Étude du comportement non-verbal d'un soignant en interaction avec un Patient-Virtuel

Jean Zagdoun

THÈSE DE DOCTORAT

présentée à

Sorbonne Université

par

Jean Zagdoun

pour obtenir le diplôme de

Doctorat de Sorbonne Université

Spécialité : Informatique

Étude du comportement non-verbal d'un Soignant en interaction avec un Patient-Virtuel

---

Soutenue le 14 Janvier 2022

JURY

Dominique Vaufreydaz	MCF HDR	Université de Grenoble	Rapporteur
Dan Istrate	MCF HDR	Université technologique de Compiègne	Rapporteur
Catherine Pelachaud	DR	Sorbonne Université	Examinatrice
Jean-Claude Martin	PR	Université Paris-Saclay	Examinateur
Giovanna Varni	MCF	Institut polytechnique de Paris	Membre invité
Mohamed Chetouani	PR	Sorbonne Université	Directeur de thèse
Laurence Chaby	MCF HDR	Université de Paris	Co-directrice de thèse

Remerciements

Mes premiers remerciements vont naturellement à mes encadrants Mohamed Chetouani et Laurence Chaby sans qui cette thèse n'existerait pas.

Je tiens ensuite à remercier tous les membres du projet VirtuAlZ pour l'avoir porté et pour s'y être investit. Il s'agit notamment des équipes du LIMSI de l'université Paris-Saclay, de l'hôpital Broca, de Cirel de l'université de Lille et de l'entreprise Sim4Health. Parmi tous les protagonistes de ce projet, je remercie tout particulièrement Amine Benamara pour ses discussions fructueuses et son apport non-négligeable.
Je ne remercierais jamais assez les personnes qui m'ont formé et apprit à cultiver un goût et une curiosité scientifique. Il s'agit de Justin Salez, mais aussi, entre autres de Stéphane Boucheron, Mathieu Merle, Thierry Meyre, Noufel Frikha, Cyrille Lucas et tant d'autres. J'aimerais aussi remercier Mario grâce à qui j'ai trouvé la motivation de poursuivre mes études et qui m'a prouvé que travailler avec acharnement portait ses fruits.
Un aspect non-négligeable du travail que j'ai effectué s'appuit sur des logiciels OpenSource et des articles en lecture libre. Pour le reste, je tiens à remercier Sci-Hub et la librairie Genesis.
Aussi, ma famille qui a été là dans des moments compliqués : mes frères ma mère mes tantes et mon père. J'adresse un remerciement spécial à ma mère sans qui cette thèse comprendrait beaucoup plus de fautes d'orthographes et dont les précieux talents ont rendu cette thèse plus lisible.
Également, mes amis qui m'ont accompagné et soutenu. À tous ceux cela un grand merci : Rus, Simon, Yohan, Mario, Yacine, Yanisse, Layla, Anya, Manon, Jacques, Pichot, Romain, Jalel, Sothea, Wail, Juliette, Félix, Manuel (ordre de cette liste tiré au hasard) et tous ceux que je n'ai pas cité mais qui comptent quand même pour moi (par exemple Lyes Kheloufi).
Je remercie aussi tous les doctorants qui ramèrent pour la plupart dans la même galère (et parfois pire) : Yohan, Thomas, Sothea, Amine, Felix, Manuel, Lucien, Wail, Jalel, Tanvi, Ramona, Victor, Sooraj, Sera, Karen, Sarah, Maira-Jose, Léo, et tous les autres que je n'ai pas cité.
Enfin, merci Émilie.

Introduction et contexte

Résumé

La formation professionnelle consiste en l'acquisition de connaissances grâce à un expert ou une source vérifiée. Cette acquisition s'accompagne souvent de tests qui portent sur des applications dans un milieu contrôlable. En effet, l'acquisition de connaissances, en elle-même, est difficilement évaluable car elle repose principalement sur l'apprenant, son attention et ses connaissances passées. Les évaluations sont alors un bon moyen de vérifier si l'apprentissage a été correct et à quel point.
L'importance d'un bon comportement du soignant vis-à-vis du patient est primordial. Ceci est d'autant plus vrai que le soin délivré traite des maladies neuro-dégénératives comme la maladie d'Alzheimer. Malheureusement, la formation des soignants sur la manière d'interagir avec des patients Alzheimer est délicate et coûteuse en temps et en personnel. Cependant, cette formation est nécessaire, car le comportement des patients Alzheimer est atypique et requiert un comportement maîtrisé de la part du soignant.
Cette thèse porte sur la formation du comportement de soignants par le biais d'interactions avec un Patient-Virtuel Alzheimer et, plus particulièrement, sur l'étude du comportement non-verbal du soignant. Automatiser une partie de cette formation apporte son lot d'avantages : les soignants appliquent leur savoir dans un environnement contrôlé et sans risque de blesser des patients, ce processus est répétable, peu cher et ne requiert pas de professionnels de santé qui sont une ressource rare ces temps-ci.
De ce fait, nous avons créé un logiciel de formation de soignants par interaction avec des Patient-Virtuels. La contribution de cette thèse consiste alors, en plus d'une partie de l'implémentation du logiciel, en trois propositions majeures :

- Une modélisation du comportement non-verbal des soignants en interaction avec un Patient-Virtuel. Cette modélisation repose sur une identification a priori des comportements non-verbaux (langage corporel, expressions faciales) pertinents et de leurs détections automatiques lors de l'interaction.
- Une analyse des passations de soignants (aide-soignants, infirmiers, médecins, psychologues) sur notre logiciel et grâce à notre modélisation.
- Nous apportons aussi des éléments de réponse dans la direction d'une modélisation plus fine et complémentaire du comportement non-verbal du soignant qui ne reposerait sur aucun comportement prédéfini.

Mots clefs : Comportement non-verbal, Traitement du Signal Social, Réseaux de neurones, Clustering, Patient-Virtuel

Table des matières

Table des figures

Définitions

- Symboles mathématiques

Symboles	Description
$\mathbb{R}$	l'ensemble des réels
$\mathbb{N}$	l'ensemble des entiers
$[\![i, j]\!]$	équivalent à $\{i, i+1, \ldots, j-1, j\}$ où $i, j \in \mathbb{N}$, $i < j$

- Définitions

Symboles	Description
Interpersonnel	Qui concerne au moins deux personnes ; exemple : échanges interpersonnel, échanges entre deux personnes ou plus.
Intrapersonnel	Qui concerne une seule personne ; exemple : signaux sociaux intrapersonnel, signaux sociaux émis par une seule personne.
Soignant	Personnel médical administrant des soins. Dans ce document il s'agit principalement de médecins, infirmiers, aide-soignants et psychologues.

- Jargon technique

Mots	Description
Epoch	Représente la totalité du jeu de données considéré. Un algorithme entraîné sur 2 *epoch* d'entraînement aura vu deux fois la totalité de la base de données d'entraînement.
Frame	Image spécifique d'une vidéo. Par exemple, quand nous pausons une vidéo, ce que nous voyons est une frame.
Feature	Caractéristique d'un point de données. Souvent utilisée pour prédire un label
Learning rate	Taux d'apprentissage de l'algorithme d'optimisation.
Point de données	*Sample* en anglais. Correspond à un individu d'une base de données.
Supervisé/ non-supervisé	Relatif à l'apprentissage d'un algorithme. Si cet apprentissage doit prédire ou classifier une caractéristiques des points de données (label) ou si l'apprentissage soit trouver par lui même des caractéristiques.

Introduction

Ouverture

Aristote disait de l'être Humain qu'il est un animal social. Notre langage et la plupart de nos compétences cognitives s'acquièrent, en effet, part le biais d'interactions avec d'autres humains. On retrouve aussi, en partie, cette notion dans l'allégorie de la caverne de Platon qui établit une relation de causalité entre les interactions avec notre environnement et notre perception de la réalité.
Cependant, si certaines interactions peuvent être source de transmission d'information et de savoir, d'autres, si elles ne sont pas maîtrisées correctement, peuvent aussi aggraver la peine et le mal-être de personnes vulnérables, fragilisées par une maladie. Ainsi, et dans le cadre de maladies neuro-dégénératives comme la maladie d'Alzheimer, la qualité de l'interaction entre soignant et patient impact fortement la qualité des soins prodigués. Dans ce cadre, la formation des soignants doit aussi porter sur leur comportement non-verbal.
Cette thèse porte sur l'analyse automatique des comportements non-verbaux de professionnels de la santé en interaction avec un Patient-Virtuel Alzheimer. Plus précisément, la problématique qui servira de colonne vertébrale au reste de la thèse est la suivante :

$\mathcal{Q}$: *Dans une interaction entre un Soignant et un Patient-Virtuel, comment quantifier automatiquement la qualité de la communication non-verbale du soignant ?*

Nous introduisons ici le contexte générale de la thèse dans la sections . Nous listons ensuite les questions de recherches auxquelles cette thèse répond dans la section . Enfin nous parlons des limitations de notre travail dans la section 2.
Cette thèse a contribué à la publication des articles suivants :

Liste des publications

- **Zagdoun, J.**, Chaby, L., Benamara, A., Urbiolla Gallegos, M. J., and Chetouani, M. (2021). Non-verbal behaviors analysis of healthcare professionals engaged with a virtual-patient. in *Socially-Informed AI for Healthcare*, ICMI'21, October 18–22, 2021, Montréal, Canada

- Benamara, A., Martin, J.C., Prigent, E., Chaby, L., Chetouani, M., **Zagdoun, J.**, Dacunha, S., Ravenet, B., (accepté) COPALZ : A Computational Model of Pathological Appraisal Biases for an Interactive Virtual Alzheimer's patient, AAMAS 2022.

- Becerril-Ortega, R., Vanderstichel, H., Petit, L., Schoch, J., Dacunha, S., Benamara, A., Ravenet, B., **Zagdoun, J.**, Chaby, L. (sous presse). Didactical conception process of a virtual simulation environment for the training of healthcare professionals in geriatrics. In Simulation Training through the Lens of Experience and Activity Analysis, Springer, Berlin.

- Petit, L., Vandertichel, H., Urbiolagallegos, M-J., Benamara, A., **Zagdoun, J.**, Chaby, L., Becerril-Ortega, R., (2021). Recherche, interdisciplinarité et conception de formations. Le projet comme analyseur de l'interdisciplinarité au sein d'une activité collective. *Actes du 55ème colloque de la SELF, Paris.* ergonomie-self.org/wp-content/uploads/2021/01/SELF-2020-actes.pdf

Contexte général

La transmission d'informations et de savoir est primordiale à notre survie et à notre bon développement. Dans ce cadre, la formation des individus joue un rôle capital au sein de nos sociétés. On observe, en France par exemple, que le budget alloué à la formation professionnelle continue et l'apprentissage est en constante hausse [3].
Bien qu'indispensable, la formation est pour certains domaines très peu présente. Il est vrai, comme nous le démontrons lors du chapitre 1, que la formation des soignants sur le comportement non-verbal à adopter lors d'une interaction avec un patient Alzheimer est compliquée à mettre en œuvre. En France, cette formation est donc principalement empirique (Becerril-Ortega et al., 2020).
Le projet VirtuAlZ, auquel cette thèse contribue, a pour but d'aider à la formation des soignants en les faisant interagir avec un Patient-Virtuel, c'est-à-dire avec un Agent-Virtuel qui présente les symptômes de la maladie d'Alzheimer.
La formation par le biais d'interactions avec des Agent-Virtuels offre de nombreux avantages. Le premier étant de permettre à l'utilisateur de s'imaginer en interaction avec un être humain et donc le stimule davantage que s'il interagissait juste avec une machine. Par exemple, il a été démontré que, sauf vallée de l'étrange (Mori, 1970), plus l'Agent-Virtuel possède des caractéristiques humaine plus l'interaction se rapproche d'une interaction avec un vrai être humain (Heyselaar et al., 2015). En effet, ces Agents-Virtuels sont censés imiter l'illusion de la vie, et aider à la "suspension d'incrédulité" (Bates, 1995), (Lester and Stone, 1997), (Wooldridge and Veloso, 1999).
Un deuxième avantage à l'utilisation des Patient-Virtuels dans la formation de personnel soignant

3. https://www.senat.fr/rap/r06-365-1/r06-365-158.html

est qu'ils procurent un environnement sûr dans lequel la pratique n'a pas d'effets dommageables. Se tromper, pour le participant, n'impact pas directement la santé et l'intégrité de vrais patients Alzheimer.
Le projet VirtuAlZ contribue à la création d'un environnement numérique où les participants peuvent apprendre en pratiquant avec un Patient-Virtuel. Au sein de ce projet, et en plus d'une partie de l'élaboration de l'environnement numérique, notre objectif porte sur l'analyse du comportement non-verbal du soignant.

Objectifs principaux et questions de recherche

La première question à laquelle nous devons répondre est la suivante : Est-il possible de détecter automatiquement un bon comportement non-verbal de la part d'un soignant en interaction avec un patient Alzheimer ?
Nous devons donc voir s'il y a dans la littérature scientifique des comportements spécifiques à adopter ou à éviter lorsqu'un soignant interagit avec un patient Alzheimer. De nombreuses études s'intéressent à cette question, parmi lesquelles celles de (Collins et al., 2011) et (Mast, 2007) qui affirment que les comportements suivants sont significativement liés à la qualité d'une interaction dans le cadre de test de performance ethno-gériatrique [4] :

- Maintenir une expression faciale adéquate.
- Avoir une gestuelle contrôlée.
- Limiter les mouvements inutiles.
- Limiter les mouvements de la main.

Cependant, nous observons trois problèmes majeurs avec cette stratégie :
Premièrement, les signaux considérés sont très spécifiques et complexes. De plus, certains signaux tels que la prosodie ou le toucher facial ne sont pas pris en compte dans ces études. En conséquence, ces signaux sont durs à détecter automatiquement et nous passons possiblement à côté d'autres informations sur le comportements non-verbal.
Deuxièmement, dans ces études, le comportement non-verbal est annoté manuellement, il est donc considéré en contexte, par rapport au comportement du patient et à un environnement spécifique. Il ne nous est pas garanti qu'il en va de même pour notre cadre d'étude.
Enfin, même dans l'hypothèse où nous pourrions connaître les comportements non-verbaux qu'il faut adopter pour chaque soignant, donner un retour automatisé peut être mal perçu.
C'est pour cela que dans le cadre du projet VirtuAlZ, nous avons préféré que ce soit le soignant qui évalue lui-même son comportement. Notre rôle est alors de l'aider dans cette démarche avec des outils d'analyse automatique.

4. *OSCE :Objective Structured Clinical Examination*, un test de performances pour mesurer les compétences cliniques des soignants

Notre but se divise alors en deux objectifs :

$\mathcal{O}1$: *Établir des clusters de profil type de comportement non-verbal de soignants.*

$\mathcal{O}2$: *Identifier si des caractéristiques personnelles des soignants jouent un rôle dans leur comportement non-verbal.*

Ces deux objectifs sont complémentaires et permettent, à terme, d'analyser automatiquement la passation d'un participant et de lui donner des retours de haut niveau sur son profil comportemental. C'est-à-dire, informer le participant des différences et similitudes entre son comportement et celui d'autres participants.
Pour ce faire, nous développons une modélisation du comportement non-verbal des participants inspirée du traitement du langage naturel, une analyse d'une première passation de 29 soignants et une contribution vers une modélisation du comportement de manière non-supervisée grâce à des réseaux de neurones siamois.

Limitations

Notre travail établit une modélisation du comportement non-verbal et répond aux deux objectifs de la section précédente ; à savoir l'objectif $\mathcal{O}1$ et l'objectif $\mathcal{O}2$. Toutefois, nous n'avons pas pu pousser cette analyse jusqu'à l'identification de cluster concret et informatif sur un comportement global des soignants en général. Notre analyse s'intéresse à 29 soignants qui ont accepté de participer à notre expérience. Cependant, le comportement de 29 participants ne permet pas de dégager des profils généraux. Il s'en suit que notre analyse, bien que pertinente, n'est pas assez riche pour pouvoir aider les participants suivants.
Deux possibilités de travaux futurs s'offrent alors. D'une part, faire plus de passation et notamment avec des participants qui n'ont aucune expérience avec le milieu médical. D'autre part, annoter manuellement les comportements des soignants pour donner plus de pertinences aux clusters comportementaux trouvés.

Au niveau de la contribution sur la modélisation non-supervisée du comportement non-verbal des soignants, nous limitons notre contribution à une preuve de concept. Nous prouvons qu'il est possible de découper de manière non-supervisée la vidéo d'un humain en différentes actions et d'identifier ces dernières. Nous utilisons pour cela des réseaux de neurones siamois. Notre contribution, bien que pertinente, ne permet pas encore l'élaboration d'un outil d'analyse automatique à proprement parler.
De plus, les actions que nous découpons et identifions ne sont pas des comportements non-verbaux qui seraient intéressant de détecter dans les passations des participants mais ce sont des actions concrètes : sauter, marcher, se gratter la tête, jeter une balle, etc. Ceci est due au fait que ce sont les bases de données les plus facilement accessibles et annotées.

Organisation du document

Cette thèse est organisée comme suit : le chapitre 1 présente un état de l'art des applications de la simulation dans le milieu médical. En plus de cela, nous abordons dans cette thèse différents domaines bien spécifiques qui nécessitent tous un état de l'art plus ou moins développé. Nous listons ici les différent thèmes de revues de la littérature incorporés dans cette thèse :

- Utilisation des N-grams dans la modélisation de comportement Humain.
- Algorithme de détection de nouveauté pour la modélisation de comportement Humain.
- Réseaux de neurones siamois.

