



MASTER RECHERCHE STAGE



RAPPORT BIBLIOGRAPHIQUE

Etat de l'art de la Robotique Sociale : Interactions, Initiation et Engagement avec un Robot Mobile Socialement Conscient

Domaine : Interaction Humain-Robot - Robotique Sociale

Auteur:
Paco DUPONT

Superviseur:
Elisabetta BEVACQUA
Matthieu COURGEON
ENIB - Lab-STICC



Résumé: De nombreux travaux ont été réalisés depuis les années 2000 sur le comportement social que doit adopter un robot social humanoïde en présence d’humains et sur l’engagement dans une interaction humain-robot. Nous souhaitons étudier les mécanismes de comportement social que doit adopter le robot pour se déplacer, détecter l’intention d’interagir et initier proactivement une interaction dans un contexte mobile en présence d’humains. Ce document explore les travaux sur la gestion des signaux sociaux (émission et réception) dans un contexte mobile, afin d’identifier les verrous scientifiques du domaine en vue de proposer un modèle intégrant l’interprétation et la production des signaux sociaux. Ce modèle visera à améliorer l’engagement dans les interactions humain-robot pour offrir un comportement social plus cohérent et naturel dans un contexte mobile. Le domaine étant très vaste nous nous limitons à la gestion de l’espace dans un lieu public et l’initiation d’une interaction humain-robot. Nous aborderons donc les thématiques suivantes : la manière dont des individus se répartissent l’espace public, l’interprétation informatique des signaux sociaux et les architectures de la navigation pour robots sociaux.

Table des matières

1	Introduction	1
2	État de l’art de la Robotique Sociale : Interactions, Initiation et Engagement avec un Robot Mobile Socialement Conscient	3
2.1	Comportements sociaux et répartition de l’espace dans un environnement public . . .	3
2.2	Notion d’engagement dans une interaction humain-robot	6
2.3	Interpréter les signaux sociaux	8
2.4	Architecture pour la navigation des robots sociaux	10
3	Conclusion et travaux futurs	12

1 Introduction

A l’heure actuelle, l’immense majorité des robots en service sont conçus pour des applications industrielles ou militaires et représente un marché de plusieurs milliards de dollars. Néanmoins, la robotique personnelle et de service, qui représente aujourd’hui un marché moindre, se développe rapidement. Des exemples d’utilisation de robots d’accueil commencent à apparaître, une chaîne de supermarchés a par exemple déployé des robots Pepper (SoftBank Robotics, anciennement Aldébaran) dans ses enseignes pour “enquêter sur la satisfaction client, divertir les enfants et offrir des renseignements sur la carte de fidélité”. Cette démarche est motivée par la volonté de “créer une relation engageante avec le client et un service de recommandation”. Au delà de l’aspect événementiel que cette attraction a créé, il reste à étudier scientifiquement l’acceptabilité et l’impact de ces technologies sur le public. Dans ce document nous nous concentrons sur un sous-domaine de la robotique appelé robots socialement intelligents (*socially intelligent robotics*). Fong et Dautenhahn [5] définissent en 2003 ces robots par des agents qui montrent des aspects de l’intelligence sociale humaine, basés sur des modèles profonds de la cognition humaine et de la compétence sociale. L’étude des interactions humain-robot a donné naissance à un vaste domaine appelé HRI (*Human-Robot Interaction*).

L’objectif de nos travaux futurs est d’étudier les différentes approches existantes dans les HRI et la navigation robotique afin de proposer un robot capable d’évoluer dans un environnement en présence d’humains (tel qu’un salon ou l’accueil d’un magasin) et de proposer une interaction fluide et naturelle basée principalement sur le comportement non-verbal du robot. Les travaux seront réalisés en collaboration avec le Lab-STICC et le Crédit Mutuel ARKEA, une banque d’investissement ouverte sur les nouvelles technologies. Le travail réalisé sera implémenté sur le robot Pepper (Fig. 1) conçu par la société SoftBank Robotics. Pepper est un robot humanoïde équipé de nombreux capteurs (caméras, sonars, lasers, capteurs RGB-D etc.) et de trois roues omnidirectionnelles lui permettant de réaliser des déplacements latéraux et des rotations (robot omni-directionnel ou holonome).

La robotique a émergée vers la fin du 20ème siècle mais la notion d’entités semblables à des robots est présente dans l’esprit des hommes depuis des siècles dans la fiction, la mythologie, la philosophie et la religion. La première apparition du mot “robot” vient du mot “*robota*” en Tchécoslovaque qui signifie travail, labeur. Nicolas Tesla fut le premier à imaginer un robot radio-contrôlé sous la forme d’un bateau (Fig. 2) qu’il généralisa ensuite dans un brevet *Method and Apparatus for Controlling Mechanism of Moving Vessels* [35]. L’invention fut utilisée comme une arme navale mais Tesla avait déjà émit l’hypothèse suivante :

“*Vous ne voyez pas là une torpille sans fil (...) vous voyez là le premier d’une race de robots, des hommes mécaniques qui feront le laborieux travail de la race humaine.*”¹

Les premiers essais de chien robot (*Electric Dog*) par l’armée en 1923 donnent lieu à la création



FIGURE 1 – Robot Pepper produit par la société SoftBank Robotics.

1. Citation originale de Tesla : “*You do not see there a wireless torpedo (...) you see there the first of a race of robots, mechanical men which will do the laborious work of the human race.*”

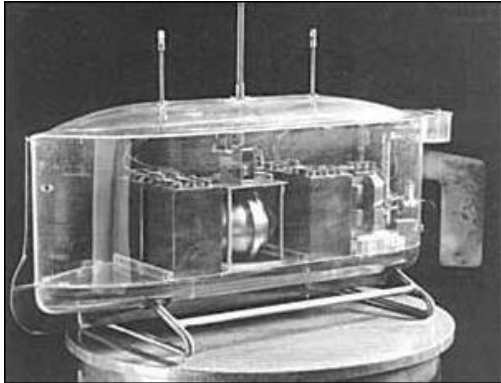


FIGURE 2 – Bateau radio contrôlé de Tesla.



FIGURE 3 – Le robot Shakey et Charles Rosen responsable du projet.

des premiers robots destinés à imiter la nature pour donner une apparence de vie. D'un autre côté l'émergence des premières techniques d'intelligence artificielle donne naissance à des robots totalement autonomes pour l'exploration sous marine ou divers milieux inaccessibles à l'homme. L'exemple le plus cité reste le robot Shakey créé en 1966 marquant une étape pour la robotique mobile (Fig. 3), un des résultats notables de ce projet fut l'algorithme A* créé par Peter Hart et son équipe à Stanford Research Institute [10, 9].

