

Humains Virtuels et Eye-Tracking : Regard Partagé et Attention Conjointe

Matthieu Courgeon¹, Gilles Rautureau², Jean-Claude Martin³ and Ouriel Grynszpan²

¹ Lab-STICC et Université Bretagne Sud, matthieu.courgeon@univ-ubs.fr

² CNRS USR 3246 et Université Pierre et Marie Curie, gilles.rautureau@upmc.fr ouriel.grynszpan@upmc.fr

³ LIMSI-CNRS et Université Paris Sud, martin@limsi.fr

Résumé — Le regard occupe une place fondamentale dans le développement de la communication et le déroulement des interactions sociales. L'attention conjointe est un mécanisme de synchronisation de l'attention entre deux individus. Son dysfonctionnement est par ailleurs l'un des principaux signes précoces dans le développement de l'autisme. Dans cet article, nous présentons une plateforme logicielle permettant la simulation de l'attention conjointe par des humains virtuels en utilisant le regard de l'utilisateur comme point d'attention. Trois études sont présentées. Nous y explorons la notion de conscience de son propre regard et des conséquences de son regard. Les objectifs de cette recherche sont en partie à visée clinique, mais nos travaux permettent également de mener des études en « affective computing » sur les mécanismes d'interaction agent-humain et de la relation sociale associée. La principale visée clinique est de préciser le déficit attribué à l'autisme concernant l'attention conjointe. A ce titre, le système développé pourra servir dans un cadre d'évaluations cliniques des troubles du spectre autistique. Cette plateforme servira également pour explorer le mimétisme, comportement spontané est reconnu comme pro-social. Nous escomptons qu'un tel système puisse être utile pour l'engagement social des patients.

Mots clé — Attention conjointe, Humain Virtuel, Eye-tracking, Autisme, Mimétisme

1. INTRODUCTION

L'appréciation de l'importance de l'influence du regard sur les interactions sociales découle de notre expérience quotidienne. Les expressions courantes comme «regarder de haut» ou «regard fuyant» illustrent le rôle social du regard. Les recherches dans le domaine de la psychologie emploient le terme de «langage des yeux» (p. 311) [1] pour décrire la fonction communicative du regard. Des expériences ont démontré que la région des yeux, seule, peut transmettre des émotions subtiles [1].

1.1 Spécificités du regard social humain

Les études sur les primates mettent en évidence le lien entre la capacité d'utiliser le regard pour envoyer des signaux aux autres individus et la sophistication de l'espèce en termes d'échanges sociaux [2]. Kobayashi et Kohshima [3] ont comparé la morphologie des yeux chez 874 primates, soit presque la moitié de toutes les espèces connues. Les auteurs ont constaté que les yeux humains sont dotés de caractéristiques exceptionnelles de leur apparence extérieure. Ils portent 1) une sclérotique blanche, dépourvues de pigmentation, contrastant avec un iris sombre, 2) le plus grand rapport de sclérotique visible et 3) un contour allongé horizontalement. Par ces caractéristiques, l'œil humain n'est pas aussi adéquat pour le camouflage comparé à celui d'autres espèces. Cependant, il est extrêmement bien adapté pour permettre de détecter la direction des yeux d'une personne. Les auteurs postulent que ces caractéristiques spécifiques de l'œil humain représentent une adaptation évolutive qui optimise la capacité d'utiliser le regard pour la communication sociale.

1.2 Le rôle central du regard dans les interactions sociales

Les yeux sont communément appelés la «fenêtre de l'âme». Dès l'âge de 3 mois, les bébés commencent à démontrer une préfé-

rence visuelle pour la région des yeux des adultes qui interagissent avec eux [4]. La simple direction des yeux d'une personne est connue pour influencer les processus cognitifs sociaux d'une autre personne qui l'observe. Des études comportementales montrent que l'identification d'une personne est facilitée lorsque les yeux de cette personne sont dirigés vers l'observateur. Par exemple, Macrae et al. [5] ont comparé les temps de réponse dans une tâche où les participants ont eu à distinguer le sexe des visages présentés, soit avec regard direct, soit avec des yeux fermés ou regard de côté. Les participants étaient plus rapides dans la tâche avec les visages dont le regard était orienté vers l'observateur. Hood et al. [6] soutiennent que les visages dont les yeux sont vers l'observateur sont plus efficacement mémorisés que lorsque les yeux sont de côté. Von Grünau et Anston [7] ont introduit le terme «effet regard-dans-la-foule» (*stare-in-the-crowd effect*) (p. 1297) pour illustrer notre capacité à détecter rapidement une personne qui nous regarde dans un groupe. Pour tester cet effet, ils ont conçu une tâche de recherche visuelle où les participants devaient identifier aussi vite que possible une photographie du visage cible ayant regard direct parmi un groupe de visages au regard dévié. Leur expérience a été reproduite par des équipes différentes [8], [9].

Senju et Johnson [10] emploient le terme générique «effet de contact visuel» (*eye contact effect*) (p. 127) pour englober les différents phénomènes d'amélioration cognitives produits par le regard direct. Basé des études en neuro-imagerie, ils ont conçu un cadre théorique, appelé le modèle *Fast Track Modulator*, selon lequel le contact visuel façonne le développement de la cognition sociale au moyen d'une voie sous-corticale rapide qui module le traitement des stimuli faciaux. Cette voie sous-corticale est densément connectée aux régions du cerveau social. Leur théorie considère le contact visuel comme un ingrédient essentiel de la capacité humaine à développer des interac-

tions sociales efficaces.

1.3 Attention conjointe

La communication par le regard a été largement étudiée dans le contexte de l'attention conjointe. L'attention conjointe apparaît au cours de la première année de vie et commence à se manifester autour de six mois [11]. Mundy et Newell [12] décrivent l'attention conjointe comme «l'expression de la capacité humaine extrêmement agile à coordonner son attention avec un partenaire social, fondamentale pour notre aptitude à l'apprentissage, au langage et aux compétences sociales sophistiquées de la vie» (p. 269). Ils soulignent la distinction entre Répondre à l'attention conjointe (*RJA*), ce qui signifie que son attention est influencée et dirigée suite à l'orientation du regard d'autrui, versus Initialisation de l'attention conjointe (*IJA*), où le regard de l'individu dirige l'attention d'une autre personne.

Selon Emery [2], l'attention conjointe peut être interprétée comme la somme de deux processus attentionnels séquentiels. Suivre le regard d'une autre personne, puis se concentrer sur le même objet. Le premier est connu comme le «suivi du regard», c'est à dire la direction des yeux d'une autre personne sert de repère pour orienter son attention. Le suivi du regard a été étudié en utilisant le paradigme de Posner [13] pour démontrer l'existence d'une tendance automatique à orienter son attention dans la direction des yeux de quelqu'un d'autre. Ces expérimentations montrent que les participants sont plus rapides à identifier une cible lorsque celle-ci a été précédemment indiquée par la direction des yeux d'un visage, même si les participants étaient prévenus qu'il n'y avait aucun lien entre les yeux du visage et la localisation ultérieure de la cible [14], [15]. La direction du regard est donc un indicateur puissant de l'objet de l'attention d'une autre personne. Il est peut donc contribuer à déduire les états mentaux d'autrui [16].

1.4 Altération des capacités de communication par le regard dans l'Autisme

Selon Senju et Johnson [17], une défaillance dans le processus de maturation du "contact visuel" pourrait entraîner des déficiences sociales, tels que celles attribuées aux troubles du spectre autistique (TSA) [17]. D'autres théories sur l'autisme soulignent le rôle crucial joué par les signaux du regard. Le modèle «d'hyperactivité de l'amygdale» postule que l'autisme est associé à une tendance à éviter de regarder les yeux des autres, en raison d'une sensibilité émotionnelle exacerbée du système limbique [18], [19]. L'hypothèse de «l'esprit énaïf» suggère qu'une attention réduite au regard de l'autre, pendant les premières années de vie, représenterait un élément crucial dans le développement menant à l'autisme [20], [21]. Des données appuyant ce point de vue ont été proposées par Jones et al. [21] qui font état d'une absence de la préférence visuelle typiquement observée pour la région des yeux chez de jeunes enfants atteints d'autisme. Dans une étude qui a eu un impact profond sur la recherche sur l'autisme, Klin et al. [22] a montré que le temps passé à fixer les yeux est également réduit chez des adultes atteints d'autisme qui avaient un QI moyen ou au-dessus de la moyenne.

L'autisme est considéré comme fortement associée à une déficience de l'attention conjointe qui joueraient un rôle crucial dans le développement de la maladie [11]. Des analyses rétrospectives de vidéos prises à domicile et fournies par les familles révèlent des altérations précoces de l'attention conjointe au

cours des deux premières années de vie [23]. Les déficiences de l'attention conjointe chez les nourrissons avec autisme prédisent l'acquisition tardive du langage et des capacités communicatives [11]. Baron-Cohen et al. [16] ont montré que les enfants atteints d'autisme avaient des difficultés à d'inférer des états mentaux d'autrui à partir de la direction de leur regard.

Explorer les principes qui sous-tendent et régulent l'attention conjointe devrait donc contribuer à affiner la symptomatologie de l'autisme. Les difficultés les plus saillantes de l'autisme semblent associées avec le mode IJA, plutôt que de le mode RJA [12]. Peu de données ont été rapporté concernant la déficience dans l'autisme de la détection de la direction regard [24], bien que les altérations de l'attention conjointe soit indubitable. À ce jour, plusieurs études ont exploré le mode RJA [25], [26], [27], mais à notre connaissance, une seule équipe a examiné le mode IJA [28]. La raison derrière cette pénurie de recherche peut être expliquée par l'absence de technologie adéquate pour implémenter les paradigmes de contrôle du regard par le sujet pendant l'expérimentation.