Le chapitre 2 présente la première collecte de données, dans le cadre du projet VirtuAlZ, effectuée à l'hôpital Broca, à Paris. Nous présentons aussi la modélisation du comportement non-verbal des participants.
Dans le chapitre 3, nous analysons les données du chapitre précédent. Ceci est fait en deux parties, en prenant en compte les caractéristiques personnelles des participants et d'autre part, de manière non-supervisée.
Enfin, au chapitre 4 nous établissons une preuve de concept sur la modélisation du comportement non-verbal de manière non-supervisée.

Chapitre 1

Simulation dans le domaine de la santé : État de l'art

Sommaire

Introduction

Le comportement [1] du soignant joue un rôle important lors de l'administration de soins à un patient. De nombreuses études ont ainsi démontré l'importance du comportement du soignant en milieu médical et les effets bénéfiques qu'il pouvait avoir : (Ha and Longnecker, 2010), (Hall et al., 2019), (Timmermann et al., 2017), (Mast, 2007). Par exemple, le travail de (Howe and Leibowitz, 2019) a récemment montré que l'effet placebo pouvait être renforcé par des démonstrations de chaleur humaine et de compétence de la part du soignant. Une meilleure relation de confiance et une meilleure adhérence au traitement peuvent aussi venir d'une communication non-verbale adéquate (Khullar, 2019). Le comportement des soignants est d'autant plus important que les soins consistent en la prise en charge et l'accompagnement du patient à travers une maladie incurable.
De ce fait, dans le cadre de patients atteints de la maladie d'Alzheimer, la qualité des soins ne se limite pas seulement à la connaissance de la maladie : adopter un comportement social adéquat avec les patients est primordial. En effet, la maladie d'Alzheimer est une maladie neuro-dégénérative, ce qui implique que le comportement des patients est très différent d'un comportement typique (Weaver Moore et al., 1993), (Thomas et al., 2017). Ainsi, une personne non-formée peut aggraver la peine d'un patient s'il ne sait pas réagir à une crise, ou à un comportement agressif de manière adéquate.
Cependant, les contraintes de temps, de personnel et l'isolement des patients ne facilitent pas son apprentissage. C'est dans l'objectif de répondre à ce manque que le projet VirtuAlZ a été initié. L'idée étant de former le personnel médical à interagir avec des patients Alzheimer avec un Agent-Virtuel. Un des buts principaux de cette thèse est d'analyser les bonnes pratiques et les bons comportements à adopter dans une interaction entre un soignant et un patient Alzheimer. Cette analyse se restreint au cadre où le patient est simulé par un Agent-Virtuel. L'Agent-Virtuel est alors appelé Patient-Virtuel.
Aussi, notre analyse ne s'est focalisée que sur les comportements non-verbaux du soignant, le sujet étant assez complexe pour que cette étude reste pertinente bien que limitée. La suite de cette partie est un aperçu de la littérature scientifique qui offre un semblant de réponse à notre problématique $\mathscr{2}$ ou à une variante de cette dernière. En effet, les études scientifiques françaises portant sur les interactions entre Soignant et Patient-Virtuel ne sont pas assez nombreuses pour que notre travail repose uniquement dessus. Cependant, nous pouvons alléger certaines des contraintes de la problématique $\mathscr{2}$ pour avoir un point de vue plus global. Nos contraintes relaxées sont alors que l'interaction doit être en rapport avec le milieu médical et/ou incorporer une forme de simulation. Pour garder une cohérence dans ce chapitre, il ne sera pas fait mention de l'analyse non-verbale des Humains en général ni des travaux relatant d'Agents-Virtuel s'ils ne sont pas en rapport avec un soignant. Ces points seront cependant abordés plus loin dans la thèse lorsque cela sera pertinent.

Présentation des différentes parties

La section 1 définit le concept de simulation et son spectre d'application dans le milieu médical. La section 2 établit l'état de l'art en listant les différents travaux que nous avons retenus dans notre étude. La section 3 développe cet état de l'art en extrayant les conclusions et les différentes métriques utilisées. La section 4 dresse les limitations de ces études.

1. Le terme "comportement" désigne les actions d'un être vivant observable de façon externe (Piéron, 1931). Au cours de ce chapitre, si aucune précision n'est donnée, le terme comportement renvoie aussi bien au comportement verbal qu'au comportement non-verbal.

1 La simulation dans le milieu médical

Dans cette section, nous nous intéressons à définir la simulation de manière générale dans le milieu médical. Nous relaxons alors les contraintes liées à : l'analyse exclusive du non-verbal du soignant, l'utilisation d'un avatar incarné et réaliste et, enfin, nous élargissons cette revue de la littérature aux compétences techniques en plus des compétences de communications lorsque cela est pertinent.
Cette section établit un état des lieux de la simulation dans le milieu médical, qu'importe le but de la simulation (geste technique, communication) et qu'importe la technologie employée (application web, Avatar, Réalité Virtuelle [RV]).

1.1 Pourquoi la simulation ?

Une Simulation peut être définie comme : "*Une représentation artificielle d'un fonctionnement, d'un processus.*" (*Le Robert*). La simulation qui nous intéresse ici est celle qui concerne la simulation d'un individu dans une interaction avec un soignant. La modélisation de flux de patients, l'anticipation du nombre de médicaments et de lits disponibles ne sont pas prises en compte. Un lecteur intéressé par ces thématiques pourra cependant se référer à (Roberts and England, 1980).
La simulation s'est d'abord développée dans la pratique de gestes techniques, et, notamment, en formation de compétences cliniques (chirurgie, anesthésie...) (Spencer and Eiseman, 1962), (Esogbue, 1979). Le concept de compétence clinique est défini comme la capacité d'établir une évaluation complète, développer et implémenter une stratégie de soin et d'en évaluer les conséquences sur des patients (Coyne et al., 2021). Une erreur dans ce domaine peut avoir des conséquences sommaires et brutales. C'est dans ce même état d'esprit qu'un pilote apprend à faire atterrir son avion sur un simulateur avant de s'essayer sur un Boeing 777. Ainsi, la simulation est un lieu propice au développement de l'agilité et de la dextérité nécessaires à la réalisation d'un geste complexe. C'est aussi un bon moyen d'acquérir de la confiance en soi par la pratique et de diminuer le stress et l'anxiété des premières fois où l'on traite un patient.
Cependant, même si un geste raté lors d'une opération à cœur ouvert peut sembler catastrophique, la plupart des accidents dans le milieu médical proviennent en fait du facteur humain (Marano et al., 2005). Il est alors normal de constater que les performances chirurgicales sont impactées par le facteur humain comme rapporté par (Louise et al., 2011) et (Flin et al., 2017). Dès lors, les gestes techniques les plus travaillés en simulation sont ceux en rapport avec les relations sociales (Marano et al., 2005), (Bracq et al., 2019).
Ainsi, et au vu de l'importance d'un bon comportement lorsqu'un soignant prodigue des soins, il est logique que la simulation ait atteint des domaines moins techniques du monde médical, ce qu'on appelle des gestes non-techniques. Les gestes non-techniques sont définis comme : "*Les compétences cognitives, sociales et personnelles qui complètent les compétences techniques et contribuent à l'exécution sûre et efficace des tâches*" (Bracq et al., 2019). Ils incluent alors compétences cognitives et compétences sociales interpersonnelles. Ces gestes sont pour la plupart liés à la communication, au travail en équipe, à la prise de décision, à l'évaluation de la situation et au commandement.
La simulation développe alors le raisonnement et la prise de décision des étudiants en santé (Henderson et al., 2016), et, en plus d'être répétable, répond au problème de temps et de coût

des formations classiques, propose de multiples scénarios et niveaux de difficultée (Coyne et al., 2021). De surcroît, cette pratique répond a certains problèmes éthiques car elle permet d'éviter de s'entraîner en interagissant directement avec un patient, et favorise la règle du "*rester le plus loin possible du patient les premières fois*" (Okuda et al., 2009).
Les bénéfices de la simulation pour les gestes non-techniques seront détaillés dans la section 3.2. Ils sont tirés de la revue de la littérature faite dans la section 2.

1.2 Les différentes formes de simulation

Chaque catégorie de type de simulation est suivie d'abréviations entre parenthèses, ces abréviations sont notamment utilisées dans le tableau 1.1 de la Section 2. Ici, et par souci de lisibilité, la catégorisation que nous construisons sur les différentes formes de simulation ne citera pas l'ensemble des travaux dont elle est issue. Ces travaux sont cependant listés dans le tableau 1.1.
La simulation peut, tout d'abord, être virtuelle ou non. La simulation non-virtuelle peut consister en un acteur, ou un expert(AcT) simulant un comportement atypique, celui d'un patient que le soignant doit gérer. Elle peut aussi être faite avec un mannequin(MnQ) sur lequel les soignants devront s'exercer. Dans le cas d'acteurs(AcT), cela peut être fait en groupe de soignants qui vont interagir chacun à leur tour, en binôme ou, dans certains cas, le soignant peut lui-même jouer le rôle du patient. Enfin, il existe aussi des cas où l'acteur est dans un milieu médical et le soignant ne sait pas que ce n'est pas un vrai patient. Tous ces scénarios possibles font appel au jeu de rôle (role-play), c'est à dire incarner une personnalité différente de la notre et la jouer à la manière d'un acteur.
L'utilisation de mannequins (MnQ) est, elle, principalement appliquée à la pratique de gestes techniques.
La simulation virtuelle peut prendre plusieurs formes :

- Application web (WeB) : des données sur le Patient sont disponibles (fiche, pathologie, historique des opérations, etc.), et c'est alors au soignant de décider des choix à adopter (prise de médicament, radio, scanner, etc.).
- Jeux Vidéo (JV) : notamment tiré de Second Life ®. Selon le même principe que pour les applications web, en plus réaliste car dans un environnement virtuel plus riche. Cependant, cette forme requiert une certaine prise en main de l'outil.
- Patient-Virtuel (PV) qui peut être en ligne (ON) ou scripté (OFF).
    - Scripté (OFF), le Patient-Virtuel suit alors un arbre de conversation et de comportement écrit à l'avance. Les branches du scénario étant parcourues en fonction des choix du soignant. Les comportements du Patient-Virtuel sont alors tous prédéfinis et ne requièrent pas de génération à la volée. Le Patient-Virtuel peut donc aussi bien être modélisé en 2D, 3D ou même être composé de vidéo d'acteurs. Le taux de perception d'Agent-Virtuel n'étant pas significativement éloigné de ceux d'acteurs filmés (Georgescu et al., 2014).
    - En ligne (ON), Le Patient-Virtuel est alors modélisé en 2D ou en 3D et incorpore un outil de génération comportementale. Son comportement est alors généré par un Humain (souvent un expert ou quelqu'un de formé) qu'on appelle Magicien d'Oz (MoZ). La simulation en ligne est une amélioration de la simulation scriptée sur de nombreux points (Przyrembel et al., 2012).

2 La simulation pour l'entraînement dans le milieu médical

Dans cette section, nous listons les différentes études qui s'inscrivent dans un cadre similaire au nôtre. Nous nous intéressons en particulier aux buts finaux de ces études et aussi à la mise en place de leurs processus de simulation. Ce faisant, nous pouvons alors comparer les différentes techniques d'analyses habituellement réalisées et voir pourquoi, dans notre cas, elles ne sont pas applicables.

2.1 Description des différentes études

L'ensemble des travaux relativement proches du nôtre est listé dans le tableau 1.1. Les conclusions de chacune de ces études ne sont pas ajoutées dans le tableau, toujours par soucis de lisibilité, mais font l'objet d'une partie en elles-même (voir Section 3.2). Chaque ligne du tableau résume quatre caractéristiques principales de chaque étude listée :

- # : le nombre de participants dans l'étude.
- **Type de participants** : quelles sont les caractéristiques de sélection des participants. S'ils font des études, lesquelles, et sinon, quel est leur domaine d'activité médicale.
- **Type de Simulation et spécificité/pathologie** : quelle technologie est employée pour la simulation et si le Patient-Virtuel a des caractéristiques spécifiques (pathologie, appartenance à un milieu social, symptômes quelconques, etc.).
- **Évaluation** : Comment est évaluée l'expérience.

Le but de ce tableau est de donner une vision d'ensemble des travaux pris en compte lors de notre étude. L'ensemble de ces travaux permet alors, en plus de nous situer au sein de la littérature scientifique, d'identifier les caractéristiques principales de ces études.

Tableau 1.1. Tableau listant les études faites sur l'utilisation de la simulation dans le milieu médical. La première colonne indique l'étude, la deuxième le nombre de participants, la troisième le type de participants, la quatrième le type de patient simulé (voir 1.2) avec la spécificité de ce patient quand elle a été précisée, et la cinquième la manière d'évaluer l'étude. RV : Réalité Virtuelle

Étude	#	Type de participants	Type de Simulation et Spécificité/pathologie	Évaluation
(Lehr and Kaplan, 2013)	54	Étudiants en soins infirmiers	MnQ/ troubles mentaux	Questionnaire sur l'anxiété pré- et post-test, post-test METI ®

	(Jack et al., 2014)	154	Étudiants en soins infirmiers	AcT (les participants ne le savent pas)	Retour des étudiants, debriefing avec les étudiants	
	(Doolen et al., 2014)	94	Étudiants en soins infirmiers	AcT entraînés/ trouble bipolaire, anxiété, schizophrénie	Questionnaire et groupe de contrôle	
	(Menzel et al., 2014a)	51	Étudiants en soins infirmiers	OFF : Avatar, JV (Second life)/ Pauvreté	Pré- et post-test, debriefing	
	(Agrawal et al., 2015)	83	Étudiant en médecine	MnQ/femme enceinte	Pré- et post-test (OSCE)	
	(Oudshoorn and Sinclair, 2015)	56	Étudiants en soins infirmiers	AcT avec scripts/ troubles mentaux	Retour des instructeurs et des étudiants	
	(Kron et al., 2016)	421	Étudiant en médecine	ON + RV + interprofessionel/ leucémie	Test OSCE	
	(GOH et al., 2016)	95	Étudiants en soins infirmiers	AcT/ trouble mentaux(sucidaire)	Pré- et post-test	
	(Fossen and Stoeckel, 2016)	40	Étudiants en soins infirmiers	Patient-Virtuel audio puis role-play développé par les auteurs (où les étudiants jouent a la fois le rôle du soignant et celui du patient).	Questionnaires, analyse thématique (revue par des chercheurs hors de l'expérience)	

	(Jacobs and Venter, 2017)	33	Étudiants en soins infirmiers	AcT, étudiants en master de théâtre/ troubles mentaux	Questionnaire avec 5 questions ouvertes (≠ QCM) et analyse thématique des réponses	
	(Wright et al., 2017)	103	Étudiants en soins infirmiers	WeB (Vsim)	Post-test (one-way ANOVA non significative) et retour des participants	
	(Stepan et al., 2017)	66	Étudiant en médecine	OFF, imagerie médicale (IR, computed tomography) rendues en 3D/ patients sains	Questionnaires pré- et post-intervention, retours des participants	
	(Borg Sapiano et al., 2017)	166	Étudiant en médecine	OFF, vidéo d'acteurs/ détérioration rapide (cardiaque, respiratoire, choc)	Pré- et post-test	
	(Sankaranar et al., 2018)	20	Gynécologue et obstétricien ou chirurgien	OFF (avatar) vs mannequin/ sur le bloc opératoire	Pré- et post-questionnaire	
	(Padilha et al., 2018)	426	Étudiants en soins infirmiers	OFF (avatar) puis application web	Questionnaires	
	(Alexander et al., 2018)	34	Étudiants en soins infirmiers	AcT	Analyse thématique des retours des participants	

	(Witt et al., 2018)	32	Étudiants en soins infirmiers	AcT professionnels + scripts/ troubles mentaux	Pré- et post-test, sondages de satisfaction, liste des comportements à adopter	
	(Liaw and Wu, 2019)	207	Étudiants en soins infirmiers, pharmacologie, infirmiers, physiothérapie, thérapie occupationelle, et travail social	OFF + RV	2 annotateurs, note de 1 à 10 basée sur une liste de 10 comportements visibles	
	(Liaw et al., 2020b)	16	Étudiants en soins infirmiers, pharmacologie, infirmiers, physiothérapie, thérapie occupationelle, et travail social	OFF + RV	2 annotateurs, analyse de thème sur le verbatim des interactions	
	(Verkuyl et al., 2020)	19	Étudiants en soins infirmiers	OFF (vidéo d'acteur)/ client prénatal dans un centre de soins ambulatoires	Analyse thématique, débriefs transcrits en verbatim er analysés par les auteurs de l'étude	

2.2 Les objectifs des différentes études

Une des données manquante au Tableau 1.1 est le but de l'étude en question : à quelles questions les auteurs ont-ils voulu répondre ?
Cette section liste de manière thématique l'ensemble des questions posées par les études du Tableau 1.1. Les articles ne sont pas cités par souci de lisibilité. Les métriques d'évaluations et les réponses aux questions posées sont, elles, présentées dans la section 3.

En ce qui concerne les gestes techniques, la simulation est surtout utilisée pour améliorer deux compétences du soignant :

- Entraînement, répétition : réussite du soin prodigué, dextérité, agilité, entraînement pratique, etc. (Grover et al., 2015), (Khan et al., 2017)

- Collaboration interéquipes, intersoignants

Pour les gestes techniques, les questions qui se posent sont naturelles et l'évaluation de la réussite du geste également. En effet, la plupart des gestes techniques consistent en la réalisation d'une tâche bien spécifique (recoudre une plaie, ablation d'une tumeur, etc.) et la réussite de cette tâche est souvent liée à la survie et la guérison du patient. Toutefois, on note l'importance de l'évaluation des techniques de communication lors de la réalisation des gestes techniques (voir Section 1.1). Dans ce cas, on cherche à répondre aux mêmes questions que lors de l'entraînement à des gestes non-techniques, qui sont :

- Quelle a été la qualité des soins prodigués ?
- L'interaction était-elle authentique ?