L'apparition des premiers robots possédant des attributs et qualités sociales, comme les tortues robotiques créées par William Grey Walter dans la fin des années 1940, va donner naissance au domaine des robots sociaux. Dans les années 1980 de grandes avancées sont réalisées grâce aux publications de Brooks et Arkin qui introduisent le paradigme de *behavior-based robotics*. La notion de *behavior-based robotics* consiste en la création de plusieurs modules autonomes responsables de comportements basiques afin d'obtenir l'émergence d'un comportement complexe. Le domaine des HRI a commencé à émerger dans le milieu des années 1990 et début 2000, une définition d'un robot social est par ailleurs proposée en 1999 par K. Dautenhahn et A. Billard [4] :

*“Les robots sociaux sont des agents incarnés qui font partie d'un groupe hétérogène : une société de robots ou d'humains. Ils sont capables de se reconnaître et de s'engager dans des interactions sociales, ils possèdent des histoires (perçoivent et interprètent le monde en fonction de leur propre expérience), et communiquent explicitement et apprennent les uns des autres.”*²

Le vaste domaine des HRI est un domaine multidisciplinaire impliquant l'analyse, la conception, la modélisation, la mise en œuvre et l'évaluation de robots à usage humain [5]. Ce domaine multidisciplinaire, allie les sciences sociales, l'informatique, l'ingénierie et la psychologie. En effet la robotique sociale et la robotique en générale contribuent et bénéficient à de nombreux domaines de recherche comme les sciences cognitives et sociales en abordant des sujets tels que la recon-

2. Citation originale de K. Dautenhahn et A. Billard : *Social robots are embodied agents that are part of a heterogeneous group : a society of robots or humans. They are able to recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other.*

naissance et la production des émotions, l’organisation sociale et le partage de l’espace dans un environnement public ou l’interaction avec un humain. Tous ces aspects incluent de nombreux comportements verbaux et non verbaux qui composent les interactions sociales complexes définies par la culture humaine dans laquelle le robot évolue. Les HRI nous obligent à mieux définir et formaliser les différentes composantes verbales et non-verbales, nombreuses et complexes, produites par l’humain lorsqu’il évolue dans un espace public en présence d’autres agents (humains ou robots). De nombreux travaux ont été réalisés depuis les années 2000 sur l’engagement dans une interaction humain-robot, ainsi que sur le comportement social que doit adopter un robot en présence d’un humain. Les avancées technologiques nous permettent de doter les robots mobiles d’une meilleure perception afin de mieux comprendre l’environnement que les entoure. Nous souhaitons étudier le comportement social que doit adopter le robot afin de se déplacer, de détecter l’intention d’engagement et d’initier proactivement une interaction dans un contexte mobile en présence d’humains. L’intérêt de nos travaux résidera dans la gestion des signaux sociaux (émission et réception) dans un contexte mobile afin d’améliorer l’engagement dans les HRI pour offrir une interaction plus fluide et naturelle.

Ce document présente les différents domaines liés à la réalisation d’un robot social mobile, dans un premier temps nous présentons les différentes études dans les sciences sociales qui ont permis la formalisation de la répartition de l’espace dans un environnement public et les normes qui s’y appliquent (Section 2.1). Nous définissons ensuite la notion d’engagement dans une interaction dyadique (Section 2.2) et présentons les métriques et méthodes permettant d’analyser et d’interpréter le comportement social non verbal d’un être humain (Section 2.3). Enfin, afin de permettre au robot de se déplacer et d’évoluer dans un environnement social, nous présentons les différentes architectures pour la navigation robotique à l’état de l’art permettant d’implémenter les différents comportements et règles sociales présentés précédemment (Section 2.4).

2 État de l’art de la Robotique Sociale : Interactions, Initiation et Engagement avec un Robot Mobile Socialement Conscient

Afin d’obtenir un robot capable d’évoluer dans un environnement peuplé d’humain et d’interagir avec ceux-ci, il est nécessaire d’étudier et de clairement définir les différentes règles et comportements établis lors d’interactions sociales entre humains, le domaine étant très vaste nous nous limitons à la répartition de l’espace dans un lieu public et les différents signaux sociaux émis lors d’une interaction humain-robot.

2.1 Comportements sociaux et répartition de l’espace dans un environnement public

L’homme adopte de nombreux comportements conscients et inconscients lorsqu’il évolue dans un espace public en fonction de ses propres buts mais aussi des différents éléments mobiles ou immobiles présents autour de lui. Dans notre cas un lieu public est défini par un lieu où plusieurs individus évoluent dans un espace dit “partagé” entre les différents agents qui l’occupent. Un robot social étant voué à évoluer dans ce genre d’espace public, un bureau par exemple [17], afin d’interagir avec les humains, il est nécessaire de lui permettre de percevoir et de comprendre l’organisation sociale.

Plusieurs études dans les sciences sociales [8, 7, 12, 23, 32] ont permis la formalisation et la mise en place de métriques pour étudier le comportement humain. Hall propose une représentation psy-

chosociale des différents espaces personnels d'un individu, définis culturellement. A partir de cette représentation, Hall introduit la notion de proxémique [8]. Il définit ce concept en proposant une segmentation des distances physiques dans les interactions dyadiques, définissant les quatre espaces suivants (Fig. 4) : la sphère publique, la sphère sociale, la sphère personnelle et la sphère intime. Il propose un modèle psychophysique basé sur la représentation des expériences sensorielles provoquées par les stimuli sociaux : l'expérience visuelle, l'expérience auditive, l'expérience kinesthésique, l'expérience olfactive, l'expérience thermique et l'expérience tactile. Grâce à cette représentation proxémique Hall explique comment et pourquoi sont gérées ces distances interpersonnelles : La sphère intime est réservée aux personnes proches avec qui nous partageons une relation intime car les expériences sensorielles olfactives, thermiques et kinesthésiques sont présentes, ces expériences sensorielles prolongées sont caractéristiques des relations intimes [7]. En revanche dans une relation amicale les expériences olfactives et thermiques n'étant pas forcément désirables les personnes se postent dans la sphère personnelle (voir sociale). A cette distance les expériences kinesthésiques restent possibles, par exemple de brèves embrassades sont communes dans les relations amicales. Deux inconnus en revanche se tiennent dans la sphère sociale, hors de portée d'une possible agression. Ces distances sont, par exemple dans la culture américaine et européenne, inconsciemment mesurées lors d'une poignée de main.