1.5 Regard et attention chez les humains virtuels

L'émergence des humains virtuels dotés de comportement attentionnel réaliste a progressivement permis de conduire des expériences contrôlées en psychologie. Le regard humain transmet une part importante de la communication non verbale et d'indices expressifs. La modélisation de ces signaux de communication chez les agents virtuels et les agents robotiques peut donc contribuer à établir une communication non verbale crédible.

Selon Bailly et al. [53], un agent virtuel conversationnel doit pouvoir convaincre le partenaire humain que ses réponses ont été produites par une intelligence humaine, non seulement par les contenus linguistique et paralinguistique, mais aussi par la production de signaux multimodaux appropriés. Parmi ces signaux, les auteurs identifient le jeu de regard mutuel comme l'un des plus difficiles à simuler. En analysant des dialogues en face-à-face par ordinateur et en utilisant de l'eye-tracking, les auteurs ont pu identifier les dynamiques temporelles et spatiales associées aux regards des participants. Ils ont confirmé les résultats issus de la littérature, à savoir que lors de la communication en face-à-face, notre système perceptif se concentre principalement sur la bouche et les yeux. Grâce à ces données, ils ont conçu un modèle génératif capable de doter les agents virtuels de comportement du regard crédible.

Peters et Sullivan [52] ont exploré les déplacements d'attention et sa simulation chez les agents virtuels. Ils soutiennent que les clignements des yeux sont fortement liés aux phases de changement d'attention. Les clignements « induit par le regard » commencent généralement en même temps que le début du mouvement de la tête et des yeux. De plus, la probabilité d'un clignement d'œil augmente avec la longueur du déplacement du regard. Grâce à leurs observations et mesures par eye-tracking, ils ont créé un modèle de regard déplacement incluant les clignements induit par le regard.

1.6 Attention conjointe et agents virtuels

Seules quelques tentatives ont été effectuées pour simuler l'attention conjointe avec des personnages virtuels. La plupart de ces recherches attribuent à l'agent virtuel le contrôle de l'attention de l'utilisateur (sur un élément ou une tâche spécifique) [45]. Ainsi, l'utilisateur ne peut que se comporter en mode RJA.

Peters et al. [29] ont créé un système à l'aide d'une estima-

tion simple de la direction du regard à l'aide d'une webcam. Ce système est limité par sa précision. Il peut uniquement déduire la direction du regard comme un vecteur 2D en estimant rotation du visage de l'utilisateur et ses yeux à partir de l'image de la webcam. Afin de surmonter cette limitation, les auteurs utilisent le paradigme de « l'objet de l'attention virtuelle » (VAO). Ces objets d'attention sont des objets connus dans l'environnement virtuel qui sont des candidats probables pour être le centre d'attention de l'utilisateur. Le système évalue alors le niveau d'engagement de l'utilisateur dans la tâche par l'estimation de l'attention conjointe.

Wilms et al. [30] ont conçu un système de regard conjoint pour étudier IJA. Dans leur procédure expérimentale, les participants ont été placés devant un écran représentant un humain virtuel et trois objets virtuels. Un oculomètre a été utilisé pour détecter les fixations du regard des participants sur l'écran. Lorsqu'une fixation est détectée sur l'un des trois objets, la direction du regard de l'humain virtuel était dirigée vers l'objet regardé. D'autre part, leur système de tracking oculaire présente l'intérêt de pouvoir être utilisé au sein d'un scanner IRM. Ce dispositif expérimental a ainsi permis d'identifier les régions du cerveau impliquées dans les mécanismes de comportements attentionnels pendant les épisodes d'attention conjointe [31].

2. SELF-MONITORING DU REGARD

2.1 Contrôle volontaire VS involontaire du regard

Étant donné le rôle crucial du regard dans les interactions sociales, être capable de contrôler son propre regard semble capital. La surveillance de son propre comportement est un élément clé de la prise de décision. Telle qu'elle est conçue par Henderson [32], " la vision est un processus actif dans lequel le spectateur cherche des informations visuelles pertinente pour la tâche qu'il réalise » (p. 498). L'exploration visuelle pendant l'observation de scènes écologiques dépend de ce que l'individu est en train d'effectuer [33]. Cela est particulièrement vrai pour l'attention conjointe, qui nécessite la coordination de notre regard avec celui les autres. Bien que le contrôle du regard puisse être volontaire, la plupart des comportements visuels dans la vie sociale quotidienne ne nécessitent pas de contrôle délibéré. En règle générale, les comportements non-délibérés orientés vers un but sont considérés comme plus efficace en termes de vitesse, comme le montrent les expériences sur des tâches de recherche visuelle [34]. Pourtant, le terme « non-délibéré » n'équivaut pas à une absence de contrôle attentionnel. Les études montrent que le temps de décision représente la plus grande partie de la latence des saccades, impliquant des processus attentionnels "bottom-up" aussi bien que "top-down" [35]. Même si nous n'en sommes pas toujours conscients, nous adaptons notre comportement visuel très rapidement lors de nos échanges sociaux pour saisir l'information visuelle pertinente et émettre les signaux sociaux adaptés. Toutefois, la dynamique d'une relation sociale peut conduire l'un des participants à devenir conscient de son comportement visuel et éventuellement passer à un contrôle volontaire de son regard. Cela soulève la question de ce qui déclenche la prise de conscience de son propre regard.

La contribution du contrôle conscient VS automatique du regard dans l'attention conjointe a été largement négligée jusqu'à présent. Pourtant, les changements dans le comportement

d'une autre personne lors d'un épisode d'attention conjointe pourraient représenter un déclencheur potentiel pour la conscience de soi et du contrôle de son regard. En effet, quand un individu A suit l'attention d'un individu B, le regard de l'individu A se conforme au regard de l'individu B. Ainsi, A fourni à B un retour visuel sur son propre regard. La question devient alors: le retour visuel de l'orientation de son regard pourrait-il susciter la prise de conscience de ce regard? Cette question est extrêmement pertinente pour la conception d'agents virtuels simulant l'attention conjointe. En effet, dans de tels cas, la rétroaction peut potentiellement être plus systématique et plus précise, augmentant ainsi la possibilité que l'utilisateur détecte qu'il a le contrôle, avec ses yeux, de l'agent virtuel.

Dans une série d'expériences préliminaires, nous avons cherché à explorer l'émergence de la conscience du regard en fournissant aux participants des informations visuelles sur l'orientation de leur regard. Le reste de cette section décrit deux expériences. Dans une première étude, les participants ont eu à écouter un humain virtuel racontant une histoire tandis que la position de leur regard était affichée en temps réel. Dans une deuxième étude, la même rétroaction de regard a été utilisée, mais l'affichage visuel n'impliquait pas d'humain virtuel. Cette recherche a été prospectivement examinée et approuvée par le comité régional d'éthique. Un consentement éclairé a été obtenu de chaque participant.

2.2 Expérimentation 1

Cette première expérience a pour but de tester l'émergence de la conscience du regard dans le cadre d'interaction sociales face-à-face interactions [36]. Les participants étaient assis devant un écran d'ordinateur sur lequel s'affiche un humain virtuel réaliste. Un oculomètre (modèle D6-HS de chez Applied Science Laboratories) était situé sous l'écran. Il mesurait à distance l'orientation du regard, sans contraindre les mouvements de tête et sans nécessiter de casque. L'humain virtuel s'adressait aux participants, en leur racontant un événement récent de la vie quotidienne. Pour contourner défis posés par la conception d'interactions séquentielles entre un utilisateur et un humain virtuel [37], nous avons créé des animations courtes (18 secondes en moyenne) qui simulent la phase de début d'une rencontre sociale, sans que l'utilisateur n'ait à répondre à l'humain virtuel. Tout en parlant, l'humain virtuel produit des expressions faciales. Le message énoncé par l'humain virtuel doit être interprété correctement, ce qui exige de regarder ses expressions faciales: par exemple, l'humain virtuel peut dire: «Génial, nous avons des épinards à midi" et afficher simultanément une expression faciale de dégoût pour transmettre de l'ironie. La tâche des participants était de comprendre le message de l'humain virtuel et de sélectionner l'interprétation correcte dans une question à choix multiple. Pour fournir une rétroaction aux participants sur la direction de leur regard, nous avons utilisé le paradigme du "fenêtre de visualisation" [38]. Nous avons conçu un système asservi au regard qui floute l'intégralité de l'écran en temps réel, à l'exception d'une fenêtre de visualisation centrée sur le point focal des participants (Fig. 1).

La fenêtre de visualisation est une zone rectangulaire avec des angles arrondis mesurant 233×106 pixels sur un écran 19 pouces avec une résolution de 1280×1024 pixels. Les dimensions de la fenêtre de visualisation en angles visuels étaient $7^\circ \times 3^\circ$. La fenêtre de visualisation englobe la région de la fovéa

où la perception des contours et des couleurs est la plus précise. Elle était assez grande pour voir les yeux de l'humain virtuel et ses sourcils simultanément.