De manière générale, ces deux questions se sous-divisent en quatre autres :

- Est-ce que la simulation est efficace ?
- Est-ce que l'apprentissage est cohérent ? La communication du Soignant s'est-elle améliorée ?
- Est-ce que la simulation est bien perçue, agréable d'utilisation ?

- Est-ce que la technologie employée est cohérente ? quelle est la meilleure ? et dans quel cas ?

On note alors le manque d'étude sur l'analyse du comportement non-verbal dans les interactions avec un Patient-Virtuel atteint de troubles mentaux. Ces études se focalisent principalement sur les stratégies de communication. En effet, parmi les travaux cités, un seul s'intéresse à des comportements non-verbaux précis et listés (Kron et al., 2016). Cette étude a mis en avant l'apprentissage de comportements non-verbaux des participant en leurs permettant de voir leurs actions enregistrées par vidéo et, en plus, en leur montrant certains comportements non-verbaux spécifiques détectés comme les hochement de tête, les mouvements de sourcils et le sourire. Ceci permettant *in fine* un meilleur apprentissage de la communication non-verbale des participants.

3 L'évaluation de la simulation

3.1 Les méthodologies d'évaluation

Dans cette section nous ne citons pas tout les articles Au niveau de l'évaluation des expériences, une grande majorité des études repose sur des analyses de haut niveau des interactions. Les évaluations s'appuient sur ces procédures :

- Pré- et post-questionnaires/tests : il s'agit de remplir des questionnaires pertinents sur les compétences requises avant et après l'interaction, et de voir s'il y a eu une augmentation quelconque du score entre-temps. Ce procédé peut aussi être réalisé avec des groupes de contrôle et couplé avec des test de significativité. Aussi, on note que certaines études proposent seulement des tests post-expérience, mais ont alors recours à des groupes de contrôle.
- Analyse thématique : il s'agit de prendre en compte les retours des participants sur des questionnaires plus ouverts (voire parfois sans questionnaire) et de voir si des thèmes principaux se dégagent des réponses.

Dans les deux cas, ces types d'analyse, bien que pertinents, sont très difficilement automatisables et donc ne répondent pas à notre problématique $\mathcal{Q}$.
Enfin, on note aussi que l'évaluation de la qualité de l'interaction peut utiliser des techniques d'imagerie médicale. Grâce à des IRM, il est alors possible de mesurer quelles parties du cerveau sont stimulées lors d'interactions avec des AV, ce qui permet de vérifier si elles correspondent aux mêmes parties du cerveau utilisées lors de vraies interactions sociales (Georgescu et al., 2014). Bien que très pertinente, cette technologie n'est pas compatible avec l'étude du comportement non-verbal du soignant. Ceci est principalement dû à son côté très entravant qui limite les mouvements. Cependant, l'utilisation de cette technologie reste cohérente lorsqu'il s'agit d'étudier la question de l'authenticité de la perception du Patient-Virtuel par le soignant.

3.2 Les apports des études considérées

Nous listons dans le tableau 1.2 les conclusions des études considérées lors de la section 2. Nous voyons ici, en détail, que les questions de recherche de ces études sont principalement celles de la section 2.2. Le détail des conclusions des études permet aussi de mettre en valeur l'intérêt de l'utilisation de la simulation dans le milieu médical.
Cet intérêt porte alors principalement sur :

- l'acquisition de compétences techniques, c'est à dire, une quantification concrète de l'amélioration de la communication des soignants.
- L'apprentissage par la pratique, les retours des participants et leurs points de vue sur leur propre apprentissage.
- l'utilisabilité, la comparaison de technologie, quel technologie performe le mieux, laquelle est la plus pertinente et dans quel contexte ?
- La réduction des altérations négatives, quel impacte positif sur les participants ?

De ce tableau nous pouvons en déduire que l'utilisation de la simulation dans le milieu médical a beaucoup d'intérêt en soi. Ceci renforce l'apport de cette thèse et surtout de la mise en place

d'un dispositif de simulation dans le cadre médical.

Tableau 1.2. Retours des différentes études listées dans le tableau 1.1. Les conclusions de ces études sont listées dans ce tableau. La première colonne représente une des quatres grandes catégories de conclusions auxquelles répondent les études (deuxième colonne). La dernière colonne précise les conclusions.

Questions	**Études**	**Détails**
Acquisition de compétences techniques (amélioration communication évaluation, confiance)	(Liaw and Wu, 2019)	Meilleure performance sur le score annoté du groupe qui a eu la formation avant l'expérience.
	(Witt et al., 2018)	Amélioration significative des scores post-tests en moyenne.
	(Jacobs and Venter, 2017)	Expérience agréable et aide a intégrer la théorie et la pratique, développer des compétences de communication.
	(Doolen et al., 2014)	Améliore la confiance et les compétences de communication.
	(Borg Sapiano et al., 2017)	Efficace dans dans l'acquisition de compétences.
	(Agrawal et al., 2015)	Classes virtuelles efficace pour l'apprentissage de connaissance.
Apprendre par la pratique, améliorer l'apprentissage	(Liaw et al., 2020b)	Permet de voir "le patient comme un tout" ainsi que de comprendre le rôle des autres personnels soignants.
	(Oudshoorn and Sinclair, 2015)	La formation a eu un impact positif mais coût du PV important, d'autres formes de PV sont alors envisagées.
	(Jack et al., 2014)	Retours positif des étudiants.
	(GOH et al., 2016)	Il est efficace d'utiliser un PV, améliore l'apprentissage de compétences.
	(Fossen and Stoeckel, 2016)	Améliore la connaissance des étudiants en médecine et la pratique des soins apportés.
	(Kron et al., 2016)	Meilleur résultats pour les participants qui ont utilisé la simulation.
	(Kneebone et al., 2016)	Les conclusions sont faites par les utilisateurs sur les améliorations et les bénéfices qu'ont ces séances de simulation. Ici, le but principal est le partage d'expérience par la pratique et la démonstration.

Utilisabilité, ergonomie, comparaison de technologies	(Stepan et al., 2017)	Étude des différences d'apprentissage entre PV+RV et livres en ligne. L'apprentissage a été le même dans les deux groupes (pas de différences dans les performances aux examens), mais ceux qui ont utilisé la RV ont trouvé l'expérience plus agréable, engageante, et utile.
	(Padilha et al., 2018)	Utile à l'apprentissage, facile d'utilisation.
	(Sankaranarayanan et al., 2018)	Étude des différences entre utilisation Mannequin et avatar. Pas de différences significatives sur les tests, cependant une meilleure réaction à l'incendie pour les personnes dans la simulation que ceux du groupe contrôle, réduction des altérations négatives.
Réduction des altérations négatives	(Lehr and Kaplan, 2013)	Baisse de l'anxiété des participants.
	(Alexander et al., 2018)	Aide à réduire l'anxiété des étudiants et permet d'ôter certains de leurs préjugés.

4 Les limitations

4.1 Les limitations de la simulation dans le milieu médical

Dans leur article, (May et al., 2009) analysent 69 études publiées de 1996 à 2005 sur l'utilisation de la simulation dans le milieu médical pour des étudiants. Une de leurs conclusions est alors qu'une majorité des travaux manquaient de rigueur dans le design de leur recherche.
De notre analyse, nous pouvons en effet établir que beaucoup d'études prises en considération dans le Tableau 1.1 mettent en effet l'accent sur l'apprentissage par la pratique comme une finalité. Alors, la question qui est majoritairement posée est la suivante : Quels avantages apporte l'utilisation de la simulation de patients lors de l'apprentissage ? Cette question se subdivise en une multitude d'autres questions et problématiques qui sont listées dans la section 2.2 et dont les réponses se trouvent dans le Tableau 1.2. On note alors l'écart entre ces problématiques et la nôtre ($\mathscr{2}$) qui, elle, s'intéresse à automatiser la mesure de la qualité du comportement du soignant en interaction avec un Patient-Virtuel réaliste.
Notre problématique, qui se conjugue en relation avec la génération automatique du comportement du Patient-Virtuel, répond au problème notamment soulevé par (Liaw et al., 2020a). Les auteurs reconnaissent les avantages de l'utilisation d'un Patient-Virtuel, mais soulignent les limites du contrôle de l'avatar par un Humain et la nécessité de l'automatiser.
Le besoin, après les séances de simulations virtuelles, d'un compte-rendu, est aussi considéré comme une des limitations technologiques (Coyne et al., 2021), (Menzel et al., 2014b), (Foronda et al., 2013).

Une autre limitation également, est celle de l'utilisation d'avatars peu réalistes (Kron et al., 2016).

4.2 Les limitations de la simulation dans le cadre du projet VirtuAlZ

Un autre problème accentue la difficulté de l'analyse automatique des comportements des soignants en interaction avec des patients. Ce problème est celui du cadre de l'étude, comme rapporté par (Gauffman, 1974). En effet, la plupart des expériences sont conçues et analysées avec un cadre primaire ; cependant, la réalité est beaucoup plus complexe et s'étend à travers les frontières culturelles et sociales des individus. Par exemple la perception et l'évaluation du Patient-Virtuel dépendent de caractéristiques personnelles (âge, expérience, sexe, métier) (Rosenthal-von der Pütten et al., 2010). Ainsi, le nombre de variables à différencier pour étudier la qualité du comportement est assez grand si tant est que l'on puisse extraire cette qualité. Ce qui est un tout autre problème.
Dans le cadre très spécifique de la maladie d'Alzheimer, en effet, les résultats de (Harwood et al., 2018), qui compilent 26 études, n'offrent aucun moyen d'identifier clairement une stratégie de communication qui puisse être utilisée dans l'apprentissage de professionnels de la santé pour surmonter les problèmes de démence dans les soins quotidiens. Une stratégie communicative doit alors être adoptée en fonction de chaque patient et soignants (Belzil and Vézina, 2015), (Becerril Ortega et al., 2019). Il est alors impossible de savoir a priori quels comportements seront les mieux à adopter de la part du soignant de manière générale.
Dans le cadre de la France, il y a un manque concret de documentation précise sur le comportement exact à adopter lors d'une interaction avec un patient Alzheimer, qui s'explique entre autre par la grande variabilité des comportements Humain et de leur interprétation. En effet, mêmes les recomandations faites par l'(H.A.S., 2009) sur ce cas précisent : "*Les attitudes de communication suivantes sont données pour exemple et sont à adapter à chaque cas*". Enfin, de notre étude (Becerril-Ortega et al., 2020), il ressort que la formation des infirmiers et aides-soignants se fait principalement de manière empirique.
Il est alors impossible par l'analyse automatique de distinguer un bon ou un mauvais comportement de par la spécificité de la maladie d'Alzheimer et de par la diversité des participants. De plus, il faut aussi considérer le problème éthique de "noter" le comportement du soignant automatiquement par une machine. Cependant, nous pouvons tout de même accompagner et aider l'entraînement du participant par des retours pertinents.

Enfin, et pour conclure sur les limitations, nous citons une partie de l'article de (Coyne et al., 2021) :
"[Leur] *analyse intégrative a permis d'identifier plusieurs lacunes dans la littérature existante, notamment l'absence d'études multidisciplinaires sur la détérioration des patients, le manque de stratégies rigoureuses qui permettraient l'utilisation, par exemple, du deep learning pour le débriefing et l'analyse finale et, enfin, la nécessité d'une interaction synchrone dans le cadre de la simulation virtuelle. Les études incluses ont montré un rapport limité sur la validité et la fiabilité des mesures d'évaluation, ce qui souligne la nécessité de développer des outils fiables pour évaluer les compétences cliniques dans les simulations virtuelles.*" [2]
Les contributions de cette thèse portent donc sur la mise en place d'un dispositif de simula-

2. Traduit en partie avec www.DeepL.com/Translator (version gratuite)

tion utilisant un Patient-Virtuel et sur l'implémentation d'outils d'évaluation automatique du comportement non-verbal du soignant.

Conclusion

Dans ce chapitre, nous avons présenté une vue d'ensemble des études pertinentes en rapport avec la simulation dans le milieu médical. Ces études offrent un point de vue global de la recherche et permettent de mieux positionner cette thèse par rapport aux travaux faits précédemment.
Une des grandes différences entre cette thèse et les études citées est dans la finalité. Le but de la plupart des études est de justifier l'utilisation de la simulation dans une certaine branche du domaine de la santé, ou bien, de former directement du personnel médical grâce à des protocoles déjà établis. Même si notre but est aussi de former des soignants, nous avons des contraintes supplémentaires qui nous empêchent de procéder comme la plupart des travaux présentés. Entre autre, la contrainte d'automatisation du Patient-Virtuel couplée à la complexité de la maladie d'Alzheimer nous empêche de procéder à des questionnaires de performance ou à des analyses thématiques des retours des participants.
La stratégie employée diffère et s'inspire alors du domaine de l'Interaction Humain-Machine et de l'apprentissage automatique. Elle consiste en l'accompagnement de l'apprentissage et l'aide à l'auto-formation par des retours qui ne sont en aucun cas connotés.

Chapitre 2

Le Projet VirtuAlZ

Sommaire

Introduction

La mission du projet **VirtuAlz** est d'aider la formation des soignants novices en les faisant interagir avec un Patient-Virtuel. Ce faisant, les soignants peuvent alors pratiquer une interaction difficile dans un milieu contrôlé sans risquer de blesser de vrais patients. À cette fin, nous avons développé un protocole expérimental en collaboration avec les différents partenaires du projet.
Au sein de ce projet, notre objectif est d'analyser le comportement du soignant en interaction avec le Patient-Virtuel et de donner des retours pertinents de manière automatique. De ce fait, le fonctionnement du Patient-Virtuel et son automatisation ne seront pas détaillés, car cela ne dépend pas de nous.
Les différents protagonistes du projet sont :

- Broca, hôpital gériatrique : Lieu de formation des soignants.
- LIMSI, université Paris-Saclay : Création et automatisation du Patient-Virtuel.
- ISIR, Sorbonne-Université : Analyse du comportement du soignant en interaction avec le Patient-Virtuel.
- Sim4Health, entreprise privée : Rendu graphique du Patient-Virtuel.
- Cirel, université de Lille : Scénario du déroulement de l'interaction.

Présentation des différentes parties
Ce chapitre décrit la mise en place du protocole expérimental, d'une première collecte de données et de la modélisation du comportement non-verbal des soignants. La section 1 décrit le protocole expérimental : le concept de l'expérience, les transmissions des différents flux de données et les logiciels utilisés. Dans la section 3, nous détaillons le processus de modélisation du comportement non-verbal du soignant.

1 Description du dispositif mis en place

1.1 Aperçu du dispositif de collecte de données

Une vue simplifiée du dispositif expérimental est présentée dans la figure 2.1. Cette figure représente l'interaction entre le soignant d'une part et le Patient-Virtuel contrôlé par un Magicien d'Oz [1] d'autre part, la captation audio et vidéo du comportement du soignant et la génération du comportement du Patient-Virtuel.

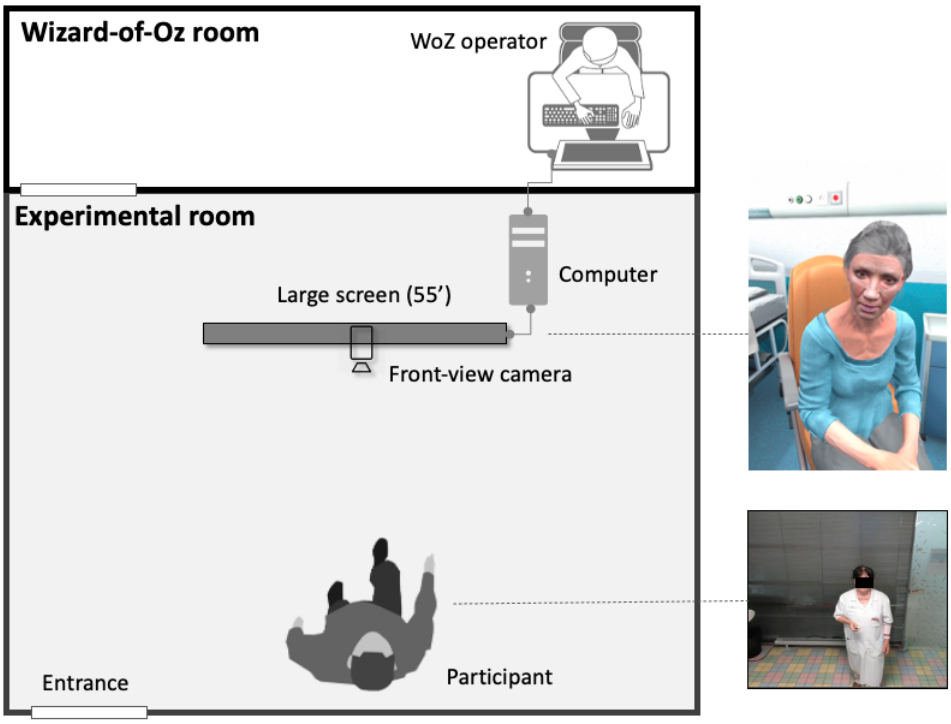

Figure 2.1. Description du déroulement de l'expérience VirtuAlZ. Le participant est dans une pièce face à l'écran du Patient-Virtuel ; son comportement est enregistré grâce à une caméra et à un micro. Le comportement du Patient-Virtuel est, lui, généré par le Magicien d'Oz qui, situé dans une autre pièce, peut suivre le déroulement de l'expérience grâce à la vidéo du participant retransmise en temps réel.

Nous présentons dans cette section le déroulement typique d'une passation avec un participant quelconque. Nous détaillons ensuite la gestion des flux de données et la communication entre les différentes machines. Cette dernière partie étant plus technique, nous invitons un lecteur peu intéressé à ne pas la prendre en considération.

