



HAL
open science

Mesure de l'engagement envers un assistant conversationnel en contexte professionnel : mise en place d'un protocole expérimental

Océane Granier, Roxane Bertrand, Kevin Gravouil, Laurent Prevot

► To cite this version:

Océane Granier, Roxane Bertrand, Kevin Gravouil, Laurent Prevot. Mesure de l'engagement envers un assistant conversationnel en contexte professionnel : mise en place d'un protocole expérimental. WACAI 2024 (Workshop Affects, Compagnons Artificiels et Interactions), Jun 2024, Bordeaux, France. hal-04870740

HAL Id: hal-04870740

<https://hal.science/hal-04870740v1>

Submitted on 7 Jan 2025

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

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

Mesure de l'engagement envers un assistant conversationnel en contexte professionnel : mise en place d'un protocole expérimental

Océane Granier^{1,2} Roxane Bertrand¹ Kévin Gravouil² Laurent Prévot³

(1) Aix-Marseille Univ, CNRS, LPL, 5 avenue Pasteur, 13100 Aix-en-Provence, France

(2) Airudit, 2 rue Marc Sangnier, 33130 Bègles, France

(3) Centre d'études français sur la Chine (CEFC), 128 Academia Rd. Sec. 2, Taipei, 11529, Taïwan

RÉSUMÉ

Les agents artificiels peuvent tenir des rôles très divers dans le cadre professionnel mais il faut que ces agents soient adaptés aux contraintes spécifiques du contexte professionnel. Les utilisateurs ne doivent pas être gênés dans leur tâche principale de travail. Un agent désincarné permet d'éviter la surcharge visuelle tandis que l'utilisation de la modalité vocale plutôt que textuelle permet de garder les mains libres. Malgré les avantages clairs de l'utilisation d'agents conversationnels comme la réduction de la charge cognitive liée à la saisie, les utilisateurs sont réticents à l'utilisation de tels outils. Pour pallier ces réticences, nous proposons de rendre l'agent conversationnel plus engageant, c'est-à-dire adapté aux besoins de l'utilisateur et donnant envie d'être utilisé. Cela implique de définir une métrique de l'engagement basée sur les paramètres prosodiques et phonétiques qui ont été peu étudiés jusqu'ici et seulement dans les interactions humain-humain, relativement aux paramètres visuels ou interactionnels. Nous proposons d'étudier l'engagement au travers de paramètres prosodiques et phonétiques aujourd'hui uniquement étudiés dans les interactions entre humains.

CCS CONCEPTS

• Engagement ; • Dialogue Humain-Machine ; • Agent Artificiel ; • Assistant conversationnel ; • Travail ;

KEYWORDS

Engagement, Agent conversationnel, Agent artificiel, Contexte professionnel

1 INTRODUCTION

L'intelligence artificielle est aujourd'hui partout que ce soit dans nos téléphones ou bien sur notre lieu de travail. L'adoption de l'IA par les entreprises est passée de 20 à 50 % depuis 2017 [Lécureuil 2023]. Les agents artificiels peuvent être d'une grande aide dans le milieu professionnel et occuper diverses fonctions comme un rôle d'outil amélioré qui extrait les bonnes données, de recherche d'information dans la documentation ou encore de collègue virtuel qui diminue la solitude et prévient la fatigue [Bittner et al. 2019; Feng et al. 2020; Lester et al. 2004]. Cette opportunité d'accompagnement se heurte cependant à l'acceptabilité de ces outils par les professionnels et notamment d'un manque de confiance envers les entités virtuelles [Cai et al. 2022], ou le sentiment d'étrangeté qu'elles peuvent provoquer [Mori et al. 2012]. Réduire ces réticences requiert d'identifier et comprendre les paramètres du comportement d'un agent conversationnel (AC) qui pourraient avoir

un impact sur le niveau d'acceptabilité des professionnels amenés à utiliser cet agent. Pour évaluer cette acceptabilité, nous faisons l'hypothèse que rendre une machine plus engageante, en tentant de mieux comprendre ce que recouvre cette notion, pourrait favoriser la relation entre utilisateur et machine.