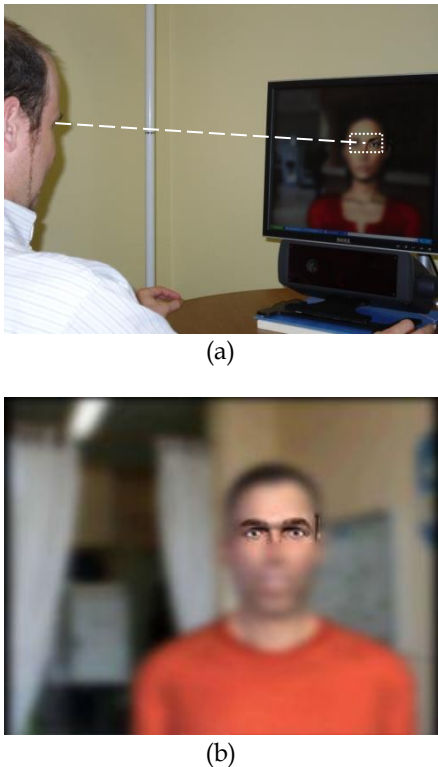


Fig. 1. (a) Le système d'asservi au regard. L'eye-tracker non intrusif est placé sous l'écran et détecte en temps réel la direction du regard de l'utilisateur. La surface d'affichage est entièrement floutée, sauf pour la zone rectangulaire centrée sur le point où regard le participant. (b) Un aperçu de l'affichage visuel avec la fenêtre de visualisation de l'expérience 1. Les participants ont été invités à interpréter le message multimodal communiqué par l'humain virtuel.

Nous avons créé deux modèles d'humains virtuels, un homme et une femme, à l'aide du logiciel Poser Pro (Smith Micro Software, Inc.). Leurs expressions faciales ont été conçues selon les spécifications d'Ekman [39] pour les émotions de dégoût, de joie, de peur, de colère et de tristesse. Les voix des personnages ont été réalisées par synthèse vocale, en utilisant la version du logiciel Virtual Speaker 2.2 en français de la société Acapela (www.acapela-group.com). Les personnages virtuels ont été intégrés dans des vidéos filmées en situation réelle pour permettre un contexte écologique et naturel. Plus d'informations techniques détaillées peuvent être trouvées dans [36].

L'expérience a été passée par 18 adolescents et adultes (10 hommes, 8 femmes) allant de 17 à 40 ans, qui étaient exempt de symptômes psychiatriques ou neurologiques connus. La procédure expérimentale est basée sur un design ABA, avec des blocs de pré et post-test où l'affichage visuel était n'était pas flouté. Le paradigme de la fenêtre de visualisation a été mis en œuvre que dans le bloc du milieu. Chaque bloc contient 20 essais. Les participants ont été informés avant l'expérience que, lors de certains essais, l'affichage visuel contiendrait des zones floues et des zones claires. Cependant, ils n'étaient pas informés du fait que les zones claires étaient asservies à leur regard. Ils savaient en revanche que l'orientation de leur regard serait

suivie tout au long de l'expérience. En outre, ils devaient subir une procédure d'étalonnage standard pour l'oculomètre avant de commencer l'expérience. A la fin de l'expérience, la question suivante leur a été posée: «Vous avez remarqué que dans certaines vidéos, il y avait des parties floues et des parties claires. Qu'est-ce qui provoquent les parties claires? » Leurs réponses ont été analysées par deux juges indépendants afin d'évaluer si oui ou non les participants avaient remarqué que leurs yeux causaient les mouvements de la zone claire (c'est à dire la fenêtre de visualisation). Un accord complet a été obtenu entre les juges. Étonnamment, seule la moitié des participants ont réalisé qu'ils avaient le contrôle de la fenêtre de visualisation [36]. Ceux qui ne s'étaient pas rendu compte qu'ils contrôlaient la fenêtre de visualisation ont communément attribué les mouvements de la fenêtre de visualisation à l'ordinateur. Par exemple, deux participants ont répondu que la fenêtre d'affichage indiquait où chercher l'information émotionnelle. Fait intéressant, trois participants ont déclaré que leur regard suivait la fenêtre de visualisation, sans avoir remarqué qu'ils la contrôlaient.

Le protocole expérimental a également été passé par un échantillon de 13 adultes et adolescents atteints d'autisme [40]. Leur QI était dans la moyenne ou au-dessus et ils avaient une maîtrise linguistique suffisante pour effectuer la tâche. Leur âge moyen était de 20,19 ans (écart type = 4,73) et l'échantillon comprenait une femme et 12 hommes (le ratio hommes-femmes est estimé à 1 femme pour 5 hommes dans le sous-groupe sans retard intellectuel du spectre autistique [41]). L'analyse de leurs réponses a montré que seulement l'un d'entre eux avait pris conscience d'être en contrôle de la fenêtre de visualisation. En outre, les données de l'eye-tracking ont révélé des altérations dans leur capacité à monitorer leur regard [40]. Fait intéressant, leurs performances dans la compréhension du message de l'humain virtuel étaient corrélées au temps passé à regarder le visage à travers la fenêtre de visualisation. Partant de ce constat, le système de fenêtre de visualisation asservie au regard que nous avons conçu pourrait être utilisé pour évaluer leur attention visuelle dans le cadre d'une interaction sociale.

2.3 Expérimentation 2

Dans une seconde expérience, nous avons cherché à trouver une explication possible pour le faible nombre de participants qui ont réalisés qu'ils contrôlaient la fenêtre de visualisation. La tâche de la première expérience nécessitait d'interpréter différentes sources multimodales d'information : la voix et les expressions faciales. Le protocole impliquait aussi une implication importante pour l'analyse du contenu pragmatique et le contexte social du message produit par l'humain virtuel. Il est donc possible que les participants aient pu être monopolisés par la tâche et ne pas avoir accordé suffisamment d'attention à la fenêtre de visualisation. Pour tester cette hypothèse, nous avons conçu une autre tâche qui ne comportait pas de traitement de l'information sociale et ne nécessitait pas la fusion d'informations provenant de différentes modalités.

La tâche dans la deuxième expérience consistait à mémoriser des transformations géométriques. De petites formes géométriques ont été dispersées dans une zone prédéfinie en forme de sablier (Fig. 2). Nous avons choisi cette forme de "sablier" pour sa similarité avec la zone balayée du regard par les participants de la première expérience. En effet, la zone en question, reliant les yeux, le nez et la bouche, correspond à la forme d'un sablier.

Les petites formes géométriques composant le sablier étaient toutes identiques et modifiées simultanément. Chaque essai se composait de quatre transformations de ces petites formes. Les formes pouvant être adoptées sont : un cercle, un triangle, un carré ou une croix. Par exemple, les éléments étaient au départ en forme de triangle, puis se changeaient simultanément en cercles, puis en carrés et enfin forme de croix. Le rythme des transformations était suffisamment lent pour permettre une exploration visuelle fluide. Chaque transformation a duré 6 secondes. Les participants ont été invités à se rappeler l'ordre des transformations, sur lequel ils ont été interrogés après chaque essai. Les formes ont été intégrées dans les mêmes vidéos d'environnement que ceux utilisés dans l'expérience 1 afin que les deux expériences soient comparables. La tâche en expérience 2 n'impliquait aucun signal audio. En dehors de la tâche, le protocole expérimental était identique à celui de l'expérience 1, y compris la question finale.



Fig. 2. Capture d'écran de la seconde expérimentation. Les participants avaient pour consigne d'observer les petites formes géométriques turquoise qui changeaient de forme.

Dix-huit adultes ont participé à l'expérience 2. Ils étaient exempts de tout symptôme psychiatrique ou neurologique connu. Aucun d'entre eux n'avait participé à l'expérience 1. Sur les 18 sujets, seuls quatre participants ont réalisé qu'ils contrôlaient la fenêtre de visualisation, soit 22% de l'échantillon, une proportion plus faible que dans l'expérience 1. Ce résultat infirme donc l'hypothèse selon laquelle les participants n'avaient pas pris conscience de contrôler la fenêtre de visualisation dans l'expérience 1 en raison de la complexité de la tâche. Lorsque nous avons réduit le nombre de sources d'information perceptive et supprimé la nécessité d'interprétation de signaux sociaux, l'émergence de contrôle volontaire de la fenêtre de visualisation n'a pas été facilitée. On pourrait au contraire soutenir que le contexte social semble favoriser la prise de conscience du regard. En effet, notre regard est avant tout perceptif. Nous n'utilisons pas naturellement notre regard pour contrôler les éléments de notre environnement, sauf dans le cas spécifique des interactions sociales, où le regard prend alors un rôle actif et n'a plus seulement une fonction perceptive.

Nous avons également fait passer cette seconde expérience à deux adolescents atteints d'autisme dans le cadre d'une étude pilote de faisabilité. Notre objectif était de tester si la tâche de l'expérience 2 était adéquate pour entraîner des adolescents atteints de TSA à contrôler leur regard en utilisant la fenêtre de visualisation. Le premier adolescent avait 12 ans et le second de

11 ans. Tous deux avaient le niveau requis de compétences verbales de comprendre les instructions et les questions utilisées dans l'expérience. Leur QI et leurs capacités non verbales étaient dans la moyenne. Ils ont terminé avec succès l'expérience et se sont révélés être en mesure de mener à bien la tâche lorsque leur vision était limitée à la fenêtre de visualisation. Les résultats des deux expériences 1 et 2 sont actuellement utilisés pour concevoir une méthode éducative spécifique à l'autisme pour l'entraînement au contrôle du regard social fondé sur le paradigme de la fenêtre de visualisation [47].

Les deux expériences qui sont décrits dans cette section ont pour but d'examiner l'émergence de la conscience de soi du regard causée par un retour visuel sous la forme d'une fenêtre de visualisation. Bien que le paradigme de la fenêtre de visualisation donne un marqueur visible du regard qui peut difficilement être ignoré, elle ne concerne que les études en laboratoire car il ne peut être utilisé en situation réelle. Pour étudier les effets du regard en situation écologique, nous proposons de simuler l'attention conjointe avec des humains virtuels. La conception de ces humains virtuels est décrite dans la section suivante.