1. L'utilisation d'un Magicien d'Oz consiste en un humain qui contrôle le comportement de l'Agent-Virtuel et le déroulement du scénario.

1.1.1 Déroulement de l'expérience

Avant d'interagir avec le Patient-Virtuel, le Magicien d'Oz explique au participant le déroulement de l'expérience et les questionnaires qu'il doit remplir. Le participant est ensuite amené dans la salle expérimentale (*experimental room* dans la figure 2.1) pendant que le Magicien d'Oz démarre l'expérience depuis un ordinateur situé en dehors.
Le participant fait face à un grand écran de 140 cm de haut avec lequel il interagit grâce à une télécommande. Un tutoriel sur l'utilisation de cette télécommande démarre alors et le participant apprend à s'en servir pour parcourir la fiche patient du Patient-Virtuel et pour la sélection des dialogues et des actions lors de l'interaction.
Une fois ce tutoriel fini, le participant lit la fiche patient du Patient-Virtuel, et prend connaissance de sa condition. L'objectif de cette interaction est de donner un médicament au Patient-Virtuel. L'interaction commence lorsque la fiche patient a été lue et se déroule en trois phases principales :

- L'approche patiente : la manière de rentrer dans la chambre et de discuter ou non avec le Patient-Virtuel.
- Le geste technique : donner le médicament au Patient-Virtuel.
- La sortie de la chambre.

L'enchaînement des phases et la progression dans l'arbre du scénario sont contrôlés par le Magicien d'Oz grâce à une interface graphique représentée dans la figure 2.2. Tous les participants ne parcourent pas le scénario de la même manière et certains n'ont pas accès à certaines phases à cause des choix qu'ils ont fait et des décisions prises par le Magicien d'Oz.

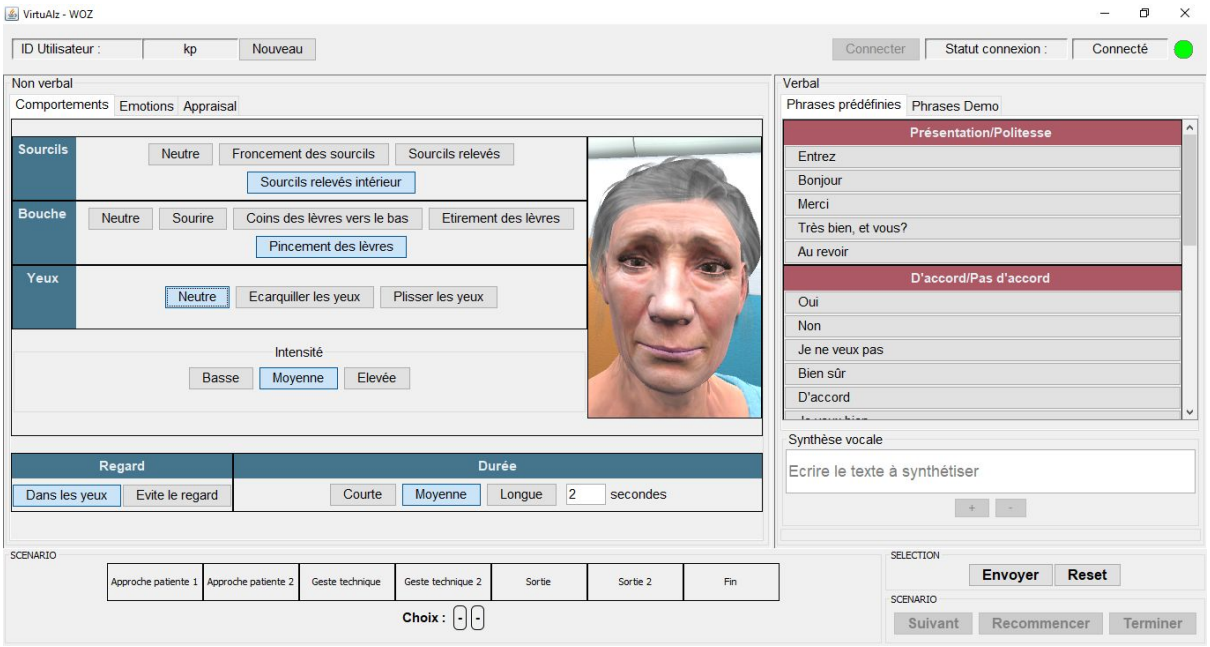

Figure 2.2. Interface de contrôle du Patient-Virtuel. Le magicien d'Oz peut décider des comportements non-verbaux du Patient-Virtuel. Il peut aussi contrôler son comportement verbal grâce à des phrases pré-enregistrées ou grâce à un synthétiseur vocal. Enfin, cette interface permet également de contrôler l'avancement dans le scénario global.

L'interaction se déroule de la manière suivante : le participant choisi un dialogue et/ou une action parmi une sélection de choix possibles comme présenté dans la figure 2.3. Une fois la sélection faite, le participant fait alors son choix. Les dialogues proposés sont généraux et laissent place à la libre interprétation du participant.

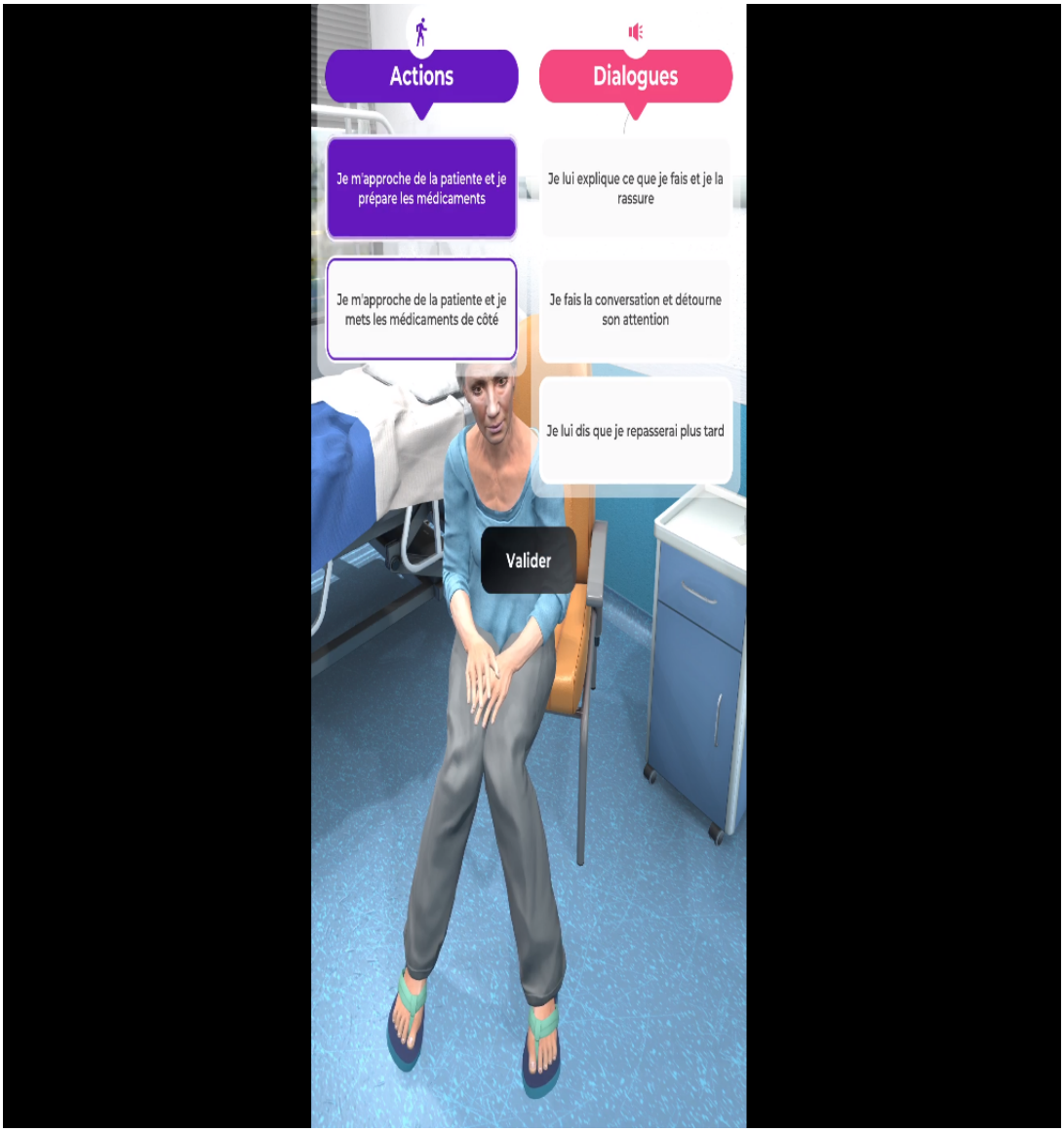

Figure 2.3. Fenêtre de sélection des actions et dialogues sélectionnables par le participant. Une fois cette sélection faite grâce à une télécommande, le participant doit les jouer.

En même temps que le participant joue ces choix, le Magicien d'Oz contrôle le comportement du Patient Virtuel. Il choisit notamment ses expressions faciales, la direction du regard, la durée de maintient des expressions faciales sélectionnées et son verbal. Le Magicien d'Oz a alors le choix de sélectionner des phrases pré-enregistrées ou bien d'en générer à la volée grâce à un synthétiseur vocal intégré depuis Windows dans l'interface graphique du Magicien d'Oz (figure 2.2).

Figure 2.4. Interface du logiciel Anvil. Permet de parcourir la passation du participant et de voir en même temps son comportement et celui du Patient-Virtuel qui sont fusionnés en une seule vidéo. (Fenêtre d'affichage de la vidéo). Grâce à la Fenêtre des pistes d'annotations, nous pouvons aussi voir clairement quels comportements ont été envoyés et générés au niveau du Patient-Virtuel. Cette fenêtre permet également de naviguer directement dans la vidéo grâce à un clique de souris sur une frame (trait rouge).

Une fois l'interaction terminée, la captation audio et vidéo s'arrête.
Le participant rejoint le Magicien d'Oz pour un debrieffing et pour remplir des questionnaires sur l'ergonomie, l'utilisabilité, l'acceptabilité et l'appréciation du dispositif, mais aussi sur la perception du Patient-Virtuel (questionnaires : godspeed (Bartneck et al., 2008), System Usability Scale (Grier et al., 2013), Acceptability e-scale (Tariman et al., 2011), KirkPatrick (Smidt et al., 2009)). Le choix et l'utilisation de ces questionnaires est exclusivement fait par nos partenaires. Le debrieffing se fait grâce à l'outil Anvil [2] (Kipp, 2001) présenté dans la figure 2.4. Cet outil permet au participant de voir sa passation, de naviguer dedans et de connaître les choix pris et les comportements sélectionnés par le Magicien d'Oz.

2. http://www.anvil-software.org/#

2 Description des données collectées

La collecte de données a été réalisée à l'hôpital Broca, à Paris. Le protocole d'étude "Outil de simulation de Patient-Virtuel pour l'entraînement de personnel médical travaillant avec des patient Alzheimer ou ayant des pathologies similaires" a été approuvé par le comité d'éthique de l'université de Paris (numéro de référence IRB : 2019-67). Tous les participants ont été informés de la procédure avant l'expérience et ont fourni un consentement écrit.
Il y a eu 32 passations qui ont duré en moyenne une heure chacune, le temps moyen de l'interaction avec le Patient-Virtuel étant de 10 minutes (minimum : 5 minutes, maximum : 18 minutes, écart-type : 3.7).
Au niveau des soignants, la répartition des professions est décrite dans le tableau 2.1.

Tableau 2.1. caractéristiques démographiques des 29 participants du corpus VirtuAlZ.

Contexte	**Détail**	**Nombre**
Profession	Aide-soignant	6
	Infirmier	8
	Médecin	9
	Psychologue	6
Expérience	<5 ans	5
	entre 5 et 15 ans	13
	>15 ans	11
Âge	25-34	7
	35-44	8
	45-54	9
	55-65	5
Sexe	Femme	24
	Homme	5

L'ensemble des passations s'est déroulé comme décrit dans la section 1. Nous analysons en priorité le comportement non-verbal des participants. Cette analyse est décrite dans le chapitre 3. Avant cela, nous présentons la manière dont nous modélisons le comportement non-verbal des participants.

3 Modélisation du comportement non-verbal

Dans cette section, nous décrivons notre approche pour la modélisation du comportement non-verbal des soignants. Notre but est de créer une représentation qui est à la fois capable de capturer des signaux multimodaux non-verbaux complexes et, aussi, de permettre de générer des retours pour les apprenants. Dans cette optique, nous avons eu recours à une modélisation symbolique discrétisée, fondée sur des modèles de traitement du langage naturel. Un des buts d'un tel modèle est de permettre l'interprétabilité des comportements non-verbaux. Un ensemble de **symboles** est alors regroupé en un **état** multimodal du comportement du soignant à un instant donné. Cet ensemble d'états est alors groupé en ***N*-grammes** pour capturer une partie de

la dynamique.
Pour ce faire, nous utilisons des logiciels d'extraction de *features* qui nous permettent de considérer directement le comportement Humain. Les *features* que nous extrayons sont des points d'intérêts sur le corps et le visage des soignants.

3.1 Détection automatique des signaux non-verbaux

3.1.1 Détection automatique de la pose du corps

Dans l'optique d'extraire des informations sur le comportement non-verbal corporel des soignants, nous avons fait le choix d'utiliser le logiciel OpenPose (Cao et al., 2018; 2017; Simon et al., 2017; Wei et al., 2016), qui extrait en temps réel les coordonnées des points d'intérêts [3] du corps, du visage et des mains des participants. Nous utilisons ce logiciel pour les raisons suivantes :

- Extraction de signaux facilitée par cet outil (voir sec :symbols).
- Traitement des données en temps réel, ce qui permet donc de transmettre les signaux extraits également en temps réel.
- Le dispositif est prévu pour une seule personne et donc l'identification de personne n'est pas un problème.

Les points d'intérêt extraits par OpenPose sont représentés sur la figure 2.9. Le système est conçu pour fonctionner avec une et une seule personne présente dans la pièce. Bien que ce logiciel puisse fonctionner en temps réel, ici, et dans l'optique des analyses faites lors de cette thèse, nous extrayons ces points d'intérêts pour chaque frame de chaque vidéo.

3. *keypoints, landmarks* en anglais, ce sont des points charnière du physique humain qui sont universels (sauf exception).

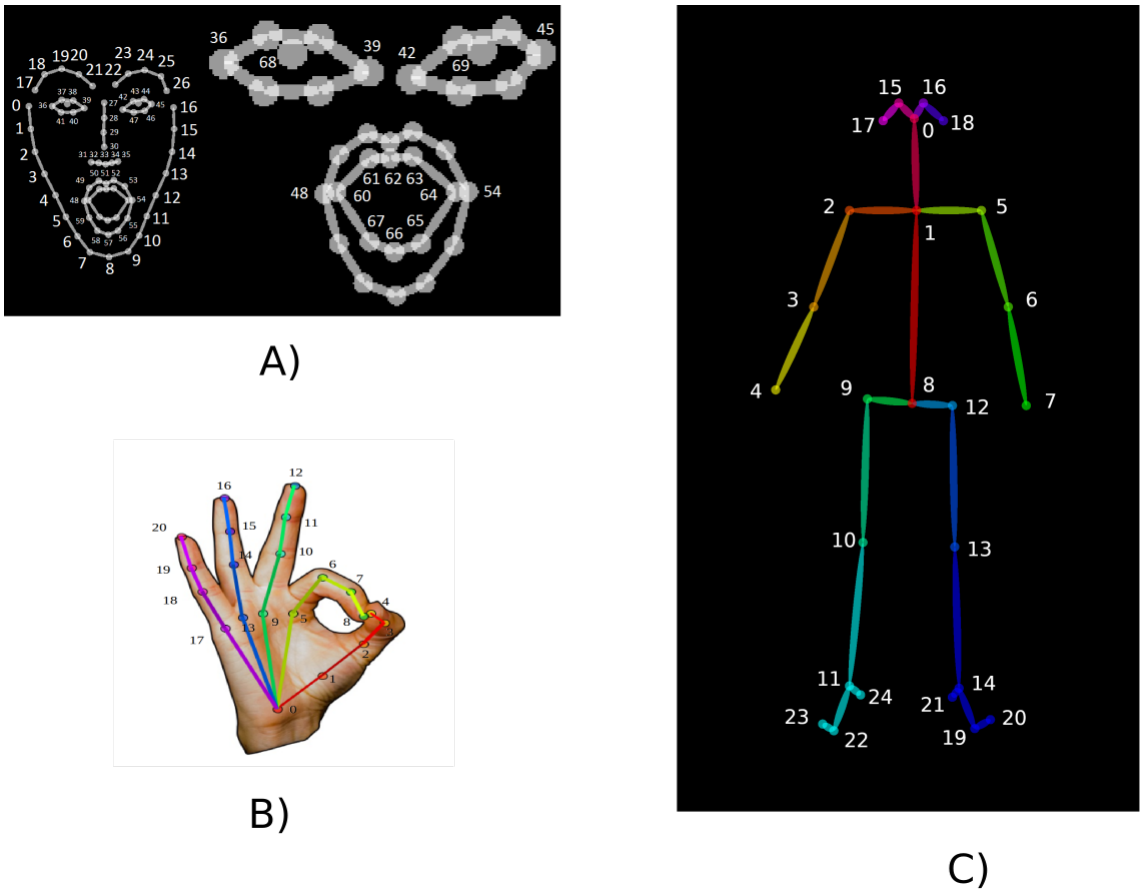

Figure 2.5. Liste des points d'intérêt (*landmarks*) d'OpenPose, pour chaque landmark détecté, le logiciel nous donne sa position sur l'écran (x, y) et la probabilité qu'il se trouve bien à cet endroit. La figure est prise du github officiel d'OpenPose [4]. Figure A) : points d'intérêts de la tête. Figure B) : points d'intérêts de la main (x2 dans le cas ou le soignant a deux mains). Figure C) : points d'intérêts du corps.

