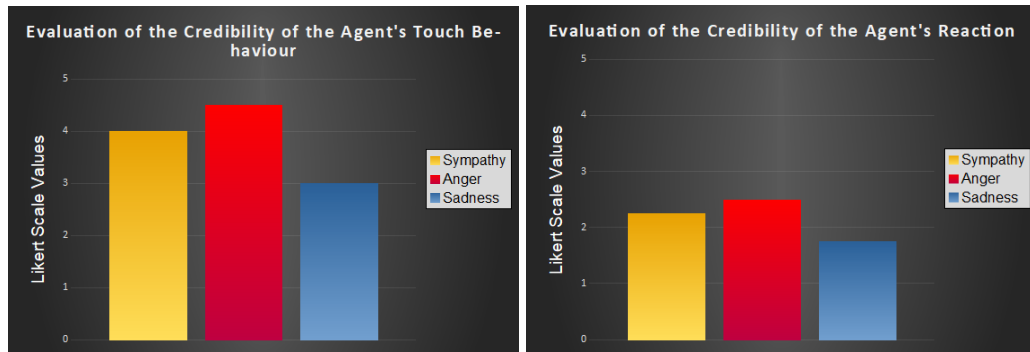


Figure 7.2 – Credibility of the virtual agent’s behaviour according to participants.

7.1.4.2 Overall Credibility of the Touch Interaction and the Agent’s Behaviour

The results shown in Figure 7.2 indicate that the agent appeared as more credible when it touched the participants to express anger (red column) and sympathy (yellow column), with the participants rating its credibility around or above 4 on average. The agent was however perceived as much less credible when it reacted to being touched or when it tried to express sadness. In their answers to the questionnaire, participants remarked that facial reactions were hardly noticeable when they touched the agent, which can explain the low credibility of the agent when it was being touched since facial expressions were the only prepared reaction in this setup.

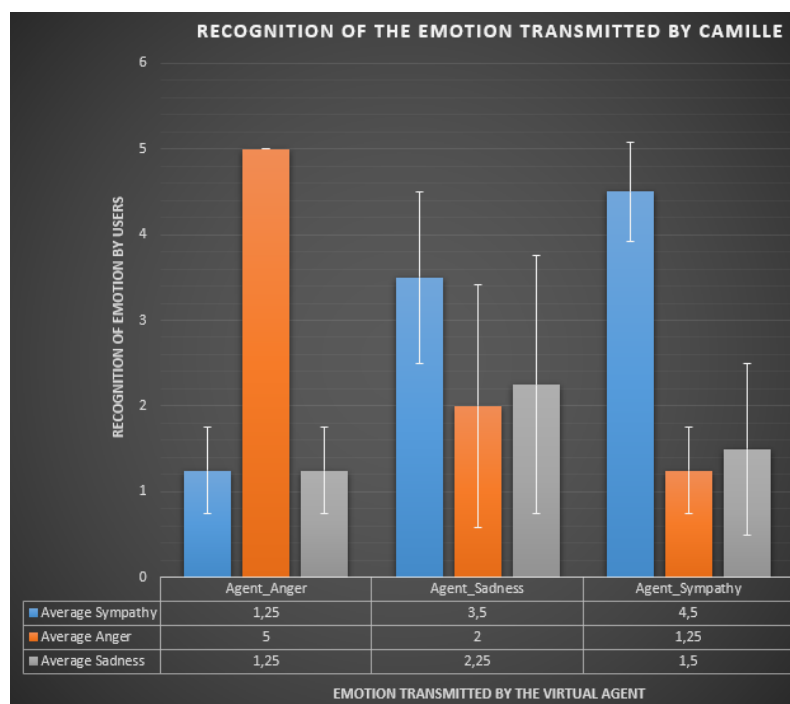


Figure 7.3 – Recognition rate of the emotion transmitted by the agent.

The results shown in Figure 7.3 indicate that the emotion transmitted by the agent was correctly recognized as anger in group C1 and as sympathy in group C3 by almost all the participants, but that the group that was confronted to sadness had a lot more trouble to correctly identify the emotion. We can add that half of the participants from group C2 said that the agent was trying to comfort them or to be compassionate instead of expressing sadness.

7.1.4.3 Discussion

We believe that our results overall support the idea that social touch is a viable modality to enhance emotional communication between human and agent. It notably shows that under favorable conditions agents using touch to express emotions can be considered as credible by humans. On the other hand, participants found the credibility of the reaction of the agent to touch extremely lacking. In particular, the agent was apparently not perceived as having noticed the touch performed by the participants. The fact that participants massively expressed not having noticed the facial reactions of the agent hints that the non-verbal behaviour of the agent can be massively improved. Furthermore, there was minimal context to the interaction and setting up a better narrative environment should go a long way in making the agent more active. The implementation of a computational model of emotion would similarly improve the responsiveness and adaptive abilities of the agent, thus allowing it to participate in an actual interactive loop, instead of the controlled situation used here.

Another interesting result is that sadness was both poorly recognized and hard to transmit through touch. When asked about it, participants said that when they feel sad they usually aren't looking to go touch someone, they are instead expecting to be touched (in order to be comforted or shown empathy). It thus appears that sadness is an emotion that prompts touch from others but is not directly expressed via touch. Sorry-for and compassion thus appear as more appropriate in a social touch context, especially in answer to sadness expressed by others.

It is also noteworthy that even though results were very encouraging about the recognition rate of the types of touch simulated with the sleeve, all the participants expressed their concern that vibrations didn't feel like an appropriate feedback for imitating the natural touch sensations. When asked what kind of perceptive substitution they would have preferred, participants emphasized how much they would like to experience some form of resistance when touching the agent, and described force-feedback devices. Such devices could indeed give a more realistic sensation of touching something with a physical presence, but are less suited to VR environment such as an immersive room. We believed that iterating over the haptic sleeve with better actuators would improve the feeling of the agent's touch, which prompted us to work on Softly later on.

Despite this, participants didn't hesitate to use unexpected types of touch that we had thought inadequate in the context of virtual reality, such as hugging or pushing, all of those being types of touch requiring some kind of physical resistance from the object being touched. While participants absolutely realized that only their hand was detected and received haptic feedback, they still tried to use the types of touch that seemed the most natural to them to express the emotion they had to express. Furthermore, when asked about this, participants didn't express uneasy feelings from the lack of haptic feedback in that instance.

Among the other possibilities that can be explored, one of the participant remarked that the vibratory sensation might have seemed less surprising and more credible if there had been some sort of mediation of the touch and the vibratory feedback through some kind of physical tool, such as a HTC Vive controller or any other command device of this kind, instead of the glove. It seemed to the participant that such a proxy would have made the vibrations feel less dissonant, since it would have used a tool that doesn't look like it aims at perfectly imitating the sensation of natural touch. This idea seemed particularly interesting to us considering that social touch is overall a rarely used social interaction modality in our daily-lives (at least outside ritualistic usages and more intimate relationships), but is, on the other hand, our main modality of interaction with technical objects and tools. [Leroi-Gourhan \(1964\)](#) has shown how by becoming bipeds and thus freeing their hands, our main touching organs, the first humans have been able to develop themselves technically and cognitively through the handling of external tools. Using some kind of proxy to mediate our touch in a virtual environment could therefore be a relevant and interesting way to produce a credible social touch sensation even with a sensory feedback very different from the actual sensation of touch. In the context of virtual reality, such a mediation coupled with pseudo-haptics could greatly enhance the quality of the perceptive substitution. The question remains as to what kind of mediation tool could be relevant in the context of virtual reality. How using such a proxy would influence the behavior of the human towards the agent also needs to be studied with more attention, as it could potentially put distance between them. We however feel this is a research topic in itself and outside of the scope of our work for now.

Regarding the perception module of the agent, it can be noted that no particular results have been reported here so far. Although the detection of touch did happen correctly in that the agent knew when it was touched and could react to the touch instantaneously, the rest of the raw data recorded by the perception module highlighted its drawbacks. The most obvious one was that the resolution of the tactile cells grid was much too low to allow a good enough localization of the touch. Most importantly, when a touch was happening right at the junction between two tactile cells, we found that the data recorded by the module would end up being interpreted as a movement oscillating between one and the other cell based on the idle movements of the agent's body. Similarly, where the experimenter had observed stroking behaviours from the human, we could not find

matching data in what was recorded by the module, most likely because the movement had happened inside a single tactile cell. Furthermore, considering the number of times the human participants had tried to hug, squeeze or shake the agent, we felt that having some idea (no matter however much estimated considering the immersive context) of the force exerted on the agent by the human would be invaluable for our work.

Based on those results, we felt that we could improve upon both our tactile sleeve and our perception module. For the perception module we realized that we could use the experimenter's observations of the humans' behaviours to map raw values with simple classification of values (when is a velocity "fast" as opposed to "slow" for example). We also realized the limitations of our resolution-based tactile cells grid and the lack of information regarding the force exerted on the agent. Those observations led to our development of the second version of the perception module presented in Section 4.1.3. Because of the pandemic we haven't been able to validate this second version of the module yet.

As for the tactile sleeve, we felt that we could especially improve on the finesse of the vibrating sensations. This was done by developing another prototype of tactile sleeve. This new version of the sleeve led to the development of SOFTLY as described in Section 6.2, which hardware and software allow for a much wider variety of haptic patterns. To validate the fundamental psycho-physical and qualitative properties of SOFTLY we designed another study which will be presented in Section 7.2.

7.2 Validation of the Haptic Sleeve Softly

In order to validate the sensations produced by our new haptic sleeve Softly described in Section 6.2, we collaborated with Maxime Grandidier to conduct a study about how actual participants perceive and qualify its haptic feedback. Most importantly, our goal in this evaluation of the device is to determine under which modalities variations in the haptic feedback is perceptible by human users and whether the haptic illusion of apparent continuous movement is properly working. By determining the perceptive thresholds between two haptic patterns we will be able to produce nuanced haptic patterns to use in actual interaction.

7.2.1 Experimental Protocol: Qualification of haptic properties

As the intensity of a touch is an important interpretative factor in social touch, the first goal of this study is to determine for which degrees of intensity the haptic feedback is perceived by humans. This will allow us to validate our designs of strong and soft touches for our haptic patterns. The first part of the experiment is thus divided in two tasks: one for the definition of the absolute intensity threshold (for which degree of intensity is a haptic feedback perceptible by a human user?) and one for the relative intensity threshold

(what minimum difference in the degrees of two intensities is necessary for a human user to perceive an increase or decrease in intensity between two feedbacks?).

It is in a second part of the experiment that users will be exposed to haptic patterns that implement haptic illusions. To determine if those illusions are properly perceived by the users, there will again be two tasks: one is to determine the localisation and number of actuators used in a haptic pattern (so that we can check whether movements and the phantom actuator illusion described in Section 6.2 can be perceived), and the other is to check for the actual apparent continuous motion perception by sending various static and moving haptic patterns to the participant.

7.2.1.1 Absolute intensity threshold

In the first task of the study, the user receives a haptic stimulation via Softly either on the side of the sleeve closest to the elbow or on the one closest to the wrist. The participant is then asked on which side of the sleeve the stimulation happened. If the participant gets the wrong answer (indicating the elbow when it was in fact closer to the wrist for example), the next stimulation will have a one step *increase* in its intensity. If the participant gets two correct answers in a row however, the intensity will be *decreased* by one step in the next stimulation.

We count how many times the participant goes from one correct answer to an incorrect one and from an incorrect answer to a correct answer (which we will call an "inversion") and we use this information to control the experiment. After the first three inversions, we will decrease the intensity by two steps instead of one, so that the intensity converges faster towards the threshold value where the stimulation is barely perceptible but still perceived. The task stops once twelve inversions have happened and we take the mean of the last ten stimuli for our final threshold value for a participant. This way we have a measure of the intensity at which a user goes from not feeling the sensation to feeling the sensation. This minimal value can then be generalized from the results of all participants and will become our standard for soft touches. It can be noted that this task is based on the assumption that as long as the stimulus is perceived, the user can't mistake one side of the sleeve for the other.

7.2.1.2 Relative intensity threshold

The second task of the study focuses on the definition of the minimal threshold necessary for a human user to differentiate between two degrees of intensity. The task is very close to the one described for the absolute intensity threshold. This time however, we send a series of three stimuli (with a pause after each stimulus) to Softly. Two of those stimuli have the same intensity A and the third stimulus has an intensity $A + e$. If the participant correctly identifies the stimulus with the highest intensity twice in a row, we decrease the value of e , but if the participant makes a mistake, we increase the value of e . This

task has the same general principle as the previous one: after the first three inversions, e is decreased by more than the initial decrease value to make the value converge faster towards the actual minimal threshold. After twelve inversions, the task is stopped and we take the mean of the last ten e values as our final difference threshold.

7.2.1.3 Localization and number of actuators

As we discussed it in Section 6.2, our haptic patterns make use of a few haptic illusions to simulate touch patterns and hide the static nature of the actuators. The third task of the study is aimed at testing the phantom actuator illusion: when two actuators are activated at the same time with a certain intensity and distance between them, humans feel a single haptic sensation between the actuators instead of two stimuli at the locations of the actuators.

In order to test the precision with which people can locate the stimuli, we send a stimulus produced by either one or two actuators. The participant is then asked to indicate on a graphical grid (which represents the sleeve) where they felt the sensation. They can choose up to four cells from the grid for each stimulus. In terms of results, we compare the actual position(s) of the actuator(s) used in the stimulus with the cells chosen by the participant. The participant's score is evaluated as the number of actuators found and we further calculate the difference between how many cells the participant chose and the actual number of activated actuators in the stimulus. This difference can allow us to determine whether the perception area is bigger than that of the actuators.

7.2.1.4 Qualitative properties

The last part of the evaluation focuses on the illusion of continuous motion. This one is however much harder to evaluate and we thus chose to directly ask participants about the qualitative properties of a selection of haptic patterns designed to simulate our four touch types selected in Section 5.5 (hit, caress, tap and sustained touch). Since some of those touch types are static and others have continuous motions, we still get information about the illusion we are interested in. Regarding this property, the participants were asked to rate how continuous the movement appeared to them on a 7-item semantic differential scale (discontinuous (1) to continuous (7)).

7.2.2 Results

The study was conducted with a total of 14 participants, with seven men and seven women. All the participants were between 15 and 50 years old with a mean of 27.6 and a standard deviation of 8.7. Due to technical problems however, not all the participants could complete all of the tasks.

For the first task regarding the absolute intensity threshold, 13 participants could complete the task. Their results are summarized in Table 7.1. The mean intensity across participants of 0.25 and a standard deviation of 0.23.

Mean intensity of the ten last stimulations
0.42
0.11
0.32
0.12
0.45
0.0074
0.21
0.020
0.020
0.75
0.09
0.45
0.038

Table 7.1 – Results of the first task

Only 9 participants could complete the second task regarding the relative intensity threshold. Again, their results are compiled in Table 7.2, with a mean difference threshold of 0.33 and a standard deviation of 0.12 across participants.

Mean of the ten last recorded differences of intensity
0.41
0.38
0.42
0.32
0.52
0.20
0.16
0.23
0.28

Table 7.2 – Results of the second task

All of the 14 participants could complete the third task with the results displayed in Table 7.3. The mean of the difference between the actual number of actuators and the number of cells selected by the participants is 1.72 and the mean for the number of actuators correctly identified is 0.25, with respective standard deviations of 0.65 and 0.19.

Finally, 11 participants could complete the last task. Their results are summarized in Table 7.4, with a mean of 4.7 across participants for the patterns corresponding to a caress and a mean of 1.5 for stimuli based on a tap. The respective standard deviations were 1.8 and 0.7.

7.2. VALIDATION OF THE HAPTIC SLEEVE SOFTLY

Mean difference real actuators/cells selected	Mean number of correct actuators identified
0.64	0.071
1.6	0.4
0.60	0.1
0.86	0.066
2.5	0.125
1.23	0.076
1.85	0.35
2.15	0.61
2.35	0.5
2.21	0.071
2.00	0.40
2.25	0.33
2.00	0.14
0.73	0.13

Table 7.3 – Results of the third task

Caress continuity perception	Tap continuity perception
7	1
1.5	3
3.4	2.3
5.6	1
2.5	1.25
5	2.2
5	1
5.5	1
3.2	1.25
6.7	1
6.3	1.42

Table 7.4 – Results of the fourth task

Considering the exploratory nature of this study, we did not perform significance tests with our results.

7.2.3 Discussion

The first thing we could verify was that the sleeve was indeed wearable, can be operated at a distance without wires and is robust over multiple sessions.

The results from the first task regarding the absolute intensity threshold allowed us to determine our minimal intensity as a baseline for the design of soft haptic patterns. Similarly, the results from the second task gave us a baseline regarding the minimal intensity increase or decrease to respect for different haptic stimuli. It should be noted however that the units we are using for the intensities are relative to our implementation of the

device and thus cannot be easily generalized. Further studies of the device should try to use more indicative types of units.

The third task's results on the other hand highlighted the fact that participants' perception appears very imprecise, which could be explained by a low density of discriminating mechanoreceptors in the forearm. Participants perceive stimuli in larger areas than those occupied by the actuators and also on places where no actuators are present, which somewhat comforts the idea that the phantom actuator illusion is perceived, but the results are neither precise nor decisive enough on that point. Furthermore, the type of signal (wave form) used might have an influence on how easy and precisely the sensation is to be felt and further exploratory experiments should be led on the possibilities offered by Softly.

From the results regarding the qualitative properties of our haptic patterns, we can observe a large difference between the continuous perception of a caress opposed to that of a tap. Apparent tactile motion thus seems to be indeed perceived by participants to some extent.