3. ATTENTION CONJOINTE SIMULEE PAR EYE-TRACKING ET HUMAINS VIRTUELS

3.1 Eye tracking

Plusieurs technologies eye-tracking coexistent et sont utilisées dans différents domaines d'application. Les dispositifs fixes sur la tête de l'utilisateur, souvent sous forme de lunettes, peuvent être utilisés pour suivre le regard de l'utilisateur pendant les activités quotidiennes dans un cadre écologique [48]. Cependant, ces technologies mobiles nécessitent une analyse post hoc des images filmées sur lesquelles se superposent les traces du regard. Ils ne sont donc pas utilisables pour les interactions en temps réel avec un système informatique. En revanche, les eyetracker « fixes », posés sous un écran, permettent de détecter et de traiter le regard de l'utilisateur en temps réel. Ils utilisent une caméra infrarouge située à proximité de l'écran de l'utilisateur. Cette caméra détecte la cornée et a pupille et en déduit la rotation des globes oculaires. La position du regard de l'utilisateur sur l'écran est ensuite calculée à partir de l'orientation des yeux.

Deux des caractéristiques cruciales pour un dispositif de suivi du regard sont la fréquence et la précision. Nous avons utilisé la caméra EYE-TRAC D6 d'Applied Science Laboratories (www.asleyetracking.com). Ce dispositif fonctionne à une fréquence de 50 Hz avec une précision de 0,5°. Les participants ont été placés à environ 57 cm de l'écran. La taille de l'écran est de 19 pouces (377 × 302 mm) avec une résolution de 1280 × 1024 pixels (environ 37° × 30° à angles visuels). Ainsi, la précision du suivi du regard sur l'écran est d'environ 0,5 cm, soit 17 pixels. Le temps de latence séparant la détection et la prise en compte du regard par l'humain virtuel a été fixé à environ 500 millisecondes. Cette latence a été conçue pour laisser suffisamment de temps à l'utilisateur de regarder les yeux de l'humain virtuel avant qu'ils ne changent de direction. Afin de réduire les mouvements erratiques du personnage virtuel (en raison de l'imprécision de la mesure ou de microsaccades autour du point de fixation), les positions du regard sont moyennées sur une fenêtre temporelle glissante de 100

millisecondes. Cet intervalle de temps a été choisi car il correspond au seuil communément utilisé pour définir la durée minimum d'une fixation [49].

3.2 MARC

La plateforme MARC (Multimodal Affective and Reactive Characters) est un ensemble d'outils dédiés à l'animation d'humain virtuel en temps réel, et utilisant un rendu de haute qualité. La plateforme inclue plusieurs modèles computationnels des émotions et plusieurs modalités d'interaction [50].

Grâce à ces éditeurs dédiés et de la possibilité d'importer des nouveaux modèles 3D, MARC permet la création d'humains virtuels personnalisés. L'utilisation de personnages dédiés peut contribuer à la validité écologique de nos expérimentations. Le travail présenté dans le document a nécessité la création de deux nouveaux personnages virtuels. Les personnages ont été conçus avec Blender 3D (www.blender.org), et sont basés sur des personnes réelles. La modélisation et les textures sont basées sur des photos de leurs visages. Nous avons opté pour des personnes réelles plutôt que des humains imaginaires importées de logiciels de conception graphique 3D comme Poser (Smith Micro Software, poser.smithmicro.com), car ils sont dotés de caractéristiques physiques trop « parfaites » et qui peuvent être trop perturbantes pour l'utilisateur. En effet, les participants de nos expériences précédentes ont fait des commentaires sur les aspects intimidants des personnages virtuels que nous avons précédemment employés. Étant donné que notre travail est destiné à être utilisé avec des personnes ayant une déficience sociale, nous avons préféré l'utilisation d'humains virtuels naturels et ordinaires. La figure 3 montre les visages de ces deux humains virtuels. Les deux humains virtuels utilisent le même modèle de corps de sorte qu'ils soient équivalents dans tous leurs aspects, à l'exception de leurs visages.



Fig. 3. Les deux humains virtuels intégrés dans MARC qui ont été utilisés pour simuler l'attention conjointe. Ils ont été conçus en se basant sur les photos de personnes réelles.

Le système présenté dans cet article utilise la capacité des personnages de MARC à suivre une cible du regard [51]. Cependant les humains virtuels sont affichés dans leur environnement

3D, alors que le point de regard du participant est détecté dans le plan à deux dimensions de l'écran physique. Simuler l'attention conjointe est donc envisageable en utilisant plusieurs paradigmes d'approximations de l'espace à deux dimensions dans le monde en trois dimensions.

3.3 Simulation de l'attention conjointe

La simulation de l'attention conjointe peut être réalisée par différentes approches. Dans notre système, la cible de l'attention est contrôlée par l'utilisateur. Il est donc important que (1) la direction du regard de l'utilisateur soit correctement détecté, et que (2) l'utilisateur perçoive que l'agent virtuel suit son regard.

Ce dernier point n'est pas évident. En effet, le regard de l'utilisateur est détecté dans un espace 2D (l'écran), tandis que le regard de l'agent virtuel est exprimé en trois dimensions. Nous devons donc créer une correspondance entre les deux espaces qui génère l'effet perceptif approprié. Pour ce faire, nous pouvons explorer les méthodes suivantes.

3.3.1 Approche 2D

L'approche la plus simple est d'utiliser un espace partagé en deux dimensions entre le monde réel et le monde virtuel. Ceci peut être réalisé par la création d'une zone plane dans le monde virtuel qui représente l'affichage réel (fig. 4).

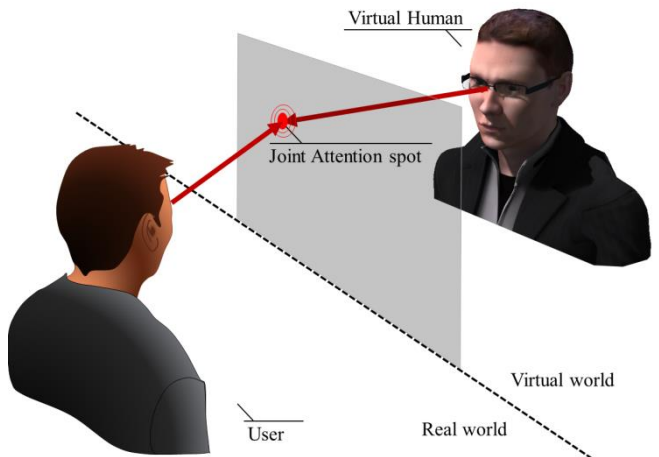


Fig. 4. Illustration d'approche 2D de plan de partagé. Monde virtuel et monde réel sont « alignés » sur un plan 2D qui représente la surface d'affichage de l'écran.

Cette approche est très utile lorsque l'attention conjointe est portée sur des images 2D qui sont affichées sur l'écran de l'utilisateur, sans faire partie de la scène 3D virtuelle. L'espace 2D, partagé entre l'utilisateur et l'agent virtuel peut être considéré comme une « fenêtre augmentée » entre les mondes réels et virtuels, capables d'afficher des informations "entre" les deux mondes, sur la fenêtre elle-même.

3.3.2 Approche 3D continue

Au lieu d'utiliser l'espace partagé 2D, il est aussi possible de considérer que le regard de l'utilisateur est un vecteur dans l'espace virtuel 3D. En traçant un rayon de sélection le long de ce vecteur, nous pouvons récupérer les coordonnées 3D du point exact que l'utilisateur regarde dans le monde virtuel. Il est alors possible de calculer les angles de rotation à appliquer au

regard de personnage virtuel (Fig. 5). Cette approche est pertinente pour être mise en œuvre dans un environnement 3D complet, lorsque l'écran est rempli intégralement avec du contenu 3D. Cependant, en études contrôlées, où l'agent est souvent affiché sur un environnement vide, la direction du regard de l'utilisateur sera susceptible passer sur le fond, où aucune coordonnée 3D ne pourra être calculé.

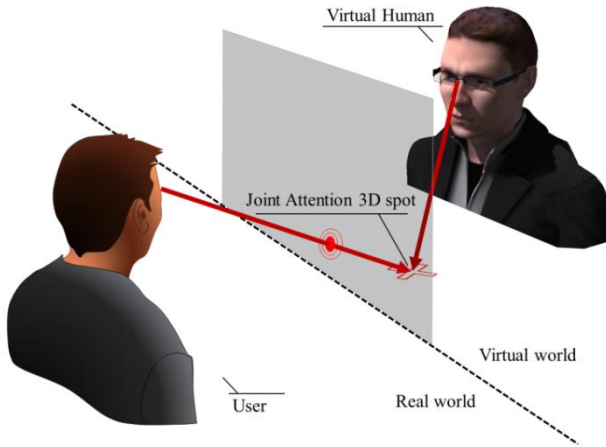


Fig. 5. Illustration de l'approche par coordonnées 3D. Nous utilisons le regard de l'utilisateur pour lancer un rayon dans le monde virtuel et faire en sorte que le personnage virtuel regarde ce point exact.