Les assistants conversationnels déployés ont pour rôle d'assister l'opérateur dans sa tâche. L'objectif étant de faciliter des tâches annexes fastidieuses ou prenantes physiquement. Ce dernier va donc utiliser l'assistant en restant centré sur sa mission et par conséquent en étant, la plupart du temps, occupé visuellement par cette mission. Les agents désincarnés présentent en particulier l'intérêt de ne pas exploiter le canal visuel. Cela limite cependant considérablement l'anthropomorphisme de l'AC tel qu'il est généralement étudié (expressions faciales, gestes...) alors que ces paramètres sont tenus pour être des facteurs favorisant l'engagement des utilisateurs dans une interaction avec un agent artificiel [Kim et al. 2018]. Dans ce contexte professionnel, nous allons nous demander quels agents artificiels sont favorisés et comment évaluer l'engagement de leurs utilisateurs.

Notre étude se place dans le cadre d'une thèse CIFRE en partenariat avec l'entreprise Airudit qui propose depuis plus de 5 ans des assistants conversationnels appliqués au contexte professionnel avec lesquels les différents utilisateurs peuvent converser quel que soit leur métier. Ce cadre va nous permettre d'appliquer directement le résultat de cette recherche à un contexte réel.

Notre objectif est de créer un agent conversationnel capable de s'adapter à l'engagement de son utilisateur. La première étape est de définir des métriques de l'engagement compatibles avec le contexte professionnel. L'objectif de la première expérimentation sera d'évaluer les paramètres de l'engagement en fonction de leur compatibilité avec le contexte et de leur significativité.

2 LES AGENTS ARTIFICIELS DANS LE CONTEXTE PROFESSIONNEL

D'après Guirdham and Guirdham [2017], la communication est essentielle en milieu professionnel pour la coordination et l'organisation des activités. Guo and Sanchez [2005] montrent la complexité des interactions professionnelles malgré leur nature *task-oriented* et un *common ground* assez important, c'est-à-dire un contexte et de nombreuses références communes. Il faut d'une part choisir le bon moyen de communication (verbal/non verbal, écrit/parole) et d'autre part tenir compte des barrières individuelles et contextuelles (statut, relations de pouvoir, hiérarchie). Par ailleurs elle est aussi plus complexe du fait des enjeux élevés tant au niveau individuel qu'au niveau de l'entreprise, les motivations conflictuelles,

le besoin de travailler en groupe ainsi que la continuité des interactions. Le contexte professionnel consiste à partager une compréhension des tâches des connaissances techniques entre collègues tout en tenant compte du rôle de chacun dans l'entreprise.

Le rôle des AC en contexte professionnel est multiple. D'une part ils ont une fonction utilitaire et d'autre part ils favorisent les interactions et la communication.

Si la voix peut s'avérer compliquée dans quelques cas (utilisateurs multiples, environnements bruyants et bruités [Cambre et al. 2019]), elle est le plus souvent une plus value incontestable. Dans des tâches prenantes comme le pilotage d'avion, elle évite l'encombrement visuel et permet de gagner en efficacité [Quigley et al. 2004]. La voix améliore l'efficacité des outils et diminue la charge cognitive des utilisateurs [Draper et al. 2003 ; Martín-Barrio et al. 2020]. Elle est également plus intuitive, notamment avec la possibilité de l'utilisation du langage naturel par opposition à l'utilisation de mots clés [Martín-Barrio et al. 2020].

Les études qui s'intéressent aux échanges langagiers restent cependant rares. Ces derniers sont généralement réservés aux agents artificiels incarnés, qu'ils soient physiques (comme les robots) ou virtuels, et qui peuvent donc faire preuve de comportements « humains » (cligner des yeux, hocher la tête). De plus, les relations sociales entre humain et agent artificiel sont généralement étudiées dans des cadres médicaux [Liu and Sundar 2018] ou sociaux [Von der Pütten et al. 2010] tandis que les études qui portent sur le milieu professionnel sont davantage axées sur l'efficacité, la productivité ou la charge cognitive (par exemple le contexte militaire avec l'étude de Quigley et al. [2004]). Dans notre cas l'objectif est de se concentrer sur les apports concrets d'un assistant conversationnel pour l'utilisateur dans le contexte professionnel. Dans cette optique, nous cherchons à améliorer l'acceptabilité de l'assistant via ses capacités communicatives et à l'évaluer grâce à l'engagement en tant que « variable régulant la dynamique de l'échange » [Pellet-Rostaing et al. 2023]. L'engagement témoigne de l'implication des participants dans l'interaction.