Overall we are satisfied by the results of Softly, and we believe that the results are sufficient to pursue the use of the sleeve with the VR environment and agent. In-context evaluation is however necessary for an actual assessment of the social touch haptic patterns, particularly regarding the emotional interpretation of the patterns. The protocol for this next study is ready and implemented, it features the presentation of our multiple haptic patterns mixed with our various emotional facial expressions (or none). This should allow us to determine whether the facial expression or the haptic pattern feel coherent with each other and which modality drives the emotional interpretation the most. The study will therefore allow us to validate whether our haptic patterns are suited to meaningful autonomous human-agent touch interactions.

7.3 Use Case Simulation of the decision model

Because of the COVID-19 pandemic, it was very difficult to have human participants to come in the immersive room for the evaluation of our framework. As discussed in Sections 7.1 and 7.2, this notably prevented us from validating the second version of the perception module and the in-context properties of SOFTLY. To evaluate our work we nonetheless had to find alternative ways to test the decision model, so that we could discuss the current abilities and limitation of our entire framework.

The first evaluation of the decision model we conducted took the form of a simple use case simulation of the model where we simulated all the inputs we fed to the model instead of having an actual human participant. The aim of this use case was to look at a complete instance of interaction and the decisions produced by the model so that we could have a first idea of what seems coherent and what seems out of place in our design of the model and its decision rules.

The implementation of the decision model used here is anterior to the one described in Section 5.6, but the details of this implementation can be found in (Boucaud et al. (2021)). Most notably, values may be different than those of the rules formerly described (Section 5.6), continuous actions could not be computed by the agent, and this version of the model does not use emotional octants to describe the emotional state of the human. Finally, the modality of the behaviour was chosen at the same time as the communicative intention.

This simulation situation however allowed us a larger variety of inputs, so we designed a task-oriented scenario (in the form of a game). Since we couldn't validate nor use our interfaces for the perception and generation of touch, we choose to focus on evaluating the decision making process and especially the decision of touching or not for the agent. Directly switching to an online setup with participants remotely interacting with our agent did not appear as a viable option for us as there are important perceptual differences between what we prepared with the VR setup and what would happen with a desktop monitor. We therefore simulate all the inputs of the "participant" and try different values of attentiveness, emotional state and static touch avoidance as initial conditions. In the scenario, the simulated participant *P* must reproduce a sequence of colours and input the colours in the order indicated by the agent *A*. *P* faces a screen with five numbered and colour-filled boxes; *A* stands next to them. To complete the task, *P* must remember which box is filled with which colour, and then *A* indicates in which order each box must be filled. This task is performed twice, with a different random sequence of colours each time. In this scenario, the goals of *A* are to build an above-average level of rapport and help *P* to accomplish the task.

The interaction is divided in the following steps:

1. The participant *P* is greeted by the agent *A* and asked how they feel today. Their answer sets an emotional baseline.
2. *A* decides how to react to *P*'s choice and indicates how to perform the first task.
3. *P* realizes the first task.
4. *A* reacts to their performance and prompts them to try again if the task was not a success.
5. Once *P* has succeeded, *A* reacts to the performance and gives indications for a second and last task.

We model *A* so that it believes that the emotional state of *P* becomes more negative when failing to perform the task; an interpretation rule of *A* also states that a failure is an indication of a low attentiveness from *P*. We thus hypothesize that *P* is emotionally affected when making mistakes and that it is a mark of their lack of attentiveness. We set the influence of the game success rate on *P*'s state of mind as follows. The task can

either be perfectly done with the right colours in the right boxes and the right input order (which leads to an increase in the perceived emotional state and in the estimated attentiveness), be done correctly in terms of colours but not in terms of input order (which *A* does not comment about, but evaluates as improving *P*'s emotional state and attention), be done with one error in terms of colours (decrease in *P*'s estimated emotional state and attention) or be completely failed (decrease in *P*'s estimated emotional state and an even larger decrease in attention). Before the beginning of the interaction, the static touch avoidance estimated for *P* is set directly in *A*'s beliefs. For this simulation, this is also done for the initial level of attentiveness and emotional state. To simulate different types of participants, we present three simulations where the participants differ in those initial levels of attentiveness, touch avoidance and emotional state.

In our first simulation, the initial inputs are the following: $Attentiveness(P) = 6$, $MoodValence(P) = 1$ and $StaticTouchAvoidance(P) = High$, which indicates a high level of attentiveness, a very negative emotional state but little appreciation for touch in general. The results of this simulation are as follows:

1. The equation 5.1 presented in section 4.2 is used to compute the level of rapport, which is estimated as: $Rapport(P) < 50$. Since $MoodValence(P) = 1$, the decision model outputs: $Speak(Step1, Step1, Comfort, Gesture)$, with no touch considering the high level of touch avoidance. A second decision is made to make the scenario progress: $Speak(Step1, Task1, Inform, Gesture)$. The model decides to give the information regarding the task, again with no touch modality.
2. Result of the task: *P* succeeds, only making a mistake about the input order: as a consequence, *A*'s belief regarding *P*'s mood has become more positive; since *P* has performed well in the task, *P* may have been attentive; the agent thus perceives an increase in its rapport with the participant: $MoodValence(P) = 2$, $Attentiveness(P) = 7$, $Rapport(P) = 60$
3. This is positively appraised by *A*, which generates a positive improvement of its own mood. Based on those values, the decision model outputs: $Speak(Task1, Step2, Congratulate, Gesture)$, with no touch since rapport is not high enough to counterbalance the high touch avoidance.
4. Even though the estimated emotional state of *P* is still quite low, the high level of attentiveness and the fact that *P* succeeded in the task make up for it, and *A* chooses to directly give the next information regarding the second task, instead of comforting again, without using the touch modality.

The second simulation uses the following variables: $Attentiveness(P) = 2$, $MoodValence(P) = 5$, $StaticTouchAvoidance(P) = Medium$, which indicates a low level of attentiveness, a positive emotional state and no strong feelings against the use of social touch in general.

7.3. USE CASE SIMULATION OF THE DECISION MODEL

1. $Rapport(P) = 60$ prompts the following output from the decision model: *Speak(Step1, Task1, Inform, Gesture)*, with no touch since the level of rapport is average and this is not a communicative intention where touch has priority.
2. Task: P has made a mistake on one colour, leading to a drop in the estimated variables. $MoodValence(P) = 4$, $Attentiveness(P) = 1$, $Rapport(P) < 50$
3. This is negatively appraised by A . Based on those values, the decision model outputs: *Speak(Fail1, Fail1, GetAttention, Touch)*, with touch since it is an appropriate modality for this communicative intention, rapport is not too low and the static touch avoidance is not high. P reacts to the touch by looking at A , which improves the attentiveness. Then A invites the participant to try again. This time P manages to find the right sequence, but not the right input order: $Attentiveness(P) = 4$, $MoodValence(P) = 5$, $Rapport(P) = 75$.
4. Since rapport is now high, the decision model first outputs a congratulation with the gesture modality, and then decides to directly give the information regarding the last task: *Speak(Step2, Task2, Inform, Touch)* where touch is used for its function of maintaining rapport.

For the third simulation, we use the following variables: $Attentiveness(P) = 3$, $MoodValence(Human) = 1$, $StaticTouchAvoidance(P) = Low$. This indicates a low-medium level of attentiveness, a very negative emotional state but a general appreciation for tactile interactions.

1. $Rapport(P) < 25$ prompts the following output from the decision model: *Speak(Step1, Step1, Comfort, Touch)*. Despite low rapport, touch is selected because the touch avoidance is low and touch is considered very adequate for comforting.
2. Task: P completely fails, which brings all variables to lowest values. $MoodValence(P) = 1$, $Attentiveness(P) = 1$.
3. This is negatively appraised by the agent. Based on those values, the decision model outputs: *Speak(Fail1, Fail1, Comfort, Gesture)*. This time touch is not selected because touch had already been attempted before and was immediately followed by a negative result. A invites P to try again with a gesture, which leads to a partial success.
4. $Attentiveness(P) = 2$, $MoodValence(P) = 2$. The decision process outputs congratulations without touch, because rapport is still far below the mean value. Then, it outputs: *Speak(Step2, Step2, Comfort, Touch)*. Here, a comforting touch is attempted again because of the low static touch avoidance. Finally, A informs P of the last task without touch.

Overall, the outputs of the model are mostly coherent with what we expected based on the literature. The selected communicative intentions fit the interaction context, and rapport and touch avoidance are taken into account in the selection of the modality. For this specific scenario we had added a rule that was selected in the third simulation and led the agent to not touch the human again as it believed the human could have failed the task because of the agent's touch. However long-term temporality of actions is still hard to take into account with our decision model, and thus the agent still ends up touching the human later in the interaction. In that regard some of the touching decisions may appear incoherent. Furthermore, this use case simulation highlighted the importance of temporality in those kinds of interactions and prompted us to work on the modelling of continuous actions, as described in Section 5.6.

While those results are encouraging, only an evaluation with actual human participants can allow us to determine whether those rules are actually coherent and acceptable in context from a "naive" perspective.

7.4 Coherence and Acceptability: Online evaluation of the model

With the outbreak of COVID-19 in France and the lockdowns and restrictions that followed, and considering the time constraints relative to the PhD, it soon appeared that it would not be possible to achieve the multiple validation studies that would have been required for the perception module, the haptic sleeve (in-context) and the decision model. Although it was a very difficult and disappointing choice, we thus decided to abandon the VR setup and haptic sleeve for our final evaluation. We are aware that this will make us unable to actually conclude on many of our original research questions. More specifically, this will prevent us from making any decisive statements regarding the actual effect of our framework and agent on an immersed user. We believe however that this was the best choice considering our situation and that the study and results we will present in this Section will still be very valuable contributions to our research field and will help us to provide avenues of improvement and guidelines for future works.

To be able to have a sizable number of human participants evaluate our decision model and part of our framework, we therefore decided to design and conduct an online study. Since we feel that we cannot hope to actually evaluate whether the touch modality would improve the agent's emotional and relational abilities towards a human participant (our original research question) with such a remote study, we will instead focus on a different aspect of the interaction. What we are most interested in at this point is the degree of coherence and acceptability of the modality choices of the functional decision model in an emotional interaction. Are our fundamental choices regarding the modeling of social touch decisions through communicative intentions and rapport correct? How does the agent's touching behaviour appear to a remote observer?

7.4.1 Design of the study

To make the study available to as many participants as possible, we developed an online evaluation based on videos and questionnaires to watch and fill directly in a web browser. Since there are massive perceptual differences between the VR setup we had worked on (level of immersion, perception module, haptic feedback) and a desktop based interaction, we felt that allowing participants to directly interact with the agent online could end up extremely misleading in terms of results as there is little evidence that the participants would actually feel involved in the interaction and would understand the touches (only displayed visually) as directed at themselves. The model would also have been even more severely deprived in terms of inputs.

Instead of having participants directly interacting with the agent, we thus decided to design a third-person validation study. There is indeed a sizable literature regarding the effects of observation of social touch interactions, which have shown that we can expect reactions similar to those of people directly touching or being touched even in vicarious touch situations (Keyser and Gazzola (2009), Schaefer et al. (2012), Walker et al. (2017)). In studies on social touch human-human interactions, vicarious protocols involving participants answering questions about videos of other people touching or being touched are indeed quite common and produce results in line with those produced with direct interaction protocols. We thus decided to adopt this form of study for our evaluation of the decision model.

To build video examples of the decisions produced by the model, we chose to develop a basic emotional scenario and three alternative sequence of inputs for a simulated human participant (represented by a virtual avatar) to interact with our agent (equipped with the functional decision model). As we want to focus on the performance of the decision model regarding emotional communication, we designed a non task-oriented situation, in which the simulated participant and the virtual agent discuss their mutual experiences of the COVID-19 lockdowns and restriction. Since we prepared the experiment at the time of the early 2021 lockdown, we used our own experiences of the situation to design the dialog and we thought that the online participants would not have any trouble to relate to the discussion.

The emotional scenario consists in three main parts:

- *Part 1* (Introduction) takes the form of a familiarization phase: the agent (*Camille*) initiates the interaction by greeting the simulated human participant (*Laura*) and introducing itself and the environment. *Laura* is then asked about how she is feeling, her familiarity with immersive environments and is finally prompted to tell *Camille* as soon as she feels ready to proceed with the actual interaction.
- *Part 2* (Pandemic2020) involves several turns of social chat with *Camille* first describing its experience of the first French lockdown of 2020 and then asking *Laura* about her own.

- *Part 3 Pandemic(2021)* similarly involves several turns of social chat but this time *Camille* and then *Laura* express their feelings about their experiences of the latest lockdown (early 2021)

We chose to use two agents with female appearances based on the literature on social touch (Suvilehto et al. (2015)) that shows that women are generally more accepting of touches. This necessarily represents a risk of perpetuating social norms that may not be ultimately desirable, but this also allows us to put aside the gender variable from this first study of the decision model. Similarly, we consider our agents to be of a western culture and don't focus on this variable either.

To test the adaptation abilities of the decision model we use three different sets of attitudes (attentiveness, initial emotional state, familiarity with virtual environment, valence of the lockdown experiences) for *Laura* which will constitute three conditions for our study. Once the inputs of *Laura* were prepared for each condition (see below for the specific characteristics of each condition), we set up a virtual avatar in our Unity3D environment that we could control to move and interact in the environment with the autonomous touching agent. *Laura* however does not display non-verbal behaviour outside of touch gestures, gaze direction and movements in the environment. By directly controlling this virtual avatar in the virtual environment (in the way we control a video game character for example), we play out the entire interaction and use the perception module to perceive proxemics and approximated gaze direction of *Laura*'s avatar to feed to the decision model, which then produces its decision and sends it to Unity and GRETA for the movement, animation and speech. The dialog actions of *Laura* were prepared in advance and we activated them at the appropriate time during the interaction. Each condition thus led to a video filmed live in the application with real-time inputs and decisions of the model.

Figure 7.4 showcases examples of shots extracted from the videos with the virtual avatar controlled by us and the autonomous touching agent. In the end, we obtained three videos of about 5 minutes each with the following characteristics:

- *Video 1 (V1): Laura* is manipulated to appear highly attentive and in a good mood (high mood valence): we made her look at the agent and state that she feels well. Her static touch avoidance is fixed as low in the beliefs of the agent: *Camille* believes that *Laura* has a low touch avoidance. *Laura* is accustomed to virtual environments and expresses having experienced difficulties during the first lockdown but far less during the latest one.
- *Video 2 (V2): Laura* is manipulated to appear very inattentive and in a somewhat negative mood (low mood valence): we made her not look much at the agent and state that she does not feel great. Her static touch avoidance is fixed as medium in the beliefs of the model. *Laura* has not much prior experience with virtual environments and expresses having experienced difficulties during both lockdowns.

7.4. COHERENCE AND ACCEPTABILITY: ONLINE EVALUATION OF THE MODEL

- *Video 3 (V3):* *Laura* is manipulated to appear very inattentive and in a good mood (high mood valence). Her static touch avoidance is fixed as low in the beliefs of the model. *Laura* is accustomed to virtual environments, and expresses not having experienced too much difficulties with the lockdowns.



Figure 7.4 – Shots extracted from the videos of the online study of the model's decisions, with the agent (Camille) in a blue shirt and the simulated participant (Laura) in a teal one.

Finally, in order to facilitate the examination of the videos for the online participants, we decided to split each video in three, with one subvideo for each part of the interaction described above. This gave us 9 videos in total: V1-Introduction, V1-Pandemic2020, V1-Pandemic2021, etc.

7.4.2 Research Questions & Measures

With this study we aim to evaluate whether the decisions of our model regarding when to touch are coherent and acceptable from an actual human perspective. This implies both evaluating the fundamental assumptions of our decision rules and more generally determining the conditions in which touches are perceived as the most coherent and acceptable.

The main research questions can thus be summarized as follows:

- (Q1.1) are the emotional state, attentiveness and rapport of *Laura* estimated similarly to the model's estimations? (Q1.2) Is the perceived level of rapport between *Laura* and *Camille* correlated with the ratings of emotional state and attentiveness of the human?
- (Q2) is the perceived level of rapport correlated with the ratings of the "correctness" (coherence and acceptability) of a touch?
- (Q3) how do the ratings of the "correctness" of touch vary within the three conditions?

- (Q4.1) are the communicative intentions associated with the touches (their functions) well understood? (Q4.2) Is there any correlation between specific intentions and the ratings of the "correctness" of a touch?
- (Q5) after participating in the study, do participants see touching agents in a positive way?

To answer those questions, we used the following measurements. In addition to a few standard demographics questions, we collected information about the profiles of the online participants regarding their familiarity with VR and medias in general, their attitudes and a priori regarding virtual agents, and their touch receptivity profile. Those were respectively assessed via a 2-item scale for the media habits, a selection of items from the Negative Attitudes towards Robots Scale (NARS) adapted for virtual agents and translated in French, and a slightly adapted, and translated in French, Touch Receptivity Questionnaire from [Bickmore et al. \(2010\)](#).

After watching a video, the participant is first asked to evaluate the valence of the mood, the degree of arousal and the level of attentiveness of Laura using a 7-point Likert scale item. This allows us to compare the estimation calculated by the decision model's rules with the assessment of the participants. Similarly, we then ask the participant to estimate the state of the rapport between Laura and Camille on a similar scale.

Regarding touch, for each subvideo (each part of the interaction), the online participant is first asked if Camille touched Laura. If it did, we then follow up with a question about the function of the touch they observed with 6 fixed functions proposed (those considered by our model: attention-getting, comfort, encouragement, turn management, emotional emphasis and calm) and an "Other" answer which must be specified with free text input from the online participant. In order to evaluate how adapted to the situation the touch appeared to the online participants, we built a 4-questions Likert scale with the following questions (translated from French): "based on the state of the situation did you feel like the touch was coherent?" ; "based on the level of rapport between them did you feel like the touch was adapted?" ; "did Laura seem to accept Camille's touch?" ; "had you been the one interacting with Camille, would you have accepted the touch she performed?" Respectively, the first question assesses the coherence of the touch in terms of its function in the scenario, the second question is interested in the adequacy of the touch considering the level of rapport between Laura and Camille, and the last two questions assess the level of acceptability of the touch. Each question is scored as a 7-point Likert scale item. The final scores obtained by the addition of the results of these items can be interpreted as the level of "correctness" of the touch decision featured in the video. The individual results of the items can further be studied to determine whether there is an effect of the first two questions on the answers to the last two questions (is acceptability indeed influenced by the rapport between Camille and Laura and by the coherence of the touch?). Finally, there are two final free text questions: one asking for a description of the

touch gesture (evaluation of the quality of the animation) and the other being a qualitative assessment of whether the online participant think they would also have used touch in this situation, had they been in the shoes of Camille.

