

Synthèse en français

1. Problématique

La communication multimodale, qui est primordiale dans les relations interpersonnelles, reste encore très limitée dans les interfaces homme-machine actuelles. Parmi les différentes modalités qui ont été adoptées par les recherches en interaction homme-machine, la modalité posturale a été moins explorée que d'autres modalités comme la parole ou les expressions faciales. Les postures corporelles sont pourtant indispensables pour interpréter et situer l'interaction entre deux personnes, que ce soit en termes de contexte spatial ou de contexte social. Les briques de base que sont les dispositifs et les modalités d'interaction (par exemple la Kinect ou les avatars) sont pourtant disponibles. Il manque cependant des modèles informatiques reliant ces médias et modalités aux fonctions de communication pertinentes dans les interactions interpersonnelles comme celles liées à l'espace ou aux émotions.

Les agents virtuels interactifs utilisent plusieurs modalités de la communication humaine comme la parole, et les comportements nonverbaux (les gestes, les postures, les expressions faciales). Une telle interface vise donc une interaction homme-machine intuitive car elle utilise des modalités que l'on utilise tous les jours. Ils sont potentiellement intéressants dans des interactions affectives (peu de modèles existants considèrent les expressions corporelles et des comportements spatiaux liés aux émotions) ou des interactions ambiantes (peu d'agents ambiants prennent en compte la position des utilisateurs). Les travaux actuels sont limités en termes d'expression posturale, de référence à l'espace ou d'expressions de certaines composantes de l'émotion comme les tendances à l'action.

Interaction homme-machine et contexte spatial

L'Interaction Homme-Machine est le processus par lequel les utilisateurs humains travaillent avec des systèmes informatiques interactifs (McGraw-Hill, 2003). Le contexte d'interaction est l'un des concepts fondamentaux qui devrait être considéré lors de concevoir les interfaces homme-machine. Le contexte n'est pas simplement un état d'un environnement prédéfini avec un ensemble de ressources d'interaction fixes. Il fait partie d'un processus d'interaction avec un environnement toujours changeant composé des ressources reconfigurables, migrants, distribués, et multi-échelles (Coutaz et al. 2005). Un concept clé qui relie les êtres humains au contexte est l'espace. L'espace fait partie d'un cadre incontournable pour l'organisation systématique de nos expériences" (Randolph, 1981). L'espace dans l'interaction est l'intention d'être les domaines définis par l'orientation des mouvements des corps des participants (Rodrigues 2009). Il résulte de différents facteurs: l'engagement dans l'interaction, la taille du corps, des caractéristiques individuelles (âge, sexe, etc.), les orientations du corps, et l'affect (Ekman & Friesen, 1969; Rodrigues 2007; Holler 2007). L'expérience corporelle dans le monde réel et la conceptualisation de l'espace sont réciproquement liés entre elles. Plusieurs études ont abordé la question et défini plusieurs différents types d'espaces en interaction, comme l'espace personnel (Hall 1975; Wallbott 1995), l'espace gestuel (McNeill, 2000), et l'espace interactionnel (Sweetser 2006).

Interfaces multimodales

Le terme « modalité » consiste d'un mode de communication en fonction de signaux de communication humaine (par exemple la vision, l'audio etc.) ou de périphériques en entrée (caméra, capteurs tactiles, et microphones). Les périphériques peuvent mapper directement ou non à des indices de la communication humaine (par exemple, clavier, souris, etc.) (Bellik 1995). Les interfaces multimodales "processus combiné des modes d'entrée naturels, tels que stylos, toucher, la parole, gestes de la main, regard, la tête et les mouvements du corps, de manière coordonnée avec la sortie du système multimédia" (Oviatt, 1999). Systèmes multimodaux ont été observées pour augmenter la convivialité de l'interaction homme-ordinateur, car ils offrent aux utilisateurs plus d'un canal de communication (par exemple, comme la parole, le geste, l'écriture) (Oviatt 1996; Mayer 2001). Les interfaces multimodales tentent de rendre les modes d'entrée mappés à la communication humaine naturelle (Bolt 1980; Oviatt 2008).