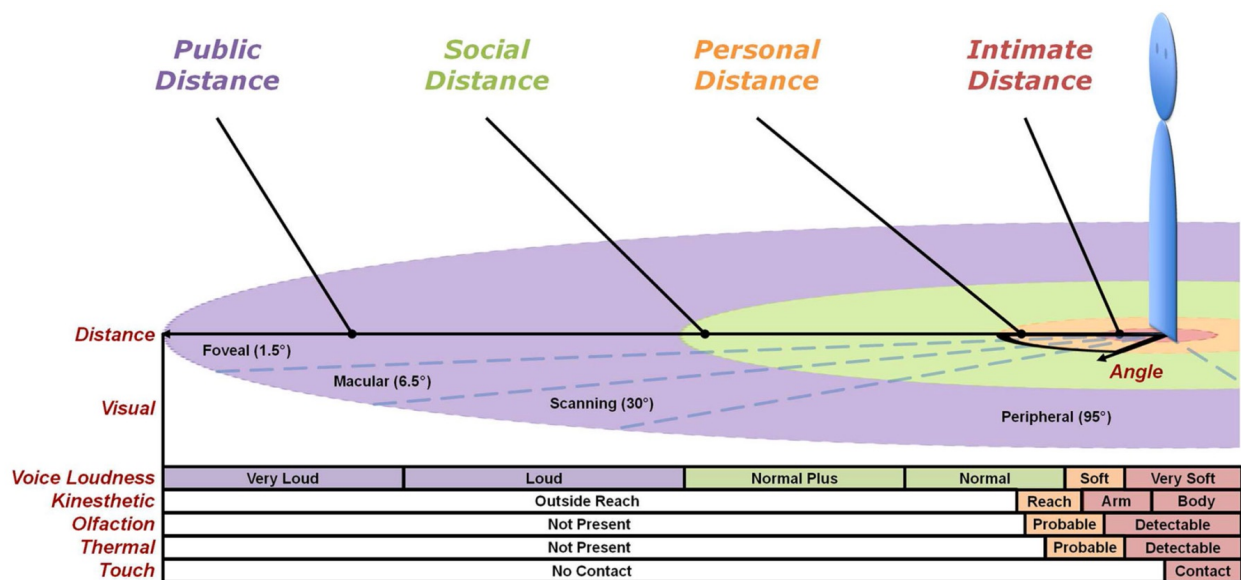


FIGURE 4 – Représentation [20] psychophysique de la proxémique proposée par Hall [8].

En addition au travail de Hall pour la représentation proxémique d'un individu, Kendon propose en 1976 [12] d'introduire une formalisation de l'organisation de l'espace social lors d'interaction avec un groupe. En effet les interactions en groupe impliquent une organisation plus complexe, de l'espace, que les interactions dyadiques et que le modèle proxémique proposé par Hall. Ces échanges non-verbaux incluent la direction du regard, la position spatiale des individus par rapport aux autres, la position du corps (épaules, torse, tête, bras) [32] et les expressions faciales. Ils offrent de nombreuses informations sur l'interdépendance des membres du groupe et la cohésion du groupe. Kendon met en avant le fait que les informations spatiales (orientation, position) sont des moyens de communications par lesquels des attentes ou des intentions peuvent être transmises [12]. Il propose

de modéliser l'organisation spatiale d'un groupe (incluant deux humains ou plus) en introduisant le paradigme des f-formations (*frequent formations*). Cette organisation divise l'espace en trois groupes (Fig. 5) :

- **o-space** : l'espace au centre du groupe
- **p-space** : l'espace où les individus se trouvent
- **r-space** : l'espace extérieur au groupe, en dehors du p-space

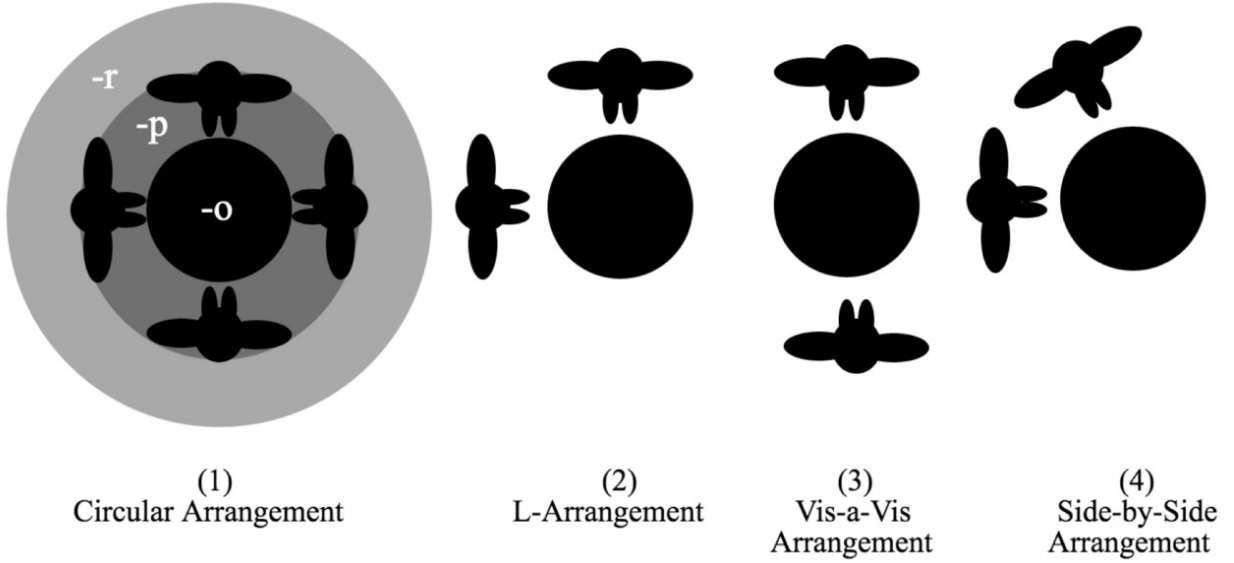


FIGURE 5 – Représentation [34] d'un groupe avec le modèle des f-formations proposé par Kendon [12].

Le concept de f-formation permet aux robots mobiles d'analyser le comportement social d'un groupe pour adapter son comportement durant l'interaction sociale. Lors de ses déplacements, ces informations permettent par exemple au robot de ne pas s'introduire dans un o-space réservé au groupe le partageant. Le p-space et la position des agents l'occupant informent le robot sur les positions possibles pour un point de rencontre afin d'initier une interaction avec un groupe.

Plusieurs études se basent sur la représentation physique de la proxémie proposée par Hall [22, 11, 20, 36, 24]. Cette approche permet d'éviter de passer trop près d'un humain en pénétrant dans sa sphère intime ou même sociale lorsqu'il n'y a pas lieu d'une interaction ou à l'inverse de positionner le robot à une distance socialement correcte pour interagir. Holthaus et al. [11] ont par exemple appliqué le modèle spatial proxémique à un robot immobile pour contrôler son attention. Les f-formations appliquées à la robotique peuvent permettre au robot de coopérer avec un groupe, déduire des relations, voir même de détecter les états affectifs des individus.

Dans leur étude réalisée en 2016 Tseng et al. [36] soulignent que les recherches précédentes se concentrent soit sur la représentation proxémique de Hall soit sur les f-formations proposées par Kendon, c'est pourquoi ils proposent une approche basée sur la perception des humains (reconnaissance faciale, langage, gestes) et associent le modèle psychosocial de Hall et les f-formations de Kendon pour permettre au robot d'inférer la nature de la situation sociale. En effet la représentation psychophysique de Hall permet de modéliser la perception et le ressenti d'un individu tandis

que le modèle proposé par Kendon permet de modéliser l’organisation de l’espace social qui émerge lorsqu’un groupe se forme pour démarrer une interaction. Lors de cette interaction le modèle de Hall reste valable, les personnes se trouvant dans le p-space respectent les distances sociales, personnelles et intimes en fonction de leurs affinités. De même que les distances publiques sont respectées par les personnes extérieures au groupe (présentes dans le r-space) et les sensations (tactiles, thermiques, kinesthésiques, etc.) modélisées par le modèle de Hall sont toujours présentes.