3.3.3 Les « Virtual Attention Objects »

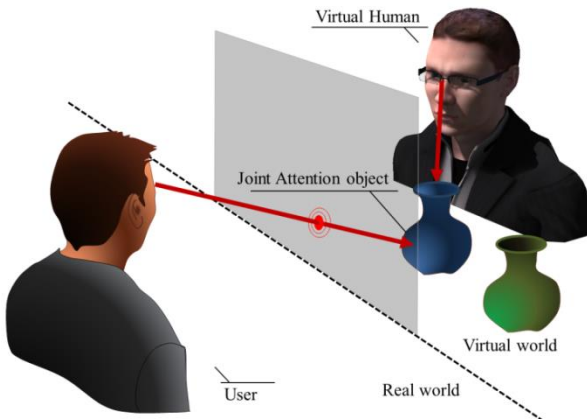


Fig. 6. Illustration de l'approche «Virtual Attention Objects ». Nous utilisons le regard de l'utilisateur de lancer un rayon dans le monde virtuel et détecter l'objet que l'utilisateur regarde. Nous faisons alors en sorte que le personnage virtuel regarde ce même objet (mais pas sur le même point).

Plutôt que de regarder exactement au même endroit, le système perceptif humain est plus susceptible de se concentrer sur les objets pertinents de l'environnement. Sur la base de cette hypothèse, nous pouvons concevoir un système qui utilise les coordonnées 3D issues du lancer de rayons dans l'environnement virtuel afin de détecter quel objet que l'utilisateur regarde. Ensuite, plutôt que de diriger le regard du personnage virtuel exactement au même endroit, nous pouvons le diriger sur le même objet. Cette approche est inspirée des recherches antérieures [34] utilisant les VAO. Cependant, plutôt que d'utiliser le regard de l'agent pour diriger l'attention de l'utilisateur, l'objet

est désigné par le regard de l'utilisateur et l'attention de l'agent virtuel va vers celui-ci (figure 6).

Bien que cette approche puisse sembler être une version dégradée de l'approche 3D pure, elle peut être plus réaliste, imitant le système de perception humaine qui se concentre sur les objets.

3.3.4 Approche polynomiale

Au lieu de raisonner sur une base géométrique, nous pouvons également envisager une approche orientée sur la perception. Nous avons conçu une approche polynomiale dans laquelle le personnage est contrôlé en spécifiant les angles d'Euler pour les rotations du cou et des yeux. Ces angles sont calculés sur la base d'équations polynomiales en utilisant l'emplacement du regard de l'utilisateur sur l'écran comme variable d'entrée. Les polynômes ont été définis à l'aide d'une régression linéaire sur la base d'un ensemble de valeurs d'étalonnage. Ces valeurs d'étalonnage ont été définies pour aligner perceptivement les positions de l'écran et la rotation des yeux et de la tête de l'agent virtuel. Les équations qui en résultent (figure 5) sont donc fortement en fonction de l'emplacement de l'humain virtuel à l'écran. Utilisation de la perception comme une référence pour l'alignement du regard peut vraisemblablement conduire à un sentiment plus fort de l'attention conjointe que les approches purement géométriques.

$$\begin{aligned} \text{Eyes}_{\text{Vertical}} &= 5.72 \cdot 10^{-6} \times \text{Screen}_y^2 + 2.44 \cdot 10^{-2} \times \text{Screen}_y + 0.9 \\ \text{Eyes}_{\text{Horizontal}} &= -2.38 \cdot 10^{-5} \times \text{Screen}_x^2 + 6.01 \cdot 10^{-2} \times \text{Screen}_y - 18.0 \\ \text{Neck}_{\text{Vertical}} &= -4.85 \cdot 10^{-10} \times \text{Screen}_y^3 \\ &\quad + 1.32 \cdot 10^{-6} \times \text{Screen}_y^2 + 1.23 \cdot 10^{-3} \times \text{Screen}_y \\ &\quad + 0.2 + 7.81 \cdot 10^{-5} \times \text{Screen}_x \times \\ &\quad \quad \quad (-2.29 \cdot 10^{-6} \times \text{Screen}_y^2 + 2.15 \cdot 10^{-3} \times \text{Screen}_y + 0.5) \\ \text{Neck}_{\text{Horizontal}} &= 8.93 \cdot 10^{-7} \times \text{Screen}_x^2 - 2.12 \cdot 10^{-3} \times \text{Screen}_y + 0.499 \end{aligned}$$

Fig. 7. Formule obtenue à partir des régressions linéaires selon les données perceptives obtenues lors de l'alignement de la direction du regard du personnage avec des points prédéfinis sur une grille de calibrage. Les variables d'entrée représentent les positions du regard sur l'écran. Les variables de sortie représentent des angles de rotation du cou et les yeux.

3.3.5 Comparaison et possibilités

La pertinence de ces différentes approches dépend de la finalité de l'application. L'approche 2D et l'approche VAO sont bien adaptés à l'attention conjointe sur les objets non-3D affichés à l'écran (comme dans [34]). Cependant, ils introduisent des limites. L'approche 2D ne permet pas de différenciation de la profondeur du point d'attention, alors que l'objet VAO limite l'attention partagée à une liste finie d'objet prédéfini. Toutes les autres approches sont continues dans l'espace et ne nécessitent aucune sorte de connaissance spécifique des objets visibles.

L'approche 3D est l'approche la plus riche puisque le personnage virtuel peut regarder exactement au même point que l'utilisateur. Pourtant, il introduit la nécessité de bien distinguer les points de fixation des saccades en temps réel. En effet, alors

que le regard de l'utilisateur est le passage entre deux points de fixation, il est susceptible de se déplacer sur des éléments de fond, éventuellement derrière le personnage. Ces mouvements doivent être ignorés, sinon le personnage peut essayer de regarder en arrière sur un temps très court, générant des mouvements ambigus et erratique qui peuvent briser l'illusion de l'attention conjointe. Il existe différentes façons de définir fixations du regard et saccades, selon le type d'oculomètre utilisé. Certains algorithmes sont plus adaptés pour la détection en temps réel que d'autres. Par exemple, les fixations du regard sont généralement caractérisés par des seuils de positions du regard en restant dans un rayon de 1° pendant au moins 100 ms [49]. Un tel algorithme simple est facile à mettre en œuvre en temps réel, mais implique une latence supplémentaire d'au moins 100 ms.

L'évaluation suivante a été réalisée en utilisant l'approche polynomiale. Cette évaluation se concentre sur la perception de l'attention conjointe. Utiliser une approche basée sur le perceptif semble donc pertinent. Toutefois, des évaluations similaires seront menées avec d'autres approches pour comparaison.

4. EVALUATION

Nous avons poursuivi la même ligne de pensée que dans les deux expériences décrites ci-dessus qui ont utilisé le paradigme de la fenêtre de visualisation. Notre objectif est maintenant d'examiner l'émergence de la conscience de son regard lorsque l'utilisateur est confronté à un humain virtuel suivant son regard. Cette évaluation a pour but de vérifier si les utilisateurs se rendent compte qu'ils contrôlent la direction du regard de l'humain virtuel. En effet, les humains virtuels qui affichent des comportements stéréotypés d'attention conjointe peuvent sans doute attirer l'attention de l'utilisateur. Si c'est le cas, alors l'interaction avec l'humain virtuel pourrait être considérée comme moins réaliste puisque l'utilisateur aurait tendance à contrôler son regard délibérément. Dans le cas opposé, l'interaction pourrait être considérée plus naturelle.

Le dispositif expérimental compare deux conditions: un humain virtuel qui suit le regard du participant et un humain virtuel dont le regard est autonome (condition contrôle). Les deux humains virtuels sont affichés côte à côte (Fig. 3). Trois éléments 2D ont été placés en face d'eux sur l'écran sur une surface intermédiaire entre eux et le participant. On a demandé aux participants de dire quel est leur objet préféré parmi les trois, et de deviner quel objet chacun des deux humains virtuels a préféré. Ces instructions étaient destinées à diriger l'attention des participants sur la direction du regard de chaque humain virtuel. Les deux humains virtuels ne produisaient pas de mouvements des yeux et du cou autre que pour diriger leur regard. Le comportement du regard autonome de la condition contrôle a été généré en enregistrant le regard d'une personne réelle avant l'expérience. Les participants ont passé quatre blocs de douze essais chacun. Chaque essai a duré 10 secondes. Le premier bloc a servi de test où les participants devaient s'entraîner à la tâche. Dans ce bloc, les deux humains virtuels affichaient des comportements regard autonomes. Le système de regard contingent a été activé dans les deuxième et troisième blocs. Dans le deuxième bloc, l'un des humains virtuels suivait le regard du participant, tandis que le regard de l'autre humain virtuel restait autonome. Dans le troisième bloc, les comportements des deux humains virtuels ont été échangés: L'humain virtuel qui avait suivi le regard du participant dans le bloc 2

adoptait un comportement de regard autonome dans le bloc 3, et de l'humain virtuel qui s'était comporté de manière autonome dans le bloc 2 suivait le regard de l'utilisateur dans le bloc 3. Les rôles des humains virtuels dans le bloc 2 et 3 ont été contrebalancés entre les participants afin de neutraliser tout effet dû à des différences dans les apparences physiques des humains virtuels. Le bloc 4 était un point de référence final similaire au bloc 1, où les comportements du regard des deux humains virtuels étaient autonomes. Ce bloc a été conçu pour vérifier si les participants remarquaient ou non un changement lorsque les comportements d'attention conjointe des humains virtuels étaient supprimées. A la fin de l'expérience, les participants ont été invités à décrire comment ils pensaient que les regards des humains virtuels avaient été contrôlés. Nous leur avons ensuite expliqué le vrai fonctionnement du regard des humains virtuels et leur avons demandé s'ils s'en étaient rendus compte.