Concernant la production des systèmes multimodaux, les chercheurs étudient les modes de présentation qui sont les plus appropriés pour les différentes situations et comment les différents modes de présentation doivent être combinés. Présentation de l'information multimodale explore comment un morceau de l'information peut être présentée à travers plusieurs modes de présentation, comme une combinaison de texte, etc parole (Hooijdonk 2008).

Agents virtuels

Un agent virtuel (Cassell et al. 2000) est une interface homme-machine qui combine des gestes, des expressions faciales et de la parole pour permettre la communication face-à-face avec les utilisateurs. Cet agent virtuel est censé être doté de certaines capacités conversationnelles inspirées par la communication humaine telles que la capacité de gérer la prise de parole (Cassell et al. 1994). Différents termes sont utilisés selon le modèle sous-jacent et les capacités de génération: «Embodied Conversational Agent» (CEA), «life-like character » (Kruppa et al 2005.), «l'agent animé" (Cassell et al, 2000.), et "l'agent virtuel" (André et Rist, 2003). Ces différents termes sont basés des significations différentes incluant les animations pré-calculées, la génération dynamique en temps réel, les agents contrôlés, et les agents autonomes (Cassell et al. 2001).

Les agents virtuels pourraient permettre la communication souple et naturelle et l'optimisation de la charge cognitive de l'utilisateur, parce qu'ils reposent sur les modes de communication que nous utilisons au quotidien pour exprimer de différentes fonctions communicatives. Par exemple, les agents virtuels expriment des émotions au cours des interactions affectives. Cependant, cela soulève plusieurs questions de recherche: Comment doter les agents virtuels avec une possibilité d'exprimer des émotions complexes à travers de multiples modalités? Comment l'utilisateur perçoit ces expressions multimodales? Plusieurs études ont exploré comment les agents virtuels pouvaient exprimer des émotions par la voix, les expressions faciales et des gestes. Peu d'études ont considéré la façon dont l'ensemble des expressions corporelles pourraient communiquer des émotions (Bianchi-Berthouze et al. 2003).

Les recherches en agents virtuels sont censées de considérer la correspondance entre les signaux de communication dans de multiples modalités (par exemple, les expressions du visage, le regard, la geste, la posture) et des fonctions communicatives (par exemple, les émotions, l'accent, etc) (Poggi 2000, Pelachaud 2005). La conception

des agents virtuels nécessite d'informer le mieux possible ces relations entre les fonctions de communication et de leurs expressions multimodales. Malgré que de nombreuses études déjà existent dans la littérature, ils ne sont pas spécifiques au contexte et il est difficile d'appliquer leurs résultats dans un contexte différent de leur contexte d'application. Ces limites nécessitent d'autres études empiriques sur les comportements non verbaux pour recueillir des données détaillées comportementales.

Multimodal Corpora

L'approche du corpus multimodale utilise les annotations manuelles et / ou automatique (Kipp et al. 2009). Les corpus multimodaux peuvent être recueillis dans un contexte d'interaction humain-humain (enregistrements de laboratoire, de la télévision, films) ainsi que pendant les interactions homme-machine (utilisant un système fonctionnel ou un système simulé avec un protocole de Magicien d'Oz).

Les études sur la communication nonverbale initialement impliquent l'utilisation des d'annotation manuelles et textuelles. Les outils d'annotation tels que l'Anvil par Kipp (2003) consistent à aider les codeurs humains dans leur annotations manuelle des comportements non verbaux à partir des vidéos. Ces annotations sont ancrées dans le temps sur plusieurs couches. Les différents systèmes de représentation doivent être définis à différents niveaux d'abstraction pour représenter les comportements multimodaux. La cohérence entre les annotations faites par les différents codeurs doit être évalué au moins sur un sous-ensemble des données pour valider le schéma de codage et le protocole d'annotation. Les chercheurs ont également développé des systèmes de transcription automatisée pour capturer les comportements non verbaux en utilisant le traitement d'image ou le system de la capture de mouvement.