Aujourd'hui l'accent est davantage mis sur des agents avec un corps et même un visage artificiel mais on s'aperçoit que la majorité des AC utilisés dans le cadre du travail ne sont pas incarnés. Nous allons donc nous concentrer sur ces derniers.

Un design anthropomorphique est cependant nécessaire pour Araujo [2018] : cela permet plus d'engagement et de confort dans la conversation. Dans le cas d'agent non incarné, cet anthropomorphisme est d'autant plus important qu'aucune donnée visuelle ne peut être utilisée. Une solution est de se concentrer sur le type de conversation qui doit se rapprocher de conversations humaines. L'expression de la vulnérabilité peut également améliorer l'impression d'humanité. Il ne faut cependant pas causer de déception vis-à-vis des capacités des intelligences artificielles qui peuvent être surestimées par les utilisateurs et créer des réticences. D'après Clark et al. [2021], on tombe alors dans la vallée de l'étrange (uncanny valley en anglais). C'est une théorie qui explique pourquoi on a une impression désagréable d'inconfort, de peur ou de répulsion en observant des humanoïdes ayant des apparences ou comportements assez proches des humains. Nous savons que lorsque la voix d'un robot est humaine cette sensation apparaît [Mitchell et al. 2011]. Kusz and Lemoine [2022] ont également montré qu'une voix

très humaine dans un agent conversationnel désincarné est moins agréable qu'une voix clairement artificielle.

Il est important d'étudier les nouvelles dynamiques qui se créent notamment au niveau interactionnel. Dans le contexte professionnel, l'humain est d'abord un utilisateur de l'AC. Celui-ci doit être engageant, c'est-à-dire adapté à ses besoins et donnant envie d'être utilisé. Dans l'idéal, une voix artificielle engageante serait une voix agréable qui se rapproche de la voix humaine et qui s'adapte à l'utilisateur selon son engagement. L'engagement est une notion centrale puisque c'est lui qui témoignera des capacités de l'agent à répondre aux besoins de l'utilisateur.

3 L'ENGAGEMENT

3.1 Définir

Corrigan et al. [2016] définissent l'engagement comme un phénomène complexe qui comprend une part cognitive (attention et concentration) et une part affective (plaisir). Cette définition se détache de la conception assez répandue de l'engagement à travers une interprétation triple : cognitive, émotionnelle et comportementale [Attfield et al. 2011 ; Bouta and Retalis 2013 ; Doherty and Doherty 2018 ; Zyngier 2008]. La conception de Corrigan et al. [2016] est plus en accord avec le contexte de notre étude dans le sens où l'objectif premier de l'interaction est la tâche professionnelle qui demande une attention visuelle et vers laquelle sont dirigés les comportements physiques. Ils ne sont donc pas exploitables.

Au niveau **émotionnel**, si l'engagement est généralement corrélié à des émotions positives [Bakker and Oerlemans 2011 ; Cowie et al. 2001], il n'en fait pas forcément partie. Cowie et al. [2001] le considère comme une tendance à l'action lié à certaines émotions. En psychologie, l'engagement est placé sur une échelle d'état d'éveil ou d'attention par opposition à l'ennui [Churchill 2010]. On retrouve notamment l'opposition entre ennui et engagement dans la difficulté des jeux [Chanel et al. 2008 ; Westgate 2020]. Si le niveau d'un jeu est adéquat, le joueur sera engagé, à l'inverse si le jeu est trop simple il va s'ennuyer. Au niveau **cognitif**, l'engagement serait défini non pas par un niveau d'attention global mais plutôt par le niveau de concentration, d'attention sélective [Corrigan et al. 2016].

Selon la définition de Sidner and Dzikovska [2002], sur laquelle nous allons nous appuyer, l'engagement est un « processus par lequel deux participants (ou plus) établissent, maintiennent et mettent fin à la connexion perçue entre eux pendant une interaction ». Elle présente un processus composé de trois phases : le départ de l'interaction, une période d'attention soutenue et la fin de l'interaction. Dans notre contexte, si commencer une interaction avec l'agent conversationnel peut être un choix, elle se termine obligatoirement avec la réussite (ou l'échec) de la tâche. Il ne peut pas y avoir de désengagement complet avant la fin de celle-ci. Ainsi, l'engagement de l'utilisateur n'est étudiable que pendant l'interaction. C'est cette partie que nous allons étudier à travers cette double conception émotionnelle et cognitive définissant l'engagement comme le degré d'attention soutenue et visée entre deux interactants dans la durée d'une interaction.