Ces deux modèles sont donc deux sources d’informations que peut utiliser le robot pour inférer des informations afin de mieux adapter son comportement durant l’interaction. Le modèle des f-formations permettent au robot de détecter des groupes pour ensuite étudier leurs interactions interpersonnelles grâce au modèle psychosocial de Hall et ainsi inférer la situation sociale pour permettre au robot de savoir s’il peut proactivement initier une interaction. L’initiation d’une interaction soulève une autre problématique, liée à la notion d’engagement lors d’une interaction dyadique, et abordée dans la section suivante.

2.2 Notion d’engagement dans une interaction humain-robot

Nous venons de décrire l’organisation de l’espace lors d’interactions entre plusieurs individus, afin de pouvoir prendre part à ces interactions un robot doit être capable de reconnaître l’initiation ou la fin d’une interaction et l’expression de l’intention d’engager une interaction (l’intentionnalité). En 2002, Bierman [33] souligne le fait que lorsque les gens parlent, ils maintiennent une connexion psychologique consciencieuse les uns avec les autres où chacun essaye de mutuellement maintenir le contact au travers de gestes, pauses ou regards.³ Il souligne le fait que “les machines” ne prennent pas en compte cette notion d’engagement et qu’il s’agit d’un domaine de recherche à part entière. Le rapport final du DARPA/NSF (Defense Advanced Research Projects Agency/National Science Foundation) sur les HRI recommande d’améliorer les relations humain-robot et de travailler en particulier sur le problème de l’intentionnalité dans une HRI [2].

Sidner et al., en 2005, [33] soulignent que les comportements conversationnels comme le regard ou les tours de parole ont été étudiés dans les sciences sociales et psychologiques dans le cadre général de la communication. Ils réalisent une étude sur l’interaction humain-robot mettant en avant l’aspect central de la notion d’*engagement*. Ils définissent l’engagement par le processus par lequel deux (ou plus) participants établissent, maintiennent et mettent fin à leur connexion perçue durant l’interaction qu’ils entreprennent conjointement. Ils proposent de modéliser l’engagement en trois étapes : (1) initiation de l’interaction, (2) maintien de l’interaction, (3) désengagement. Ils étudient pour la première fois les effets des différents comportements d’un robot sur l’engagement d’un humain dans une HRI et montrent une augmentation de l’engagement lorsque le robot produit des gestes pendant l’interaction. Pour l’humain l’engagement est un processus implicite et majoritairement inconscient et ces informations sont émises verbalement et non-verbalement lorsqu’une personne souhaite interagir. Rich et al. étendent ces travaux en proposant un modèle informatique pour reconnaître l’engagement à partir de signaux sociaux appelés événements de connexion (*connection events*) [29]. Au travers de la littérature sociale, ils formalisent quatre événements de connexion pour mesurer

3. Citation originale de Bierman : “*The point is that when people talk, they maintain conscientious psychological connection with each other and each will not let the other person go. When one is finished speaking, there is an acceptable pause and then the other must return something. We have this set of unspoken rules that we all know unconsciously but we all use in every interaction. If there is an unacceptable pause, an unacceptable gaze into space, an unacceptable gesture, the cooperating person will change strategy and try to re-establish contact. Machines do none of the above, and it will be a whole research area when people get around to working on it*”.

l'engagement et ils proposent un modèle permettant de reconnaître ces différents éléments lors d'une interaction (Fig. 6) :

1. **Regard dirigé (*directed gaze*)** : Une personne (*the initiator*) regarde et optionnellement pointe du doigt un élément de l'environnement, après quoi l'interlocuteur (*the responder*) regarde l'élément introduit (Fig. 6 (a)).
2. **Regard facial mutuel (*mutual facial gaze*)** : L'initiateur regarde le visage de l'interlocuteur, après un délais (possiblement très bref) l'interlocuteur regarde le visage de l'initiateur ce qui débute la période du regard facial mutuel (Fig. 6 (b)).
3. **Enoncés conversationnels contigus (*conversational adjacency pairs*)** : En linguistique une *conversational adjacency pair* consiste en deux énonciations par deux orateurs, avec un délai minimal entre les deux, de telle sorte que la première énonciation provoque la seconde. Un exemple classique de *conversational adjacency pairs* est la question-réponse (Fig. 6 (c)).
4. **Retroactions (*backchannels*)** : L'interlocuteur (*the responder*) émet une brève communication verbale ou gestuelle vers l'initiateur pendant qu'il communique, par exemple un hochement de tête (Fig. 6 (d)).

Argyle et Cook [1] relèvent l'évidence du manque d'intérêt et d'attention dans le cas d'un échec lors d'un échange de regard et précisent que le regard est une source d'information pour mesurer l'attention d'un individu.

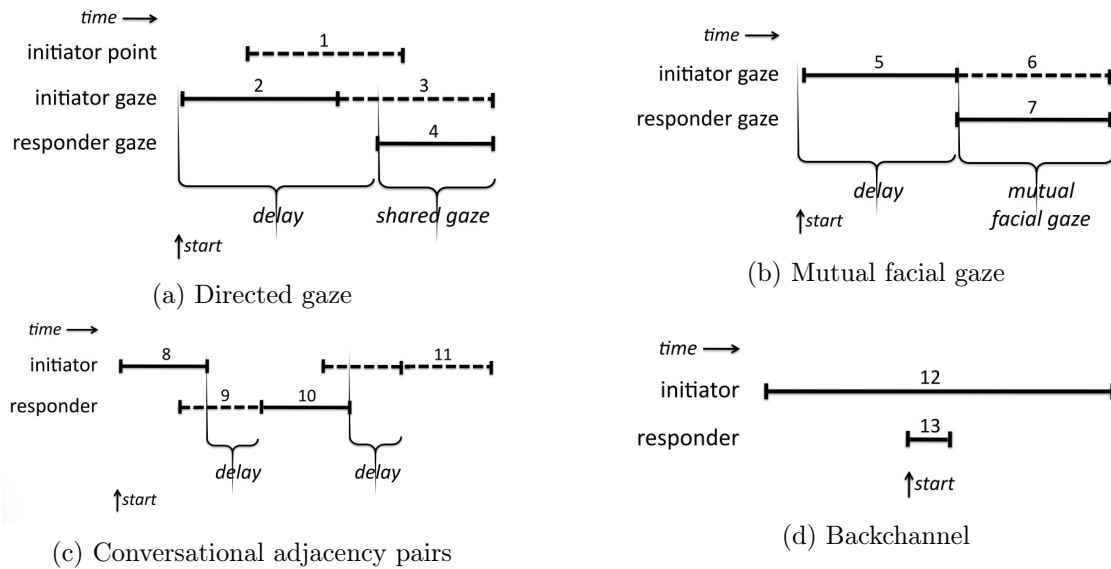


FIGURE 6 – Formalisation des événements de connexion par Rich et al. [29]