4. https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/02_output.md

3.1.2 Détection automatique des Action Units du visage

Nous extrayons aussi des Actions Units du soignant grâce au logiciel OpenFace (Baltrusaitis et al., 2018). Ce logiciel calcule des points d'intérêts en 3D et des Actions Units. Ces Actions Units correspondent à des activations musculaires faciales qui sont à la bases des expressions faciales et de certaines émotions. Une illustration des Actions Units est représentée dans la figure 2.9.

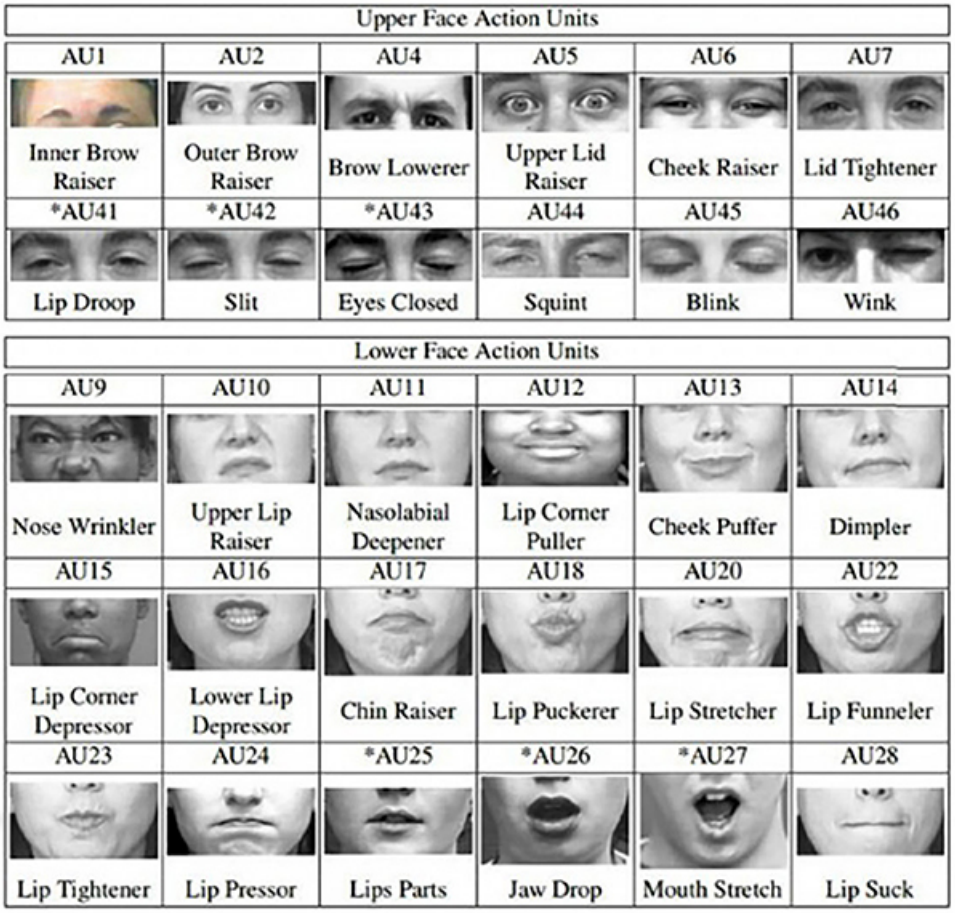

Figure 2.6. Liste des Action Units. Pour chaque Action Unit détectée, le logiciel nous informe si elle est activée ou non ainsi que la probabilité d'activation. La figure est issue du travail de (Jia et al., 2021; De la Torre et al., 2015) Copyright © 2015, IEEE.

3.2 Du signal au symbole

Nous définissons un symbole comme un signal ou un groupe de signaux qui possèdent une information incompressible et directement identifiable par un être humain.
Dans la suite de cette thèse, chaque symbole est un nombre binaire : 1 si le symbole est détecté, sinon 0.
Ces symboles sont exclusivement extraits des vidéos des participants. Nous nous intéressons alors à deux groupes de symboles qui sont ceux liés aux expressions faciales d'une part et au langage corporel d'autre part.

3.2.1 Liste des symboles

Comme nous l'avons indiqué dans la section 3.1, les signaux sur lesquels nous construisons nos symboles sont des points d'intérêts du corps des participants et les activations musculaires de sa tête. Ce faisant, la construction de symboles est alors facilitée et repose sur deux méthodes principales. Pour les symboles liés aux expressions faciales nous monitorons l'activation d'Action Units spécifiques grâce à OpenFace. Pour les symboles liés à la pose du corps, un travail supplémentaire est nécessaire, mais repose principalement sur la construction d'une distance à partir des données d'OpenPose (OPCS : *"OpenPose Custom Score"*). La liste des symboles pris en compte dans notre modélisation est alors représentée dans le tableau 2.2.

Tableau 2.2. Liste des différents symboles d'intérêt pris en compte dans notre analyse. La première colonne représente le symbole en question, la dernière la méthode utilisée pour détecter ce symbole, OPCS signifiant *"OpenPose Custom Score"*. La construction de ce score est alors spécifique pour chaque symbole lié à la pose du corps.

Symbole	**Détection**
Sourire	AU12
Froncement de sourcils	Au4
Sourcils relevés	Au5 Au1
Coins de la bouche vers le bas	Au15
Touché de la tête	OPCS
Proximité envers le Patient-Virtuel	OPCS
Ouverture des bras	OPCS

3.2.2 Symboles faciaux

Le comportement facial non-verbal correspond à l'activation d'Action Units spécifiques. Nous en extrayons des symboles pertinents, notamment ceux liés à l'activation de la bouche et des sourcils. Dans des configurations similaires à la nôtre, il a été rapporté que ces symboles jouent un rôle important (Collins et al., 2011; Biancardi, 2019).
Nous détectons les symboles faciaux grâce à OpenFace en observant l'activation du muscle

concerné sur le visage du participant. Nous surveillons, parmi les expressions faciales, l'activation du sourire, le mouvement des sourcils (froncés ou levés) et le coin de la bouche vers le bas.

3.2.3 Symboles corporels

En ce qui concerne la construction des symboles liés au corps du participant, nous travaillons grâce aux 25 points d'intérêts corporels (voir la section 3.1.1). C'est à dire 25 coordonnées x, y. À partir de ces coordonnées, nous construisons des distances qui nous permettent de détecter l'activation ou non des symboles d'intérêt.

Proximité : La proximité du soignant envers un patient est considérée comme importante dans la satisfaction du patient (Mast, 2007). Nous construisons ce symbole comme décrit dans la figure 2.7. Nous suivons la distance entre les épaules du participant, qui augmente au fur et à mesure que ce dernier se rapproche de la caméra. La construction du symbole est faite de la manière suivante :

- Extraction de la distance entre les épaules.
- Normalisation pour prendre en compte les différents types de morphologie.
- Lissage grâce à des ondelettes.
- Palier pour avoir un score binaire.

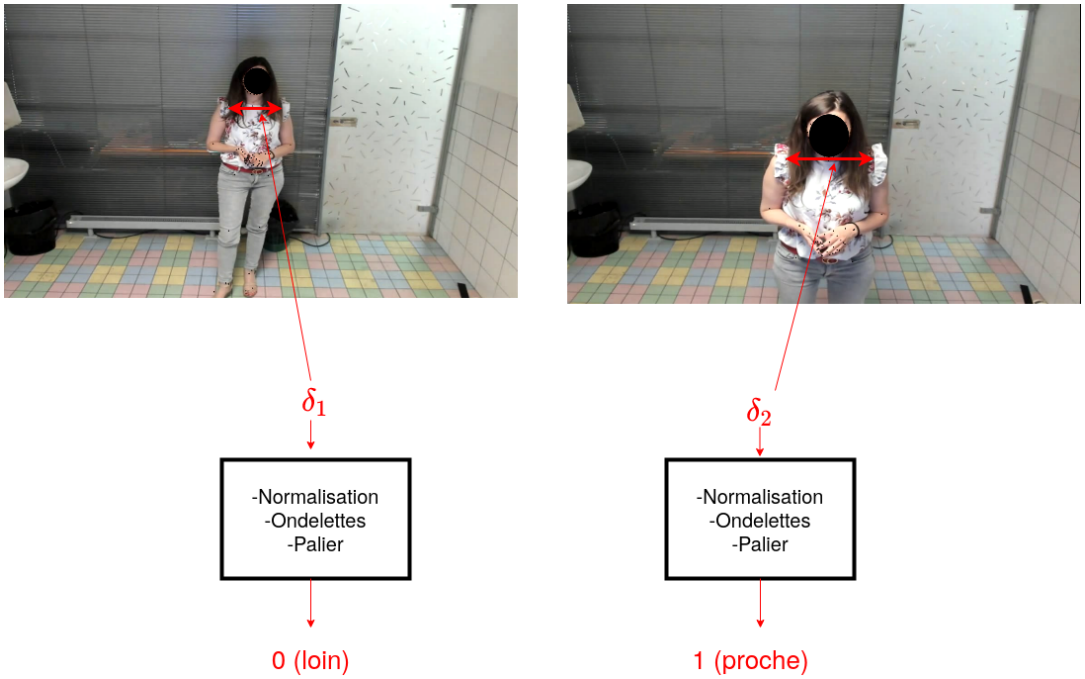

Figure 2.7. Détection du symbole de proximité. Nous monitorons l'écart entre les deux épaules du participant et si ce dernier se rapproche de la caméra, cet écart augmente. Cette distance nécessite cependant 3 phases de nettoyage et de traitement avant d'être un symbole interprétable. Une phase de normalisation pour faire fi des différentes morphologie, une phase de nettoyage grâce à des ondelettes et une phase de palier pour la binarisation.

Toucher facial : Le toucher facial est entre autre lié au stress et au traitement de l'information (Harrigan, 1985). Comme montré dans la figure 2.8, nous observons les distances entre le visage et l'une des deux mains. Si la coordonnée y d'une des deux mains dépasse celle du visage, nous considérons ce symbole comme actif.

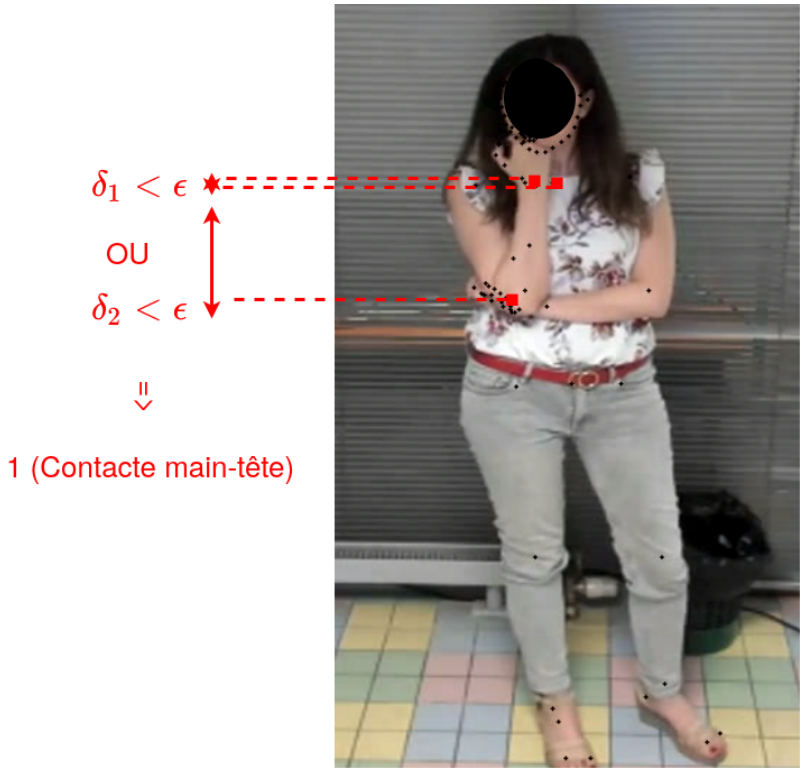

Figure 2.8. Détection du symbole lié au toucher facial. Ce symbole est activé si une des deux mains du participant entre en contact avec sa tête. La distance δ_1 représente la distance de la coordonnée y de la main droite et celle du cou, la distance δ_2 représente la même distance mais à gauche.

Ouverture des bras : L'ouverture des bras peut être liée à une expression de chaleur humaine ou bien à l'agitation. Nous calculons ce symbole en observant la distance entre les poignets et les épaules. Si la coordonnée x du poignet gauche est supérieur à la coordonnée x de l'épaule gauche alors le symbole est actif, et inversement pour le poignet droit. Ceci est représenté dans la figure 2.9.

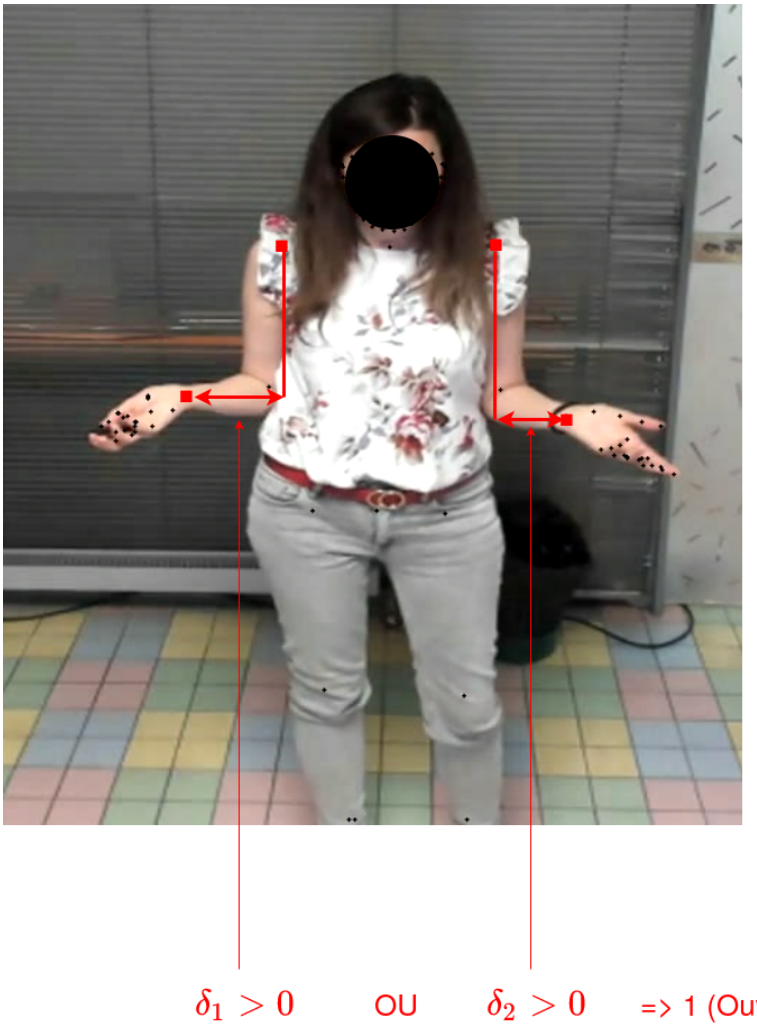

Figure 2.9. Détection de l'ouverture des bras. La distance δ_1 représente l'écart entre la coordonnée x du poignet droit et celle de l'épaule droite, δ_2 représente la même valeur mais à gauche et inversée (multipliée par -1).

3.3 Des symboles aux états

Des symboles de la section précédente, nous construisons un état pour chaque frame de la vidéo du participant.
Un état est une collection de symboles et est uniquement identifié par un entier comme présenté dans la figure 2.10. Alors, le comportement non-verbal du participant est modélisé par une séquence d'états discrets et du temps passé dans ces états comme décrit dans la figure 2.11.

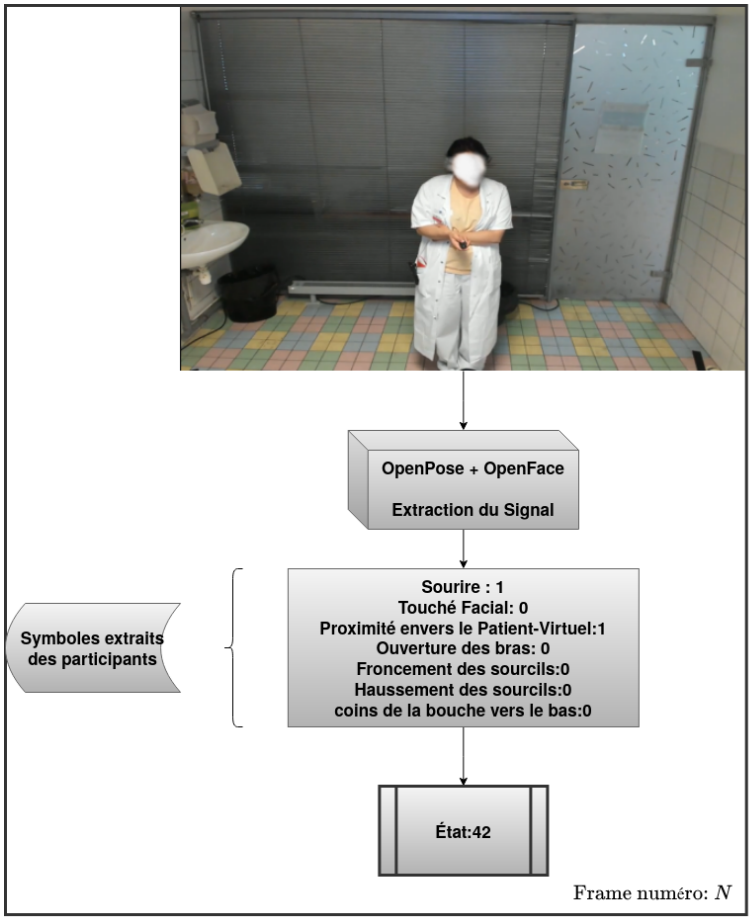

Figure 2.10. Capture d'écran anonymisée de la modélisation du comportement non-verbal d'un participant où l'on peut voir qu'il est loin, se touche la tête et ne sourit pas. Cet ensemble d'activation de symboles représente un état spécifique et est uniquement identifié par le numéro 42.