After answering the questions regarding the touch of Camille, a shorter section is dedicated to questions regarding the touch of Laura. If the participant did observe Laura touch Camille, they are asked about the function, coherence and adequacy of the touch with the same items as the ones used for Camille's touch. One additional question was a qualitative free text assessment of how the online participant would have reacted had they been in the shoes of Camille. This last question allows us to better understand how to model the decision rules regarding how to react to a touch.

Finally, after all the subvideos have been watched and their respective questionnaires answered, there are three final questions: one asks the participant whether they would like to interact with our agent (Camille) in the future (to assess what the participant felt about its behaviour), one that asks more generally if they find the idea of being touched by virtual agents interesting, and one that asks more generally if they find the idea of being able to touch virtual agent interesting.

7.4.3 Procedure

The study is entirely hosted online. Upon accessing the webpage, the participant is asked for their consent to participate in the study. Then, each participant is assigned to one of the three conditions V1, V2, V3. Participants are told that the videos they will see are all part of the same interaction and that this interaction involved a real human participant (Laura) controlling a virtual avatar and interacting with our agent. We showed pictures of the agents to make clear to the participants which character was the human subject (Laura) and which was the agent (Camille). To reinforce this impression that Laura was in fact a real subject, we had an actual human dub her dialog lines. All the online participants see the three subvideos (parts) of their video condition (V1-Introduction, V1-Pandemic2020 and V1-Pandemic2021 for the V1 group for example) but each subvideo is presented separately from the others (on its own page) with its own questionnaire, as detailed in Section 7.4.2. Once all the subvideos and their questionnaires have been seen and answered, the three final questions regarding the appreciation of Camille and of the idea of touching and being touched by an agent are displayed on a separate page.

7.4.4 Results

7.4.4.1 Participants profiles and distribution across conditions

67 subjects completed the study, with 36 males, 28 females and 3 other. Participants were aged between 18 and 73 with a mean of 37.5 and a standard deviation of 13.7. All of them

were francophone and 94% were of the French nationality (with the remaining exceptions still being of an occidental cultural background).

Cronbach's alpha for the 10-item touch receptivity questionnaire was 0.857 which indicates a good internal consistency. The cronbach's alpha of the attitude towards virtual agents 7-item questionnaire was 0.67 however, which indicates a questionable internal consistency. There were no significant differences in the scores of touch receptivity across genders (mean of 4.13 for males, 4.09 for females, 4 for other on a 1-to-7 scale) nor in the scores of attitude towards virtual agents.

Condition V1 had 22 participants, condition V2 had 23 participants and condition V3 had 22 participants. Touch receptivity scores means (and standard deviations) for the V1, V2 and V3 conditions were 4 (1.13), 4.1 (1.25) and 4 (1.14) respectively. A one-way ANOVA showed that there were no significant differences among those means $F(2,64) = 0.049$, $p > 0.1$. Similarly, means (and standard deviations) of the NARS scores for the V1, V2 and V3 conditions were 3.32 (0.89), 3.1 (0.76) and 3.39 (0.97) respectively, with no significant differences. Participants were thus homogeneously distributed across conditions in those regards.

Using their results on those two scales, we further divided the participants between high and low groups for touch receptivity and positive and negative groups for their a priori towards agents. The classification was made based on the median values with participants with a lower score than 4.2 being classified as having a low touch receptivity and the rest as having a high touch receptivity. On the other hand, since the second scale's polarity is reversed, participants with a lower score than 3.1 were classified as having positive a priori about virtual agents and the rest as having negative views of virtual agents.

7.4.4.2 Emotional state, attentiveness and rapport

In the following sections we will address the three conditions as V1, V2, V3, referring to the complete video interactions. For the subvideos we will add a 1 (Introduction), 2 (Pandemic2020) or 3 (Pandemic2021) to the name of the discussed raw variables when necessary. For the ratings of mood for example, Mood1 will thus refer to the scores rated for the first subvideo (Introduction), Mood2 to the second, etc.

Figures 7.5 and 7.6 summarize the evolution of the ratings of mood and attentiveness across conditions and parts of video. Figure 7.7 does the same for the ratings of Rapport. Normality tests showed that most of the distributions of those ratings were not normal, so we decided to use non-parametric tests to check for the statistical significance of our results. In the first (Introduction) subvideos we find the following:

- Mood was rated significantly higher in condition V1 (mean = 4.7, sd = 0.75) compared to both V2 (mean = 3.5, sd = 1.03) and V3 (mean = 3.5, sd = 1.18). This was confirmed by a Kruskal-Wallis test (chi-squared = 18.1776, df = 2, p-value < 0.01)

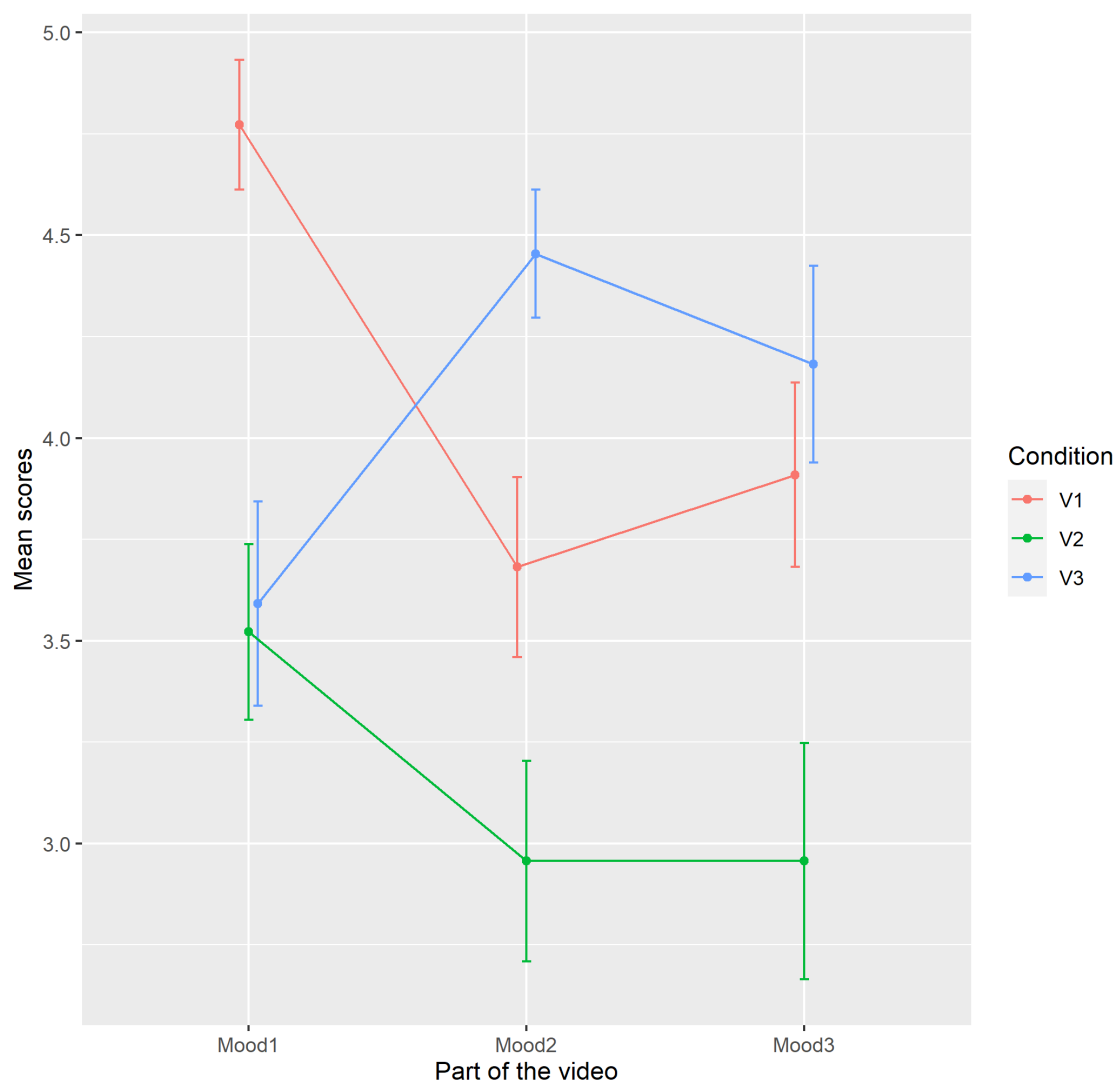


Figure 7.5 – Evolution of the ratings of Laura’s mood across conditions and subvideos.

and multiple comparison of the groups using the Conover-Iman test ($p < 0.001$ for V1 - V2 and V1 - V3, p values adjusted with the Bonferroni method).

- Similarly, Attentiveness was rated significantly higher in V1 (mean = 4.7, sd = 1.07) compared to both V2 (mean = 3.30, sd = 1.63) and V3 (mean = 3.31, sd = 1.46), confirmed by a Kruskal-Wallis test (chi-squared = 12.7521, df = 2, $p < 0.01$) and multiple comparison of the groups using the Conover-Iman test ($p < 0.01$ for V1 - V2 and V1 - V3, p values adjusted with the Bonferroni method).
- Again, the same thing is observed for Rapport with V1 (mean = 4.5, sd = 0.59), V2 (mean = 3.65, sd = 0.83) and V3 (mean = 3.68, sd = 0.89), confirmed by the Kruskal-Wallis test (chi-squared = 16.8725, df = 2, $p < 0.001$) and multiple

comparison of the conditions ($p < 0.001$ for V1 - V2 and V1 - V3, p values adjusted with the Bonferroni method).

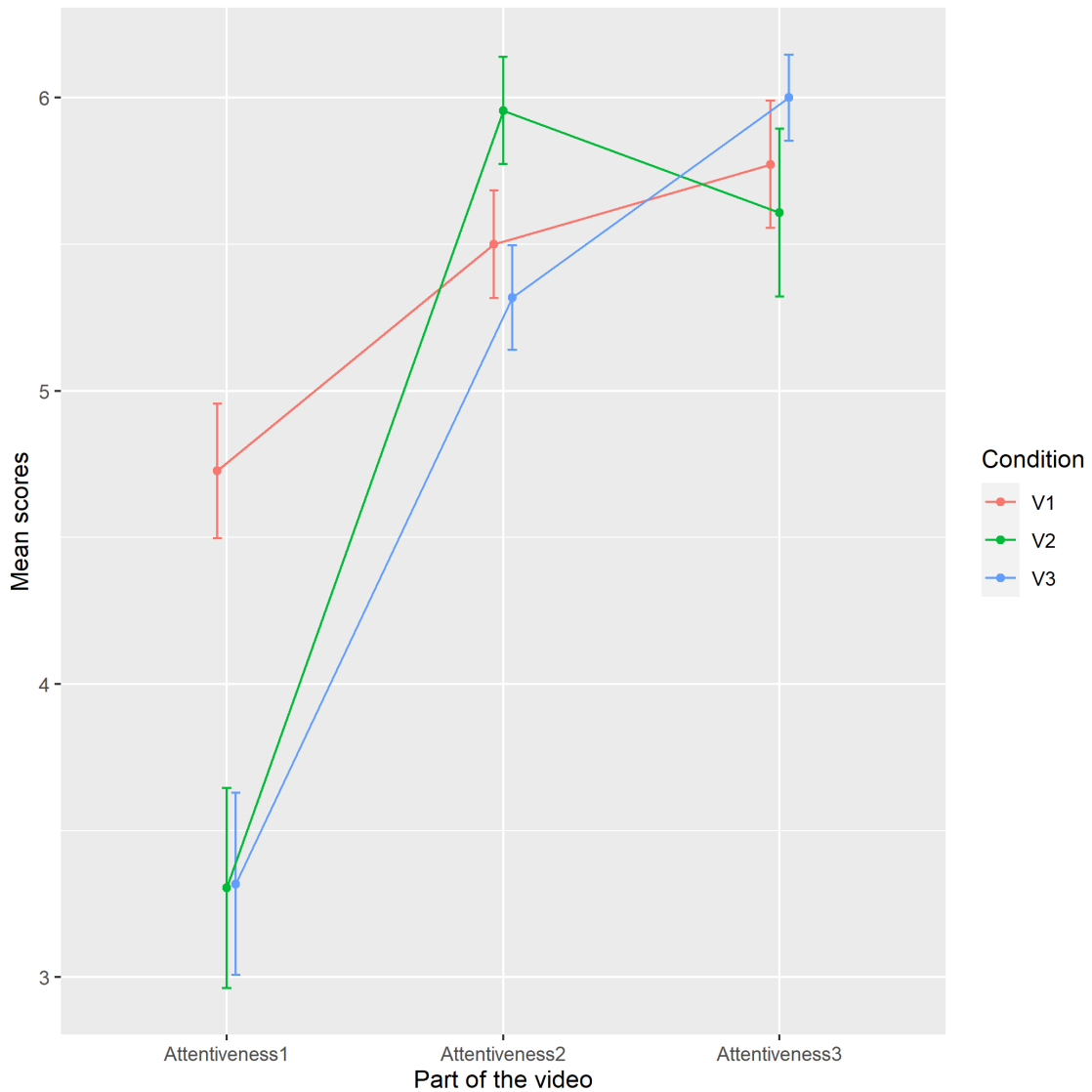


Figure 7.6 – Evolution of the ratings of Laura's attentiveness across conditions and sub-videos.

In the second (Pandemic2020) subvideos we observe the following:

- The ratings of the mood of Laura were V1 (mean = 3.68, sd = 1.04), V2 (mean = 2.95, sd = 1.18), V3 (mean = 4.45, sd = 0.73). A Kruskal-Wallis test showed that there were significant differences across the conditions (chi-squared = 20.565, df = 2, $p < 0.001$) and multiple comparison of the groups using the Conover-Iman test showed that both V1 and V2 had significantly different ratings from V3, but that the ratings between V1 and V2 were only approaching significance ($p = 0.06$ for V1 - V2 and $p < 0.05$ for V1 - V3, p values adjusted with the Bonferroni method).

7.4. COHERENCE AND ACCEPTABILITY: ONLINE EVALUATION OF THE MODEL

- Attentiveness was also rated with a significant difference among the conditions with V1 (mean = 5.5, sd = 0.85), V2 (mean = 5.95, sd = 0.87), V3 (mean = 5.31, sd = 0.83) based on a Kruskal-Wallis test (chi-squared = 6.2676, df = 2, $p < 0.05$), but multiple comparison of the conditions using the Conover-Iman test only highlighted that this significance only applied for the ratings of V3 compared to V2 ($p < 0.05$, p value adjusted with the Bonferroni method).
- As for Rapport, we had V1 (mean = 4.86, sd = 0.77), V2 (mean = 5.21, sd = 0.9), V3 (mean = 4.81, sd = 0.73) and although we can observe slightly higher ratings for condition V2 the Kruskal-Wallis test showed that this difference was not significant (chi-squared = 2.8605, df = 2, $p > 0.1$).

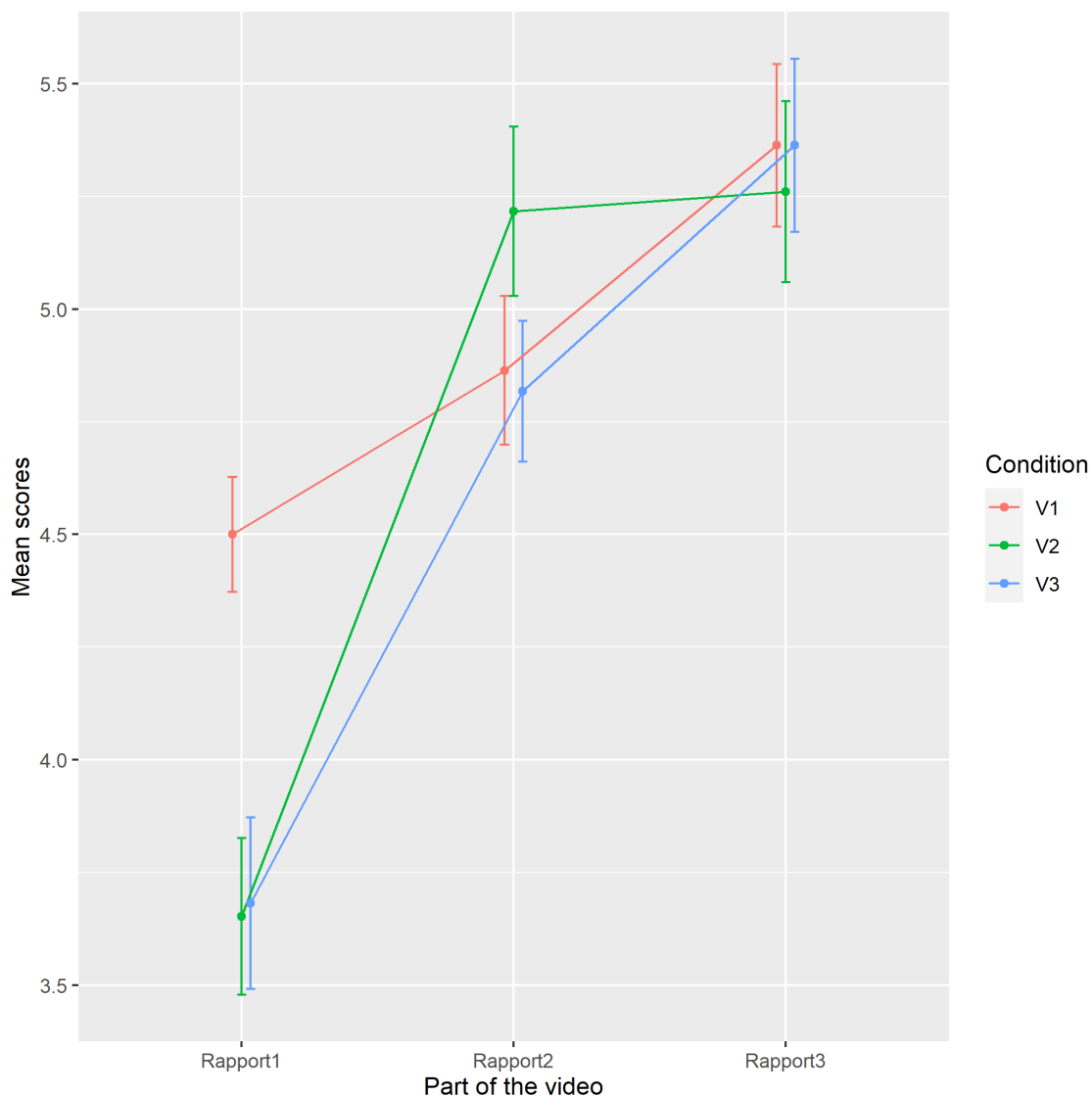


Figure 7.7 – Evolution of the ratings of the level of rapport between Laura and Camille across conditions and subvideos.