3.2 Mesurer

Les types de paramètres mesurés dans les interactions entre humains et humains-machines sont différents. Dans les interactions humain-machine, les paramètres étudiés sont des paramètres principalement comportementaux, interactionnels ou visuels. Bien que moins étudiées, les interactions humain-humain le sont majoritairement sous le prisme de la prosodie et de la phonétique. Les mesures physiologiques sont les seules que l'on retrouve dans les deux types d'interaction.

Comme dit plus haut, les paramètres comportementaux et individuels sont inexploitablement dans notre contexte. Les seuls paramètres applicables au contexte professionnel ayant déjà été étudiés en interaction humain-machine sont les paramètres interactionnels.

Certains paramètres sont étudiables indépendamment de l'agent ou du contexte, ce sont les paramètres liés à la voix de l'utilisateur tels que les paramètres prosodiques et phonétiques. Les paramètres prosodiques étudiés peuvent être assez généraux comme la moyenne et la hauteur maximale de la fréquence fondamentale¹ [Wrede and Shriberg 2003] mais aussi très précis comme la prééminence prosodique et les variations de la fréquence fondamentale [Buxó-Lugo et al. 2018; Gustafson and Neiberg 2010]. Il serait intéressant de voir si ces paramètres étudiés dans les interactions humain-humain s'appliquent également aux interactions humain-machine.

Les paramètres physiologiques ont déjà été étudiés dans les deux types d'interaction. En contexte professionnel, les méthodes d'évaluation nécessitant une installation invasive sont impossibles. C'est le cas de l'électro-encéphalogramme [Arapakis et al. 2017; Szafer and Mutlu 2012] qui est une méthode trop invasive. L'activité électro-dermale [Perugia et al. 2017; Rani and Sarkar 2005] et la fréquence cardiaque [Monkaresi et al. 2016; Rani and Sarkar 2005] sont des méthodes moins invasives grâce au Biopac et donc plus adaptées. Les mesures peuvent être prises sans gêner la tâche principale de l'utilisateur. Cependant, étant donné la complexité de la mise en place de tels dispositifs, ces paramètres pourront être utilisés pour confirmer certaines mesures expérimentales mais ne pourront pas être utilisés par les travailleurs au quotidien.

La distorsion du temps, corrélée à la charge cognitive d'une tâche, permet d'évaluer l'engagement d'une personne dans une tâche [?]. Cette mesure est principalement étudiée dans le cadre de l'engagement utilisateur [Attfield et al. 2011; O'Brien and Toms 2008]. Plus une personne sera engagée dans la tâche, plus le temps sera perçu comme raccourci. À l'inverse, plus une personne va s'ennuyer, plus le temps sera perçu comme rallongé. Nous utiliserons cette mesure afin d'évaluer l'engagement des participants.

4 APPORT EXPÉRIMENTAL

Dans cette optique, nous allons mettre en place une étude qui permettra d'évaluer l'engagement en tenant compte des contraintes liées au contexte. L'objectif de cette étude est de s'inspirer de cas et problématiques professionnelles concrètes afin d'accompagner les utilisateurs dans un plus grand confort de travail. Nous allons nous appuyer sur les outils utilisés par Airudit pour mettre en place une expérimentation intégrée au contexte professionnel.

4.1 Mesures étudiées

Nous allons nous concentrer sur des mesures prosodiques et phonétiques de l'engagement. Pour cela, nous allons d'une part considérer les paramètres de la fréquence fondamentale comme la hauteur moyenne et maximale [Wrede and Shriberg 2003], les écarts de variations et l'intensité [Gustafson and Neiberg 2010; Oertel et al. 2011]. D'autre part nous allons nous intéresser à la convergence phonétique² [Biro et al. 2022; Smith 2000]. Ces mesures n'ont été étudiées pour le moment que dans des interactions humain-humain à notre connaissance. Nous allons également nous appuyer sur des méthodes de mesures physiologiques non invasives (battements du cœur et réponse électrodermale), déjà étudiées dans les deux conditions, ainsi que sur l'étude de la perception du temps par les participants. Ces mesures nous permettront de s'assurer de l'engagement des participants pendant notre étude.