Le recrutement des participants à l'expérience est toujours en cours. Nous présentons ici les résultats préliminaires obtenus avec 10 participants. Ils étaient exempts de tout symptôme psychiatrique ou neurologique connus. Aucun d'entre eux n'a perçu que les humains virtuels suivaient leur regard, même s'ils étaient parfaitement conscients que leur regard était suivi par un oculomètre durant toute l'expérience. Des données qualitatives ont été recueillies après chaque essai en demandant aux participants d'expliquer les suppositions qu'ils avaient fait sur les éléments préférés par les deux humains virtuels. Leurs réponses ont montré que les participants se fiaient au regard des humains virtuels: Les participants ont mentionné l'"orientation de la tête", ils ont utilisé des expressions telles que «les avatars ont regardé l'objet du milieu» ou «ils ont scruté les 3 articles ». Les participants avaient tendance à attribuer des états mentaux des humains virtuels en fonction de leur comportement d'observation. Par exemple, un participant a répondu que «les avatars semblaient trouver l'inspiration en regardant vers le haut". Un autre participant a interprété les mouvements oscillatoires de la tête de l'humain virtuel comme un signe de tête signalant que l'humain virtuel avait fait son choix. Deux participants ont effectivement affirmé qu'ils avaient vu les humains virtuels sourire en hochant la tête. Si le hochement de tête peut être dû aux oscillations du regard de l'utilisateur, aucune expression faciale n'a été générée durant ce test.

Fait intéressant, les participants ont déclaré que les deux humains virtuels s'influençaient mutuellement. Par exemple, un participant a expliqué que «l'avatar à droite a choisi le chapeau parce que l'avatar de gauche l'a regardé plus longuement". Un autre participant a déclaré que les humains virtuels s'échangeaient de signaux de la tête. Un troisième participant a indiqué que l'un des humains virtuels suivait le regard de l'autre humain virtuel, alors qu'il suivait en fait, à ce moment-là, le regard de l'utilisateur. Nous présumons que le participant suivait lui-même le regard de l'autre humain virtuel, ce qui forçait le premier humain virtuel à suivre le regard du second. Étonnamment, le participant a donc remarqué le comportement de suivi du regard de l'humain virtuel qu'il contrôlait, mais n'a pas remarqué la similitude avec son propre comportement visuel. Un quatrième participant a mentionné dès le deuxième essai du bloc 2 (première occurrence du comportement de suivi) que l'un des humains virtuels prêtait beaucoup d'attention à l'autre humain virtuel. Celui-ci était effectivement contrôlé par le regard du participant. Après le premier essai du bloc 3, le même parti-

cipant a souligné que le rôle des humains virtuels avait été «renversé» et, en effet, le fonctionnement de leur regard venait d'être échangé. Puis, sur le deuxième essai du bloc 4, lorsque les humains virtuels étaient revenus à un comportement autonome, ce participant a noté que les deux avatars étaient désormais «plutôt calme», et que cela contrastait avec leurs précédents «comportements hystériques» d'«alternance entre surveillance de l'autre avatar et de choix entre les objets». Ce participant semble avoir été très sensible à l'évolution des comportements visuels des humains virtuels. Il a rapidement perçu les contingences de regard, mais n'a pas réussi à percevoir que son propre regard en était la source.

5. DISCUSSION

La simulation de l'attention conjointe offre de nouvelles directions de recherche en interaction homme-machine, et qui sont également pertinentes pour la psychologie, avec des applications potentiellement utiles en psychopathologie. Les progrès récents de la technologie eye tracking permettent désormais un suivi rapide et précis du regard de l'utilisateur, sans nécessiter de dispositifs invasifs et contraignants. Associé à la réalité virtuelle, cette technologie ouvre de nouvelles possibilités pour étudier le comportement humain. Nous avons créé une plateforme pour l'étude de l'attention conjointe qui fournit validité écologique et faisabilité expérimentale. Un expérimentateur humain ne peut pas réussir à suivre le regard d'un participant avec la même précision et la même régularité que l'humain virtuel dans notre plate-forme. En outre, notre plate-forme affiche des personnages virtuels réalistes qui ressemblent à des individus de la vie quotidienne. Ces personnages ont en effet été conçus à partir de photographies de personnes réelles et ils sont capables de produire des mouvements de tête et de regard similaires à ceux de vrais êtres humains.

Pour évaluer l'attention conjointe avec des personnages virtuels, nous avons mis en place une tâche qui exige la sélection d'objets préférés. Des tâches similaires ont déjà été employées pour évoquer le désir mimétique [52], [53]. Le désir mimétique est un effet de préférence sur l'objet qui est étroitement liée à l'attention conjointe: les gens ont tendance à évaluer plus positivement les éléments qui sont regardés par d'autres. Dans une expérience classique du désir mimétique réalisée par Bayliss et al. [52], les participants ont été exposés à un visage qui soit regardait un objet, soit regardait dans la direction opposée. Les objets qui étaient systématiquement évités par le visage différaient de ceux qui étaient systématiquement regardés uniquement par leur couleur. Les participants ont ensuite été invités à attribuer des notes à chaque objet. Les résultats ont montré qu'ils ont évalué plus positivement les objets qui avaient été regardés par le visage. Le désir mimétique souligne l'importance de l'attention conjointe dans la façon dont l'attention des autres influence nos choix personnels. Que les effets du désir mimétique soit implicite ou explicite reste sujet à débat. En d'autres termes, la tendance à évaluer favorablement un objet regardé par d'autres pourrait dépendre de notre capacité à prendre conscience de l'association objet-regard [53]. Dans les tests d'évaluation de notre plate-forme, les participants semblaient porter une attention explicite à l'orientation du regard des humains virtuels pour identifier leurs objets préférés. Les observations qualitatives que nous avons recueillies indiquent que les participants ont porté des jugements conscients basés sur les asso-

ciations objet-regard.

Les résultats les plus intrigants des tests d'évaluation présentés ici est que les participants n'ont pas compris que leur regard influençait les humains virtuels. La conscience de soi n'est pas apparue en dépit de la haute-fidélité du système de suivi du regard. Ce manque de prise de conscience a été précédemment décrit dans des expériences sur l'«effet caméléon» [54]. Ce terme désigne les phénomènes liés à l'imitation: Des études ont montré que (1) on a tendance à imiter spontanément les attitudes de la personne avec qui nous interagissons; (2) nous évaluons plus favorablement une personne qui imite notre comportement. Un être humain virtuel suivant le regard de l'utilisateur peut en fait être considéré comme imitant l'utilisateur. Notre plate-forme est donc aussi pertinente pour l'étude du mimétisme du regard, ce qui représente une autre façon d'appréhender le suivi du regard. Une hypothèse plausible est que le suivi du regard par les humains virtuels induit un effet de mimétisme et ils peuvent donc être évalués plus positivement par les utilisateurs.

Enfin, les résultats présentés ici fournissent des informations précieuses concernant les interfaces informatiques utilisant le suivi du regard. Les deux premières expériences décrites dans cet article montrent que les utilisateurs ne sont pas nécessairement spontanément conscients des actions qu'ils causent par leurs mouvements oculaires sur une interface graphique, même lorsque cet effet est clairement visible. Ce fut le cas lorsque les utilisateurs ont effectué des tâches sociales et non-sociales. L'évaluation préliminaire de notre plate-forme donne à penser qu'il est encore plus difficile pour l'utilisateur de prendre conscience de l'influence de son regard dans le contexte de l'attention conjointe. Sur la base de ces observations, on ne peut pré-supposer que les utilisateurs d'interfaces graphiques asservies au regard comprennent spontanément le fonctionnement de l'interface, puisqu'il se pourrait qu'ils ne se rendent même pas compte qu'ils contrôlent l'interface. Cela signifie que ces types d'interfaces pourraient être beaucoup moins intuitifs que les autres formes d'interactions homme-machine. Les conséquences sont notables pour l'industrie informatique compte tenu de la tendance actuelle dans le développement de fonctionnalités eye-tracking (par exemple dans les appareils mobiles tel que le Samsung Galaxy S4). Les considérations éthiques doivent aussi être soulignées. Par exemple, si le regard est utilisé pour contrôler un navigateur Internet, l'utilisateur pourrait accéder à du contenu non autorisé en raison d'un manque de contrôle volontaire de ses mouvements oculaires.

6. PERSPECTIVES

6.1 Interaction avec des humains virtuels

Selon Richardson et al. [55], la coordination est un élément essentiel pour construire des interactions conversationnelles qui se déroulent facilement et de manière intelligible. Ces travaux montrent que le regard mutuel et l'attention conjointe portent un rôle important dans cette coordination. Les auteurs se réfèrent au couplage des mouvements oculaires pour désigner la tendance d'une deux individus à regarder le même objet lors d'une conversation.

L'effet de couplage se fait à plusieurs niveaux et plusieurs modalités [56]. Krueger et Michael [57] ont identifié des mécanismes de couplage sur des gestes et de la cognition sociale.

Selon eux, le couplage est un phénomène neuronal qui facilite l'interaction sociale et la compréhension mutuelle. Permettre un couplage multi-niveaux entre un utilisateur et un humain virtuel pourrait ainsi contribuer à une interaction plus fluide et à l'augmentation de l'empathie envers l'humain virtuel. Jusqu'à présent, un petit nombre d'études ont été menées pour explorer le couplage avec des personnages virtuels selon plusieurs modalités : les gestes, la prosodie de la parole, et la conversation en face-à-face [58].

Kopp [58] définit les agents conversationnels « résonants » comme des agents virtuels capables de partager et de coordonner leur comportement avec l'utilisateur. Il soutient que les agents résonants peuvent conduire à une meilleure interaction et de meilleures relations sociales entre l'agent et l'utilisateur. Cependant, son travail se concentre principalement sur les *backchannels* et les comportements gestuels.