Finally for the last (Pandemic2021) subvideos:

- Mood ratings of condition V2 (mean = 2.95, sd = 1.39) were shown to be significantly different from those of V1 (mean = 3.9, sd = 1.06) and V3 (mean = 4.18, sd = 1.13) by a Kruskal-Wallis test (chi-squared = 11.3253, df = 2, p-value < 0.01) and multiple comparison of the conditions using the Conover-Iman test (V2 - V1 (p < 0.05) and V2 - V3 (p < 0.01)).
- Attentiveness' ratings were V1 (mean = 5.77, sd = 1.02), V2 (mean = 5.6, sd = 1.37) and V3 (mean = 6, sd = 0.69) with no significant differences according to the Kruskal-Wallis test (chi-squared = 0.6011, df = 2, p > 0.1).
- Similarly to Attentiveness, Rapport was overall rated as follows: V1 (mean = 5.36, sd = 0.84), V2 (mean = 5.26, sd = 0.96), V3 (mean = 5.36, sd = 0.9) and the Kruskal-Wallis test didn't highlight any significant differences between the conditions (chi-squared = 0.1647, df = 2, p > 0.1).

To answer question Q1.2 we investigated the effects and interactions of the ratings of attentiveness and mood on the ratings of rapport. Considering the ordinal nature of all those likert items ratings we decided to use ordinal logistic regression and found the following over the 3 conditions:

- For the Introduction subvideos, only the ratings of Mood had a significant influence on the ratings of Rapport (t-value = 1.97804, p < 0.05).
- For the Pandemic2020 subvideos, the ratings of Mood and Attentiveness both had an influence on the ratings of Rapport (t-value = 1.96289, p < 0.05 for Mood and t-value = 2.89093, p < 0.01) but there was no interaction between Mood and Attentiveness.
- For the Pandemic2021 subvideos, only Mood had a statistically significant influence on the ratings of Rapport with (t-value = -2.18532, p < 0.05), but there was an interaction between Mood and Attentiveness (t-value = 2.4067744, p < 0.05).
- On the other hand, there was no effect of NARS nor Touch Receptivity groups on the evaluation of Rapport.

7.4.4.3 Functions of Touch: recognition of communicative intentions

Regarding the interpretation of the touches by the online participant, Table 7.5 shows the distribution of the participants choices regarding the functions of the agent's touches they observed for each subvideo and each condition. Again there were 22 participants in conditions V1 and V3 and 23 participants in condition V2.

We can note that in the Introduction subvideo of condition V1 Encouragement and Emotional Emphasis were the most selected functions (with 10 and 6 choices respectively),

7.4. COHERENCE AND ACCEPTABILITY: ONLINE EVALUATION OF THE MODEL

Touch Functions	Introduction			Pandemic2020			Pandemic2021		
	V1	V2	V3	V1	V2	V3	V1	V2	V3
GetAttention	3	0	2	1	0	0	0	0	1
Comfort	0	0	7	0	21	0	19	18	16
Encouragement	10	0	0	2	0	0	0	4	2
Turn Management	0	0	0	9	0	0	0	0	0
Emotion Emphasis	6	0	1	3	0	0	2	0	1
Calm	1	0	2	1	1	0	1	0	0
Other	2	0	1	0	0	0	0	1	1
No Touch Observed	0	23	9	6	1	22	0	0	1

Table 7.5 – Distribution of the participants’ choices regarding the function of the agent’s touches they observed for each subvideo of each condition.

while it was Comfort that was the most selected in condition V3 (7 choices). There were no touches in the Introduction subvideo of condition V2 and accordingly participants didn’t answer the questions related to touch for this subvideo. In condition V3 however, we can note that although there were touches initiated by the agent, 9 of the 23 participants of this condition declared not having observed any touch.

In the Pandemic2020 subvideos, there were no touches initiated by the agent in condition V3 and no participant of V3 declared having observed a touch. However there were 6 participants of V1 that declared not having observed a touch from the agent despite there having one. Among the remaining participants of V1, the most selected function was Turn management (9 votes). In condition V2, 21 out of the 22 participants identified the touches’ function as Comfort, with the last one identifying it as Calming.

Finally in the Pandemic2021 subvideos there were touches in all conditions and all the participants except one from condition V3 declared having observed it. There, all participants across all conditions overwhelmingly identified the function of the touches as Comfort (19 votes in V1, 18 in V2 and 16 in V3).

As for Laura, she only touched in the Pandemic2020 and Pandemic2021 subvideos, and this was only consistently observed for the Pandemic2021 subvideos. Participants overwhelmingly chose Comfort as the function of the touch (49 participants out of the 63 that observed a touch by Laura) and no other function was chosen by more than 4 participants.

7.4.4.4 "Correct" touch decisions: Coherence & Acceptability

As explained in Section 7.4.2 we had four questions relating to the "correctness" of the touch modality decisions with one item being related to coherence, one to relationship adequacy, and two to acceptability. We obtain the overall "Correctness of the touch modality decision" score with the mean of those four items scores.

Looking at the Correctness scores across the three subvideos and across the conditions (summarized in Figure 7.8), we can observe the following:

- In the Introduction subvideos, the correctness score for condition V1 is rather low (mean = 3.4, sd = 0.8) and somewhat higher in condition V3 (mean = 3.98, sd = 1.18) (no touch observed in condition V2), however an independent samples t-test showed that this difference was only approaching significance ($p = 0.06$).
- In the Pandemic2020 subvideos, correctness scores for V1 (mean = 3.98, sd = 1.18) and V2 (mean = 4.27, sd = 1.08) were again quite close. Since the ratings were not normally distributed, a Kruskal-Wallis test was used and showed no significant differences between those scores (chi-squared = 0.6686, $df = 1$, $p > 0.1$). We can note an increase in the score for condition V1 compared to the Introduction score but again this didn't prove significant from the statistical point of view.
- In the Pandemic2021 subvideos finally, the correctness scores were the following for V1 (mean = 5.1, sd = 1.01), V2 (mean = 4.28, sd = 1.12), V3 (mean = 4.5, sd = 1.13). A one-way ANOVA confirmed there was a significant difference between those values ($F(2, 64) = 3.3073$, $p < 0.05$) and a multiple comparison of means using tukey contrasts confirmed it to be between V1 and V2.

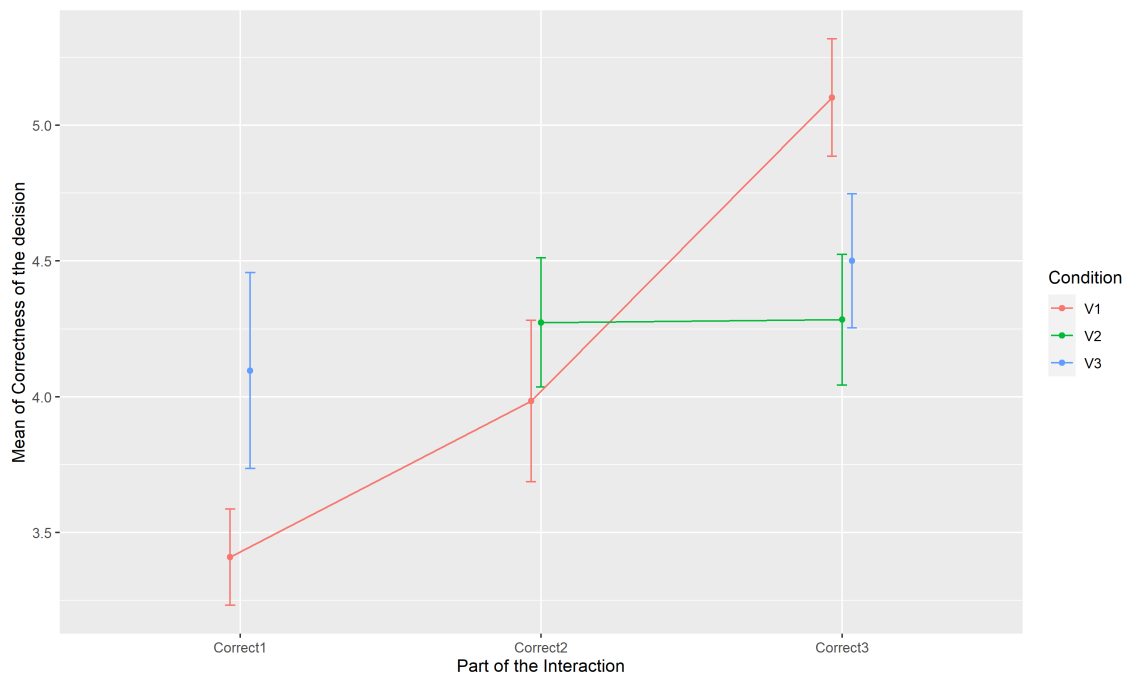


Figure 7.8 – Evolution of the ratings of the degree of "correctness" of Camille's touch decision across conditions and subvideos.

Considering our research questions Q2 and Q4.2, we further investigated the correlations between touch functions and correctness score as well as between rapport and

correctness scores. Because of the uneven frequency numbers we have for the touch intentions (see Table 7.5) it is difficult to investigate this correlation rigorously from the statistical point of view with the current dataset. Figure 7.9 showcases the correctness ratings grouped by touch function identified in the Introduction subvideos (conditions taken all together). We can overall observe lower ratings for Emotion Emphasis and the "Other" category and the highest average rating for the Comfort intention. Those observations stay pretty much the same for the other subvideos, although there are much less ratings for non-Comfort functions.

Pearson's correlation test hasn't shown any correlation between the Rapport and Correctness scores in subvideos Introduction and Pandemic2020. Pandemic2021 scores did lead to a significant correlation coefficient ($t=4.7976$, $df = 63$, $p < 0.001$), but this should be taken with a grain of salt considering that values are reliably high across both variables in those subvideos. We also used a factorial ANOVA to investigate the influence of the ratings of the questions about the coherence and the relation adequacy of the touch on the ratings of the questions about the acceptability of the touch (the first two and last two questions that compose the Correctness score), but the only significant result was for the Pandemic2020 subvideo where the relation adequacy ratings did appear correlated with the acceptability ratings (sum squared = 8.4904, $df = 5$, F value = 3.4643, $p < 0.05$). Furthermore, there were no significant differences between the ratings of Correctness for High and Low touch receptivity, nor between those of negative and positive a priori attitudes towards virtual agents.

As for Laura's touches, a Wilcoxon signed rank test with continuity correction showed no significant differences between the ratings of its correctness and that of the agent in the Pandemic2021 subvideo, which is the only one where there were enough ratings to make a meaningful comparison (Laura having used far less touches in the videos than Camille).

7.4.4.5 Appreciation of Touching Agents

Regarding our question Q5 finally, we overall observed that when asked whether they would like to have the opportunity to interact with our touching agent in the future (with a 1-to-7 scale of "not at all interested" to "extremely interested") participants scored a mean of 3.4 ($sd = 1.81$, median = 4). On the other hand, when asked if they found the idea of touching a virtual agent interesting they scored a mean of 4.2 ($sd = 2.08$, median = 5, mode = 5), and a mean of 4.19 ($sd = 1.91$, median = 5, mode = 5) for the idea of being touched by a virtual agent.

We did not find significant differences between conditions for those ratings, but interestingly condition V2 showed the highest average ratings for all the questions and V3 showed the lowest average ratings for all the questions. Kruskal-Wallis tests showed no significant differences between the ratings of High and Low touch receptivity groups, but there were significant differences between those of the Negative and Positive a priori atti-

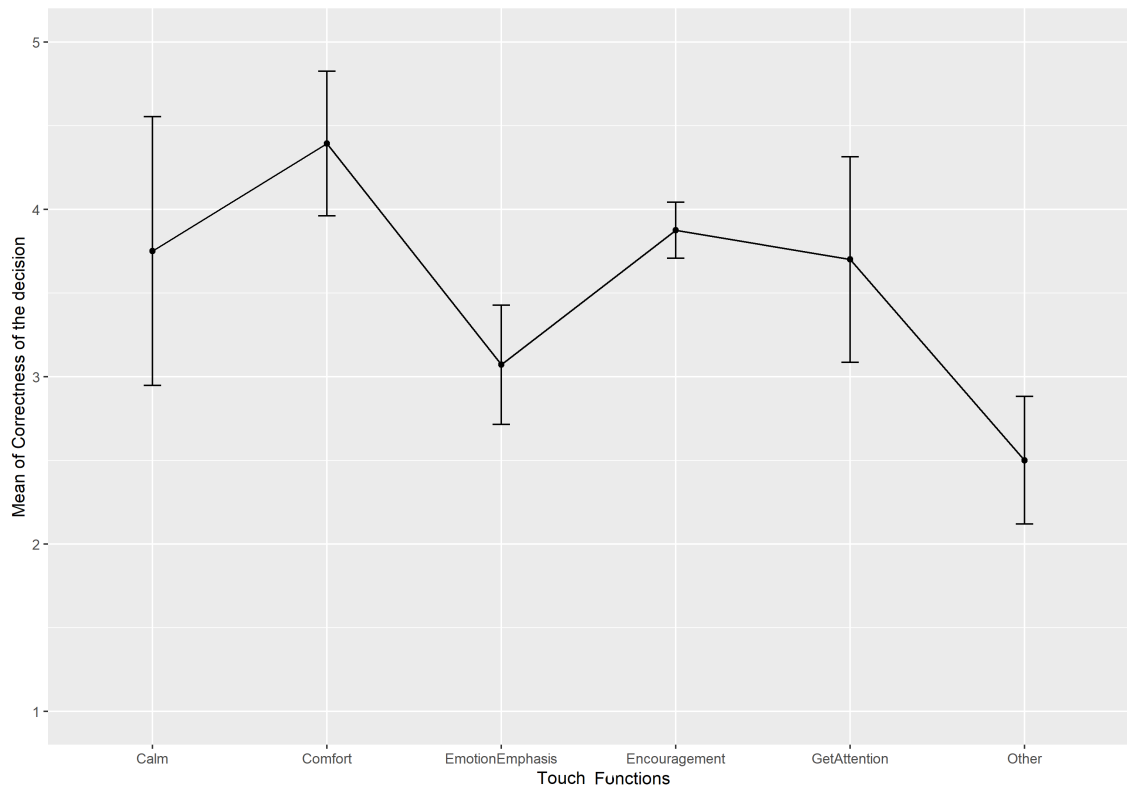


Figure 7.9 – Mean correctness ratings for each identified touch function in the Introduction subvideos

tudes towards virtual agents groups. Results of the test for the question about interacting with Camille in the future were $\chi^2 = 4.948$, $df = 1$, $p < 0.05$, for the question about the interest of agents that we could touch they were $\chi^2 = 12.161$, $df = 1$, $p\text{-value} < 0.001$ and for the question about the interest of agents that could touch humans they were $\chi^2 = 12.422$, $df = 1$, $p < 0.001$. Participants classified as having a positive attitude towards virtual agents significantly rated higher on all those questions.

7.4.4.6 Qualitative Results

As described in Section 7.4.2, the questionnaires left some opportunities of qualitative assessments with free text fields. For Camille's touch, participants were asked to describe the touches they had seen and to assess whether, had they been in the same situation, they would have touched Laura too. For Laura's touch, the question was about how the participants thought they would have reacted had they been the recipient of Laura's touch, instead of Camille, in this situation. All the following words and quotes are translated from French and the quotes are accompanied by the TR (touch receptivity) and NARS (a priori attitude towards virtual agents) profile of the person that said it.

For the description of the touches question, our initial intention was to collect information about the touch animations' appreciation of human observers. Interestingly, par-

7.4. COHERENCE AND ACCEPTABILITY: ONLINE EVALUATION OF THE MODEL

ticipants not only answered the question in that regard but also took the opportunity to include general comments regarding their impression of the acceptability and function of the touch. What was most mentioned was the artificiality of the touch animations, often with single word answers (or a collection of adjectives), and the words that were most used in that regard were: “artificial”, “fake”, “mechanical”, “blunt”, “jerky”, “too fast” and “violent”. For the participants that expressed a positive impression of the touches, there were almost no single word answers and the adjectives that were used were not the antonyms of the former ones, but instead words that related to the intention or emotion of the gesture: “Empathic”, “Like a gesture aimed at comforting the person and showing understanding, encouraging them to keep on talking.” (Low TR, Negative NARS), “Positive [impression], the gesture is related to an expression of comfort and support.” (High TR, Negative NARS), “Comforting movement aimed at expressing their empathy.” (Low TR, Positive NARS), “Simple, adapted to the situation.” (Low TR, Negative NARS).