4.2 Contexte professionnel

L'expérimentation ne peut pas se dérouler directement dans une entreprise pour des questions de sécurité et d'accueil du public. Nous allons donc recréer un contexte professionnel à partir de trois éléments. D'une part nous allons utiliser le bras robotique DOFBOT [Yahboom 2024] déjà déployé au sein d'Airudit. D'autre part nous allons lui associer un assistant conversationnel permettant de le commander à distance. Enfin nous allons établir un contexte professionnel crédible autour de ce bras robotique et de ses accessoires en s'inspirant de cas réels. Le bras robotique est composé de plusieurs parties toutes manipulables par des moteurs individuels lui conférant un grand panel d'actions (6 degrés de liberté). Au bout du bras, une pince permet de saisir des objets. Un assistant conversationnel avec lequel on peut interagir en langage naturel permettant une interaction vocale entre le participant et le bras robotique. Nous allons présenter le bras comme une miniaturisation d'un bras utilisé pour déplacer des conteneurs dans une usine de tri de déchets. Un tapis fourni avec le bras représentant des zones de tris et des cubes en bois représentant les conteneurs compléteront le matériel.

L'AC utilisé correspond au type d'agent utilisé en contexte professionnel d'après l'état de l'art. C'est une commande vocale en langage naturel qui permet la diminution de la charge cognitive (plus besoin de saisie textuelle, pas de nécessité de retenir des mots clés). Il est désincarné pour éviter une pollution visuelle de l'environnement de travail.

4.3 Précisions techniques

Nous allons mettre en place un système de *Wizard of Oz* (ou magicien d'Oz) pour simuler une interaction entre le bras robotique et les participants à l'étude (voir Fig.1). L'expérimentateur sera dans un espace différent de l'espace expérimental. Le bras sera piloté à distance grâce à une manette de console de jeu. La production vocale marchera en pressant des boutons correspondant à des phrases pré-enregistrées. Une interface développée pour l'occasion permettra à la fois de manipuler le bras, de choisir les boutons à presser et d'observer l'espace d'expérimentation en temps réel.