Dans nos travaux, nous souhaitons nous intéresser plus particulièrement à l'initiation de l'interaction. De plus les travaux réalisés par Sidner et al. sont présentés dans un contexte statique du point de vue du robot et l'humain est le seul responsable de la gestion des distances sociales. Dans un contexte mobile un robot ne peut pas seulement attendre que l'humain s'approche de lui pour initier une conversation. Afin que le robot ait la possibilité d'apporter de l'aide à une personne qui en a potentiellement besoin, il lui est nécessaire d'apporter un service de manière proactive. Le robot doit alors être capable de respecter l'organisation de l'espace social évoqué plus haut et d'émettre

des signaux sociaux afin d'exprimer ses intentions. C'est le cas étudié par Satake et al. [31, 30] au début des années 2010. Ils réalisent pour la première fois une étude sur l'approche proactive d'un robot vers un humain dans le contexte d'un centre commercial. Des capteurs sont positionnés dans une allée, limitant l'espace d'interaction du robot, afin de suivre précisément la positions des passants et leur comportement. Ils étudient aussi un nouveau point important qui consiste à doter le robot de la capacité d'indiquer non verbalement l'intention d'initier une conversation. Dans cette étude les informations sur la position des épaules, la direction de la tête et sa taille ainsi que la vitesse de l'humain sont collectées afin de générer un engagement proactif tout en respectant le modèle proxémique proposé par Hall.

Knight et al. [13] identifient l'importance de l'intentionnalité pour les HRI dans les robots de théâtre et suggèrent que la production de gestes relationnels peut clarifier les objectifs et intentions du robot afin d'améliorer l'interaction. Vaufreydaz et al. [37] choisissent d'étudier le processus durant lequel les humains expriment inconsciemment leurs intentions d'interagir. Ils réalisent une étude destinée à identifier les caractéristiques les plus pertinentes pour détecter la volonté d'un humain à initier une interaction même dans un environnement restreint, tel qu'une maison, dans le but de réaliser une fusion multimodale des signaux sociaux sur un robot. L'étude reprend le modèle en trois étapes de l'engagement proposé par Sidner et al. [33]. Ils relèvent que la distance est un critère couramment utilisé pour détecter l'intention d'interagir, mais dans certains cas, par exemple dans des espaces restreints ou dans une foule, la proximité n'indique pas forcément l'intension d'interagir, nous détaillons leur résultats dans la section suivante (Section 2.3).

Les évolutions technologiques nous permettent aujourd'hui de mieux équiper les robots sociaux afin de percevoir leur environnement. Pour donner au robot la capacité d'analyser et de comprendre des concepts sociaux tels que l'engagement, la perception des différents éléments non-verbaux exprimé par les humains devient une notion centrale dans la robotique sociale. Il est donc nécessaire de doter le robot social d'outils lui permettant de détecter les éléments clés du corps humain tel que la position des bras, des épaules, du regard etc. dans le but d'interpréter les signaux émis durant l'interaction sociale. C'est ce que nous développons dans la section suivante.

2.3 Interpréter les signaux sociaux

Les signaux sociaux sont les expressions d'une attitude envers une situation sociale [34]. En effet la gestuelle, les positions du corps et la proxémique sont des repères comportementaux aujourd'hui détectables par un robot mobile et s'avèrent être une source d'informations très importante pour la compréhension des groupes sociaux. L'homme utilise des signaux proxémiques tels que la distance, la posture, l'orientation des hanches et des épaules, le regard, pour communiquer leur intention d'initier, accepter, maintenir, terminer ou éviter une interaction sociale [19]. Afin de mesurer ces différents paramètres et d'analyser le comportement humain, des métriques sont communément utilisées dans la robotique sociale, les métriques de Schegloff, McNeill, Mehrabian ainsi que la proxémique de Hall et les f-formations présentées en Section 2.1.

Schegloff en 1998 [32] souligne l'importance de distinguer la partie supérieure de la partie inférieure du corps humain. Il écrit que lorsqu'une partie du corps dévie de sa position initiale (i.e. 0° par rapport à la partie inférieure), la déviation ne dure pas longtemps et un comportement compensatoire se produit de la part de la partie du corps qui a émis la déviation (e.g. si la tête réalise une déviation, elle retrouvera sa position d'origine dans un temps réduit). En revanche si le phénomène de compensation est produit par une autre partie du corps il est probable que cela indique un changement dans l'engagement (e.g. la tête tourne mais ce sont les épaules qui réalisent

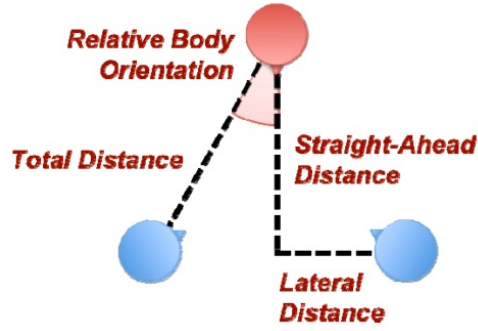


FIGURE 7 – Représentation des métriques de Mehrabian [19]

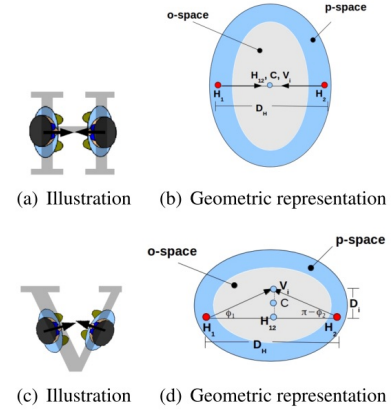


FIGURE 8 – Représentation [25] géométrique des o-spaces.

le mouvement compensatoire). Schegloff définit ce phénomène par le nom *body torque*. Les métriques proposées par Schegloff sont la position et l'angle des chevilles, des hanches, du torse et des épaules. McNeill [18] avance que des gestes attentionnels peuvent aussi servir pour manipuler l'espace social dans des interactions sociales pour introduire un objet d'intérêt mutuel. Les métriques de McNeill incluent la position du bras droit et gauche ainsi que leur référence déictique. Pour l'analyse des comportements proxémiques, Mehrabian [23] propose des métriques basées sur les distances et les orientations interpersonnelles (Fig. 7).

Transposé au domaine de l'informatique, la perception et le traitement de ces informations représentent un domaine à part entière appelé *Social Signal Processing* (qui s'appuie sur les recherches du domaine *Computer Vision*). Les technologies récentes, comme les capteurs RGB-D (Kinect), ont permis de grandes avancées dans le domaine en permettant d'extraire le squelette d'un humain afin de récolter des informations sur les gestes et les postures de l'interlocuteur du robot. La fusion multimodale permet de traiter ces données afin d'inférer des états de l'engagement, de l'attention et des gestes produits par les humains [15, 6]. Des études ont réalisé la fusion multimodale des informations audio et vidéo afin d'inférer le contexte social [27, 26, 37]. L'état de l'art détaillé, publié fin 2016 par Taylor et al. [34], présente les travaux sur la perception d'un groupe par un robot social et souligne que du travail reste à faire pour comprendre simultanément la gestuelle de plusieurs humains. Plusieurs travaux utilisent aussi la géométrie des o-spaces pour étudier le comportement d'un groupe. Narayanan et al. en 2015 proposent la représentation géométrique suivante (Fig. 8) pour permettre à un robot non-holonyme de planifier son trajet lors de l'approche d'un groupe.