Au cours de la conversation spontanée, les modalités comme la parole, le geste, la posture et le regard sont combinées de façon sophistiquée (Loehr 2003). Pourtant, les études sur la façon dont les différentes modalités non verbales interagissent restent manquantes, en partie à cause du manque de ressources accessibles et pertinentes. Par exemple, les corpus multimodaux qui sont actuellement disponibles pour la communauté des chercheurs français sont très limités en termes de nombre de modalités, de l'accessibilité et de la spontanéité. En outre, les études sur les relations entre ces modalités exigent la définition de schémas de codage fiable, décrivant plusieurs niveaux incluant les niveaux phonologique, gestuelle et posturale.

L'interaction ambiante

L'Intelligence Ambiante (Fontana et al 2003) est caractérisée par trois concepts fondamentaux: 1) l'informatique ubiquitaire: les microprocesseurs sont intégrés dans les objets du quotidien qui, traditionnellement, n'ont aucune capacité de l'informatique (comme les livres ou les vêtements), 2) la communication omniprésente: les objets sont dotés d'une capacité de communication sans fil, s'appuyant sur des sources d'énergie qui leur fournissent l'autonomie. Les objets sont capables de l'interopérabilité avec d'autres objets de façon spontanée. 3) les interfaces d'utilisateur intelligentes: les utilisateurs humains sont capables d'interagir avec ces objets de façon naturelle (par exemple en utilisant des gestes), et les objets doit prendre en compte à la fois les préférences des utilisateurs et le contexte. Dans un système d'intelligence ambiante, le calcul intelligent devraient être intégrés dans les environnements quotidiens grâce à une infrastructure omniprésente transparente (composé d'une multitude de capteurs,

actionneurs, processeurs et réseaux) qui est capable de reconnaître, répondre et s'adapter à des individus de manière fluide et discret (Ducatel 2001). Le contexte de l'interaction ambiante se compose du contexte physique (objets de l'environnement), du contexte perceptif (événements ou des actions des autres utilisateurs, projeté / attendu des actions des utilisateurs), du contexte conversationnel (courant rôles conversationnels des participants) et du contexte social (les rôles sociaux des usagers, visiteurs, par exemple, un employé). Les informations contextuelles contiennent à la fois des informations statiques (qui ne change pas pendant l'interaction) et des facteurs dynamiques (qui changent au cours de l'interaction) (Pfleger & Lockelt 2006). Il est nécessaire de cartographier la dynamique d'information contextuelle sur les modalités en sortie pertinentes. Les modalités en sortie qui représentent l'information ambiante devrait ainsi permettre de gérer l'espace partagé entre les utilisateurs et les agents intelligents embarqués.

Les tentatives qui ont été faites pour résoudre ce problème sont limitées à des comportements spatiaux tels que la gestion de proximité. Par exemple, le MirrorSpace (Roussel et al. 2004) consiste d'un écran pour la communication vidéo. Il est capable de la sensibilisation ambiante et utilise la proximité physique pour contrôler le flou de l'image affichée. Par exemple, l'image devient net quand l'utilisateur qui se rapproche de l'écran, et devient floue quand l'utilisateur s'en va. Le concept est basé sur celui de l'espace personnel (Hall 1966), plus proche l'utilisateur se tient à l'appareil, plus l'engagement qu'ils attendent de l'interlocuteur par la communication vidéo.

Les systèmes ambiants ont besoin des protocoles d'évaluation fiables parce que les utilisateurs pourraient éventuellement s'inquiéter de la perte de contrôle du système ambiant, de la complexité de l'installation et la maintenance de ces systèmes, ou des attaques de sécurité qui donnent l'accès à des données concernant leur vie privée. Les chercheurs considèrent plusieurs critères lors de la conception des systèmes pour répondre à ces questions, par exemple la capacité d'information, le niveau de notification, de la fidélité de représentation et de l'accent esthétique (Pousman et Stasko, 2006).

Les agents virtuels permettrait de promouvoir la présence sociale d'une entité vivante dans un contexte d'interaction ambiante (Bailenson et al, 2001; Blascovich, 2002; McQuiggan et al, 2008). Cela pose plusieurs questions: Comment doter les agents virtuels d'une capacité de présence en termes des adaptations dynamiques aux comportements de l'utilisateur au sein de l'environnement ambiant? Comment les utilisateurs se sentent lorsqu'ils sont en présence d'un agent virtuel dans l'environnement ambiant?