In some cases, the quality of the animation negatively impacted the impression of the participants despite their understanding of the intention: “Through the intonation and the timing of the gesture, we understand the intention, but the rendering of the gesture is way too fast and violent in my opinion.” (High TR, Positive NARS). Negative impressions of the touches were not only influenced by the quality of the animations however and participants also expressed concerns regarding the appropriateness of the touches: “Not very natural (with regards to the animation, but I don’t find it realistic for real life either (except maybe in a lesbian romance))” (High TR, Positive NARS), “It looks mechanical, unnatural, and they are too close to each other for a professional situation. When facing Laura, I would have taken a step back.” (Low TR, Negative NARS). Another negative point expressed by many participants was the number of touches made by the agent, which they felt was way too high: “Mechanical. Too many of them. A comforting rub would be better than those bad taps of sympathy.” (High TR, Negative NARS).

For the question regarding whether they would also have used touch at this point in the interaction, participants mostly answered that they wouldn’t have for all conditions and subvideos as displayed in Table 7.6. It can however be observed that it was in the Pandemic2021 subvideos that there were the most "yes" answers, with almost half of "yes" in condition V1.

Conditions	Introduction			Pandemic2020			Pandemic2021		
	Yes	No	Unsure	Yes	No	Unsure	Yes	No	Unsure
V1	1	19	1	2	13	1	9	11	2
V2	0	0	0	4	17	0	6	13	3
V3	1	10	2	0	0	0	4	15	2

Table 7.6 – Distribution of the participants’ answers about whether they would also have used touch in this particular situation had they been in the shoes of Camille, the agent.

For the Introduction subvideos, almost all the “no” answers were accompanied by a justification that related to their absence of relationship (“We don’t know each other”, “I only touch people that I’m close to or that I would like to flirt with”) or to their own personal preference (“I am not a tactile person”, “I don’t touch people”). The only “yes” answers explained that it would have been to reassure Laura about the immersive experience and to install some proximity and mutual trust.

The same justifications were used in the Pandemic2020 and Pandemic2021 subvideos, although there were also mentions of their culture not being a tactile one. The participants that would have touched expressed however that it would have been to show empathy, understanding and emotional connection because of the topic of the discussion, and also that they would have used touch to avoid interrupting Laura’s speech. Some participants that said they wouldn’t have touched in that situation expressed that it would require an extreme situation of emotional distress for them to use comforting touches with someone they don’t really know and one participant further expressed that they wouldn’t have touched because Laura didn’t seem to have had a worse experience than themselves.

When asked how they would have reacted to the touch of Laura had they been in the shoes of Camille, participants once again displayed a variety of reactions based on their understanding of the interaction: “It depends on my relationship with Laura. If I consider that my relationship started at the beginning of this set of videos, I would have been surprised but I would have understood the intention, so I would have accepted it. In general, when somebody touches me, I pay more attention to their verbal and non-verbal communication.” (High TR, Positive NARS), “It would have interrupted me in my discourse.” (Low TR, Positive NARS), “I prefer not to be touched, but I understand the gesture” (Low TR, Positive NARS). Context also seemed to play a big part in their interpretation of the interaction as some participants declared that: “I would have found it strange considering the context at the moment of the interaction between Laura and Camille.” (High TR, Positive NARS). Finally, some participants expressed that the immersive setup might have helped them feel better about the exchange of touches, with one of them saying: “I would have felt less ill-at-ease than if it had happened in real life” (Low TR, Negative NARS).

7.4.5 Discussion

It is interesting to note that with this protocol the online participants were basically in the same situation as the decision model in terms of the inputs they could base their estimations on. The simulated participant being represented by a virtual avatar, there were indeed no facial expressions nor additional gestures that online participants could rely on to interpret the interaction, in the same way that our agent would not be able to rely on those pieces of information in an actual interaction with the immersive setup because of the current limitations of the perception module. A drawback however is that it is not possible to ascertain that the online participants did consider Laura as an actual

human during the whole interaction despite our efforts in the dubbing and control of Laura's movements.

Going over our initial research questions in the light of the current results we can first notice that participants seem to have correctly identified our initial settings of the conditions (Q1.1). V1 ratings described a rather attentive and content Laura, while V2 and V3 described an initially less attentive Laura, not in a great mood.

The evolution of the ratings over the course of the different parts of the interaction also follow the estimations made by the decision model (see Figures 7.5 and 7.6): in V1 the ratings for Laura's mood start decreasing with the second subvideo where she shares negative experiences of the COVID lockdown ; in V2, the initial negative mood of Laura decreases even more in the subsequent subvideos where she also shares her negative experience of the lockdowns ; in V3, the initial negative mood improves when she shares positive experiences of the lockdowns. Those variations mirror those estimated by the decision model very closely. In that regard we can highlight a fundamental ambiguity in social interactions: what we feel might be different from what we express in our verbal and non-verbal behaviours. While Laura does express positive experiences of the lockdowns, her initial mood may not really change over the course of her discourse despite the estimations of the agent and human observers. While such "mistake" in terms of estimation would certainly also happen with human interpretations, we believe that it is of great importance, when working on an intimate modality such as touch, to consider the addition of doubt mechanisms so that the agent may still try to verify its beliefs by directly asking its interlocutor in certain situations. In that regard, modelling uncertainty in the decision model rules and beliefs could go a long way towards improving the impression that the agent gives to human interlocutors.

For the attentiveness, we observe a much higher estimation at the beginning of the interaction for V1 followed by a general increase for all conditions in the second and last subvideos where the ratings are very close to each other across all conditions. In that regard, our decision model showed more variations in its estimation of the attentiveness in condition V3, because the perception module is very sensible to the gaze direction of the human. When Laura moves around to put some space between herself and the agent, the agent makes several low estimation of the attentiveness because Laura temporarily turns her back on Camille. For the online participants, this behaviour would most likely not influence their ratings because Laura was not expressing inattentive behaviour *per se*, she simply repositioned herself while still listening to Camille. Since the interaction was a simple discussion with no particular task to accomplish that could have distracted her, it is rather logical that all the online participants felt that Laura was attentive once the actual conversation had started. This highlights the need for the decision model to take the temporality of the interaction into account in a better way: a short change in the gaze direction should not instantly bring the estimation of attentiveness to zero, the decrease should instead happen in a gradual way based on elapsed time.

As for the estimation of Rapport (see Figure 7.7), we can observe that the ratings of Rapport in the initial subvideo mirror those of mood and attentiveness, and there are very similar curves between the evolution of Rapport and that of Attentiveness. From the statistical point of view, we did not find a correlation between Rapport ratings and Attentiveness ratings for all the subvideos, but we did find an influence of Mood ratings for most. Overall, we believe that this shows that our model of Rapport (see Section 5.4 and Equation 5.1) are mostly consistent with actual human estimations in a third-party setting and the lack of statistical correlation between Attentiveness and Rapport is probably due to the interactive situation that did not feature much variations in Attentiveness ratings. Again, only actual in-person interactions with our agent would be able to firmly validate those observations for the intended immersive setup.

Interestingly, the current statistical results don't allow us to decisively validate our fundamental hypothesis regarding Q2 that rapport influences the acceptability of a touch. We did observe however that rapport was increasing over the course of the interaction and so did the perceived "correctness" of the touching decisions of the model for conditions V1 and V3, with V2 having stable above average ratings. Furthermore, the qualitative results did feature remarkable assessments on par with our hypothesis and the literature on social touch as participants used relational and emotional lexical fields when expressing positive feelings towards the touch of the agent. Furthermore, they directly stated how relationship state and personal preferences would influence their touching behaviours. This self-awareness regarding touching behaviours was really interesting to observe: human-agent interaction seems to be able to elicit introspective behaviours regarding the participants' own attitudes. We feel like it would be interesting to reiterate the study with an additional condition in which no direct questions would be asked about the touching behaviours, to see if only witnessing the interaction would have been enough to prompt participants to remark on the subject of the touches.

With regards to questions Q4.1 and Q4.2, the assessments of the touches' functions and their influence on the "correctness" ratings of the touches, it seems clear that the study's protocol was ill-suited to measuring it in a reliable way. Multiple touches with different functions were grouped together under only three subvideos and only one question relating to it. Furthermore, the context of interaction (emotional discussion about potentially sad events) was not propitious for a wide variety of functions to be examined. While this also prevented us from being able to make decisive statistical assessments of the correlation between those variables, we still observed overall higher ratings of "correctness" for the comforting function, which was certainly the most appropriate to the context. In their qualitative assessments, participants again only expressed positive feelings about the touches when they could justify it by an adequate function. With the current results, we believe that our initial hypothesis (that adequacy between the function of the touch and the context of interaction would improve its coherence and, in turn, its acceptability) is correct but needs to be validated with a more appropriate protocol. More specifically, hav-

ing the evaluation be conducted in the originally intended immersive environment could give us more ways to have the human participant dynamically evaluate the touches in the course of the interaction, one at a time. Furthermore, only the use of more varied situations of interaction would allow us to decisively study those assumptions.

If we now come back to question Q3, Figure 7.8 did show that the "correctness" ratings were lower in condition V1 than in condition V3 for the Introduction subvideos, and they were the lowest ratings of all the subvideos and conditions. Although there was only an approaching statistical significance for this difference between V3 and V1 in the Introduction, it seems coherent with our expectations: it is the beginning of the encounter (lower Rapport) and Laura is in a good mood and attentive in condition V1, which means that there is not really a very strong reason to touch her. Furthermore, it is also V1 that showed the overall highest average ratings for the "correctness" of the touches in the Pandemic2021 subvideo. This is coherent with the idea that having estimated the human as in a positive mood at the beginning of the interaction in V1, the sad feelings expressed in the latter subvideos could be considered as stronger (by contrast), than those expressed in V2 where the mood was already negative in the beginning. As one of the qualitative assessment stated, the subjective estimation of the level of distress experienced by our interlocutor can play an important role in our decision to touch or not, and also in the acceptability of this touch. Overall, there seems to be a contradiction between this lack of statistical support for the influence of Rapport and touch functions on the "correctness" of a touch and the qualitative assessments of the participants, which are on par with our expectations. We believe that this is not really a contradiction. In Chapters 2 and 3 we have shown that literature indicates that relationship is an important factor in our acceptance of a touch, but that this notion would be hard to model in short-term interactions with an agent, which led to us adopting the notion of Rapport. Based on the literature we modeled Rapport as a construct based on attentiveness and positivity of the emotional state of the interlocutor, and this turned out to be a rather accurate way to estimate Rapport in the present study. However, we believe that it is precisely in the difference between the notions of Rapport and relationship that the above mentioned contradiction lies: our qualitative assessments show that even though participants did rate Rapport in a similar way to our model, they did not equate those ratings to their estimation of the relationship between Laura and Camille. A dimension seems to be lacking from our model of Rapport for it to more accurately give us a measure of the relational acceptance of a touch: time, or, more specifically, familiarity. There are indeed other computational models of ECAs that have used the familiarity dimension to build Rapport successfully (Novick and Gris (2014), Pecune et al. (2013), Cassell et al. (2007)) and we believe that we, too, should add this dimension to our calculation of Rapport.

As for the fact that touch receptivity scores did not influence any of the ratings of the participants regarding the correctness of the touches, this can be interpreted in two different ways. First, we can consider that touch receptivity did not in fact matter for this

specific study, because the participants were not directly interacting with the agent. Their personal preferences may indeed have been eclipsed in their answers by the third-person perspective when they actually rated the Likert items. The second interpretation, which, furthermore, does not exclude the first one, would be that as shown by research works such as the one by Lucas et al. (2017), people are more comfortable discussing personal matters with an agent than with a human. Similarly, and as expressed by one of the qualitative statements in Section 7.4.4.6, the fact that Laura was interacting with an agent may have made the online participants rate acceptability of the touches higher than they would have with an interaction between two humans.

Regarding question Q5 finally, participants had rather negative ratings of their interest regarding the idea of interacting with our agent. This can be explained by their overall assessments regarding the agent's low-quality animations and the fact that its behaviour was more often felt as unpleasant in the qualitative assessments. On the other hand however, participants consistently expressed general interest in both being able to touch and be touched by virtual agents. Somewhat expectedly, this interest was more marked for people that had positive a priori towards virtual agent, but there was no influence of the touch receptivity on those ratings. This could again be explained by the fact that participants did not feel as strongly about touch with an agent compared with touch with another human. It can also be explained by the fact that touch receptivity measures a threshold of acceptance of a touch from a stranger, but doesn't say anything about the appreciation of touch when it is performed acceptably. This comforts us in the idea that while there is still work to be done on the side of the decision model and overall animation of the agent, we are working in the correct direction and touching agents could become parts of our social life in the future. Again, it should nonetheless be noted that we had elaborated our framework and decision model for in person interaction in an immersive environment and this third-person observational study's results cannot entirely replace an actual immersive interaction that would allow us to better validate our developments.

Conclusion and Perspectives

On sait que les choses et les personnes sont toujours forcées de se cacher, déterminées à se cacher quand elles commencent. Comment en serait-il autrement ? Elles surgissent dans un ensemble qui ne les comportait pas encore, et doivent mettre en avant les caractères communs qu'elles conservent avec l'ensemble, pour ne pas être rejetées. L'essence d'une chose n'apparaît jamais au début, mais au milieu, dans le courant de son développement, quand ses forces sont affermies.

Gilles Deleuze

Contents

8.1	Summary of Contributions	120
8.1.1	Contribution 1: Review of the Social Touch literature	120
8.1.2	Contribution 2: Architecture and Design Choices for a Touching Agent	121
8.1.3	Contribution 3: A Perception Module for a Virtual Sense of Touch	121
8.1.4	Contribution 4: A Decision Model for a Touching Agent	122
8.1.5	Contribution 5: the Vibrotactile Sleeve SOFTLY	122
8.1.6	Contribution 6: Functional Framework and Evaluations	123
8.2	Current limitations	123
8.3	Perspectives for future works	125

IN this Chapter we conclude this Thesis by summarising our contributions in Section 8.1, by discussing the current limitations of this work in Section 8.2 and by sug-

gesting perspectives to improve our work and guidelines for more future works on touching agents in Section 8.3.

8.1 Summary of Contributions

Our initial research question was the following: how can we endow embodied conversational agents with social touch abilities in such a way that it would enhance their emotional and relational abilities? Our proposition throughout this Thesis was to answer by building and evaluating an ECA able to partake with a human in an interactive loop that integrates the social touch modality. This, in turn, led us to the proposition of the following contributions.

8.1.1 Contribution 1: Review of the Social Touch literature

To determine what such an ECA would need to be able to interact meaningfully via touch with a human, we first proposed in Chapter 2 an in-depth review of the literature on both social touch among humans and social touch technologies. We first highlighted the physical and anatomical properties of the sense of touch, before showing how important touch is to the well-being and socio-psychological development of humans. We further showed that social touch is particularly well suited to the expression of emotions and to the process of relational bonding. In that regard, while the literature highlights the many positive effects of touch, it also shows that touch remains ultimately an intimate communicative channel and that all touches are not equally acceptable for the recipient. More specifically, each person has a different touch avoidance which is based on both culture and personal preferences, and the level of relationship we have with others influences our acceptance of their touches.

In the field of social touch technologies, we saw how haptics is a rapidly developing field of research where more and more different types of technologies and applications appear. There is still no way to perfectly simulate our natural tactual sensations via touch for now, but there are many available alternatives that already prove to be of interest to human users. We however identified a lack of research regarding complete human-agent bidirectional interactions in the simulated social touch literature. While there are numerous studies on the effects of an agent's or robot's touch on humans and on the way we can perceive human touch and interpret its meaning via technology, there is almost no works we are aware of that implement a complete interactive loop with an adaptable agent that can touch and be touched. We chose to position ourselves in this blind spot and propose a functional framework and decision model to investigate those kinds of interactions.

8.1.2 Contribution 2: Architecture and Design Choices for a Touching Agent

To enable an interactive loop featuring social touch in an immersive environment, we proposed to adopt the classic architecture of an autonomous agent which requires perception, decision and action abilities. We also chose to use an immersive setup with an immersive room for its powerful immersion abilities and its advantage of letting the user see their own body during the interaction, thus avoiding the questions of embodiment that need to be addressed with VR HMDs. This decision of a VR setup was particularly influenced by [Huisman et al. \(2014a\)](#)'s argument that VR would benefit the study of social touch as a whole because it enables us to craft controlled but believable interactive scenarios which may help create a more ecological setting for participants.

To integrate touch into such an architecture and immersive setup, we thus proposed the development of a perception module for a virtual sense of touch that would allow the agent to detect the human's touches and their physical properties. As for the decision, we felt, based on our inspection of the literature, that it would be critical for a touching agent to be able to take emotions, coherence and acceptability into account when deciding when and how to touch. Touch is indeed a powerful and intimate modality which requires the greatest caution when using it with others. We therefore proposed to use a computational model of emotion for our touching agent to actually touch: both physically and emotionally. Finally, for the actual instantiation of the behaviour chosen by the decision model, we used the GRETA platform for the animation of the agent and developed SOFTLY, a vibrotactile sleeve for the haptic feedback.

This architecture was implemented in a software structure designed to be easily reproducible with different VR setups (HMDs for example) and haptic devices. We believe that this is especially important when using specific expensive setups such as an immersive room, so that our work can still be extended to different and more easily available setups with minimal effort required.

8.1.3 Contribution 3: A Perception Module for a Virtual Sense of Touch

In Chapter 4 we showed how we could implement a virtual sense of touch for immersive interactions and improve upon the prior existing works in the literature. We first applied the strategy proposed in the literature and used a grid of colliders to compose a virtual skin for the agent. With this approach we can detect when and where a touch is happening as well as some of the physical properties of the touch (velocity, duration, movement).