Il est possible de collecter d'autres informations avec des capteurs à placer sur l'interlocuteur, mais dans le cadre d'un robot mobile des méthodes non-intrusives sont préférables afin de privilégier le confort de l'interlocuteur en proposant une interaction naturelle, fluide et spontanée. D'autres méthodes consistent à placer des capteurs dans l'environnement où le robot évolue, lui permettant d'avoir une vue exocentrique de la scène. Dans notre cas nous privilégions le point de vue egocentrique pour l'autonomie du robot et sa capacité à s'adapter à différents environnements. L'approche egocentrique apporte tout de même des désavantages tels que le bruit des capteurs notamment généré par les comportements sociaux du robot. Par exemple, dans le cadre d'un robot humanoïde, un mouvement de la tête pour réaliser un événement de connexion, tel qu'un *directed gaze*, va modi-

fier le champ de vision du robot et va par conséquent perturber les capteurs présents sur sa tête (caméras, capteurs RGB-D, micros etc.).

La communication humaine est intrinsèquement multimodale et nous émettons donc de multiples signaux verbaux et non-verbaux pour communiquer nos intentions. En 2013 Mead et al. [20] proposent une approche basée sur la fusion des données métriques physiques proposées par Mehrabian [23], le modèle psychophysique proposé par Hall ainsi que les caractéristiques individuelles étudiées par Schegloff [32], afin d’entraîner des HMMs (Hidden Markov Models) à reconnaître des comportements spatiotemporels signifiant l’initiation et la conclusion d’une interaction. Des travaux récents tentent d’aller plus loin en donnant la capacité au robot d’inférer la nature de la situation sociale d’un groupe d’humain [36]. Dans leur étude, Tseng et al. proposent une approche basée sur la perception des humains (reconnaissance faciale, langage, gestes) et la modélisation de l’espace social pour permettre au robot d’inférer la nature de la situation sociale. Vaufreydaz et al. [37] précisent à nouveau que la proxémique n’est pas suffisante pour estimer l’engagement et que la fusion multimodale et l’utilisation des informations posturales sont une source d’informations supplémentaires. L’étude propose donc d’effectuer la classification des caractéristiques (incluant les métriques de Schegloff et les instances proxémiques de Hall) avec l’algorithme MRMR (Minimum Redundancy Maximum Relevance) afin d’extraire les données les plus pertinentes pour la fusion multimodale. Le résultat de l’étude dresse un classement des caractéristiques pour détecter la volonté d’initier une interaction (voir extrait Fig. 9).

Order	Short name	Unit	Description
1	shoulderPose_rot	<i>radian</i>	Rotation of the shoulder
2	cible_dx	<i>meter.seconde⁻¹</i>	Speed in x of pedestrian
3	cible_y	<i>meter</i>	position on Y axis of pedestrian
4	face_size	<i>pixel</i>	Size of face in the RGB frame
5	face_x	<i>pixel</i>	Lateral position of the face
6	beam	<i>radian</i>	Activated audio beam
7	angle	<i>radian</i>	Audio localization (azimut)
8	hipPose_x	<i>meter</i>	Hip X attribute
9	hipPose_y	<i>meter</i>	Hip Y attribute
10	hipPose_rot	<i>radian</i>	Hip rotation angle

FIGURE 9 – Extrait du classement, réalisé avec l’algorithme MRMR, des caractéristiques sociales pour détecter la volonté d’interaction [37]

L’ensemble des métriques présentées dans ces sections et les avancées technologiques réalisées sur les capteurs et les robots mobiles nous permettent de mieux interpréter les comportements verbaux et non-verbaux produits par les humains pour adapter le comportement social du robot. Afin d’implémenter ces connaissances et règles sociales dans un robot, des architectures sont dédiées à la navigation de robots socialement intelligents. Ces architectures permettant la navigation sont décrites dans la section suivante.

2.4 Architecture pour la navigation des robots sociaux

La navigation d’un robot dans un context social n’est plus seulement une question de trouver le plus court chemin sans entrer en collision avec un objet, mais un problème multi-objectifs nécessitant une approche sophistiquée afin de modéliser l’environnement et ses règles sociales pour atteindre un objectif avec une navigation fluide et confortable pour les individus entourant le robot. La navigation

en robotique est un domaine étudié depuis de nombreuses années. En 2010 Marder-Eppstein et al. proposent une architecture pour réaliser la navigation d'un robot dans un bureau [17]. Le framework proposé comporte quatre composants principaux :

- **Perception** : L'ensemble des capteurs du robot permettant de percevoir le monde extérieur (caméra, capteur RGB-D, laser, sonar etc.).
- **Représentation du monde** : L'ensemble des informations est stocké dans une grille de voxel (en trois dimensions), permettant au robot de se localiser et d'en extraire une carte appelée *costmap* (en deux dimensions) représentant les coûts de déplacement. Dans la *costmap* les objets et les humains sont représentés par un coût dit fatal (*lethal*), interdisant au robot de naviger dans ces zones. Il est ainsi possible de respecter et de nuancer les différents espaces sociaux en les modélisant dans la *costmap*, la sphère sociale d'un être humain n'est pas représentée par un coût *lethal* mais elle est pondérée afin d'inciter le robot à l'éviter si l'environnement le permet.
- **Planificateur global (*Global Planner*)** : Le planificateur global utilise la *costmap* pour calculer un plan de haut niveau que le robot doit suivre pour atteindre un but donné.
- **Planificateur local (*Local Planner*)** : Le planificateur local est responsable de la génération des commandes de vitesse pour le robot afin de l'acheminer de manière fluide vers un but donné par le *Global Planner*. Il prend en entrée le plan généré par le *Global Planner* et essaye de le suivre en prenant en compte la cinématique et la dynamique du robot ainsi que les informations présentes dans la *costmap*.

Cette architecture a été implémentée dans le module navigation⁴ sur la plateforme ROS⁵, un système d'exploitation *open-source* pour robot proposé par Quigley et al. [28]. Cette plateforme permet l'échange et la réutilisabilité des recherches et des architectures conçues. L'architecture proposée par Marder-Eppstein a reçu la contribution de nouvelles recherches et a été utilisée dans de nombreux travaux. Une amélioration notable fut la proposition de la *costmap* en couches (*layered costmap*) proposée par Lu et al. en 2014 [16] (Fig. 10). Cette amélioration est maintenant implémentée dans les nouvelles versions du module navigation.