Informatique affective et sociale

L'informatique affective consiste à l'étude et au développement des systèmes et des dispositifs qui sont capables de reconnaître, d'interpréter, de traiter et de simuler l'affect humaine (Picard, 1997). La machine doit être capable de s'adapter et de répondre à des états affectifs des utilisateurs, parce que cette capacité est censée être perçue comme naturelle, efficace et digne de confiance (Pantic et al. 2005). Un défi principal est de favoriser la communication avec les utilisateurs socialement située.

Les recherches en informatique affective inspirent de plusieurs approches à l'émotion qui sont développées en psychologie. Par exemple, l'approche cognitive des émotions

considère que l'émotion résulte d'une évaluation cognitive du contexte actuel et de la situation émotionnelle (Scherer 2010).

Interactions corporelles en sciences humaines

De nombreuses études en sciences humaines ont étudié les comportements non-verbaux qui se produisent lors de l'interaction sociale, par exemple, les expressions faciales (Ekman & Friesen, 1975) et les gestes (McNeill 2005, Kendon 2004). Le corps a été moins étudié par rapport à ces modalités. Les chercheurs qui travaillent pour concevoir les interfaces multimodales inspirent souvent de ces études. Par exemple Kipp (2007) a élaboré un schéma d'annotation des gestes conversationnels. Kleinsmith et Berthouze (2003) a recueilli un corpus de capture de mouvements pour étudier les expressions corporelles affectifs.

Peu d'études ont considéré les expressions corporelles. Bull (1978) a proposé un système de notation pour décrire les postures statiques et dynamiques. Il l'a appliqué à plusieurs études sur les interactions dyadiques, y compris deux personnes assises en face à face.

La modalité du corps est censé être importante pour plusieurs fonctions 1) la convergence des représentations du sens chez des interlocuteurs au cours de la conversation (Van Baaren et al sous presse, Lakin et al 2003); 2) la gestion de l'espace au cours de l'interaction sociale (McNeill 2002 ; Rodrigues 2010), et 3) l'expression des émotions (Bianchi-Berthouze et Kleinsmith 2003; de Gelder et al 2004; Coulson 2004; Kret et al 2010).

Plusieurs indices sont censés de promouvoir l'établissement de relations spatiales entre les individus au cours des interactions sociales, par exemples, le pointage, l'orientation du corps, le penchement du tronc, l'évitement / l'approche. Les chercheurs en sciences sociales classent ces signes non verbaux dans deux groupes fonctionnels: la déictique et la proxémique. Les gestes déictiques indiquent un objet, un emplacement ou une direction, ce qui est signifié par la projection d'une ligne droite à partir du point le plus éloigné du corps qui a été prolongée vers l'extérieur/l'espace qui s'étend au-delà de l'individu (Kendon 2004).

Ekman a considéré que les actions du corps pourraient fournir des informations sur l'intensité de l'émotion ressentie (Ekman, 1965). D'autres chercheurs tels que Wallbott a observé que les caractéristiques discriminantes des catégories d'émotion à la fois dans la posture et dans la qualité du mouvement (Wallbott, 1998). Néanmoins, de futures études et des corpus multimodaux seront à fournir des réponses à ces questions: Comment les gens perçoivent les expressions corporelles des autres en termes des émotions?

Interactions corporelles dans interfaces homme-machine

Les interfaces homme-machine basée sur le corps émergent ces derniers ans (Fogtmann et al 2008; Leite et al 2008). Plusieurs conceptions interactives impliquent les modalités en entrée telles que les gestes, le regard, et les mouvements du corps. Néanmoins, les interactions corporelles dans ces applications sont souvent unidirectionnelles. Les utilisateurs peuvent envoyer des informations à travers leurs corps aux systèmes, mais ne peuvent pas percevoir des signaux non verbaux et corporels depuis ces systèmes. Plusieurs tentatives sont quand même faites pour afficher les informations