The evaluation presented in Section 7.1 however showed the limitations of this implementation: the resolution of the grid was a highly limiting factor that impacted the data recorded negatively. Furthermore, the lack of estimation of the force exerted on the agent was also a critical missing piece of the implementation. We showed how we could improve upon this implementation by using a local coordinates system for the tactile cells

and by using a god-object method with a spring to have some kind of estimation of the force exerted on the agent. We complemented this perception module with gaze direction and proxemics estimations.

Finally, we implemented a simple algorithm using pre-defined fuzzy sets to have a pre-interpretation of the raw values recorded by the perception module.

8.1.4 Contribution 4: A Decision Model for a Touching Agent

In Chapter 5 we proposed a decision process that would allow us to use the inputs produced by the perception module to model beliefs that would be adapted to evaluate when a touch will be acceptable. In that regard, we chose to model a calculation of the level of rapport between the human and the agent based on the attentiveness and the emotional states of the human.

We further detailed how we could model appropriate appraisal and interpretation rules to generate the beliefs and the emotional state of the agent based on the events that it perceives. By combining the static touch avoidance of the human to those beliefs, and especially the estimation of rapport, we argued that our agent would be better able to determine when touching would be acceptable for the human. To that effect we developed multiple decision rules for different states of the interaction that may make the threshold of acceptability of a touch vary: interpersonal distance, level of rapport, touch avoidance, emotional state, situation of interaction, etc. This model was implemented in an already existing and openly available computational model of emotion, FATiMA, which we modified to suit our specific needs when necessary.

As for the meaning of the touch, we implemented a two-step decision process in which we first determine the most adequate communicative intention to be conveyed for the current situation and then determine whether touch is an adequate modality to express it. The mapping of the types of touch with their appropriate communicative intention was established based on the literature on social touch.

8.1.5 Contribution 5: the Vibrotactile Sleeve SOFTLY

Finally, in Chapter 6 we presented SOFTLY, an easily wearable vibrotactile sleeve composed of a grid of voice coils that we can use to generate an appropriate haptic feedback. The voice coil technology allows us to use many different kinds of wave forms and frequencies to produce interesting sensations. We can then combine our haptic patterns with the animation of our agent to synchronously instantiate the decision model's touch decisions.

An evaluation presented in Section 7.2 allowed us to investigate some of the psychophysical properties of the device with human users and calibrate some of our values.

8.1.6 Contribution 6: Functional Framework and Evaluations

All of those contributions taken together form both a theoretical and a practical framework, which has been developed to be easy to import in similar settings of immersive interactions with agents, even if the specific hardware differs a little from the one presented here.

While, at the moment, the COVID pandemic has made it impossible to evaluate the model in the intended setting, we still presented in Chapter 7 evaluations of the overall model, which showed promising results. Most importantly, we presented a third-person observational study conducted online where the participants were shown videos of our agent taking touching decisions. We asked them about their impressions regarding the coherence and acceptability of the agent's decisions regarding touch and we also investigated whether their estimations of the simulated human's emotional state and attentiveness were close to those of the model. The results showed that our initial assumptions regarding the design choices of the model seemed overall correct, but required some adjustments. Only an evaluation in the complete immersive setup would allow us to decisively conclude about the abilities of our model however.

We nonetheless hope to have demonstrated how interesting this research topic is and how vast the research questions in this field can be. We also hope to have demonstrated that such an ECA able to partake in a social touch-based interactive loop is indeed possible to build, although it requires facing numerous challenges that still need to be fully addressed on each level: perception, decision and feedback.

As a last contribution for this present Thesis, we now propose to summarize the current limitations of our work and give pointers for short-term improvements and guidelines for future works in this field.

8.2 Current limitations

As mentioned in Chapter 7, we did not have the opportunity to evaluate and validate our complete framework in the initially intended immersive setting. We are well aware of the fundamental limitations that this imposes on our current results. We definitely cannot conclude anything regarding our initial research question, which was to investigate whether social touch abilities would improve the abilities of an ECA for emotional communication and relational bonding in immersive environment.

As for the rest of the framework, the perception module remains entirely untested in actual conditions. There is little doubt that its outputs would appear very lackluster in comparison to the literature on social touch perception and recognition. Working on the calibration of the module with human participants and taking inspiration from the recognition algorithms that exist for tangible interfaces would be a way to validate and improve the perception of touch in the future.

Similarly, SOFTLY still isn't a very robust device for now and improving its material structure with better wires would make it easier to put on and off. As for its potential in terms of haptic feedback, while we believe that many interesting things can be done with it, it remains to be more thoroughly investigated with the design of many more types of signals and perceptual studies. We could indeed imagine trying using actual recordings of skin-to-skin contacts with the device to see whether the sensation produced is more realistic.

While this might be the most obvious and fundamental limitation of our work, we still believe that our online observational study indicated interesting avenues of improvement for our touching agent. In the current state of things, discussed in more details in Section 7.4.4, the online validation study did not statistically validate the usefulness of calculating rapport to dynamically evaluate the acceptability of a touch. We indeed showed that while the calculations of the rapport produced by our model were on par with the estimation of the human observers, this did not influence their ratings of the acceptability of the touches. We however realized that in their qualitative assessment, participants did consider relationship as one of the most important factors of the acceptability of a touch. We thus proposed that this discrepancy in our results was due to our model of rapport lacking a dimension that is present in the notion of relationship: familiarity. Adding this new dimension to our calculation of rapport could prove very impactful for the acceptability of its touching decisions.

From the statistical point of view, we could not conclude much regarding our hypothesis that the adequacy of a touch intention with the situation it happens in would improve its perceived coherence. The protocol and the interactive scenario we used in this study were both ill-suited to such an investigation. We still believe that the qualitative assessments made by the participants are supportive of our hypothesis and that it would be valuable to craft a new protocol for the immersive setup where this could be more easily investigated.

Finally, the ratings of the participants regarding their interest for our agent and such touching agents in general showed that they did not want to interact with our agent but did find the idea of interacting via touch with virtual agents interesting. This highlights both the potential of simulated social touch and the limitations of our current work. We already discussed the limitations of our work regarding the decision model, but the other limitation that most likely impacted the ratings of our agent are the animations of the agent which are quite jerky and lack a lot of polish. In that regard, the 3D models we used are also quite far from the current standards in terms of photo-realism. While we have known for a while that the uncanny valley effect may be a risk when using photo-realistic agents (although recent studies such as the one by Zibrek et al. (2018) have nuanced this fear), we believe that improving the animations of the agent would make a great difference in the appreciation of the agent. Touch being such an intimate modality

it seems important to have fluid animations to better represent soft and slow touches and caresses for example.

Outside of the limitations highlighted by our evaluation, we can add that although we used a computational model of emotion (FAtiMA), we did not have the opportunity to make use of all its abilities yet. More specifically, our use of the emotional states of the agent is very limited for now, and we could greatly improve our model by designing decision rules that would take this into account, especially for the reactions of the agent to touch. That being said, different emotional profiles would not be as suited to every type of situation and designing different types of agent should be carefully done in that regard.

As a limitation inherent to classical rule-based models, we can not avoid mentioning the fact that developing decision rules for social touch interactions, for which even the literature provides little formal expertise, represents a very tall order. In the long run, such a model will always struggle to achieve finesse and nuance in its decisions, although it makes up for it with a high level of explainability and by the fact that the decision rules produced can be used in the context of social touch research even among humans.

8.3 Perspectives for future works

While our work certainly has a good amount of limitations in its current form, those also represent avenues of investigation for its improvement. Many of those have already been touched upon in Section 7.4.4, and we now propose to expand upon them here.

First of all, as short-term perspectives, the experience of the online study will help us design an even more interesting study in the immersive setup. Notably, it highlighted the need to propose varied situations that will allow the investigation of more functions of touch. Furthermore, a pure discussion context proved to be somewhat limited in the types of touching interactions that can be explored. Finding a situation where touch fits well for a short interaction is hard however, especially since there are basically no examples from the literature on human-human social touch. We would most likely need to survey the literature on mediated social touch to find protocols more focused on the interactive context that we could draw inspiration from. In any case, having done this observational study will allow us to compare our results with those of the immersive condition and maybe find new directions for touching agents, with new hypothesis and research questions that could emerge from this comparison.

For the decision model itself, we would like to integrate the familiarity dimension to our calculation of Rapport, so as to better model the level of the relationship between the agent and the human. There are already social models for autonomous agents that have taken this dimension into account (Pecune et al. (2013), Novick and Gris (2014)) and drawing inspiration from it, we believe that much better results could be achieved with regards to the acceptability of a touch.

Again, we would like to reiterate that we believe that acceptability is an extremely crucial notion to take into account when working on social touch. While it might first appear as a necessary ethical concern to avoid hurting the potential users, we would like to bring attention to other important aspects of this question of acceptability. In the field of affective computing as it was founded by Picard (1997), the goal has always been to build agents able to sense, interpret and express emotions. More specifically, as expressed by McDuff and Czerwinski (2018), the idea is that agents able to understand human social norms and emotions will feel more natural and valuable to interact with. As those researchers however also highlight, such emotionally sensitive agents will both require sensing technologies (that have already started to be used for marketing and advertisement purposes) and adopt the social behaviours that we endow them with. The emotional data sensed by those perception technologies is inherently intimate and personal, and in the same way that we are now more and more aware of how personal data collected from our use of internet can be bought, sold and used to influence our behaviours by showing us specific items of interest, so too our emotional states could be collected and used without us even noticing. There is little doubt that this will make agents feel more natural and indeed valuable to interact with, but it might also lead to instances of soft manipulation. When presenting this Thesis work at a conference on affective touch, someone with a background far remote from human-computer interaction and artificial intelligence asked me if, considering the known midas touch effects of interpersonal touch on securing compliance, we didn't fear that our agents may ultimately be used to manipulate someone to perform crimes or bring harm to others or themselves. While it was a somewhat unexpected question and it might seem a bit ridiculous at the moment considering the limited abilities of our agent, it is a very real concern that needs to be addressed. As McDuff and Czerwinski (2018) put it, those emotional agents may eventually be an integral part of our everyday lives and, in turn, will certainly have an influence on our social behaviours at large. In science-fiction literature, the figure of the robot has always been used as a mirror of humanity, asking us "what is humanity?". This is the main concern of a book like *Do Androids Dream of Electric Sheep?* by Dick (2007), but also the subtext of Asimov (1993) and his many books about robots. Similarly, in the domain of the philosophy of technology, many philosophers have brought forward the fact that technology has a direct impact on our social and cognitive existences. When we build a social agent, we proceed to a grammatization (see Stiegler (2005) for example, for more details on this notion) of social behaviour: we transpose our knowledge and habits of social norms, which are continuous and temporal behaviours automated within us, to a discrete nature that we can manipulate, compute and import inside our agent. This acts as an externalization of our social knowledge and this can become a danger, as expressed by Stiegler (2015) with regards to the risks involved with the processes of grammatization and automation in a large scale: as machines have proletarianized workers and decision algorithms may very well start to do the same to high status executives, we can imagine a proletarianization of

social relationships and professions. By interacting on a daily basis with such social agents, McDuff and Czerwinski (2018) suggest that should they have slightly off behaviours, we may adapt to them and integrate them to an extent that we would then develop difficulties to interact with other humans. A touching agent is ultimately a grammatization of our social touch habits and norms and should we fail to address the question of acceptability, we take the risk, no matter how far away this might seem at the moment, of inducing inadequate touching behaviours in future human-human interactions too.

At the same time, we believe that if we do manage to take those concerns into account, touching agents also represent a great opportunity to further investigate our own touching behaviours. In the answers of our participants to the online study, we were very impressed by their self-awareness regarding their touching behaviour and expectations. We argue that following the principles of the interactional approach of affective computing (as proposed by the works of Boehner et al. (2007) for example) is essential for touching agents. With this approach, all the emotional data perceived and interpreted from the human should be given back to them in some way, so that they are aware of what the agent has perceived and understood. Not only this will go a long way towards transparency of the use of the personal data of the user, we believe that this will also allow the human to reflect on its own emotional state and touching behaviours. As touch and emotional expressions in general are often ambiguous, this would further help the agent to not only react to what it believes it has understood from the human, but also communicate about its uncertainties and thus negotiate the meanings of touch in an interactive paradigm.

For all those reasons, we believe that having a way to model uncertainty of the beliefs of the agent is important to produce more nuanced behaviours for the agent. With a measure of confidence in the estimations of the agent about the human, the decision model would be able to adopt verification and negotiation strategies to check whether a touch would indeed be acceptable in a certain situation by directly asking the human about their emotional state or touch receptivity. It would also be particularly important to continue our work on the notions of continuous actions. Even before the touch actually happens, we already have information about the pre-touch phase thanks to the ethereal body of the agent and this could be considered as a part of the touch too. Having a way to model a touch from its approach to its end and to manage the timing with regards to the number of touches to perform in a certain time frame would probably make the touch feel a lot more natural. In FATiMA, there isn't any real time management functions: the data stored in the memory and the decision steps are not measured with actual time units (seconds, minutes,...) but with artificial "ticks" index, that only allow a chronological ordering. Modifying the model to enable it to take actual time into account would go a long way in making it more suited to fully interactive situations where we could manage interruption of actions and reaction to pre-touch approaches. Ultimately however, a rule-based decision

model will struggle with micro gestures interaction and we should contemplate using another type of model for those types of interactions.

As for the perception module, actual calibration and validation in context is obviously a priority. There is little doubt that the performance of the module will be below those of the literature in terms of interpretation, but we believe that future works could adapt already working social touch recognition algorithm and integrate them to the immersive modality. This would probably not be easy, as we cannot get the same amount nor quality of data from colliders compared to tangible interfaces, but it might still lead to interesting results.

For the rendering of the actions of the agent, we also believe that there is still much that can be improved on and explored with SOFTLY. In terms of material improvements, due to the way the voice coils are fixed to the fabric, the device is quite noisy at the moment and thus distracting. By changing the way the voice coils are attached to the sleeve we would massively improve the quality of use of the device. Similarly, better types of thin wires would make the sleeve more robust overall. As for the haptic patterns, many types of signals and wave forms can be used with the device to produce many different kinds of sensations. We believe that it would be extremely interesting to keep testing and evaluating different types of signals and the way they are appreciated and interpreted by humans. We are most interested in trying out different types of sounds such as recorded skin-to-skin human touch gestures or cat purring which is known to have appeasing effects on humans.

The 3D models and animations of the agents themselves could certainly be improved a lot in terms of fluidity and appearance. In that regard, the fact that the agent we chose to use has a rather realist humanoid appearance led us to consider similar touch interactions as those we experience in everyday life. As it was touched upon in the discussion of the preliminary study from Section 7.1.4, strictly imitating real life interactions might not be the approach most suited to immersive interactions with virtual agents. Gilles Deleuze's quote from the epigraph of the present Chapter 8 expresses exactly this: new things all start by imitating what is already existing in order to avoid being rejected. It is later on that their singularity and particular traits express themselves. Similarly, we believe that reproducing real life interactions is a necessary first approach, but it is not the only one and it would be extremely interesting to imagine other ways to interact with a virtual agent via touch. Not only this might lead to more compelling ways to interact with agents, this might also reveal new applications for touching agents that we have yet to imagine.

As a conclusion to this present work we can wonder what an interaction with a very advanced touching agent might look like should we overcome the challenges of building such an ECA. We argue that it should be a companion able to evaluate coherence and acceptability of touch in the course of the interaction and to communicate about its estimations, especially when unsure of what the human's emotional states might be. It should be able to conduct long-term interactions and therefore remember the user it interacts

8.3. PERSPECTIVES FOR FUTURE WORKS

with from one interaction to another so as to be able to estimate not only Rapport but also the level of the relationship, as well as remember their static touch avoidance and touching habits. Finally, it should be empathic and promote self-awareness regarding touching behaviours. As we mentioned it in the introduction of this Thesis, touch is the most fundamental sense of our bodies and the body is currently still the focal point of discrimination and aggression. Careful development of touching agents that don't perpetuate those kinds of behaviours and instead interrogate users about their own behaviours could be a way to combat prejudice at large if those types of agents ever become part of our everyday social lives.