Chik et al. ont très récemment réalisé, une revue complète du domaine de la navigation *Social-Aware* [3]. Dans cette revue, quatre frameworks pour la navigation *social-aware* sont distingués pour synthétiser l'état de l'art du domaine, allant du framework le moins avancé au framework le plus complet (Fig. 11 et 12). Chik et al. introduisent une notion supplémentaire dans l'architecture proposée par Marder-Eppstein et al. :

- **Le modèle de prédiction social (*Social prediction model*)** : Un modèle dédié à prédire le déplacement des humains afin d'augmenter l'efficacité de la navigation dans un environnement peuplé et dynamique.

Ce modèle est généralement introduit dans le calcul de la navigation grâce à la *costmap*. Dans le quatrième framework proposé, Chik et al. suggèrent d'améliorer la navigation en ajoutant un

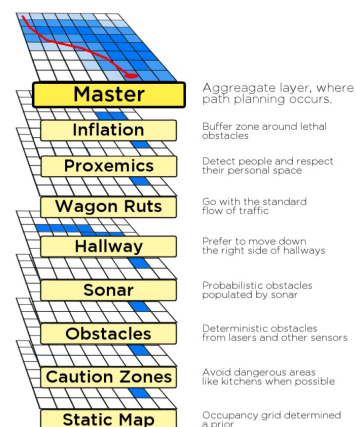


FIGURE 10 – *Layered costmap*, concept introduit par David V. Lu [16].

4. <http://wiki.ros.org/navigation>

5. <http://www.ros.org/>

module calculant les coûts sociaux (*social costs*) mis en avant par Kruse et al. [14] en 2013. Les coûts sociaux incluent l’occlusion d’objets ou les zones cachées afin, par exemple, d’éviter les zones très peuplées ou de réaliser une meilleure approche d’un humain.

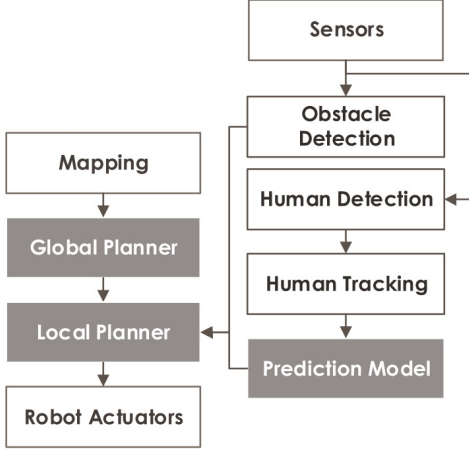


FIGURE 11 – Navigation framework associé à un modèle social prédictif, proposé par Chik et al. [3]

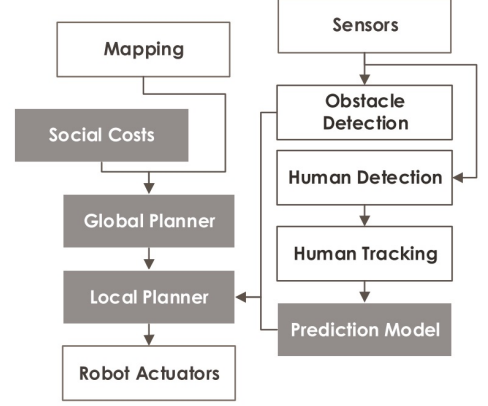


FIGURE 12 – Social-aware navigation framework associé à un modèle prédictif, proposé par Chik et al. [3]

Les différents frameworks proposés par Chik et al. se dédient exclusivement à la navigation du robot et ne prennent pas en compte les signaux sociaux que le robot devrait émettre afin d’exprimer son intention. Il est donc nécessaire d’ajouter différents modules dédiés, par exemple, à la gestion du comportement du robot, au dialogue, à l’engagement ou à la détection des émotions etc. Rich et al. [29] (dont les travaux ont été présenté en section 2.2) ont par ailleurs implémenté un module dédié à la détection de l’engagement. En 2015 Mead et al. [21] montrent l’intérêt du positionnement du robot lors de l’interaction et soulignent le lien entre les distances proxémiques et la façon de générer les signaux sociaux. En effet le robot doit adapter son comportement et son positionnement dans l’objectif de minimiser les risques d’échec lors de la communication avec l’interlocuteur. Mead et al. proposent donc en 2016 [22] un réseau bayésien afin d’estimer le meilleur positionnement du robot en prenant en compte la proxémique de l’interlocuteur et les capacités de perception du robot.

Pour un robot disposant d’un corps anthropomorphique et afin d’obtenir un robot mobile entièrement autonome et capable d’initier une interaction, il est donc nécessaire d’ajouter à ce framework un module que l’on pourrait appeler *Social Planner* chargé de prévoir et contrôler l’engagement, le comportement social et l’intentionnalité du robot à travers la partie supérieure de son corps grâce, en partie, à son regard et à sa gestuelle. Ce module serait responsable de la prise de décision du but à atteindre pour le *Global Planner*.

3 Conclusion et travaux futurs

La robotique sociale mobile est un domaine présentant de multiples problématiques incluant la navigation, la planification de trajectoire, ainsi que la perception et l’interprétation des différents signaux sociaux émis par l’être humain. Le modèle psychosocial de Hall et les f-formations permettent au robot de reconnaître la répartition de l’espace social dans des environnements publics.

Ces modèles permettent aussi d'améliorer la perception et ainsi la navigation du robot en détectant des groupes afin de modéliser l'espace social qui leur est réservé. Ces deux modèles sont donc des sources d'informations que peut utiliser le robot pour inférer des informations afin de mieux adapter son comportement durant l'interaction. Le modèle des f-formations permettent au robot de détecter des groupes pour ensuite étudier leurs interactions interpersonnelles grâce au modèle psychosocial de Hall et ainsi inférer la situation sociale pour permettre au robot de savoir s'il peut proactivement initier une interaction. L'architecture du module navigation de ROS permet d'implémenter, notamment grâce à la *costmap*, la représentation des coûts sociaux lors des déplacements du robot, offrant ainsi une navigation fluide, naturelle et socialement correcte.

Afin d'accomplir une tâche d'assistance, d'information ou de conseil un robot social mobile doit être capable de percevoir l'engagement d'un humain, de détecter l'intentionnalité d'engagement et parfois d'initier proactivement une interaction. La notion d'engagement est centrale dans les HRI, elle permet au robot d'interagir de façon naturelle avec des humains et de mieux comprendre et s'adapter aux situations sociales. L'interaction humaine étant complexe, de nombreux paramètres sont à prendre en considération pour pouvoir interpréter les signaux sociaux non-verbaux émis par l'être humain. Plusieurs métriques sociales permettent de mesurer les différents signaux sociaux émis par l'humain afin de les interpréter et de mieux adapter le comportement social du robot. L'évolution des technologies nous permet de mieux et mieux équiper les robots mobiles de capteurs embarqués, offrant un plus large panel d'informations sur les gestes, les postures et le regard de l'interlocuteur du robot.