multimodales à travers le corps des agents virtuels dans des environnements ambiants, par exemple, un danseur virtuel (Reidsma et al. 2006), un chef d'orchestre virtuel (Davis & Bobick 1998), un entraîneur virtuel (Ruttkay et al. 2006), un assistant d'anatomie virtuelle (Wiendl 2007), un assistant du chef de cuisine virtuel (Miyawaki & Sano 2008). Par exemple un agent virtuel est capable d'exprimer des backchannels aux utilisateurs utilisant les hochements de tête (Huang et al. 2011). Très peu des agents virtuels sont capables d'afficher des expressions corporelles pour exprimer les attitudes (Gillies & Ballin 2004). En bref, l'utilisation de l'ensemble du corps de l'agent virtuel pour exprimer la convergence, les backchannels ou de gérer un espace commun d'interaction avec l'utilisateur reste limitée.

2. Questions de recherches

L'objectif de cette thèse est de concevoir un premier modèle informatique permettant d'exploiter les postures dans les interactions homme-machine. Pour cela, il faut modéliser les expressions posturales et les comportements spatiaux chez les caractères virtuels et de les appliquer aux interactions ambiantes et affectives.

Cela soulève plusieurs questions de recherche : comment représenter symboliquement des postures exprimées durant des interactions interpersonnelles ? Comment spécifier les comportements posturaux des agents virtuels ? Quelles doivent être les caractéristiques d'un modèle d'interaction corporel permettant des interactions entre un agent virtuel et un utilisateur dans différents contextes physiques ou sociaux ?

3. Spécification en profondeur

Afin de surmonter ces problèmes, nous allons explorer les expressions corporelles en gardant l'esprit d'améliorer l'interaction entre les agents virtuels et les utilisateurs dans le cadre des deux domaines d'application prometteurs: l'interaction affective et l'interaction ambiante.

Nous avons pour objectif de définir, implémenter et évaluer un modèle informatique pour la sélection pertinente des expressions corporelles d'un agent virtuel en fonction de deux fonctions de communication principales et complémentaires: 1) la communication sur la gestion de l'espace d'interaction partagés par les utilisateurs et les agents virtuels, 2) l'expression des émotions du agent virtuel.

4. Positionnement des travaux

Les systèmes multimodaux actuels et des agents virtuels restent limités dans leur façon de considérer l'ensemble interactions corporelles. Par exemple, une utilisation adéquate des interactions corporelles dans des environnements ambiants devrait favoriser la prise en compte des relations spatiales entre les utilisateurs et les systèmes d'une manière située et sociale. De même, l'étude de l'expression corporelle pleine d'émotion demeure rare.

Peu d'études en sciences humaines examinent les mesures dans lesquelles les expressions corporelles sont impliquées dans la gestion de l'espace qui se fait partagée par deux personnes en interaction ou transmettent les émotions. Il ya un besoin pour de

nouveaux corpus qui permettent l'exploration de ces expressions corporelles pendant l'interaction.

Deux domaines prometteurs semblent pertinents pour l'interaction corporelle: l'interaction ambiante et interaction affective.

Plus précisément, nous mettons l'accent sur les expressions corporelles et nous n'examinons pas en détail la façon dont elles interagissent avec d'autres modalités comme les paroles ou les expressions faciales. En plus, nous utilisons les corpus vidéo et les annotations manuelles pour l'enregistrement de l'annotation des données. L'utilisation des techniques de capture de mouvements dépasse la portée de cette thèse.

5. Contributions

Notre objectif peut être décomposé en trois questions de recherches expérimentales: comment se produisent les expressions corporelles dans la conversation dyadique? Les utilisateurs perçoivent-ils les émotions exprimées par des agents virtuels à travers le corps entier? Comment les utilisateurs perçoivent les relations spatiales avec un agent virtuel?

5.1. Contribution et démarche I

L'approche proposée consiste dans un premier temps à prendre comme point de départ des corpus vidéo filmés dans différentes situations. Nous avons défini un schéma de codage pour annoter manuellement les informations posturales à différents niveaux d'abstraction et pour les différentes parties du corps. Ces représentations symboliques ont été exploitées pour effectuer des analyses des relations spatiales et temporelles entre les postures exprimées par deux interlocuteurs.