1. paramètre acoustique / phonétique qui permet de mesurer la hauteur de la voix

2. Processus par lequel un locuteur adapte sa manière de parler à son interlocuteur

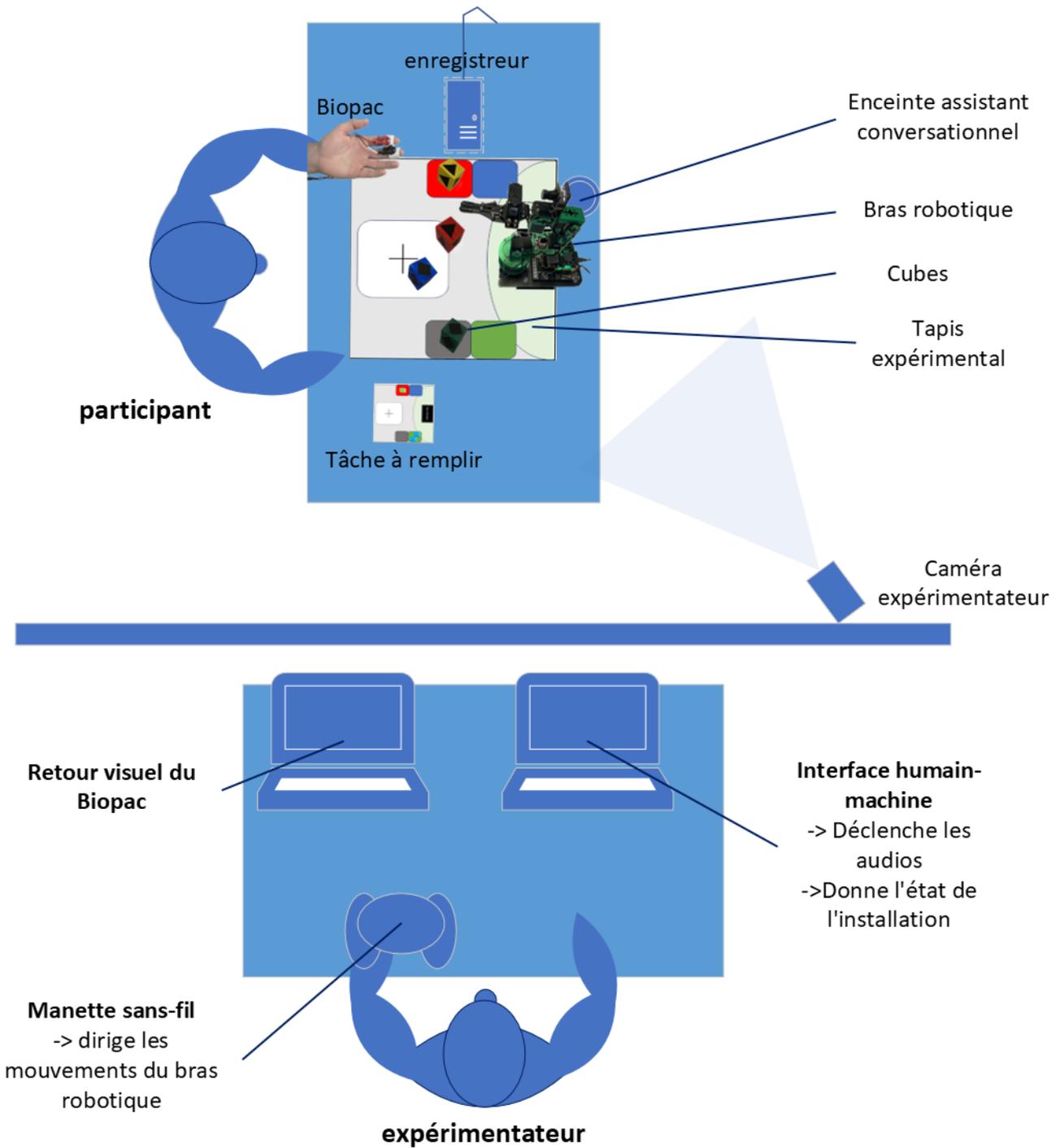


FIGURE 1 : Protocole expérimental envisagé

Pendant chaque session, le participant devra donner des instructions en langage naturel au bras robotique afin de déplacer des cubes selon des objectifs.

Il y aura deux difficultés d'objectifs : faible et élevée. La difficulté faible, considérée comme très facile devrait être ennuyeuse tandis que la difficulté élevée, considérée comme moyenne, devrait

être engageante. Le bras robotique pourra utiliser deux voix différentes :

- Une voix artificielle, proche de la voix humaine, considérée comme engageante ;
- Une voix humaine qui ferait tomber le bras robotique dans la vallée de l'étrange et serait donc moins engageante.

L'expérience durera environ 1 heure 30 durant laquelle les participants, environ 40 personnes, feront quatre sessions d'échanges de 15 minutes chacune avec le bras robotique après une session test de 5 minutes. Ils verront donc les 4 types de session possible (2 voix X 2 difficultés). Les participants seront recrutés via une plateforme du laboratoire de recherche. Pendant l'expérience, ils seront informés sur les outils dont ils seront équipés et sur l'utilisation des données récoltées. Une indemnisation de 15€ par participant est envisagé.

Le participant sera équipé du Biopac, un instrument en forme de gant, muni de capteurs, qui mesure les battements du coeur et l'activité électrodermale afin de mesurer l'engagement émotionnel. Il lui sera également demandé d'estimer le temps passé durant chaque session afin de mesurer la distorsion du temps due à l'engagement cognitif. Enfin, la voix du participant sera enregistrée via un micro pour extraire les données prosodiques et acoustiques de l'interaction. Cela nous permettra de déterminer précisément les paramètres significatif de l'engagement.

4.4 Hypothèses

Nous faisons l'hypothèse que les participants seront plus engagés dans l'interaction si la difficulté est plus élevée et ce d'autant plus s'ils échangent avec la voix artificielle plutôt qu'humaine. De même, les participants devraient être plus engagés lorsqu'ils échangent avec la voix artificielle plutôt qu'humaine et ce d'autant plus que la difficulté est élevée.