Bibliography

- Allen, J., Cang, L., Phan-Ba, M., Strang, A., and MacLean, K. (2015). Introducing the Cuddlebot: A Robot that Responds to Touch Gestures. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts - HRI'15 Extended Abstracts* (Portland, Oregon, USA: ACM Press), 295–295. doi: 10.1145/2701973.2702698
- Andersen, P. A. and Leibowitz, K. (1978). The development and nature of the construct touch avoidance. *Environmental psychology and nonverbal behavior* 3, 89–106. doi: 10.1007/BF01135607
- App, B., McIntosh, D. N., Reed, C. L., and Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: How may depend on why. *Emotion* 11, 603–617. doi: 10.1037/a0023164
- Asimov, I. (1993). *Caves of steel*. (Bantam Doubleday). OCLC: 734106107
- Bailenson, J. N. and Yee, N. (2008). Virtual interpersonal touch: Haptic interaction and copresence in collaborative virtual environments. *Multimedia Tools and Applications* 37, 5–14. doi: 10.1007/s11042-007-0171-2
- Bailenson, J. N., Yee, N., Brave, S., Merget, D., and Koslow, D. (2007). Virtual Interpersonal Touch: Expressing and Recognizing Emotions Through Haptic Devices. *Human-Computer Interaction* 22, 325–353. doi: 10.1080/07370020701493509
- Bianchi-Berthouze, N. and Tajadura-Jiménez, A. (2014). It's not just what we touch but also how we touch it. In *Workshop on Tactile User and touching objects of different shapes as a way to Experience Evaluation Methods*. 4
- Bickmore, T. and Cassell, J. (2001). Relational Agents: A Model and Implementation of Building User Trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM), CHI '01, 396–403. doi: 10.1145/365024.365304. Event-place: Seattle, Washington, USA

BIBLIOGRAPHY

- Bickmore, T. W. and Cassell, J. (1999). Small Talk and Conversational Storytelling In Embodied Conversational Interface Agents
- Bickmore, T. W., Fernando, R., Ring, L., and Schulman, D. (2010). Empathic Touch by Relational Agents. *IEEE Transactions on Affective Computing* 1, 60–71. doi: 10.1109/T-AFFC.2010.4
- Blackwell, L., Ellison, N., Elliott-Deflo, N., and Schwartz, R. (2019). Harassment in Social VR: Implications for Design. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 854–855. doi: 10.1109/VR.2019.8798165. ISSN: 2642-5254
- Block, A. E. and Kuchenbecker, K. J. (2018). Emotionally Supporting Humans Through Robot Hugs. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA: ACM), HRI '18, 293–294. doi: 10.1145/3173386.3176905. Event-place: Chicago, IL, USA
- Block, A. E. and Kuchenbecker, K. J. (2019). Softness, Warmth, and Responsiveness Improve Robot Hugs. *International Journal of Social Robotics* 11, 49–64. doi: 10.1007/s12369-018-0495-2
- Boehner, K., DePaula, R., Dourish, P., and Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 275–291. doi: 10.1016/j.ijhcs.2006.11.016
- Bonanni, L., Vaucelle, C., Lieberman, J., and Zuckerman, O. (2006). TapTap: a haptic wearable for asynchronous distributed touch therapy. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal Québec Canada: ACM), 580–585. doi: 10.1145/1125451.1125573
- Boucaud, F., Pelachaud, C., and Thouvenin, I. (2021). Decision Model for a Virtual Agent that can Touch and be Touched. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), AAMAS '21, 232–241
- Bouzbib, E., Bailly, G., Haliyo, S., and Frey, P. (2021). "Can I Touch This?": Survey of Virtual Reality Interactions via Haptic Solutions. *ArXiv* doi: 10.1145/3450522.3451323
- Bradac, J. J., O'Donnell, M., and Tardy, C. H. (1984). Another Stab at a Touchy Subject: Affective Meanings of Touch. *Women's Studies in Communication* 7, 38–51. doi: 10.1080/07491409.1984.11089661
- Brave, S., Nass, C., and Sirinian, E. (2001). Force-Feedback in computer-mediated communication. In *Towards an Information Society for All, Proceedings of {HCI} International '2001* (New Orleans, USA: Lawrence Erlbaum Associates, Inc), 145–149

BIBLIOGRAPHY

- Breazeal, C. (2011). Social robots for health applications. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 5368–5371. doi: 10.1109/IEMBS.2011.6091328
- Cafaro, A., Ravenet, B., Ochs, M., Vilhjálmsón, H. H., and Pelachaud, C. (2016). The Effects of Interpersonal Attitude of a Group of Agents on User's Presence and Proxemics Behavior. *ACM Transactions on Interactive Intelligent Systems* 6, 1–33. doi: 10.1145/2914796
- Cang, X. L., Bucci, P., Strang, A., Allen, J., MacLean, K., and Liu, H. S. (2015). Different Strokes and Different Folks: Economical Dynamic Surface Sensing and Affect-Related Touch Recognition. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15* (Seattle, Washington, USA: ACM Press), 147–154. doi: 10.1145/2818346.2820756
- Carlson, M. and Earls, F. (1997). Psychological and Neuroendocrinological Sequelae of Early Social Deprivation in Institutionalized Children in Romania. *Annals of the New York Academy of Sciences* 807, 419–428. doi: 10.1111/j.1749-6632.1997.tb51936.x
- Cascio, C. J., Moore, D., and McGlone, F. (2019). Social touch and human development. *Developmental Cognitive Neuroscience* 35, 5–11. doi: 10.1016/j.dcn.2018.04.009
- Cassell, J., Gill, A. J., and Tepper, P. A. (2007). Coordination in conversation and rapport. In *Proceedings of the Workshop on Embodied Language Processing - EmbodiedNLP '07* (Prague, Czech Republic: Association for Computational Linguistics), 41–50. doi: 10.3115/1610065.1610071
- Cassell, J. and Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13, 519–538. doi: 10.1080/088395199117360
- Cassell, J., Velkovska, J., and Relieu, M. (2020). « Tisser des liens »: L'interaction sociale chez les agents conversationnels. *Réseaux* N°220-221, 21. doi: 10.3917/res.220.0021
- Castellanos, S., Rodríguez, L.-F., Castro, L. A., and Gutierrez-Garcia, J. O. (2018). A computational model of emotion assessment influenced by cognition in autonomous agents. *Biologically Inspired Cognitive Architectures* 25, 26–36. doi: 10.1016/j.bica.2018.07.007
- Cekaite, A. and Kvist Holm, M. (2017). The Comforting Touch: Tactile Intimacy and Talk in Managing Children's Distress. *Research on Language and Social Interaction* 50, 109–127. doi: 10.1080/08351813.2017.1301293
- Chatel-Goldman, J., Congedo, M., Jutten, C., and Schwartz, J.-L. (2014). Touch increases autonomic coupling between romantic partners. *Frontiers in Behavioral Neuroscience* 8. doi: 10.3389/fnbeh.2014.00095

BIBLIOGRAPHY

- Crandall, F. M. (1897). Hospitalism. *Archives of Pediatrics* 14, 448–454
- Cranny-Francis, A. (2011). Semefulness: A social semiotics of touch. *Social Semiotics* 21, 463–481. doi: 10.1080/10350330.2011.591993
- Crusco, A. H. and Wetzel, C. G. (1984). The Midas Touch: The Effects of Interpersonal Touch on Restaurant Tipping. *Personality and Social Psychology Bulletin* 10, 512–517. doi: 10.1177/0146167284104003
- Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. (1993). Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA: ACM), SIGGRAPH '93, 135–142. doi: 10.1145/166117.166134. Event-place: Anaheim, CA
- Dangxiao, W., Yuan, G., Shiyi, L., Yuru, Z., Weiliang, X., and Jing, X. (2019). Haptic display for virtual reality: progress and challenges. *Virtual Reality & Intelligent Hardware* 1, 136. doi: 10.3724/SP.J.2096-5796.2019.0008
- Demeure, V., Niewiadomski, R., and Pelachaud, C. (2011). How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? *Presence: Teleoperators and Virtual Environments* 20, 431–448. doi: 10.1162/PRES_a_00065
- Derlega, V. J., Lewis, R. J., Harrison, S., Winstead, B. A., and Costanza, R. (1989). Gender differences in the initiation and attribution of tactile intimacy. *Journal of Nonverbal Behavior* 13, 83–96. doi: 10.1007/BF00990792
- Dias, J., Mascarenhas, S., and Paiva, A. (2014). FAtiMA Modular: Towards an Agent Architecture with a Generic Appraisal Framework. In *Emotion Modeling: Towards Pragmatic Computational Models of Affective Processes*, eds. T. Bosse, J. Broekens, J. Dias, and J. van der Zwaan (Cham: Springer International Publishing), Lecture Notes in Computer Science. 44–56. doi: 10.1007/978-3-319-12973-0_3
- Dibiase, R. and Gunnoe, J. (2004). Gender and Culture Differences in Touching Behavior. *The Journal of Social Psychology* 144, 49–62. doi: 10.3200/SOCP.144.1.49-62
- Dick, P. K. (2007). *Do androids dream of electric sheep?* (London: Gollancz). OCLC: 800259295
- Ditzen, B., Neumann, I. D., Bodenmann, G., von Dawans, B., Turner, R. A., Ehlert, U., et al. (2007). Effects of different kinds of couple interaction on cortisol and heart rate responses to stress in women. *Psychoneuroendocrinology* 32, 565–574. doi: 10.1016/j.psyneuen.2007.03.011

BIBLIOGRAPHY

- Eichhorn, E., Wettach, R., and Hornecker, E. (2008). A stroking device for spatially separated couples. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services - MobileHCI '08* (Amsterdam, The Netherlands: ACM Press), 303. doi: 10.1145/1409240.1409274
- Elbani, W. (2019). Social touch: stimuli-imitation protocol and automated recognition. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (Cambridge, United Kingdom: IEEE), 11–15. doi: 10.1109/ACIIW.2019.8925025
- Erp, J. B. F. V. and Toet, A. (2013). How to Touch Humans: Guidelines for Social Agents and Robots That Can Touch. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 780–785. doi: 10.1109/ACII.2013.145
- Field, T. (2010). Touch for socioemotional and physical well-being: A review. *Developmental Review* 30, 367–383. doi: 10.1016/j.dr.2011.01.001
- Fisher, J. D., Rytting, M., and Heslin, R. (1976). Hands Touching Hands: Affective and Evaluative Effects of an Interpersonal Touch. *Sociometry* 39, 416–421. doi: 10.2307/3033506
- Floyd, K. (2014). Relational and Health Correlates of Affection Deprivation. *Western Journal of Communication* 78, 383–403. doi: 10.1080/10570314.2014.927071
- Fribourg, R., Argelaguet, F., Lécuyer, A., and Hoyet, L. (2020). Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE transactions on visualization and computer graphics* 26, 2062–2072
- Fukuda, H., Shiomi, M., Nakagawa, K., and Ueda, K. (2012). ‘Midas touch’ in human-robot interaction: Evidence from event-related potentials during the ultimatum game. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 131–132. doi: 10.1145/2157689.2157720
- Garcia-Valle, G., Ferre, M., Brenosa, J., and Vargas, D. (2018). Evaluation of Presence in Virtual Environments: Haptic Vest and User’s Haptic Skills. *IEEE Access* 6, 7224–7233. doi: 10.1109/ACCESS.2017.2782254
- Garrod, S. and Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science* 1, 292–304. doi: 10.1111/j.1756-8765.2009.01020.x
- Gebhard, P. (2005). ALMA: A Layered Model of Affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (New York, NY, USA: ACM), AAMAS '05, 29–36. doi: 10.1145/1082473.1082478. Event-place: The Netherlands

BIBLIOGRAPHY

- Gebhard, P. and Kipp, K. H. (2006). Are Computer-Generated Emotions and Moods Plausible to Humans? In *Intelligent Virtual Agents*, eds. J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 343–356
- Gomez Jauregui, D. A., Argelaguet, F., Olivier, A.-H., Marchal, M., Multon, F., and Lecuyer, A. (2014). Toward "Pseudo-Haptic Avatars": Modifying the Visual Animation of Self-Avatar Can Simulate the Perception of Weight Lifting. *IEEE Transactions on Visualization and Computer Graphics* 20, 654–661. doi: 10.1109/TVCG.2014.45
- Gratch, J., Kang, S.-H., and Wang, N. (2013). Using Social Agents to Explore Theories of Rapport and Emotional Resonance. In *Social Emotions in Nature and Artifact*, eds. J. Gratch and S. Marsella (Oxford University Press). 181–197. doi: 10.1093/acprof:oso/9780195387643.003.0012
- Gratch, J. and Marsella, S. (2005). Evaluating a Computational Model of Emotion. *Autonomous Agents and Multi-Agent Systems* 11, 23–43. doi: 10.1007/s10458-005-1081-1
- Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., et al. (2006). Virtual Rapport. In *Intelligent Virtual Agents*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier (Berlin, Heidelberg: Springer Berlin Heidelberg), vol. 4133. 14–27. doi: 10.1007/11821830_2
- Gratch, J., Wang, N., Gerten, J., Fast, E., and Duffy, R. (2007a). Creating Rapport with Virtual Agents. In *Intelligent Virtual Agents*, eds. C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 125–138
- Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., van der Werf, R. J., et al. (2007b). Can Virtual Humans Be More Engaging Than Real Ones? In *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, ed. J. A. Jacko (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 286–297
- Guimarães, M., Mascarenhas, S., Prada, R., Santos, P. A., and Dias, J. (2019). An Accessible Toolkit for the Creation of Socio-EmotionalAgents. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), AAMAS '19, 2357–2359
- Guéguen, N. (2004). Nonverbal Encouragement of Participation in a Course: the Effect of Touching. *Social Psychology of Education* 7, 89–98. doi: 10.1023/B:SPOE.0000010691.30834.14

BIBLIOGRAPHY

- Haans, A., de Bruijn, R., and IJsselsteijn, W. A. (2014). A Virtual Midas Touch? Touch, Compliance, and Confederate Bias in Mediated Communication. *Journal of Nonverbal Behavior* 38, 301–311. doi: 10.1007/s10919-014-0184-2
- Haans, A., IJsselsteijn, W. A., Graus, M. P., and Salminen, J. A. (2008). The Virtual Midas Touch: Helping Behavior After a Mediated Social Touch. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA: ACM), CHI EA '08, 3507–3512. doi: 10.1145/1358628.1358882. Event-place: Florence, Italy
- Hall, E. T. (1990). *The hidden dimension* (New York: Anchor Books)
- Harlow, H. F. (1958). The nature of love. *American Psychologist* 13, 673–685. doi: 10.1037/h0047884
- Haruna, M., Ogino, M., and Koike-Akino, T. (2020). Proposal and Evaluation of Visual Haptics for Manipulation of Remote Machine System. *Frontiers in Robotics and AI* 7, 529040. doi: 10.3389/frobt.2020.529040
- Hashimoto, H. and Manoratkul, S. (1996). Tele-Handshake through the Internet. In *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA* (Tsukuba, Japan: IEEE), 90–95. doi: 10.1109/ROMAN.1996.568767
- Hauser, S., McIntyre, S., Israr, A., Olausson, H., and Gerling, G. (2019). Uncovering Human-to-Human Physical Interactions that Underlie Emotional and Affective Touch Communication. doi: 10.1109/WHC.2019.8816169
- Hertenstein, M. J., Holmes, R., McCullough, M., and Keltner, D. (2009). The communication of emotion via touch. *Emotion* 9, 566–573. doi: 10.1037/a0016108
- Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., and Jaskolka, A. R. (2006a). Touch communicates distinct emotions. *Emotion* 6, 528–533. doi: 10.1037/1528-3542.6.3.528
- Hertenstein, M. J., Verkamp, J. M., Kerestes, A. M., and Holmes, R. M. (2006b). The communicative functions of touch in humans, nonhuman primates, and rats: a review and synthesis of the empirical research. *Genetic, social, and general psychology monographs* 132, 5–94. doi: 10.3200/MONO.132.1.5-94
- Hieida, C., Horii, T., and Nagai, T. (2018). Deep Emotion: A Computational Model of Emotion Using Deep Neural Networks. *arXiv:1808.08447 [cs]* ArXiv: 1808.08447
- Hoppe, M., Rossmly, B., Neumann, D. P., Streuber, S., Schmidt, A., and Machulla, T.-K. (2020). A Human Touch: Social Touch Increases the Perceived Human-likeness of Agents in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA: ACM), 1–11. doi: 10.1145/3313831.3376719

BIBLIOGRAPHY

- Hudlicka, E. (2005). *A Computational Model of Emotion and Personality: Applications to Psychotherapy Research and Practice*
- Huisman, G., Bruijnes, M., Kolkmeier, J., Jung, M., Darriba Frederiks, A., and Rybarczyk, Y. (2014a). Touching Virtual Agents: Embodiment and Mind. In *Innovative and Creative Developments in Multimodal Interaction Systems*, eds. Y. Rybarczyk, T. Cardoso, J. Rosas, and L. M. Camarinha-Matos (Springer Berlin Heidelberg), IFIP Advances in Information and Communication Technology, 114–138
- Huisman, G., Frederiks, A. D., Dijk, B. V., Hevlen, D., and Kröse, B. (2013). The TaSSt: Tactile sleeve for social touch. In *2013 World Haptics Conference (WHC)*. 211–216. doi: 10.1109/WHC.2013.6548410
- Huisman, G., Frederiks, A. D., van Erp, J. B. F., and Heylen, D. K. J. (2016). Simulating Affective Touch: Using a Vibrotactile Array to Generate Pleasant Stroking Sensations. In *Haptics: Perception, Devices, Control, and Applications*, eds. F. Bello, H. Kajimoto, and Y. Visell (Springer International Publishing), Lecture Notes in Computer Science, 240–250
- Huisman, G., Kolkmeier, J., and Heylen, D. (2014b). Simulated Social Touch in a Collaborative Game. In *Haptics: Neuroscience, Devices, Modeling, and Applications*, eds. M. Au-ray and C. Duriez (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 248–256
- Huisman, G., Kolkmeier, J., and Heylen, D. (2014c). With Us or Against Us: Simulated Social Touch by Virtual Agents in a Cooperative or Competitive Setting. In *Intelligent Virtual Agents*, eds. T. Bickmore, S. Marsella, and C. Sidner (Springer International Publishing), Lecture Notes in Computer Science, 204–213
- Héron, R., Safin, S., Baker, M., and Détienne, F. (2022). The Functions of Computer-Mediated Touch at a Distance: An Interactionist Approach. In *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, eds. N. L. Black, W. P. Neumann, and I. Noy (Cham: Springer International Publishing), vol. 223. 45–53. doi: 10.1007/978-3-030-74614-8_6. Series Title: Lecture Notes in Networks and Systems
- Israr, A. and Poupyrev, I. (2011). Tactile Brush: Drawing on Skin with a Tactile Grid Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (New York, NY, USA: ACM), CHI '11, 2019–2028*. doi: 10.1145/1978942.1979235. Event-place: Vancouver, BC, Canada
- Izard, C. E. (2010). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review* 2, 363–370. doi: 10.1177/1754073910374661
- Jones, S. E. and Yarbrough, A. E. (1985). A naturalistic study of the meanings of touch. *Communication Monographs* 52, 19–56. doi: 10.1080/03637758509376094