Notre système nous permet de simuler l'attention conjointe et l'effet de couplage sur le comportement du regard durant les interactions multimodales et des conversations avec des personnages virtuels. Nous croyons qu'un tel comportement peut contribuer à améliorer le niveau de résonance de personnages virtuels, améliorant ainsi l'interactivité, l'intercompréhension et l'acceptabilité.

6.2 Affective Computing

L'attention conjointe est un mécanisme perceptif qui déclenche la compréhension mutuelle entre deux personnes (ou plus) en interaction et contribue à communiquer et échanger sur le contexte partagé. Il influe également l'évaluation des objets dans l'environnement. Le *Social Appraisal* a été définie par Manstead et Fischer [59] comme l'impact du contexte social, et plus particulièrement les évaluations des autres, sur son évaluation d'un événement ou d'un objet. Dans ce contexte, l'attention conjointe apparaît comme un mécanisme clé. Pour évaluer la compréhension des autres, il faut comprendre ce sur quoi ils portent leur attention. L'utilisateur doit être capable de percevoir ce que l'agent se concentre sur les mêmes événements ou des objets pour que l'évaluation sociale puisse émerger. La simulation attention conjointe dans un agent virtuel est donc l'un des composants nécessaires pour représenter le *social appraisal* induit par l'utilisateur sur un agent virtuel.

Nos travaux antérieurs ont exploré un modèle informatique du *social appraisal* [50, 62]. Ce modèle a été testé en utilisant deux humains virtuels. En combinant la simulation de l'attention conjointe et la reconnaissance des expressions faciales, nous visons à créer un système capable d'effectuer du *social appraisal* entre un utilisateur et un agent virtuel. Nous croyons que le système présenté dans ce document peut contribuer à une meilleure perception et compréhension des réactions sociales. Nous allons mener des expériences avec et sans simulation de l'attention conjointe pour explorer cette hypothèse.

Ce genre de comportement social, avec empathie et d'autres concepts, est crucial pour développer une relation sociale dynamique fiable entre un utilisateur et un humain virtuel / compagnon. Nous croyons que la conception d'un humain virtuel ne porte pas seulement sur les rôles d'assistants personnels purement fonctionnels, mais de compagnons de la vie quotidienne. Ces compagnons doivent donc être dotés de compétences sociales avancées et des moyens de les exprimer. Le regard et l'attention conjointe sont des signaux sociaux très

important et contribuent à la construction de la communication sociale. Comme expliqué précédemment, le regard de l'autre personne agit comme un puissant signal pour son attention. En outre, la pertinence des signaux induits par le regard d'autrui a été montrée comme capable d'influencer la façon dont autrui est jugée en termes de personnalité (mensongère contre coopérative) [60]. L'attention conjointe est donc, à notre avis, une clé pour construire de meilleurs humains virtuels qui permettront une interaction de haute qualité et à long terme.

6.3 Recherches Cliniques

Les projets présentés dans ce document ont été élaborés dans le cadre d'un programme de recherche sur l'autisme. Le système de fenêtre de visualisation contrôlée par le regard que nous avons développé est destiné à l'entraînement du contrôle du regard au cours d'interactions sociales. Nous sommes actuellement en cours de conception d'une procédure de traitement fondée sur ce système. Les technologies informatiques sont de plus en plus utilisées pour l'entraînement des personnes avec autisme. À ce jour, il n'existe pas de solutions pharmacologiques et des alternatives sont nécessaires. Un nombre important de projets impliquant des technologies innovantes ont vu le jour dans la dernière décennie et les études d'évaluation commencent à montrer des résultats positifs [61]. Nous nous attendons à ce que l'entraînement des individus avec autisme à contrôler leur regard en utilisant le paradigme de la fenêtre de visualisation les aide dans le traitement de l'information émotionnelle significative au cours des interactions sociales.

La plate-forme de simulation de l'attention conjointe décrite ici doit être utilisée dans les expériences psychologiques qui cherchent à mieux comprendre les troubles attribués à l'autisme. Comme décrit précédemment, l'attention conjointe est considérée comme un facteur majeur dans le développement des personnes autistes. La plateforme que nous avons conçue permet une analyse précise de leur capacité à initier l'attention conjointe et devrait donc contribuer à affiner les caractéristiques des altérations associées à l'autisme. À long terme, nous pensons qu'elle pourrait servir à des évaluations cliniques nécessaires pour définir le traitement le plus adéquate. De plus, compte tenu de l'appréciation positive associée à l'effet caméléon, l'imitation du regard produite avec un humain virtuel dans notre plateforme devrait donner lieu à des interactions pro-sociales qui seraient bénéfiques pour les personnes qui éprouvent des difficultés émotionnelles ou sociales. Des extensions à d'autres troubles psychologiques que l'autisme seront à explorer.

REMERCIEMENTS

Les auteurs tiennent à remercier Rachel Dupuis et Aliye Karasu pour leur assistance expérimentale. Nous sommes très reconnaissants à Jacqueline Nadel pour ses conseils sur la conception des protocoles expérimentaux. Ce travail a été soutenu par des subventions de la Fondation Orange (projet n° 71/2012, coordinateur: O. Grynszpan), La Fondation de France et la Fondation Adrienne et Pierre Sommer (projet n° 2007 005874, coordinateur: O. Grynszpan) et L'Agence Nationale de Recherche (ANR projet n°12 SAMA 011 01 coordinateur: P. Fossati).

REFERENCES

- [1] S. Baron-Cohen, S. Wheelwright, and T. Jolliffe, "Is there a 'language of the eyes'? Evidence from normal adults, and adults with autism or Asperger syndrome," *Vis. Cogn.*, vol. 4, no. 3, pp. 311–331, 1997.
- [2] A. Jaimes and N. Sebe, "Multimodal Human Computer Interaction: A Survey," in *Computer Vision in Human-Computer Interaction*, N. Sebe, M. Lew, and T. S. Huang, Eds. Springer Berlin Heidelberg, 2005, pp. 1–15.
- [3] G. Bailly, S. Raidt, and F. Elisei, "Gaze, conversational agents and face-to-face communication," *Speech Commun.*, vol. 52, no. 6, pp. 598–612, Jun. 2010.
- [4] H. Wang, M. Chignell, and M. Ishizuka, "Empathic tutoring software agents using real-time eye tracking," in *Proceedings of the 2006 symposium on Eye tracking research & applications*, 2006, pp. 73–78.
- [5] N. J. Emery, "The eyes have it: the neuroethology, function and evolution of social gaze," *Neurosci. Biobehav. Rev.*, vol. 24, no. 6, pp. 581–604, Aug. 2000.
- [6] H. Kobayashi and S. Kohshima, "Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye," *J. Hum. Evol.*, vol. 40, no. 5, pp. 419–435, May 2001.
- [7] M. M. Haith, T. Bergman, and M. J. Moore, "Eye contact and face scanning in early infancy," *Science*, vol. 198, no. 4319, pp. 853–855, Nov. 1977.
- [8] C. N. Macrae, B. M. Hood, A. B. Milne, A. C. Rowe, and M. F. Mason, "Are You Looking at Me? Eye Gaze and Person Perception," *Psychol. Sci.*, vol. 13, no. 5, pp. 460–464, 2002.
- [9] B. M. Hood, C. N. Macrae, V. Cole-Davies, and M. Dias, "Eye remember you: The effects of gaze direction on face recognition in children and adults," *Dev. Sci.*, vol. 6, no. 1, pp. 67–71, 2003.
- [10] M. von Grünau and C. Anston, "The detection of gaze direction: a stare-in-the-crowd effect," *Perception*, vol. 24, no. 11, pp. 1297–1313, 1995.
- [11] H. Doi, K. Ueda, and K. Shinohara, "Neural correlates of the stare-in-the-crowd effect," *Neuropsychologia*, vol. 47, no. 4, pp. 1053–1060, Mar. 2009.
- [12] A. Senju, Y. Kikuchi, T. Hasegawa, Y. Tojo, and H. Osanai, "Is anyone looking at me? Direct gaze detection in children with and without autism," *Brain Cogn.*, vol. 67, no. 2, pp. 127–139, Jul. 2008.
- [13] A. Senju and M. H. Johnson, "The eye contact effect: mechanisms and development," *Trends Cogn. Sci.*, vol. 13, no. 3, pp. 127–134, Mar. 2009.
- [14] T. Charman, "Why is joint attention a pivotal skill in autism?," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 358, no. 1430, pp. 315–324, Feb. 2003.
- [15] P. Mundy, "Attention, Joint Attention, and Social Cognition," *Curr. Dir. Psychol. Sci. Wiley-Blackwell*, vol. 16, no. 5, pp. 269–274, Oct. 2007.
- [16] M. I. Posner, "Orienting of attention," *Q. J. Exp. Psychol.*, vol. 32, no. 1, pp. 3–25, Feb. 1980.
- [17] J. Driver, "Gaze Perception Triggers Reflexive Visuospatial Orienting," *Vis. Cogn.*, vol. 6, no. 5, pp. 509–540, Oct. 1999.
- [18] A. Frischen, A. P. Bayliss, and S. P. Tipper, "Gaze cueing of attention: visual attention, social cognition, and individual differences," *Psychol. Bull.*, vol. 133, no. 4, pp. 694–724, Jul. 2007.
- [19] S. Baron-Cohen, R. Campbell, A. Karmiloff-Smith, J. Grant, and J. Walker, "Are children with autism blind to the mentalistic significance of the eyes?," *Br. J. Dev. Psychol.*, vol. 13, no. 4, pp. 379–398, 1995.
- [20] A. Senju and M. H. Johnson, "Atypical eye contact in autism: models, mechanisms and development," *Neurosci. Biobehav. Rev.*, vol. 33, no. 8, pp. 1204–1214, Sep. 2009.
- [21] B. Corden, R. Chilvers, and D. Skuse, "Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome," *Neuropsychologia*, vol. 46, no. 1, pp. 137–147, Jan. 2008.
- [22] K. M. Dalton, B. M. Nacewicz, T. Johnstone, H. S. Schaefer, M. A. Gernsbacher, H. H. Goldsmith, A. L. Alexander, and R. J. Davidson, "Gaze fixation and the neural circuitry of face processing in autism," *Nat. Neurosci.*, vol. 8, no. 4, pp. 519–526, Apr. 2005.
- [23] A. Klin, W. Jones, R. Schultz, and F. Volkmar, "The enactive mind, or from actions to cognition: lessons from autism," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 358, no. 1430, pp. 345–360, Feb. 2003.
- [24] W. Jones, K. Carr, and A. Klin, "Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder," *Arch. Gen. Psychiatry*, vol. 65, no. 8, pp. 946–954, Aug. 2008.
- [25] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Arch. Gen. Psychiatry*, vol. 59, no. 9, pp. 809–816, Sep. 2002.
- [26] J. L. Adrien, P. Lenoir, J. Martineau, A. Perrot, L. Hameury, C. Larmande, and D. Sauvage, "Blind ratings of early symptoms of autism based upon family home movies," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 32, no. 3, pp. 617–626, May 1993.
- [27] A. Senju, Y. Tojo, H. Dairoku, and T. Hasegawa, "Reflexive orienting in response to eye gaze and an arrow in children with and without autism," *J. Child Psychol. Psychiatry*, vol. 45, no. 3, pp. 445–458, Mar. 2004.
- [28] T. Falck-Ytter, C. von Hofsten, C. Gillberg, and E. Fernell, "Visualization and Analysis of Eye Movement Data from Children with Typical and Atypical Development," *J. Autism Dev. Disord.*, Feb. 2013.
- [29] A. Rombough and G. Iarocci, "Orienting in Response to Gaze and the Social Use of Gaze among Children with Autism Spectrum Disorder," *J. Autism Dev. Disord.*, vol. 43, no. 7, pp. 1584–1596, Jul. 2013.
- [30] M. R. Swanson, G. C. Serlin, and M. Siller, "Broad autism phenotype in typically developing children predicts performance on an eye-tracking measure of joint attention," *J. Autism Dev. Disord.*, vol. 43, no. 3, pp. 707–718, Mar. 2013.
- [31] T. Dratsch, C. Schwartz, K. Yanev, L. Schilbach, K. Voegele, and G. Bente, "Getting a grip on social gaze: control