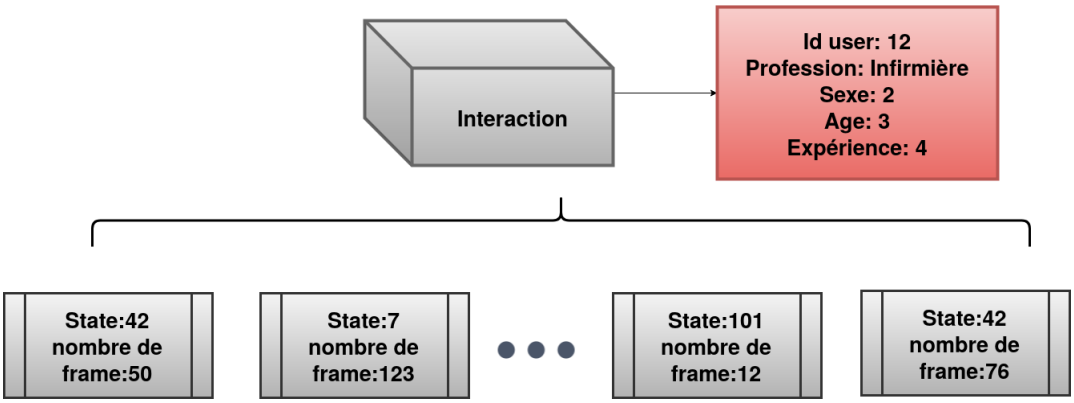

Figure 2.11. Modélisation du comportement d'un participant lors de l'interaction avec le Patient-Virtuel. Cette interaction est divisée en états qui durent plus ou moins longtemps. Chaque état est labélisé uniquement comme décrit dans la figure 2.10.

3.4 Des états aux N-grammes

Les N-grammes sont un outil simple mais efficace pour construire des caractéristiques à partir de données dans lesquelles une forme de temporalité est encodée.
Étant donné une suite de données, il s'agit d'en regrouper à chaque fois N de manière adjacente. Ceci est représenté dans la figure 2.12 où la suite de données est le comportement d'un participant modélisé par des états (voir figure 2.11).
Les N-grammes ont d'abord été pensés pour l'analyse de texte (Shannon, 1948), puis appliqués entre autre à la compression de texte (Nguyen et al., 2016) ou à l'identification de langue (Giwa and Davel, 2013). Ces applications transcendent aujourd'hui l'analyse de corpus de texte et peuvent être appliquées à la génomique [5] (Tomović et al., 2006), par exemple. Dans cette thèse, nous utilisons les N-grammes principalement pour l'étude du corpus dans le but de répondre aux objectifs $\mathcal{O}1$ et $\mathcal{O}2$.
Depuis chaque passation issu d'un corpus de participants nous pouvons alors construire un lexique de comportement grâce à la modélisation par N-grammes. Ce lexique de N-grammes nous permet d'utiliser des méthodes originalement issues du traitement du langage naturel.
Cependant, les N-grammes ont été développés pour le langage écrit ; nous serions donc en droit de questionner la pertinence de ce type d'outil pour l'étude du comportement non-verbal. D'une part, nous observons que le comportement non-verbal est souvent théorisé et modélisé comme un langage (Guerra Filho and Aloimonos, 2007; Cheng et al., 2015; Cong et al., 2011; Zhao and Li, 2011; Ren et al., 2015). De plus, les N-grammes couplés avec des méthodes de clustering ne dépendent pas du langage utilisé (Damashek, 1995a; Giwa and Davel, 2013; Aisopos et al., 2011), et donc ne requièrent aucune hypothèse sur ce même langage.
D'autre part, cette modélisation a déjà été employée avec succès sur du comportement non-verbal pour de la reconnaissance et prédiction d'actions (Thurau and Hlaváč, 2007; Takano et al., 2011). Dans leur travail, (Thurau and Hlaváč, 2007) concluent même que grâce aux N-grammes "[ils] *ont pu utiliser des méthodes conventionnelles d'analyse de séquence souvent appliquées au text mining et speech processing*".
Si ces deux études diffèrent dans la modélisation de l'action, elles reposent tout de même sur des extractions de caractéristiques très bas niveaux [6], l'utilisation de N-grammes et de méthodes de clustering (Ward (Ward, 1963) pour les deux). Leurs approches suit aussi le travail de (Saint-Georges et al., 2011), au sein duquel les interactions ont été analysées de la même manière : extraction de symboles, construction de N-grammes et clustering. La différence de cette dernière étude est que les symboles sont annotés manuellement.

5. Étude du génome Humain
6. *fine grain feature extraction*

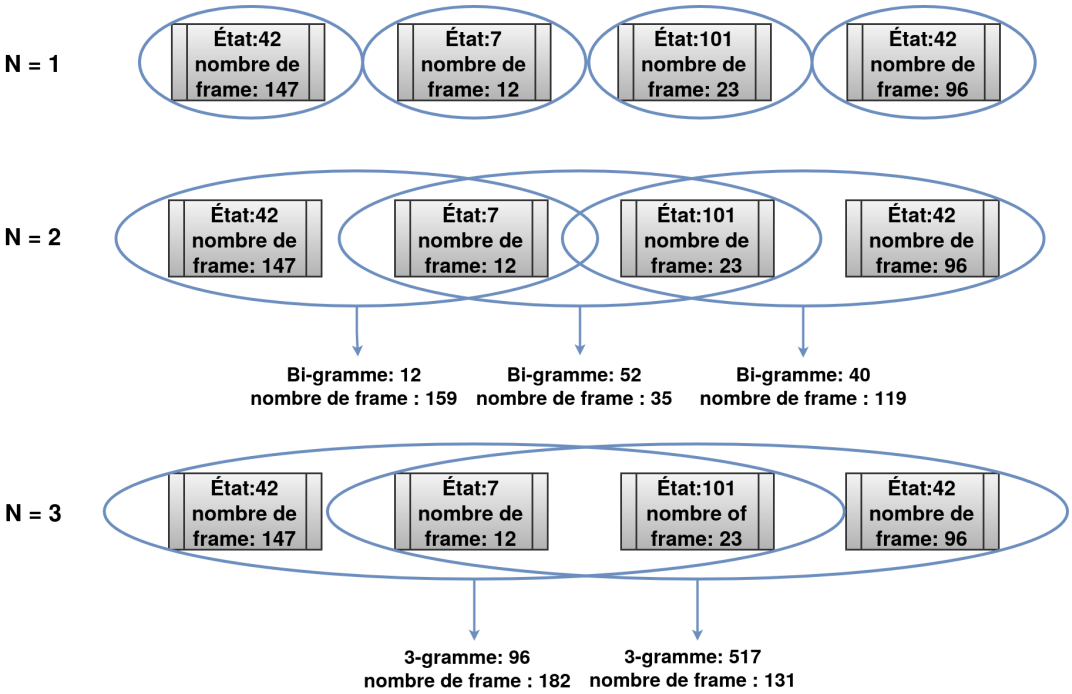

Figure 2.12. Construction de N-grammes sur des données séquentielles et pour un $N \in \{1, 2, 3\}$. Chaque N-gramme observé est labélisé par un entier unique de la même manière que l'est chaque état (voir figure 2.10).

Chapitre 3

Analyses des données de VirtuAlZ

Sommaire

1 Introduction

1.1 Rappel sur le corpus collecté

Le corpus VirtuAlZ, collecté au cours de l'été 2020 à l'hôpital Broca, à Paris, est très riche. Il contient :

- Les caractéristiques personnelles des soignants : âge, sexe, métier, expérience avec des patients Alzheimer.
- Les données multimodales (audio et vidéo) des soignants lors de leur passation.
- Les données comportementales des soignants (Action Unit et *landmarks*) extraites des vidéos.
- La progression des soignants dans le jeu sérieux (les choix pris, les différentes phases du scénario explorées, le temps passé dans chaque phase).
- Les questionnaires d'utilisabilité de VirtuAlZ remplis par les soignants.
- Le comportement du Patient-Virtuel (les Actions Units, la direction du regard).

Le protocole de collecte et de modélisation des données est détaillé dans le chapitre précédent.

1.2 La stratégie soutenant nos analyses

Nous disposons d'une quantité de données importante mais d'assez peu de participants (N=31). Ainsi, l'ensemble des analyses possiblement applicables doit être considéré avec rigueur et pragmatisme. L'ensemble des analyses faites sur ce corpus n'est pas présenté dans ce chapitre, seules le sont celles qui aboutissent à des informations pertinentes. Ce chapitre est surtout le reflet de l'article que nous avons publié (Zagdoun et al., 2021), avec quelques informations supplémentaires.
Nous rappelons que nous devons répondre à la problématique $\mathcal{Q}$ de l'introduction, à savoir qu'il nous faut un moyen de quantifier la qualité du comportement des soignants et de leur donner un retour automatisé. Or, comme nous l'avons démontré dans le chapitre 1, avant de l'avoir vu à l'œuvre, il est difficile pour un humain de juger le comportement d'un soignant avec un patient Alzheimer. Ainsi, faire des retours automatisés sur une bonne ou une mauvaise interaction avec le Patient-Virtuel est, dans l'état actuel des choses, impossible. Alors, la stratégie que nous avons adoptée est d'aider l'auto-évaluation du soignant. C'est dans ce but que nous développons des outils d'identification automatique du comportement des soignants et de clustering. Ces outils sont fortement inspirés du traitement du langage naturel et permettent alors de répondre aux objectifs $\mathcal{O}1$ et $\mathcal{O}2$.
Le but *in fine* étant de permettre au soignant qui utilise la plateforme VirtuAlZ de pouvoir aisément se comparer aux autres utilisateurs ou bien à lui-même, s'il a fait plusieurs séances.
Il pourra alors directement voir les différences et similitudes entre son comportement et ceux auxquels il se compare.
Nous avions aussi prévu de collecter des données de personnes entièrement novices au milieu médical pour faire une analyse comparative entre ces deux groupes. Les nouvelles passations de participants n'ont pas pu être réalisées à temps au vu du contexte actuel, car elles se font sous la supervision de l'hôpital Broca.

2 Présentation des outils

Dans le chapitre 2, nous développons une méthode de discrétisation du comportement non-verbal humain à l'aide de symboles, d'états, et de N-grammes. Les N-grammes permettent alors de quantifier le comportement non-verbal en plus d'incorporer des informations temporelles.
Cependant, la construction des états a un impact important sur la construction des N-grammes et sur leur interprétabilité.
En effet, la cardinalité de l'ensemble de tous les états possibles a une incidence polynomiale sur le nombre de N-grammes potentiellement constructibles. De plus, si nous augmentons le nombre de symboles que contient un état, son interprétation devient plus difficile et moins intuitive. C'est pour ces raisons que nous avons pré-selectionné des symboles d'intérêt (voir section 3.2) et que nous n'avons pas utilisé tous les symboles disponibles.
Cette section présente l'ensemble des outils d'analyse que nous déployons sur le corpus VirtuAlZ à partir de cette modélisation. Il s'agit alors d'un score de similarité basé sur des N-grammes (section 2.1.2), et de manière complémentaire, d'analyse de récurrence (section 2.2.2).

2.1 Score sur les N-grammes

La construction du score permet de répondre à la question : À quel point deux interactions sont-elles similaires ? Notre score représente alors une réponse réelle qui évolue entre 0 et 1.
Ce score peut ainsi se généraliser à des groupes d'interactions, et des études plus fines peuvent donc être faites.

2.1.1 Fixer la variable N du N-gramme

Dans le but de fixer la variable N du N-gramme, nous déployons une analyse comparative. Notre objectif est de fixer cette variable de manière à optimiser la diffusion de l'information dans notre modèle.
Si nous choisissons un N trop grand, nous observerons une répartition trop parcimonieuse [1] de l'information. Seule serait observée une toute petite quantité de N-grammes qui seraient, de surcroît, trop spécifiques à chaque interaction. Cependant, plus N est grand, plus on a d'informations temporelles intégrées dans notre modélisation.
Nous construisons alors 10 modélisations différentes de notre corpus avec 10 N-grammes différents : N variant de 1 à 10. Nous traçons ensuite la fréquence ordonnée et sa version "log-log" (c'est-à-dire en passant au logarithme sur chacun des axes) dans la figure 3.1.
Le nombre de N-grammes différents observés varie de 188 avec $N = 1$ à 26130 avec $N = 10$. Ainsi, et dans un soucis de lisibilité sur la figure de la fréquence normale, nous décidons de ne représenter que les 100 N-grammes les plus fréquents.

Nous utilisons alors pour le reste de toutes nos analyses des Bi-grammes ($N = 2$), c'est-à-dire la courbe jaune dans la figure 3.1. En effet, nous voyons dans la figure 3.1b une meilleure répartition globale sur les 50 N-grammes les plus fréquents qu'avec n'importe quel autre N.

1. "*sparse*" en anglais

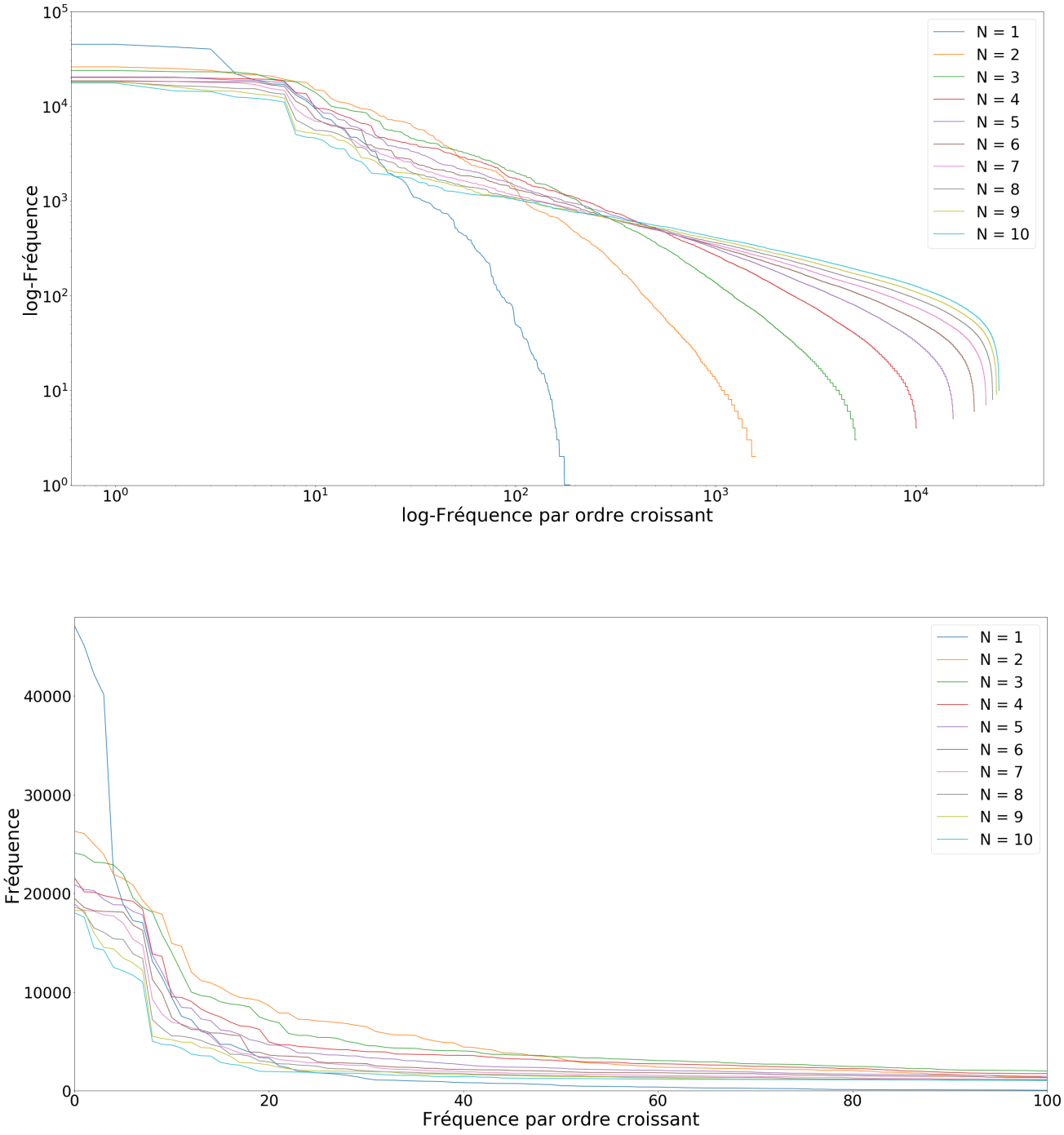

Figure 3.1. Visualisation des N-grammes. Représentation graphique de la fréquence ordonnée de chaque N-gramme dans notre corpus pour N variant de 1 à 10. Pour chaque N, les N-grammes sont ordonnés du plus fréquent au moins fréquent. La figure du dessus représente le graphique log-log (en passant au logarithme sur chaque axe).