Nous avons proposé une méthode pour l'annotation d'un corpus de conversation spontanée du français utilisant un schéma de codage pour l'annotation des postures. Deux types d'analyses ont été faites sur les annotations. L'accord inter-juges a été calculé sur les annotations de l'espace de geste. Nous avons aussi partitionné des postures en catégories (algorithme de classification non-supervisée). Ceci permet l'étude conjointe de plusieurs modalités nonverbales telles que les relations entre la sémantique des gestes et la configuration physique des postures.

Le schéma de codage et les annotations effectuées sur des vidéos ont été exploités pour mesurer la convergence posturale. Notre hypothèse a été que les comportements posturaux des écouteurs se rapprochent des comportements posturaux des locuteurs de façon régulière au cours de la conversation. Nous avons défini et implémenté un algorithme pour quantifier ce phénomène interactionnel à partir des données annotées en termes de postures et du rôle conversationnel.

L'adaptation corporelle au cours de l'interaction sociale a été étudiée. Nous avons défini la convergence posturale comme le phénomène dans lequel la posture de l'auditeur au temps t (s) s'adapte à la posture de l'orateur au temps $t-1$ (s) au sein d'un espace partagé par les deux interlocuteurs durant une conversation dyadique. Nous avons développé une méthode automatique de constater la convergence posturale dans la conversation dyadique sur la base des annotations manuelles de postures. Nous avons constaté que la convergence posturale a lieu principalement en termes de pivotement du bras

(maintenir les bras près du corps ou de lever les bras), de l'orientation du bras radial (diriger le bras à côté ou en avant), de la distance de jambe (maintenir les jambes en retrait vers l'arrière ou rester droite) et du pivotement des jambes (jambes vers l'extérieur de maintenir plié à 90 °, ou vers l'intérieur). Cette étude a confirmé que les postures peuvent permettre l'adaptation à travers des interlocuteurs dans l'interaction sociale.

5.2. Contribution et démarche II

Ces représentations symboliques de postures ont été utilisées dans un deuxième temps pour simuler des expressions corporelles d'agents virtuels. La question a été de savoir si le corps puisse fournir des modalités nonverbales pertinentes dans la transmission des émotions. En effet, l'un des enjeux de la recherche sur les agents virtuels est de spécifier l'expression multimodale des émotions complexes. Les recherches sur l'expression d'émotion se sont plus particulièrement focalisées sur certaines modalités : expressions faciales, regard ou gestes. Au contraire, les expressions posturales de l'émotion ont reçu moins d'attention, de même que les autres composantes de l'émotion, telle que les tendances à l'action.

Nous avons donc introduit une approche de conception et d'évaluation des expressions corporelles à partir d'un composant des émotions particulièrement pertinent pour les études sur les expressions corporelles : les tendances à l'action.

Des animations impliquant deux agents virtuels ont été ainsi conçues puis évaluées perceptivement.

Dans un premier temps, les formes de ces expressions corporelles émotionnelles ont été étudiées. Nous avons conçu des expressions posturales des tendances à l'action en utilisant la plate-forme d'animation des agents virtuels MARC. Ces postures ont été spécifiées à partir des données issues de la littérature et des annotations manuelles d'un corpus vidéo de tendances à l'action. Dans un second temps, 5 couples d'images d'expressions posturales ont été évalués en termes de tendances à l'action et de catégories d'émotion discrète. Les sujets ont reconnu les expressions posturales des tendances à l'action suivantes: *attend*, *disappear from view* et *exuberant*. La perception des catégories d'émotion a été également conforme aux prédictions. Ces résultats ont suggéré que les expressions posturales peuvent être utiles pour exprimer les tendances à l'action. Nous avons confirmé que les expressions corporelles pourraient fournir des informations sur les tendances à l'action des agents virtuels.

Nous avons ensuite étendu ces représentations des tendances à l'action à des animations dynamiques.

Dans un premier temps, nous avons spécifié les paramètres. De même, nous avons spécifié les expressions posturales de tendance à l'action en utilisant la plate-forme d'animation d'agents virtuels MARC.