Au niveau physiologique, l'activité phasique de l'activité électrodermale et le nombre de battements de coeur par minute du participant seront plus élevés lorsque le participant est engagé. Au niveau prosodique, la moyenne, le maximum de la fréquence fondamentale et l'écart de variations devraient être plus élevés. Au niveau phonétique, il devrait y avoir une convergence phonétique du participant vers la voix de l'agent conversationnel, notamment quant à la durée des voyelles.

Plus les participants seront engagés, plus ils sous-estimeront le temps écoulé et rempliront au moins un objectif facultatif.

5 CONCLUSION

Le contexte professionnel rend l'étude des interactions humain-agent artificiel complexe dans la mesure où les interactions avec l'agent ne doit pas entraver les tâches principales du travail. Le choix du design et de la modalité d'échange avec l'agent doit être adapté à ce contexte. Le choix d'agents désincarnés permet d'éviter la surcharge visuelle d'un robot ou d'un agent incarné. Tandis que le choix d'un AC qui utilise la parole et non uniquement du texte permet de libérer les mains et de réduire la charge cognitive liée à la saisie manuelle. Malgré les avantages non négligeables des AC en contexte professionnel, les travailleurs ont encore des réticences à les utiliser.

Pour évaluer ces réticences, la notion d'engagement est assez parlante. Bien que de nombreux articles évaluent l'engagement, le contexte professionnel complexifie l'utilisation de la plupart des mesures. Les paramètres visuels et comportementaux largement utilisés dans l'évaluation de l'engagement dans les interactions entre humain et machine ne sont pas applicables car concentrés sur la tâche professionnelle. Évaluer l'engagement grâce à des paramètres physiologiques n'est possible en contexte professionnel que dans les cas non invasifs. Nous proposons d'étudier l'engagement au travers de paramètres prosodiques et phonétiques aujourd'hui uniquement étudiés dans les interactions entre humains. Notre étude respectera les problématiques liées au contexte professionnel en tenant compte du type d'agent utilisé dans ce contexte (agent désincarné) et des contraintes de mesures.

RÉFÉRENCES

- Arapakis, I., Barreda-Angeles, M., and Pereda-Baños, A. (2017). Interest as a proxy of engagement in news reading : Spectral and entropy analyses of eeg activity patterns. *IEEE Transactions on Affective Computing*, 10(1) :100–114.
- Araujo, T. (2018). Living up to the chatbot hype : The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, 85 :183–189.
- Attfield, S., Kazay, G., Lalmas, M., and Piwowarski, B. (2011). Towards a science of user engagement. In *WSDM Workshop on User Modelling for Web Applications*, pages 09–12.
- Bakker, A. B. and Oerlemans, W. G. (2011). Subjective well-being in organizations.
- Biro, T., Toscano, J. C., and Viswanathan, N. (2022). The influence of task engagement on phonetic convergence. *Speech Communication*.
- Bittner, E., Oeste-Reiß, S., and Leimeister, J. M. (2019). Where is the bot in our team ? toward a taxonomy of design option combinations for conversational agents in collaborative work.
- Buta, H. and Retalis, S. (2013). Enhancing primary school children collaborative learning experiences in maths via a 3d virtual environment. *Education and Information Technologies*, 18 :571–596.
- Buxó-Lugo, A., Toscano, J. C., and Watson, D. G. (2018). Effects of participant engagement on prosodic prominence. *Discourse Processes*, 55(3) :305–323.
- Cai, D., Li, H., and Law, R. (2022). Anthropomorphism and ota chatbot adoption : a mixed methods study. *Journal of Travel & Tourism Marketing*, 39(2) :228–255.
- Cambre, J., Liu, Y., Taylor, R. E., and Kulkarni, C. (2019). Vitro : designing a voice assistant for the scientific lab workplace. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pages 1531–1542.
- Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2008). Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, pages 13–17.
- Churchill, E. F. (2010). Enticing engagement. *interactions*, 17(3) :82–87.
- Clark, L., Ofemile, A., and Cowan, B. R. (2021). Exploring verbal uncanny valley effects with vague language in computer speech. *Voice attractiveness : Studies on sexy, likable, and charismatic speakers*, pages 317–330.
- Corrigan, L. J., Peters, C., Küster, D., and Castellano, G. (2016). Engagement perception and generation for social robots and virtual agents. *Toward Robotic Socially Believable Behavior Systems-Volume 1 : Modeling Emotions*, pages 29–51.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1) :32–80.
- Doherty, K. and Doherty, G. (2018). Engagement in hci : conception, theory and measurement. *ACM Computing Surveys (CSUR)*, 51(5) :1–39.
- Draper, M., Calhoun, G., Ruff, H., Williamson, D., and Barry, T. (2003). Manual versus speech input for unmanned aerial vehicle control station operations. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 47, pages 109–113. SAGE Publications Sage CA : Los Angeles, CA.
- Feng, S., Buxmann, P., et al. (2020). My virtual colleague : A state-of-the-art analysis of conversational agents for the workplace. In *HICSS*, pages 1–10.
- Guirdham, M. and Guirdham, O. (2017). *Communicating across cultures at work*. Bloomsbury Publishing.
- Guo, L. C. and Sanchez, Y. (2005). Workplace communication. *Organizational behavior in health care*, 4 :77–110.
- Gustafson, J. and Neiberg, D. (2010). Prosodic cues to engagement in non-lexical response tokens in swedish. In *DiSS-LPSS Joint Workshop 2010*.
- Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., and Welch, G. F. (2018). Does a digital assistant need a body ? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In *2018 IEEE International*