Ceci est matérialisé par une courbe plus haute que toutes les autres sur les 50 premières valeurs. Après cela, pour les N-grammes ordonnés dont le rang est supérieur à 50, nous observons une convergence de l'ensemble des distributions. En résumé, en choisissant d'utiliser des Bi-grammes, nous observons une distribution moins asymétrique de l'information.
Enfin, nous observons empiriquement une distribution de Zipf, (Zipf, 1949), pour N allant jusqu'à 3. Cela indique une distribution similaire des N-grammes de notre corpus à ceux d'un corpus textuel et renforce, une fois de plus, l'utilisation des outils d'analyse de texte.

2.1.2 Score de similarité

À partir des Bi-grammes, nous construisons un score de similarité entre deux interactions de notre corpus, comme cela est fait dans (Damashek, 1995b), en définissant la distance cosinus décrite dans l'équation 3.1 :

$$Cos_{X_1,X_2} = \frac{\sum_{j=0}^{J}(X_{1j} - \mu_j)(X_{2j} - \mu_j)}{\sqrt{\sum_{j=0}^{J}(X_{1j} - \mu_j)^2 \sum_{j=0}^{J}(X_{2j} - \mu_j)^2}} \tag{3.1}$$

Où X_1 représente une interaction, X_{1j} le temps passé dans le Bi-grammes j dans l'interaction X_1 et μ_j le temps moyen passé dans le Bi-grammes j dans l'ensemble de notre corpus. Une matrice de similarité M est alors construite, où

$$M_{ij} = Cos_{X_i,X_j} \forall i, j \in [\![1, 29]\!]$$

2.2 Analyse Quantitative de la Récurrence

L'analyse quantitative de la récurrence est issue du monde de la physique, dans l'étude des systèmes dynamiques, et a été appliquée ensuite à l'analyse de comportements humains.
Un exemple d'application est celui du traitement du language naturel (Lira et al., 2018). Nous pouvons notamment nous intéresser au travaux de (Dubuisson Duplessis et al., 2021) qui listent une série de métriques liées à l'analyse quantitative de la récurrence et les appliquent pour, entre autre, améliorer l'alignement verbal d'un Agent-Virtuel en interaction avec un humain.

2.2.1 Exemple introductif : Plus Longue Sous-Suite Commune (LSSC)

La section précédente proposait un moyen de différencier et de grouper des interactions différentes. De cela, nous pouvons, par exemple, déduire quelle est la caractéristique personnelle la plus fréquente dans un groupe d'interactions. Ce type d'analyse est une analyse **macro**, car les informations extraites portent sur les interactions en elles-mêmes.
Une manière complémentaire de comparer des interactions est alors une analyse **micro**. Pour ce faire, nous cherchons quels sont les états, ou les symboles, les plus communs entre deux, ou plusieurs, interactions. Nous avons choisis d'utiliser pour cela l'algorithme de la plus Longue Sous-Suite Commune (LSSC).

LSSC est principalement utilisée pour :
- Détection de plagiat, de doublon ou de similarité dans les textes : (Elhadi and Al-Tobi, 2009), (Zhang et al., 2012), (Baba et al., 2017).
- Détection de génomes similaires ou compression d'ADN, (Beal et al., 2016), (Chowdhury, 2021), (Saha et al., 2018).
Cet algorithme est aussi utilisé dans la construction de l'analyse par quantification de récurrence de la section suivante, son utilité est donc multiple.

Comment construire LSSC

L'algorithme de la LSSC peut être implémenté de plusieurs manières, mais requiert tout de même quelques notions d'algorithmique pour éviter les débordements de mémoire et pour finir dans un temps acceptable. Sa construction repose sur de la programmation dynamique et, plus précisément, sur la mémoïsation, c'est-à-dire sur l'utilisation d'une matrice pour garder la trace des sous-états communs précédemment calculés ; voir Algorithme 1. En général, cette matrice est parcimonieuse et plus les sous-séquences communes sont présentes, moins il y a de zéros. Cela peut aussi nous donner une représentation visuelle sur la similarité entre deux interactions.

Algorithm 1 Algorithme de la plus longue sous-suite commune

$M = \{0\}^{l_1 \text{x} l_2}$; `// Matrice de zéros de taille` l_1xl_2
Max, LCSSS $= 0$, ""
for $x \in [\![1, 1 + l_1]\!]$ **do**
  for $y \in [\![1, 1 + l_1]\!]$ **do**
    if $s_1[x-1] == s_2[y-1]$ **then**
      $M[x][y] = M[x-1][y-1] + 1$
      if $M[x][y] > Max$ **then**
        Max, LCSSS $= M[x][y], s_1[x - M[x][y] : x]$
      end
      else
        $M[x][y] = 0$
      end
    end
  end
end
return LCSSS

Exemple introductif :

Soit deux suites de caractères : s_1="abbcd", s_2="effbcd", alors la plus longue sous-séquence commune est "bcd" et la matrice de similarité est la suivante :

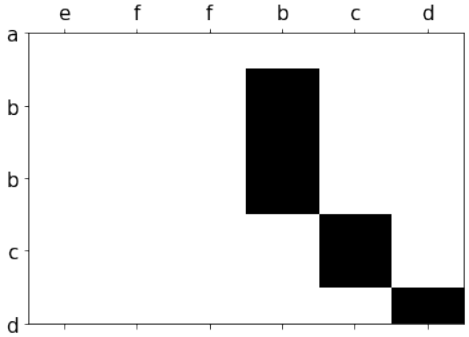

Figure 3.2. Matrice de memoïsation construite pour trouver la LSSC entre "abbcd" et "effbcd".

Dans le cas des données de VirtuAlZ, comme les interactions sont composées d'états qui sont uniquement identifiables par un entier, le même type d'analyse peut être construit et, par exemple, on peut afficher la matrice de similarité d'une interaction avec elle-même.
Cette matrice est alors très similaire à une analyse par récurrence et permet de voir apparaître des pattern de similarité entre deux interactions (voir figure 3.4), ou dans le cas de la figure 3.3, au sein d'une seule interaction.

Autre exemple d'application :

La matrice de mémoïsation est un outil visuel permettant de voir le modèle de récurrence, mais son utilité ne va pas plus loin. Cependant, dans sa construction, l'algorithme LSSC extrait toutes les sous-suites en commun entre nos interactions. De ces suites, plusieurs métriques peuvent être déduites pour permettre la quantification de la récurrence.
Par exemple, nous avons identifié deux clusters de 4 et 2 interactions chacun (voir section 4, le cluster bleu et rouge) et nous nous posons maintenant la question de savoir ce que ces clusters ont en commun et ce qui les différencie.
Nous pouvons, pour commencer, tracer les différentes fréquences d'activation des symboles de ces deux clusters comme dans la figure 3.5.

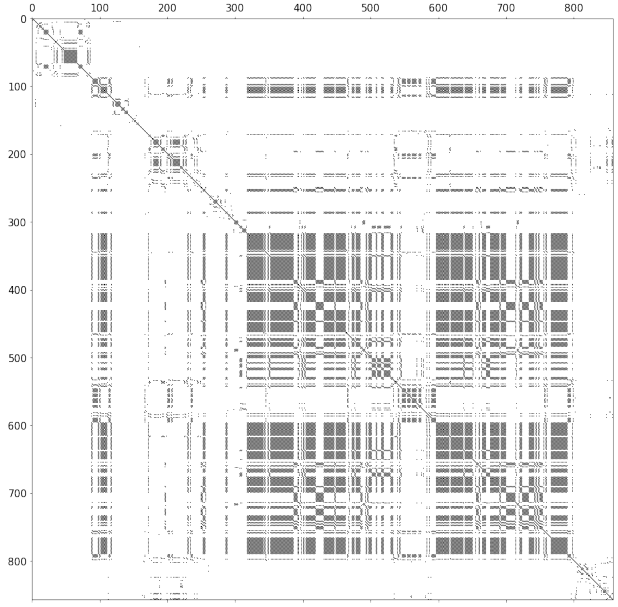

Figure 3.3. Matrice de memoïsation construite grâce au LSSC sur une interaction choisie aléatoirement au sein de notre corpus.

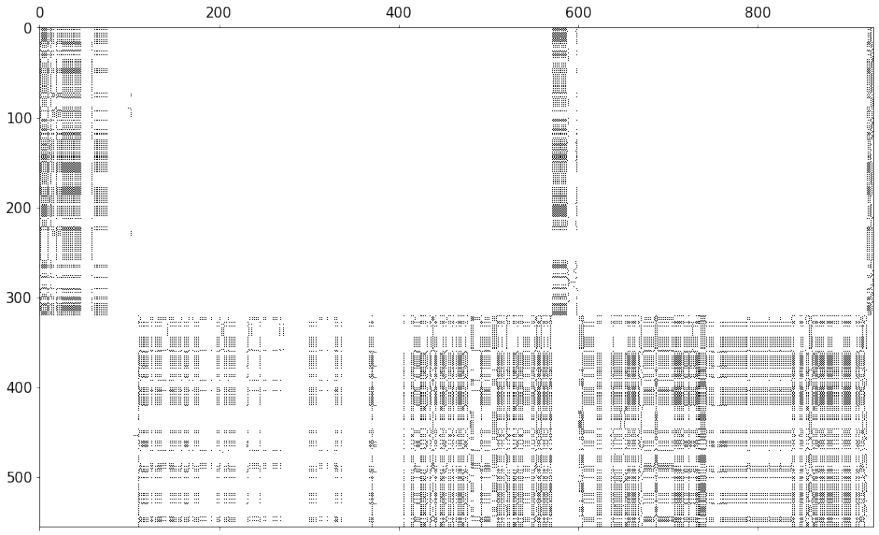

Figure 3.4. Matrice de memoïsation construite grâce au LSSC sur deux interactions choisies aléatoirement au sein de notre corpus.

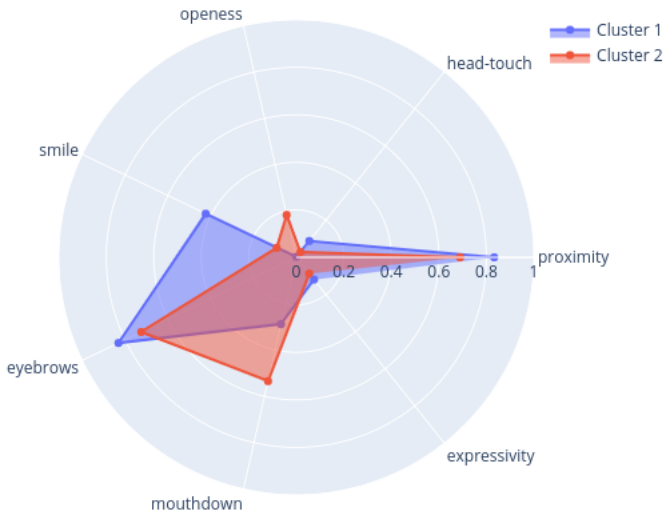

Figure 3.5. Diagramme radar sur la fréquence d'activation des différents symboles au sein de chaque cluster.

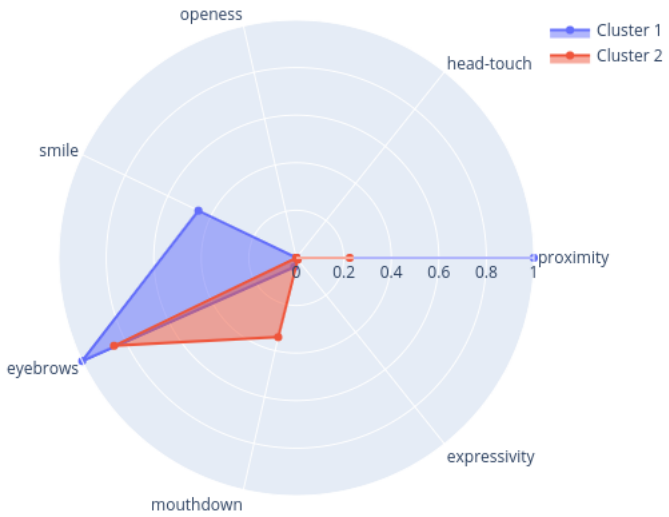

Figure 3.6. Diagramme radar sur la fréquence d'activation des différents symboles au sein de sous-suite commune de longueur d'au moins 7 au sein de chaque cluster.

La question se pose alors de savoir quelles différences sont significatives et lesquelles ne le sont pas. Grâce au LSSC, nous pouvons pousser la question un peu plus loin. Nous pouvons nous intéresser à la fréquence d'activation des symboles sur les suites d'une longueur supérieure à 7 (voir figure 3.6). Nous avons ainsi une version affinée de la figure 3.5.

La question à laquelle nous pouvons alors répondre est : parmi un groupe spécifique d'interaction et au sein des suites de comportements que ces interactions ont en commun, quels sont les symboles les plus fréquents ?
Nous voyons, par exemple, ici deux profils très distincts de comportement se dessiner. Dans le cluster 1, les gens sont proches, souriants, et haussent souvent les sourcils. Dans le cluster 2, on assiste à un comportement inverse : les soignants sont éloignés, les coins de leurs lèvres vont fréquemment vers le bas, et, eux aussi, haussent souvent les sourcils.

Enfin, nous pouvons aussi extraire les informations suivantes :
Sur le cluster 1 :

- La plus longue suite de symboles en commun est de longueur 14 et porte principalement sur des activations de sourire.
- 56 suites de symboles de longueur d'au moins 7 sont partagées par l'ensemble des interactions de ce cluster.

Sur le cluster 2 :

- La plus longue suite de symboles en commun est de longueur 15 et porte principalement sur des manifestations de coins des lèvres vers le bas.
- 45 suites de symboles de longueur d'au moins 7 sont partagées par l'ensemble des interactions de ce cluster.

Enfin, il est à noter que d'autres variantes à LSSC existent aussi.
La plus Longue Sous-Séquence Commune, cherche la plus longue séquence d'éléments en commun dans un texte. La différence entre suite et séquence est qu'une séquence peut être interrompue par des symboles qui ne rentrent pas dans la suite. Par exemple, dans le mot "abbcdghe", la séquence "abbde" est présente.
Enfin, une amélioration encore plus poussée serait d'utiliser des algorithmes de recherche de plus long pattern introduit dans (Kostakis and Papapetrou, 2015). Cet algorithme s'intéresse à la récurrence de symboles et non d'état. Il prend notamment en compte le temps d'activation chaque symbole. À partir de là, il permet de raccourcir une période d'activation d'un symbole donné, ou bien de l'omettre pour trouver des pattern en commun entre plusieurs interactions. Cependant, ce problème étant NP-complet nous l'évoquons juste en piste d'amélioration possible de notre analyse.

2.2.2 Récurrences

L'analyse par quantification de la récurrence se base sur la construction de matrices de mémoïsation présentées dans la section précédente, et réalisées grâce à l'algorithme 1. Comme nous avons pu le constater, cet algorithme est intéressant en lui-même, et nous pouvons déjà en sortir des informations pertinentes. C'est pour cela qu'il a été notamment appliqué à des domaines comme la génétique et le traitement de texte.
L'analyse par quantification de récurrence permet alors, en partant de notre algorithme 1, de quantifier les transitions de phase et le chaos au sein d'un système dynamique. Cette analyse

peut être bivariée ou univariée, comme présenté dans les figures 3.5 et 3.6 de la section précédente.
L'analyse par quantification de récurrence croisée (bivariée) est utilisée pour comparer deux séries temporelles ; elle peut être définie comme un équivalent non-linéaire de la corrélation croisée : elle quantifie la force, mais aussi la forme et la compléxité des dynamiques partagées entre deux systèmes (Fusaroli et al., 2014). L'avantage de cette analyse est alors qu'elle dépend grandement de la structure temporelle des données et qu'elle n'a pas besoin d'hypothèses sur les séries temporelles (stationarité, markov...).
Dans sa forme simple, elle offre une analyse univariée similaire. Son utilité peut être, par exemple, de voir si un participant a eu un comportement plus riche, plus complexe ou non que la moyenne.
L'analyse par quantification de récurrence approche le problème de l'analyse par l'introduction d'une série de métrique :

- Le taux de récurrence (RR) : quantifie la tendance qu'ont deux séries temporelles à visiter les mêmes états. Mesure la symétrie entre deux interactions.
- Le déterminisme (Det) : mesure à quel point une série temporelle peut être utilisée pour prédire l'autre.
- La longueur moyenne des lignes diagonales (L) : mesure la longueur moyenne des suites d'états en commun de deux interactions.
- La longueur maximale des lignes diagonales (Lmax) : mesure la plus grande période en commun entre deux interactions.
- L'entropie (ENT) : mesure la régularité de la durée des moments synchrones. Cette mesure peut être interprétée comme la richesse de l'espace d'état que deux interactions partagent : plus elles sont différentes et moins l'entropie est élevé (Mccamley et al., 2017).
- La laminarité (Lam) : quantifie la proportion qu'a une trajectoire de rester dans la même région.
- Temps de piégeage (TT) : temps moyen que les deux interactions passent dans la même région, dans un état spécifique.

Ces métriques sont, entre autres, sensibles à des paramètres tels que l'âge, le sexe, les modalités de l'interaction et sa difficulté (Fusaroli et al., 2014).