A notre connaissance aucune étude n'existe quant au comportement non-verbal social que devrait adopter un robot mobile humanoïde pour initier une interaction. Pepper est un robot humanoïde complet doté d'articulations motorisées au niveau du torse, des bras, des mains, des hanches, des genoux et de la tête. Ses caractéristiques lui donnent la possibilité d'exprimer des comportements non-verbaux destinés à émettre l'intentionnalité d'engagement à travers des signaux tels que le regard, la posture ou les gestes d'illustrations durant une HRI. Le robot Pepper dispose de nombreux capteurs nécessaires à l'extraction des informations sociales tels qu'un capteur RGB-D, des micros, des lasers et des caméras. L'ensemble de ces atouts et sa capacité à réaliser des déplacements holonomes en font l'un des robots les plus adapté à l'heure actuelle pour étudier les Interactions Humain-Robot. L'objectif de nos travaux sera de proposer un modèle pour intégrer l'interprétation et la production des signaux sociaux, afin d'améliorer l'engagement dans les HRI pour offrir une interaction sociale plus cohérente et naturelle dans un contexte mobile.

Références

- [1] Michael Argyle and Mark Cook. Gaze and mutual gaze. Cambridge University Press, 1976.
- [2] Jennifer L Burke, Robin R Murphy, Erika Rogers, Vladimir J Lumelsky, and Jean Scholtz. Final report for the darpa/nsf interdisciplinary study on human-robot interaction. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(2) :103–112, 2004.
- [3] SF Chik, CF Yeong, ELM Su, TY Lim, Y Subramaniam, and PJH Chin. A review of social-aware navigation frameworks for service robot in dynamic human environments. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 8(11) :41–50, 2016.
- [4] Kerstin Dautenhahn and Aude Billard. Bringing up robots or—the psychology of socially intelligent robots : From theory to implementation. In Proceedings of the third annual conference on Autonomous Agents, pages 366–367. ACM, 1999.

- [5] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. Robotics and autonomous systems, 42(3) :143–166, 2003.
- [6] Joseph F Grafsgaard, Kristy Elizabeth Boyer, Eric N Wiebe, and James C Lester. Analyzing posture and affect in task-oriented tutoring. In FLAIRS Conference, 2012.
- [7] Edward T Hall. A system for the notation of proxemic behavior. American anthropologist, 65(5) :1003–1026, 1963.
- [8] Edward Twitchell Hall. The hidden dimension. Doubleday & Co, 1966.
- [9] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. IEEE transactions on Systems Science and Cybernetics, 4(2) :100–107, 1968.
- [10] Peter E Hart, Nils J Nilsson, and Bertram Raphael. Correction to a formal basis for the heuristic determination of minimum cost paths. ACM SIGART Bulletin, (37) :28–29, 1972.
- [11] Patrick Holthaus, Karola Pitsch, and Sven Wachsmuth. How can i help ? International Journal of Social Robotics, 3(4) :383–393, 2011.
- [12] Adam Kendon. The f-formation system : The spatial organization of social encounters. Man-Environment Systems, 6 :291–296, 1976.
- [13] Heather Knight. Eight lessons learned about non-verbal interactions through robot theater. In International Conference on Social Robotics, pages 42–51. Springer, 2011.
- [14] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation : A survey. Robotics and Autonomous Systems, 61(12) :1726–1743, 2013.
- [15] Zhiguang Liu, Liuyang Zhou, Howard Leung, and Hubert PH Shum. Kinect posture reconstruction based on a local mixture of gaussian process models. IEEE transactions on visualization and computer graphics, 22(11) :2437–2450, 2016.
- [16] David V Lu, Dave Hershberger, and William D Smart. Layered costmaps for context-sensitive navigation. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 709–715. IEEE, 2014.
- [17] Eitan Marder-Eppstein, Eric Berger, Tully Foote, Brian Gerkey, and Kurt Konolige. The office marathon : Robust navigation in an indoor office environment. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 300–307. IEEE, 2010.
- [18] David McNeill. Gesture, gaze, and ground. In International Workshop on Machine Learning for Multimodal Interaction, pages 1–14. Springer, 2005.
- [19] Ross Mead, Amin Atrash, and Maja J Matarić. Proxemic feature recognition for interactive robots : automating metrics from the social sciences. In International conference on social robotics, pages 52–61. Springer, 2011.
- [20] Ross Mead, Amin Atrash, and Maja J Matarić. Automated proxemic feature extraction and behavior recognition : Applications in human-robot interaction. International Journal of Social Robotics, 5(3) :367–378, 2013.
- [21] Ross Mead and Maja J Matarić. Proxemics and performance : Subjective human evaluations of autonomous sociable robot distance and social signal understanding. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5984–5991. IEEE, 2015.

- [22] Ross Mead and Maja J Matarić. Autonomous human-robot proxemics : socially aware navigation based on interaction potential. Autonomous Robots, pages 1–13, 2016.
- [23] Albert Mehrabian. Nonverbal communication. Transaction Publishers, 1972.
- [24] Jonathan Mumm and Bilge Mutlu. Human-robot proxemics : physical and psychological distancing in human-robot interaction. In Proceedings of the 6th international conference on Human-robot interaction, pages 331–338. ACM, 2011.
- [25] Vishnu K Narayanan, Anne Spalanzani, François Pasteau, and Marie Babel. On equitably approaching and joining a group of interacting humans. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 4071–4077. IEEE, 2015.
- [26] Aastha Nigam and Laurel D Riek. Social context perception for mobile robots. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 3621–3627. IEEE, 2015.
- [27] Maria Francesca O’Connor and Laurel D Riek. Detecting social context : A method for social event classification using naturalistic multimodal data. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, volume 3, pages 1–7. IEEE, 2015.
- [28] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros : an open-source robot operating system. In ICRA workshop on open source software, volume 3, page 5. Kobe, 2009.
- [29] Charles Rich, Brett Ponsler, Aaron Holroyd, and Candace L Sidner. Recognizing engagement in human-robot interaction. In Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on, pages 375–382. IEEE, 2010.
- [30] Satoru Satake, Takayuki Kanda, Dylan F Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How to approach humans?-strategies for social robots to initiate interaction. In Proceeding of 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 109–116. IEEE, 2009.
- [31] Satoru Satake, Takayuki Kanda, Dylan F Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. A robot that approaches pedestrians. IEEE Transactions on Robotics, 29(2) :508–524, 2013.
- [32] Emanuel A Schegloff. Body torque. Social Research, pages 535–596, 1998.
- [33] Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. Artificial Intelligence, 166(1-2) :140–164, 2005.
- [34] Angelique Taylor and Laurel D Riek. Robot perception of human groups in the real world : State of the art. In 2016 AAAI Fall Symposium Series, 2016.
- [35] Nikola Tesla. Method of and apparatus for controlling mechanism of moving vessels or vehicles, November 8 1898. US Patent 613,809.
- [36] Shih-Huan Tseng, Yen Chao, Ching Lin, and Li-Chen Fu. Service robots : System design for tracking people through data fusion and initiating interaction with the human group by inferring social situations. Robotics and Autonomous Systems, 83 :188–202, 2016.
- [37] Dominique Vaufreydaz, Wafa Johal, and Claudine Combe. Starting engagement detection towards a companion robot using multimodal features. Robotics and Autonomous Systems, 75 :4–16, 2016.