BIBLIOGRAPHY

- Kang, S.-H. and Gratch, J. (2011). People like virtual counselors that highly-disclose about themselves. *Studies in health technology and informatics* 167, 143–8
- Keysers, C. and Gazzola, V. (2009). Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology* 19, 666–671. doi: 10.1016/j.conb.2009.10.006
- Kleinke, C. L. (1977). Compliance to requests made by gazing and touching experimenters in field settings. *Journal of Experimental Social Psychology* 13, 218–223. doi: 10.1016/0022-1031(77)90044-0
- Knapp, M. L. and Hall, J. A. (2010). *Nonverbal communication in human interaction* (Boston, MA: Wadsworth, Cengage Learning). OCLC: 244767251
- Knez, I. and Hygge, S. (2001). The circumplex structure of affect: A Swedish version. *Scandinavian Journal of Psychology* 42, 389–398. doi: 10.1111/1467-9450.00251
- Knight, H., Toscano, R., Stiehl, W. D., Chang, A., Wang, Y., and Breazeal, C. (2009). Real-time social touch gesture recognition for sensate robots. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (St. Louis, MO, USA: IEEE), 3715–3720. doi: 10.1109/IROS.2009.5354169
- Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., et al. (2006). Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In *Intelligent Virtual Agents*, eds. J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 205–217
- Kotranza, A. and Lok, B. (2008). Virtual Human + Tangible Interface = Mixed Reality Human An Initial Exploration with a Virtual Breast Exam Patient. In *2008 IEEE Virtual Reality Conference* (Reno, NV, USA: IEEE), 99–106. doi: 10.1109/VR.2008.4480757
- Kotranza, A., Lok, B., Deladisma, A., Pugh, C., and Lind, D. (2009a). Mixed Reality Humans: Evaluating Behavior, Usability, and Acceptability. *IEEE Transactions on Visualization and Computer Graphics* 15, 369–382. doi: 10.1109/TVCG.2008.195
- Kotranza, A., Lok, B., Pugh, C. M., and Lind, D. S. (2009b). Virtual Humans That Touch Back: Enhancing Nonverbal Communication with Virtual Humans through Bidirectional Touch. In *2009 IEEE Virtual Reality Conference*. 175–178. doi: 10.1109/VR.2009.4811019
- Kowalczyk, Z. and Czubenko, M. (2016). Computational Approaches to Modeling Artificial Emotion – An Overview of the Proposed Solutions. *Frontiers in Robotics and AI* 3. doi: 10.3389/frobt.2016.00021

BIBLIOGRAPHY

- Krahé, C., von Mohr, M., Gentsch, A., Guy, L., Vari, C., Nolte, T., et al. (2018). Sensitivity to CT-optimal, Affective Touch Depends on Adult Attachment Style. *Scientific Reports* 8, 14544. doi: 10.1038/s41598-018-32865-6
- LaFrance, M. (1979). Nonverbal Synchrony and Rapport: Analysis by the Cross-Lag Panel Technique. *Social Psychology Quarterly* 42, 66. doi: 10.2307/3033875
- Lazarus, R. S. (1991). *Emotion and adaptation* (New York: Oxford Univ. Press). OCLC: 23252739
- Lee, J. and Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. In *Intelligent Virtual Agents*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier (Berlin, Heidelberg: Springer Berlin Heidelberg), vol. 4133. 243–255. doi: 10.1007/11821830_20. Series Title: Lecture Notes in Computer Science
- Lenay, C. (2019). Technical innovation in human science: Examples in cognitive technologies. *THEORIA. An International Journal for Theory, History and Foundations of Science* 34, 389. doi: 10.1387/theoria.17901
- Leroi-Gourhan, A. (1964). *Le geste et la parole* (Albin Michel). Google-Books-ID: hZ_hAAAAMAAJ
- Loomis, J. M. and Lederman, S. J. (1986). Tactual perception. In *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance* (Oxford, England: John Wiley & Sons). 1–41
- Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., et al. (2017). Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Frontiers in Robotics and AI* 4. doi: 10.3389/frobt.2017.00051
- Löken, L. S., Wessberg, J., Morrison, I., McGlone, F., and Olausson, H. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nature Neuroscience* 12, 547–548. doi: 10.1038/nn.2312
- Madaio, M. A., Lasko, R., Cassell, J., and Ogan, A. (2017). Using Temporal Association Rule Mining to Predict Dyadic Rapport in Peer Tutoring. In *EDM*
- Major, B. (1981). Gender Patterns in Touching Behavior. In *Gender and Nonverbal Behavior*, eds. C. Mayo and N. M. Henley (New York, NY: Springer New York), Springer Series in Social Psychology. 15–37. doi: 10.1007/978-1-4612-5953-4_2
- Marsella, S. C. and Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10, 70–90. doi: 10.1016/j.cogsys.2008.03.005

BIBLIOGRAPHY

- Martin, M. M. and Anderson, C. M. (1993). Psychological and biological differences in touch avoidance. *Communication Research Reports* 10, 141–147. doi: 10.1080/08824099309359927
- McDuff, D. and Czerwinski, M. (2018). Designing Emotionally Sentient Agents. *Commun. ACM* 61, 74–83. doi: 10.1145/3186591
- McGlone, F., Wessberg, J., and Olausson, H. (2014). Discriminative and Affective Touch: Sensing and Feeling. *Neuron* 82, 737–755. doi: 10.1016/j.neuron.2014.05.001
- McIntyre, S., Mougou, A., Boehme, R., Isager, P. M., Lau, F., Israr, A., et al. (2019). Affective touch communication in close adult relationships. doi: 10.17605/OSF.IO/7XRWC
- Mehrabian, A. and Russell, J. A. (1974). *An approach to environmental psychology* (Cambridge, Mass.: MIT Press). OCLC: 714320
- Meyer, C., Streeck, J., and Jordan, J. S. (eds.) (2017). *Intercorporeality: emerging socialities in interaction*. Foundations of human interaction (New York, NY: Oxford University Press)
- Moerland, T. M., Broekens, J., and Jonker, C. M. (2018). Emotion in reinforcement learning agents and robots: a survey. *Machine Learning* 107, 443–480. doi: 10.1007/s10994-017-5666-0
- Montagu, A. (1986). *Touching: The Human Significance of the Skin* (HarperCollins). Google-Books-ID: XU7Z_aqCYggC
- Nguyen, N. and Wachsmuth, I. (2009). Modeling Peripersonal Action Space for Virtual Humans Using Touch and Proprioception. In *Intelligent Virtual Agents*, eds. Z. Ruttkay, M. Kipp, A. Nijholt, and H. H. Vilhjálmsón (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 63–75
- Nguyen, N. and Wachsmuth, I. (2011). From body space to interaction space - Modeling spatial cooperation for virtual humans. vol. 2, 1047–1054
- Nguyen, N., Wachsmuth, I., and Kopp, S. (2007). Touch perception and emotional appraisal for a virtual agent
- Novick, D. and Gris, I. (2014). Building Rapport between Human and ECA: A Pilot Study. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, and M. Kurosu (Cham: Springer International Publishing), vol. 8511. 472–480. doi: 10.1007/978-3-319-07230-2_45

BIBLIOGRAPHY

- Nunez, C. M., Huerta, B. N., Okamura, A. M., and Culbertson, H. (2020). Investigating Social Haptic Illusions for Tactile Stroking (SHIFTS). In *2020 IEEE Haptics Symposium (HAPTICS)* (Crystal City, VA, USA: IEEE), 629–636. doi: 10.1109/HAPTICS45997.2020.ras.HAP20.35.f631355d
- Ong, D. C., Zaki, J., and Goodman, N. D. (2019). Computational Models of Emotion Inference in Theory of Mind: A Review and Roadmap. *Topics in Cognitive Science* 11, 338–357. doi: 10.1111/tops.12371
- Ortony, A., Clore, G. L., and Collins, A. (1988). *The Cognitive Structure of Emotions* (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511571299
- Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., et al. (2004). Caring for Agents and Agents That Care: Building Empathic Relations with Synthetic Agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1* (Washington, DC, USA: IEEE Computer Society), AAMAS '04, 194–201. doi: 10.1109/AAMAS.2004.82. Event-place: New York, New York
- Pecune, F. (2013). Toward a Computational Model of Social Relations for Artificial Companions. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 677–682. doi: 10.1109/ACII.2013.118
- Pecune, F., Murali, S., Tsai, V., Matsuyama, Y., and Cassell, J. (2019). A Model of Social Explanations for a Conversational Movie Recommendation System. In *Proceedings of the 7th International Conference on Human-Agent Interaction* (Kyoto Japan: ACM), 135–143. doi: 10.1145/3349537.3351899
- Pecune, F., Ochs, M., and Pelachaud, C. (2013). A Formal Model of Social Relations for Artificial Companions
- Petersen, S., Houston, S., Qin, H., Tague, C., and Studley, J. (2016). The Utilization of Robotic Pets in Dementia Care. *Journal of Alzheimer's Disease* 55, 569–574. doi: 10.3233/JAD-160703
- Picard, R. W. (1997). *Affective computing*. Affective computing (Cambridge, MA, US: The MIT Press)
- Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* 17. doi: 10.1017/S0954579405050340
- Pusch, A. and Lécuyer, A. (2011). Pseudo-haptics: from the theoretical foundations to practical system design guidelines. In *Proceedings of the 13th international conference on multimodal interfaces - ICMI '11* (Alicante, Spain: ACM Press), 57. doi: 10.1145/2070481.2070494

BIBLIOGRAPHY

- Rantala, J., Salminen, K., Raisamo, R., and Surakka, V. (2013). Touch gestures in communicating emotional intention via vibrotactile stimulation. *International Journal of Human-Computer Studies* 71, 679–690. doi: 10.1016/j.ijhcs.2013.02.004
- Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E., et al. (2013). Computational Modeling of Emotion: Towards Improving the Inter- and Intradisciplinary Exchange. *IEEE Transactions on Affective Computing* vol. 4, pp. 246–266. doi: 10.1109/T-AFFC.2013.14
- Remland, M. S., Jones, T. S., and Brinkman, H. (1995). Interpersonal Distance, Body Orientation, and Touch: Effects of Culture, Gender, and Age. *The Journal of Social Psychology* 135, 281–297. doi: 10.1080/00224545.1995.9713958
- Romano, J. M. and Kuchenbecker, K. J. (2012). Creating Realistic Virtual Textures from Contact Acceleration Data. *IEEE Transactions on Haptics* 5, 109–119. doi: 10.1109/TOH.2011.38
- Rosis, F. d., Pelachaud, C., Poggi, I., Carofiglio, V., and Carolis, B. D. (2003). From Greta’s mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies* 59, 81–118. doi: 10.1016/S1071-5819(03)00020-X
- Sanderson, C. M. and Jorgensen, J. D. (1997). To Touch or Not To Touch: An Examination of Tactile Communication between College Student and Professor.
- Schaefer, M., Heinze, H.-J., and Rotte, M. (2012). Embodied empathy for tactile events: Interindividual differences and vicarious somatosensory responses during touch observation. *NeuroImage* 60, 952–957. doi: 10.1016/j.neuroimage.2012.01.112
- Scherer, K. R. (2000). Psychological models of emotion. In *The neuropsychology of emotion* (New York, NY, US: Oxford University Press), Series in affective science. 137–162
- Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion* 15, 123–148. doi: 10.1007/BF00995674
- Sefidgar, Y. S., MacLean, K. E., Yohanan, S., Loos, H. F. M. V. d., Croft, E. A., and Garland, E. J. (2016). Design and Evaluation of a Touch-Centered Calming Interaction with a Social Robot. *IEEE Transactions on Affective Computing* 7, 108–121. doi: 10.1109/T-AFFC.2015.2457893
- Seifi, H. and MacLean, K. E. (2013). A first look at individuals’ affective ratings of vibrations. In *2013 World Haptics Conference (WHC)* (Daejeon: IEEE), 605–610. doi: 10.1109/WHC.2013.6548477

BIBLIOGRAPHY

- Shaltout, H. A., Tooze, J. A., Rosenberger, E., and Kemper, K. J. (2012). Time, Touch, and Compassion: Effects on Autonomic Nervous System and Well-Being. *EXPLORE* 8, 177–184. doi: 10.1016/j.explore.2012.02.001
- Sidner, C. L., Lee, C., and Lesh, N. (2003). Engagement by Looking: Behaviors for Robots when Collaborating with People. In *In Kruiff-Korbayova and Kosny (eds.), Diabruck: The Proceedings of the Seventh Workshop on the Semantics and Pragmatics of Dialogue. University of Saarland.* 123–130
- Silvera-Tawil, D., Rye, D., and Velonaki, M. (2014). Interpretation of Social Touch on an Artificial Arm Covered with an EIT-based Sensitive Skin. *International Journal of Social Robotics* 6, 489–505. doi: 10.1007/s12369-013-0223-x
- Stiegler, B. (2005). Individuation et grammatisation : quand la technique fait sens... *Documentaliste-Sciences de l'Information* Vol. 42, 354–360
- Stiegler, B. (2015). *La société automatique* (Paris: Fayard)
- Stiehl, W. D. and Breazeal, C. (2005). Affective Touch for Robotic Companions. In *Affective Computing and Intelligent Interaction*, eds. J. Tao, T. Tan, and R. W. Picard (Springer Berlin Heidelberg), Lecture Notes in Computer Science, 747–754
- Stier, D. S. and Hall, J. A. (1984). Gender Differences in Touch: An Empirical and Theoretical Review. doi: 10.1037/0022-3514.47.2.440
- Suhonen, K., Müller, S., Rantala, J., Väänänen-Vainio-Mattila, K., Raisamo, R., and Lantz, V. (2012). Haptically augmented remote speech communication: A study of user practices and experiences. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (New York, NY, USA: Association for Computing Machinery), NordiCHI '12, 361–369. doi: 10.1145/2399016.2399073
- Summerhayes, D. L. and Suchner, R. W. (1978). Power implications of touch in male—Female relationships. *Sex Roles* 4, 103–110. doi: 10.1007/BF00288381
- Suvilehto, J. T., Glerean, E., Dunbar, R. I. M., Hari, R., and Nummenmaa, L. (2015). Topography of social touching depends on emotional bonds between humans. *Proceedings of the National Academy of Sciences* 112, 13811–13816. doi: 10.1073/pnas.1519231112
- Sykownik, P. and Masuch, M. (2020). The Experience of Social Touch in Multi-User Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA: Association for Computing Machinery), VRST '20, 1–11. doi: 10.1145/3385956.3418944
- Teyssier, M., Bailly, G., Pelachaud, C., and Lecolinet, E. (2018). MobiLimb: Augmenting Mobile Devices with a Robotic Limb. In *Proceedings of the 31st Annual ACM Symposium*

BIBLIOGRAPHY

- on *User Interface Software and Technology* (Berlin Germany: ACM), 53–63. doi: 10.1145/3242587.3242626
- Teyssier, M., Bailly, G., Pelachaud, C., and Lecolinet, E. (2020). Conveying Emotions Through Device-Initiated Touch. *IEEE Transactions on Affective Computing* , 1–1doi: 10.1109/TAFFC.2020.3008693
- Thompson, E. H. and Hampton, J. A. (2011). The effect of relationship status on communicating emotions through touch. *Cognition and Emotion* 25, 295–306. doi: 10.1080/02699931.2010.492957
- Tickle-Degnen, L. and Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry* 1, 285–293. doi: 10.1207/s15327965pli0104_1
- Ustek, D., Chow, K., Zhang, H., and MacLean, K. (2018). A Multimodal Illusion of Force Improves Control Perception in Above-Surface Gesture: Elastic Zed-Zoom. In *Haptics: Science, Technology, and Applications*, eds. D. Prattichizzo, H. Shinoda, H. Z. Tan, E. Ruffaldi, and A. Frisoli (Springer International Publishing), Lecture Notes in Computer Science, 295–308
- van Erp, J. B. F. and Toet, A. (2015). Social Touch in Human–Computer Interaction. *Frontiers in Digital Humanities* 2. doi: 10.3389/fdigh.2015.00002
- Walker, S. C., Trotter, P. D., Woods, A., and McGlone, F. (2017). Vicarious ratings of social touch reflect the anatomical distribution & velocity tuning of C-tactile afferents: A hedonic homunculus? *Behavioural Brain Research* 320, 91–96. doi: 10.1016/j.bbr.2016.11.046
- Webb, A. and Peck, J. (2015). Individual differences in interpersonal touch: On the development, validation, and use of the “comfort with interpersonal touch” (CIT) scale. *Journal of Consumer Psychology* 25, 60–77. doi: 10.1016/j.jcps.2014.07.002
- Wilson, G. and Brewster, S. A. (2017). Multi-moji: Combining Thermal, Vibrotactile & Visual Stimuli to Expand the Affective Range of Feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver Colorado USA: ACM), 1743–1755. doi: 10.1145/3025453.3025614
- Winnicott, D. W. (2016). *Transitional Objects and Transitional Phenomena*, vol. 1 (Oxford University Press). doi: 10.1093/med:psych/9780190271367.003.0034
- Yohanan, S. and MacLean, K. E. (2012). The Role of Affective Touch in Human-Robot Interaction: Human Intent and Expectations in Touching the Haptic Creature. *International Journal of Social Robotics* 4, 163–180. doi: 10.1007/s12369-011-0126-7

BIBLIOGRAPHY

- Zhao, R., Sinha, T., Black, A. W., and Cassell, J. (2016). Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior. In *Intelligent Virtual Agents*, eds. D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski (Cham: Springer International Publishing), vol. 10011. 218–233. doi: 10.1007/978-3-319-47665-0_20
- Zhou, R., Wu, Y., and Sareen, H. (2020). HexTouch: Affective Robot Touch for Complementary Interactions to Companion Agents in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA: Association for Computing Machinery), VRST '20, 1–2. doi: 10.1145/3385956.3422100
- Zibrek, K., Kokkinara, E., and McDonnell, R. (2018). The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? *IEEE Transactions on Visualization and Computer Graphics* 24, 1681–1690. doi: 10.1109/TVCG.2018.2794638
- Zilles, C. and Salisbury, J. (1995). A constraint-based god-object method for haptic display. In *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots* (Pittsburgh, PA, USA: IEEE Comput. Soc. Press), vol. 3, 146–151. doi: 10.1109/IROS.1995.525876