Deuxièmement, nous avons étudié la mesure dans laquelle la reconnaissance de tendances à l'action varie en fonction du temps de perception. Nous avons trouvé que 2 / 3 du temps d'affichage fait déjà les utilisateurs humains atteint leur seuil de perception de tendances à l'action. Cette étude amène plusieurs suggestions et affirmations: 1.) les utilisateurs humains aurait besoin plus de temps pour percevoir la tendance à l'action

qui se rapporte à des catégories émotionnelles négatives (telles que la tristesse, la honte et la culpabilité) que de percevoir la tendance d'action qui se rapporte à des catégories émotionnelles neutre (telle que la surprise et l'anxiété). 2) il pourrait avoir un effet *arousal* par rapport au temps de perception et la perception de la tendance d'action. Cela signifie que, plus longtemps les utilisateurs humains se mettent exposés aux expressions corporelles dynamiques d'une tendance d'action, plus d'émotions de fortes activations ils attribuaient à cette tendance l'action. 3) les utilisateurs humains seraient mieux à détecter les expressions corporelles menaçantes que celles de non menaçantes.

5.3. Contribution et démarche III

Enfin, dans un troisième temps, les expressions corporelles d'un agent virtuel ont été conçues dans une application mixte faisant intervenir un agent virtuel et un utilisateur dans un cadre d'interaction ambiante. Les postures de l'agent virtuel ont été utilisées pour aider l'utilisateur à localiser des objets dans leur monde réel afin de gérer l'espace partagé avec les utilisateurs. Nous avons défini un premier modèle d'agent qui adapte son comportement postural et spatial à la position des utilisateurs dans une pièce intelligente (iRoom, Bellik & Pruvost). La sélection des expressions corporelles s'est basée sur les informations de localisation de l'utilisateur et de l'objet présent dans la pièce intelligente sous formes de l'aide orienté-tâche.

Nous avons effectué une évaluation expérimentale dans une pièce intelligente impliquant un seul utilisateur. L'agent virtuel adaptatif a suscité chez les utilisateurs des niveaux de la présence sociale perçue, de l'adaptabilité perçue et de l'engagement d'interaction plus élevés par rapport à l'agent virtuel non-adaptatif. En outre, la performance a été moins influencée par la difficulté de la tâche.

S'appuyant sur les résultats de cette évaluation, nous avons étendu les analyses à l'interaction entre l'agent et l'utilisateur. Nous avons étudié les effets d'un tel agent virtuel sur le comportement de l'utilisateur (le regard, actions de mouvement et la tâche des actions liées prise) et les perceptions de l'utilisateur (facilité d'utilisation perçue, l'efficacité perçue, le contrôle perçu, la satisfaction émotionnelle et comportementale, amabilité, la serviabilité, perçu la présence sociale et l'adaptabilité perçue). Cela m'a permis de modéliser l'interaction en prenant en compte de nouveaux éléments : le modèle d'agent, les caractéristiques des utilisateurs, leur perception, leurs comportements. Un modèle causal statistique impliquant une médiation et une modulation a été établi : l'expertise et le sexe des utilisateurs semblent ainsi modérer l'effet que le modèle d'agent exerce sur l'engagement des utilisateurs, alors la perception de la présence sociale à l'égard de l'agent virtuel adaptatif semble médiatiser cet effet.

6. Conclusion

L'utilisation des interactions corporelles ont été supposée pour faciliter la représentation de l'espace chez les utilisateurs humains lors des interactions homme-machine. Nous nous focalisons donc sur les expressions corporelles au sein des multiples espaces d'interaction et explorons son potentiel d'utilisation dans l'interaction homme-ordinateur. Nous avons étudié les expressions corporelles qui ont lieu au sein de l'espace du monde réel, celui de l'environnement virtuel et celui de l'environnement ambiant qui combinent le monde réel et le monde virtuel. Différent domaines ont été abordés, chacun ayant une vue spécifique autour de l'étude des expressions corporelles.

Les résultats de nos expérimentations ont montré que l'espace en interaction serait déterminé par les facteurs tels que la convergence conversationnelle, l'affect et l'engagement dans l'interaction.

7. Perspectives

Cette thèse ouvre des perspectives sur des études plus spécifiques de l'interaction corporelle nécessitant par exemple des annotations automatiques via des dispositifs de capture de mouvement ou la prise en compte des différences individuelles dans l'expression posturale.