- over others' gaze helps gaze detection in high-functioning autism," *J. Autism Dev. Disord.*, vol. 43, no. 2, pp. 286–300, Feb. 2013.
- [32] C. Peters and C. O'Sullivan, "Attention-driven eye gaze and blinking for virtual humans," in *ACM SIGGRAPH 2003 Sketches & Applications*, 2003, pp. 1–1.
- [33] S. Martinez, R. J. Sloan, A. Szymkowiak, and K. Scott-Brown, "Using virtual agents to cue observer attention," in *CONTENT 2010, The Second International Conference on Creative Content Technologies*, 2010, pp. 7–12.
- [34] C. Peters, S. Asteriadis, K. Karpouzis, and E. de Sevin, "Towards a real-time gaze-based shared attention for a virtual agent," in *International Conference on Multimodal Interfaces*, 2008.
- [35] M. Wilms, L. Schilbach, U. Pfeiffer, G. Bente, G. R. Fink, and K. Vogeley, "It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience," *Soc. Cogn. Affect. Neurosci.*, vol. 5, no. 1, p. 98, 2010.
- [36] L. Schilbach, M. Wilms, S. B. Eickhoff, S. Romanzetti, R. Tepest, G. Bente, N. J. Shah, G. R. Fink, and K. Vogeley, "Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry," *J. Cogn. Neurosci.*, vol. 22, no. 12, pp. 2702–2715, Dec. 2010.
- [37] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends Cogn. Sci.*, vol. 7, no. 11, pp. 498–504, Nov. 2003.
- [38] M. S. Castelano, M. L. Mack, and J. M. Henderson, "Viewing task influences eye movement control during active scene perception," *J. Vis.*, vol. 9, no. 3, 2009.
- [39] T. S. Horowitz, J. M. Wolfe, G. A. Alvarez, M. A. Cohen, and Y. I. Kuzmova, "The speed of free will," *Q. J. Exp. Psychol.* 2006, vol. 62, no. 11, pp. 2262–2288, Nov. 2009.
- [40] S. B. Hutton, "Cognitive control of saccadic eye movements," *Brain Cogn.*, vol. 68, no. 3, pp. 327–340, 2008.
- [41] O. Grynspan, J. Simonin, J.-C. Martin, and J. Nadel, "Investigating social gaze as an action-perception online performance," *Front. Hum. Neurosci.*, vol. 6, p. 94, 2012.
- [42] V. Groom, C. Nass, T. Chen, A. Nielsen, J. K. Scarborough, and E. Robles, "Evaluating the effects of behavioral realism in embodied agents," *Int. J. Hum.-Comput. Stud.*, vol. 67, no. 10, pp. 842–849, 2009.
- [43] G. W. McConkie and K. Rayner, "The span of the effective stimulus during a fixation in reading," *Percept. Psychophys.*, vol. 17, pp. 578–586, Nov. 1975.
- [44] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial expressions*. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1975.
- [45] O. Grynspan, J. Nadel, J. C. Martin, J. Simonin, P. Bailleul, Y. Wang, D. Gepner, F. Le Barillier, and Constant, Jacques, "Self-monitoring of gaze in high functioning autism," *J. Autism Dev. Disord.*, in press.
- [46] F. Happé and U. Frith, "The neuropsychology of autism," *Brain J. Neurol.*, vol. 119 (Pt 4), pp. 1377–1400, Aug. 1996.
- [47] O. Grynspan, J. Nadel, Constant, Jacques, F. Le Barillier, N. Carbonell, J. Simonin, J. C. Martin, and M. Courgeon, "A New Virtual Environment Paradigm for High-Functioning Autism Intended to Help Attentional Disengagement in a Social Context," *J. Phys. Ther. Educ.*, vol. 25, no. 1, pp. 42–47, 2011.
- [48] B. Noris, J. Nadel, M. Barker, N. Hadjikhani, and A. Billard, "Investigating Gaze of Children with ASD in Naturalistic Settings," *Plos One*, vol. 7, no. 9, p. e44144, Sep. 2012.
- [49] B. R. Manor and E. Gordon, "Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks," *J. Neurosci. Methods*, vol. 128, no. 1–2, pp. 85–93, Sep. 2003.
- [50] M. Courgeon and C. Clavel, "MARC: a framework that features emotion models for facial animation during human-computer interaction," *J. Multimodal User Interfaces*, pp. 1–9.
- [51] N. Tan, G. Pruvost, M. Courgeon, C. Clavel, Y. Bellik, and J.-C. Martin, "A location-aware virtual character in a smart room: effects on performance, presence and adaptivity," in *Proceedings of the 16th international conference on Intelligent user interfaces*, 2011, pp. 399–402.
- [52] A. P. Bayliss, M. A. Paul, P. R. Cannon, and S. P. Tipper, "Gaze cuing and affective judgments of objects: I like what you look at," *Psychon. Bull. Rev.*, vol. 13, no. 6, pp. 1061–1066, Dec. 2006.
- [53] C. Bry, E. Treinen, O. Corneille, and V. Yzerbyt, "Eye'm lovin' it! The role of gazing awareness in mimetic desires," *J. Exp. Soc. Psychol.*, vol. 47, no. 5, pp. 987–993, Sep. 2011.
- [54] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception-behavior link and social interaction," *J. Pers. Soc. Psychol.*, vol. 76, no. 6, pp. 893–910, Jun. 1999.
- [55] D. C. Richardson, R. Dale, and N. Z. Kirkham, "The art of conversation is coordination: common ground and the coupling of eye movements during dialogue," *Psychol. Sci.*, vol. 18, no. 5, pp. 407–413, May 2007.
- [56] M. M. Louwerse, R. Dale, E. G. Bard, and P. Jeuniaux, "Behavior matching in multimodal communication is synchronized," *Cogn. Sci.*, vol. 36, no. 8, pp. 1404–1426, Dec. 2012.
- [57] J. Krueger and J. Michael, "Gestural coupling and social cognition: Möbius Syndrome as a case study," *Front. Hum. Neurosci.*, vol. 6, p. 81, 2012.
- [58] S. Kopp, "Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors," *Speech Commun.*, vol. 52, no. 6, pp. 587–597, Jun. 2010.
- [59] A. S. Manstead and A. H. Fischer, "Social appraisal: The social world as object of and influence on appraisal processes," 2001.
- [60] A. P. Bayliss and S. P. Tipper, "Predictive gaze cues and personality judgments: Should eye trust you?," *Psychol. Sci.*, vol. 17, no. 6, pp. 514–520, Jun. 2006.
- [61] O. Grynspan, P. Weiss, F. Perez-Diaz, and E. Gal, "Innovative technology based interventions for Autism Spectrum Disorders: A meta-analysis," *Autism Int. J. Res. Pr.*, in press.
- [62] Mumenthaler, C., & Sander, D. (2012). Social Appraisal Influences Recognition of Emotions. *Journal of Personality and Social Psychology*, 102(6), 1118-1135.