- Symposium on Mixed and Augmented Reality (ISMAR)*, pages 105–114. IEEE.
- Kusz, N. and Lemoine, J.-F. (2022). Influence de la voix d'un assistant vocal sur les réactions des utilisateurs. In *38ème colloque International de l'Association Française du Marketing (AFM)*.
- Lécureuil, J. (2023). 12 statistiques sur l'intelligence artificielle (ia) à connaître en 2024.
- Lester, J., Branting, K., and Mott, B. (2004). Conversational agents. *The practical handbook of internet computing*, pages 220–240.
- Liu, B. and Sundar, S. S. (2018). Should machines express sympathy and empathy? experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, 21(10) :625–636.
- Martín-Barrio, A., Roldán, J. J., Terrile, S., del Cerro, J., and Barrientos, A. (2020). Application of immersive technologies and natural language to hyper-redundant robot teleoperation. *Virtual Reality*, 24 :541–555.
- Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1) :10–12.
- Monkareisi, H., Bosch, N., Calvo, R. A., and D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, 8(1) :15–28.
- Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2) :98–100.
- O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6) :938–955.
- Oertel, C., Looze, C. D., Scherer, S., Windmann, A., Wagner, P., and Campbell, N. (2011). Towards the automatic detection of involvement in conversation. In *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pages 163–170. Springer.
- Pellet-Rostaing, A., Bertrand, R., Boudin, A., Rauzy, S., and Blache, P. (2023). A multi-modal approach for modelling engagement in conversation. *Frontiers in Computer Science*, 5 :14.
- Perugia, G., Rodríguez-Martín, D., Boladeras, M. D., Mallofré, A. C., Barakova, E., and Rauterberg, M. (2017). Electrodermal activity : Explorations in the psychophysiology of engagement with social robots in dementia. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1248–1254. IEEE.
- Quigley, M., Goodrich, M. A., and Beard, R. W. (2004). Semi-autonomous human-uav interfaces for fixed-wing mini-uavs. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2457–2462. IEEE.
- Rani, P. and Sarkar, N. (2005). Operator engagement detection and robot behavior adaptation in human-robot interaction. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2051–2056. IEEE.
- Sidner, C. and Dzikovska, M. (2002). Human-robot interaction : Engagement between humans and robots for hosting activities. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 123–128. IEEE.
- Smith, J. (2000). *GrandChair : conversational collection of grandparents' stories*. PhD thesis, Massachusetts Institute of Technology.
- Szafir, D. and Mutlu, B. (2012). Pay attention! designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 11–20.
- Von der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S.-H. (2010). "it doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior*.
- Westgate, E. C. (2020). Why boredom is interesting. *Current Directions in Psychological Science*, 29(1) :33–40.
- Wrede, B. and Shriberg, E. (2003). Spotting "hot spots" in meetings : human judgments and prosodic cues. In *INTERSPEECH*, volume 2003, pages 2805–2808.
- Yahboom (2024). Yahboom dofbot ai vision robotic arm with ros python programming for raspberry pi 4b 8gb/4gb.
- Zyngier, D. (2008). (re) conceptualising student engagement : Doing education not doing time. *Teaching and teacher education*, 24(7) :1765–1776.