2.3 Pertinence et significativité des résultats

Le principal problème auquel est confronté l'ensemble des outils que nous avons développés est la significativité des résultats. En effet, nous analysons 29 interactions ; ce qui nous intéresse est alors de voir quels sous-groupes vont se distinguer et quelles sont les relations inter/intra sous-groupes. Il nous est dès lors impossible d'utiliser des tests statistiques de significativité dans ce contexte.
Un des avantages et inconvénients tout à la fois de l'analyse par quantification de récurrence de la section 2.2.2 est qu'elle n'est pas originellement pensée pour être couplée à des tests de significativité (même si certains utilisent quand même lesdits tests).
Cependant, nous allons tout de même adopter un esprit de rigueur et utiliser des méthodes dites d'*oversampling* dans l'idée de (Ramseyer and Tschacher, 2006) et de (Delaherche and Chetouani, 2010). L'idée est de mélanger les données pour en créer de nouvelles, factices, qui n'ont a priori pas de rapport entre-elles. Nous pouvons alors comparer les relations qui définissent un sous-groupe d'intérêt aux relations qui définissent nos données factices et voir si elles sont

significativement différentes.
Plus précisément, et dans le cadre de notre score de similarité de la section 2.1.2, cela est fait en créant des pseudo-interactions qui représentent de vraies interactions, mais dont l'ordre est mélangé. Ceci est fait 100 fois par couples considérés dans l'analyse et le score est jugé significativement différent s'il s'écarte au-delà de deux fois de l'écart-type. Cette méthode permet d'augmenter artificiellement le nombre de données disponibles tout en gardant une cohérence globale de distribution des états et des symboles.
En ce qui concerne les résultats de l'analyse par quantification de récurrence, nous procédons en groupant aléatoirement des interactions et en sortant les mêmes mesures. Nous faisons cela 1000 fois et moyennons les résultats pour éviter des tirages favorables qui arrangeraient ou non nos analyses. Ainsi, nous avons un groupe d'interaction témoin auquel nous pouvons nous comparer et qui a priori n'a aucune caractéristique commune.

3 Analyse du corpus en contexte

3.1 Phases

Une des premières informations à nettoyer sur nos données concerne les phases du scénario. En effet, comme décrit au chapitre 2, l'enregistrement des comportements se lance au moment où l'utilisateur démarre le jeu. Or, ce jeu est composé de plusieurs étapes dont certaines ne concernent pas l'interaction avec le Patient-Virtuel, notamment les phases de tutoriel, la lecture de la fiche patient, l'entrée dans la chambre et la sortie. Nous pensons que nous devons nous concentrer uniquement sur les phases d'interaction, et nous allons démontrer au cours de cette section que les phases d'interactions sont les plus riches, mais aussi étudier les répartitions de symboles au sein des différentes phases.
Une manière de décrire l'activité présente dans notre corpus en fonction des différentes phases de scénario est de compter le nombre de Bi-grammes observés et, aussi, de compter le nombre de Bi-grammes uniques. ceci est représenté dans le tableau 3.1. Nous listons aussi la fréquence d'activation de chaque signal par phase au sein de l'intégralité de notre corpus dans le tableau 3.2.

Tableau 3.1. Répartition des Bi-grammes en fonction des phases. La première colonne indique le numéro identifiant la phase de l'interaction, la deuxième le nombre de Bi-grammes différents au sein de chaque phase, la troisième représentant le nombre total de Bi-grammes observés dans chaque phase, et la dernière colonne la proportion.

Phase	Bi-grammes uniques	Nombre total de Bi-grammes	Proportion
0	159	776	0.205
1	286	2881	0.099
2	622	8516	0.073
3	497	6937	0.072
4	462	5456	0.085
5	382	2998	0.127
6	144	769	0.187

Tableau 3.2. Fréquence des symboles en fonction des phases. La première colonne indique le numéro identifiant la phase de l'interaction, les autres colonnes la fréquence d'activation du symbole durant cette phase (durée d'observation du symbole sur durée totale de la phase et ce pour toutes les interactions).

Phase	Proximité	Touché	Ouverture	Sourire	Hausser [2]	Bouche [3]	Froncer [4]
0	0.486	0.004	0.005	0.13	0.207	0.036	0.019
1	0.28	0.006	0.002	0.041	0.122	0.043	0.055
2	0.398	0.015	0.01	0.089	0.114	0.057	0.039
3	0.445	0.012	0.007	0.078	0.125	0.04	0.039
4	0.459	0.005	0.002	0.087	0.148	0.039	0.037
5	0.47	0.018	0.002	0.091	0.123	0.061	0.031
6	0.452	0.002	0.001	0.107	0.124	0.087	0.038

Du tableau 3.1, il ressort que la diversité lexicale est la plus forte lors des phases 2, 3 et 4, donc les phases d'interaction avec le Patient-Virtuel. On remarque aussi que se sont les phases qui contiennent le plus de Bi-grammes et dans lesquelles les soignants ont passé le plus de temps. Au niveau de l'activation des symboles, on constate une augmentation du taux de proximité en fonction de la phase : plus on avance dans le jeu, plus les soignants ont tendance à se rapprocher de l'écran.
Ainsi, et pour la totalité du reste des analyses, nous restreignons l'analyse comportementale des interactions aux phases 2, 3 et 4.

3.2 Contexte lié au soignant

Nous construisons grâce au score de similarité de l'équation 3.1 une matrice de similarité entre tous les participants de notre corpus. Cette matrice de similarité est construite grâce aux Bi-grammes. La moyenne et la variance de ce score sur chaque sous-groupe contextuel sont représentées dans la dernière colonne du tableau 3.3.

En terme de similarité, aucune caractéristique personnelle ne joue un rôle primordial dans le comportement du soignant. En effet, aucun des scores de similarité n'est assez haut ni n'a de variance assez basse pour permettre de conclure qu'une caractéristique personnelle est responsable d'un comportement global. Ceci signifie qu'un médecin peut se comporter comme un psychologue, un aide-soignant ou un infirmier lors de notre étude. C'est aussi vrai pour l'âge, le sexe et l'expérience avec les patients Alzheimer.
Cependant, un résultat intéressant qui ressort de notre expérience est le manque significatif de similarité dans les comportements des soignants inexpérimentés. En effet, les participants avec le plus faible degré d'expertise semblent tous se comporter de manière différente. Pour confirmer ce résultat, nous étudions ces sous-groupes grâce à l'analyse par quantification de récurrence croisée présenté à la section 2.2.2. Nous moyennons alors, pour chaque couple possible d'un sous-groupe contextuel, l'ensemble des métriques liés à la quantification de récurrence ; ceci est représenté dans le tableau 3.4. À titre de comparaison, nous extrayons et moyennons les mêmes métriques univariées, ceci est représenté dans le tableau 3.5

2. Haussement de sourcils.
3. coins des lèvres vers le bas.
4. Froncement de sourcils.

Tableau 3.3. Liste des scores de similarité intra-groupe concernant chaque caractéristique contextuelle de notre corpus. La similarité intra-groupe est calculée comme la moyenne du score de similarité de l'équation 3.1. La variance est affichée à côté entre parenthèses.

Contexte	Détail	Similarité Intra-groupe
Moyenne globale	None	0.25(0.15)
Profession	Aide soignant	0.33(0.15)
	Infirmier	0.26(0.15)
	Médecin	0.22(0.13)
	Psychologue	0.16(0.11)
Expérience	Moins de 5 ans d'exp.	0.07(0.06)**
	Entre 5 et 15 ans d'exp.	0.26(0.16)
	Plus de 15 ans d'exp.	0.26(0.17)
Age	25-34	0.29(0.16)
	35-44	0.24(0.16)
	45-54	0.21(0.12)
	55-65	0.23(0.17)

Il en ressort que le sous-groupe des personnes les moins expérimentées a le plus faible taux de récurrence, de déterminisme, d'entropie et de longueur moyenne d'états en commun. On voit aussi que ce n'est pas le cas si nous nous intéressons uniquement aux métriques univariées. Ceci signifie qu'en soi, ces comportements sont aussi riches et complexes que les autres si on les considère individuellement, c'est-à-dire avec l'analyse par récurrence. Cependant, si nous nous intéressons aux dynamiques de groupe, c'est-à-dire en utilisant l'analyse par récurrence croisée, ce n'est pas le cas. En effet, à l'intérieur du groupe des personnes non-expérimentées, nous observons des comportements plus divers, chacun des participants agissant de manière différente par rapport aux autres.

Notre conclusion est alors que ni l'âge, le sexe, l'expérience ou la profession n'ont d'impact global sur le comportement du soignant en interaction avec le Patient-Virtuel. Le manque d'expérience, lui, semble être un facteur de comportements plus divers et, peut-être, moins précis. Cependant, au vu de la faible quantité de données à notre disposition, ce résultat doit néanmoins être confirmé par d'autres études. Par exemple, la passation d'un groupe de personne qui n'a jamais eu d'expérience dans le milieu médical.

Enfin, et toujours à cause du faible échantillonnage de notre jeu de données, nous ne pouvons étudier les combinaisons de sous-groupes. Par exemple, nous intéresser au comportement de médecins jeunes et peu expérimentés ou à celui de femmes psychologues âgées.

Tableau 3.4. Étude en contexte des interactions grâce à l'analyse par quantification de récurrence croisée. On voit ici que les catégories interpersonnelles n'influent pas sur un comportement global. Du moins du point de vue de l'analyse par quantification de récurrence. Valeurs minimale en bleu dans chaque colonne et maximale en rouge. r(X) représente X interactions prises au hasard. Nous répétons cette opération 1000 fois et moyennons les métriques pour éviter d'avoir une "bonne pioche" qui arrangerait nos analyses.

Catégorie	**RR**	**DET**	**Lam**	**TT**	**Lmax**	**L**	**Ent**
Médecin	0.062(0.044)	0.981(0.008)	0.993(0.004)	19.509(5.771)	161.944(83.11)	10.809(2.619)	3.08(0.279)
Infirmier	0.065(0.037)	0.967(0.024)	0.985(0.018)	21.457(9.998)	161.75(109.238)	10.747(3.127)	3.061(0.281)
Psychologue	0.053(0.036)	0.971(0.016)	0.985(0.012)	20.633(10.273)	137.0(103.65)	9.956(3.397)	2.956(0.34)
Aide-Soignant	0.07(0.02)	0.981(0.006)	0.993(0.004)	21.659(7.43)	150.0(61.307)	10.252(1.736)	3.068(0.161)
25-34 ans	0.068(0.041)	0.97(0.014)	0.99(0.006)	16.676(4.55)	110.095(39.756)	8.996(2.018)	2.909(0.245)
35-44 ans	0.078(0.036)	0.975(0.019)	0.987(0.019)	24.597(9.497)	187.929(109.624)	12.057(3.51)	3.158(0.267)
45-54 ans	0.061(0.046)	0.978(0.013)	0.991(0.008)	21.777(7.668)	172.889(102.33)	10.745(3.022)	3.065(0.302)
55-65 ans	0.059(0.036)	0.979(0.007)	0.992(0.004)	22.662(5.929)	156.6(57.375)	10.978(2.101)	3.116(0.22)
moins de 5 ans d'exp	0.029(0.015)	0.969(0.017)	0.987(0.009)	15.237(5.216)	97.2(36.166)	8.876(2.381)	2.842(0.29))
entre 5 et 15 ans d'exp	0.078(0.043)	0.974(0.02)	0.987(0.017)	21.956(9.975)	180.179(100.195)	11.39(3.506)	3.108(0.294)
plus de 15 ans d'exp	0.066(0.039)	0.98(0.008)	0.992(0.005)	22.28(8.663)	149.509(64.256)	10.724(2.695)	3.095(0.257)
r(5)	0.066(0.016)	0.976(0.007)	0.99(0.004)	21.315(3.723)	160.598(39.938)	10.798(1.529)	3.063(0.144)
r(10)	0.066(0.009)	0.976(0.004)	0.99(0.002)	21.41(2.281)	160.774(22.214)	10.836(0.906)	3.067(0.085)

Tableau 3.5. Étude en contexte des interaction grâce à l'analyse par quantification de récurrence. Nous pouvons y lire le même profil de comportement au sein des interactions du corpus. Valeurs minimale en bleu dans chaque colonne et maximale en rouge. r(X) représente X interactions prises au hasard. Nous répétons cette opération 1000 fois et moyennons les métriques pour éviter les tirages favorables qui arrangeraient nos analyses.

Catégorie	RR	TT	L	Ent
Psychologue	0.15(0.059)	26.149(7.531)	14.107(3.708)	3.325(0.3)
Infirmier	0.176(0.096)	28.684(13.833)	15.367(7.033)	3.337(0.4)
Aide-soignant	0.145(0.03)	23.789(8.487)	12.888(4.295)	3.253(0.312)
Médecin	0.183(0.055)	26.93(5.573)	14.499(2.842)	3.393(0.18)
25-34 ans	0.151(0.037)	21.17(6.305)	11.607(3.187)	3.133(0.268)
35-44 ans	0.181(0.087)	31.432(12.365)	16.782(6.294)	3.456(0.353)
45-54 ans	0.173(0.071)	27.253(8.172)	14.647(4.089)	3.388(0.266)
55-65 ans	0.154(0.057)	25.309(5.647)	13.615(2.87)	3.325(0.202)
Moins de 5 ans d'exp.	0.145(0.036)	23.969(6.176)	13.047(3.066)	3.239(0.254)
Entre 5 et 15 ans d'exp.	0.175(0.082)	27.834(11.382)	14.945(5.792)	3.344(0.358)
Plus de 15 ans d'exp.	0.167(0.059)	26.344(8.186)	14.171(4.118)	3.367(0.256)
r(10)	0.167(0.001)	26.537(6.497)	14.291(1.663)	3.332(0.007)
r(5)	0.168(0.001)	26.822(17.33)	14.436(4.438)	3.338(0.018)

3.3 Analyse du Patient-Virtuel

Au niveau du comportement du Patient-Virtuel, nous ne prenons en compte que 28 interactions à cause de perte de données en début d'expérience. Les comportements du Patient-Virtuel pris en compte sont des comportements non-verbaux faciaux. Le langage du reste du corps n'étant pas encore programmé. Nous prenons alors en compte dans notre analyse l'intégralité des symboles faciaux à disposition, c'est-à-dire :

- Haussement de sourcils (AU1+AU2).
- Froncement de sourcils (AU4).
- Écarquiller les yeux (AU5).
- Sourire (AU12).
- Pincement des lèvres (AU23).
- Coins des lèvres vers le bas (AU15).
- Plisser les yeux (AU44).
- Direction du regard.

L'ensemble du comportement du Patient-Virtuel lors de ces 28 interactions totalise 110 Bi-grammes différents. Nous considérons ici directement des Bi-grammes pour ne pas trop nous éloigner des analyses faites sur les soignants. Nous voyons à quel point le non-verbal du Patient-Virtuel est pauvre comparé au non-verbal du soignant. En effet, même en restreignant le non-verbal du soignant aux 4 signaux faciaux que sont le sourire, coin des lèvres vers le bas, froncement de sourcils, haussement de sourcils, nous trouvons au total 202 Bi-grammes différents. Or, ici, le Patient-Virtuel a 8 symboles pris en compte, c'est à dire le double. Nous rappelons aussi que le nombre de symboles à un impact polynomial sur le nombre de Bi-grammes possibles.
Ceci n'est pas une surprise et s'explique aisément car le comportement du Patient-Virtuel est généré par le Magicien d'Oz après une sélection faite à la souris.

L'analyse du comportement du Patient-Virtuel s'arrête ici. L'application des outils développés dans la section 2 ne donnant rien de pertinent à cause de la faible diversité de comportement. Nous rappelons cependant que ces outils ont été développés pour l'analyse de comportement humain car c'est le but principal de ma thèse. L'analyse du comportement du Patient-Virtuel requiert d'autres types d'analyse qui font l'objet d'une thèse à part entière, soutenue par Amine Benamara. Nous avons analysé le comportement du Patient-Virtuel pour voir si nous pouvions détecter de la synchronie ou d'autre caractéristiques qui permettraient de mieux comprendre le comportement du soignant.

4 Analyse du corpus hors contexte

La question restante porte maintenant sur la différence dans le comportement des participants indépendamment de leur profession, âge, sexe ou expérience. Nous répondons à cette question en utilisant un algorithme de clustering en plus de notre score de similarité.
La plupart des travaux connexes ((Thurau and Hlaváč, 2007), (Takano et al., 2011)) utilisent le clustering de Ward (Ward, 1963). Cependant, cette méthode de clustering est généralement utilisée pour les études à plus grande échelle (n > 100). Nous utilisons plutôt le clustering hiérarchique (Bishop and Nasrabadi, 2007) qui peut directement être appliqué à notre matrice de similarité introduite dans la section 2.1.2. Les résultats sont illustrés dans la figure 3.7.
De ce clustering, nous identifions cinq groupes différents de participants qui sont représentés par les clusters rouges, bleus, roses, jaunes et noires sur le dendogramme en haut et à gauche de la figure 3.7. Parmi ces clusters, deux se distinguent par leur taille et leur uniformité : ce sont les clusters rouges et bleus. Ces deux groupes contiennent 42% des participants de notre expérience. Le cluster rouge se distingue notamment par le fait que trois des quatre soignants sont des femmes médecins et ont entre 45 et 54 ans. Les autres participants sont soit groupés par deux, soit se comportent de manière unique.
Avec ces deux groupes, nous pouvons maintenant analyser en détail les principales différences et similitudes entre eux. Pour ce faire, nous comptons l'activation moyenne des symboles à l'intérieur de chaque groupe, comme le montre la figure 3.8. Nous pouvons notamment voir que le principal facteur discriminant entre ces deux clusters est le symbole du froncement des sourcils et celui des sourcils relevés. Le froncement de sourcils est très présent dans le cluster bleu et représente moins de 80% du temps. Au contraire, les sourcils levés sont plus fréquents dans le cluster rouge. Nous constatons également que le comportement non-verbal du visage est beaucoup plus présent que le comportement non-verbal du corps (sauf pour la proximité).

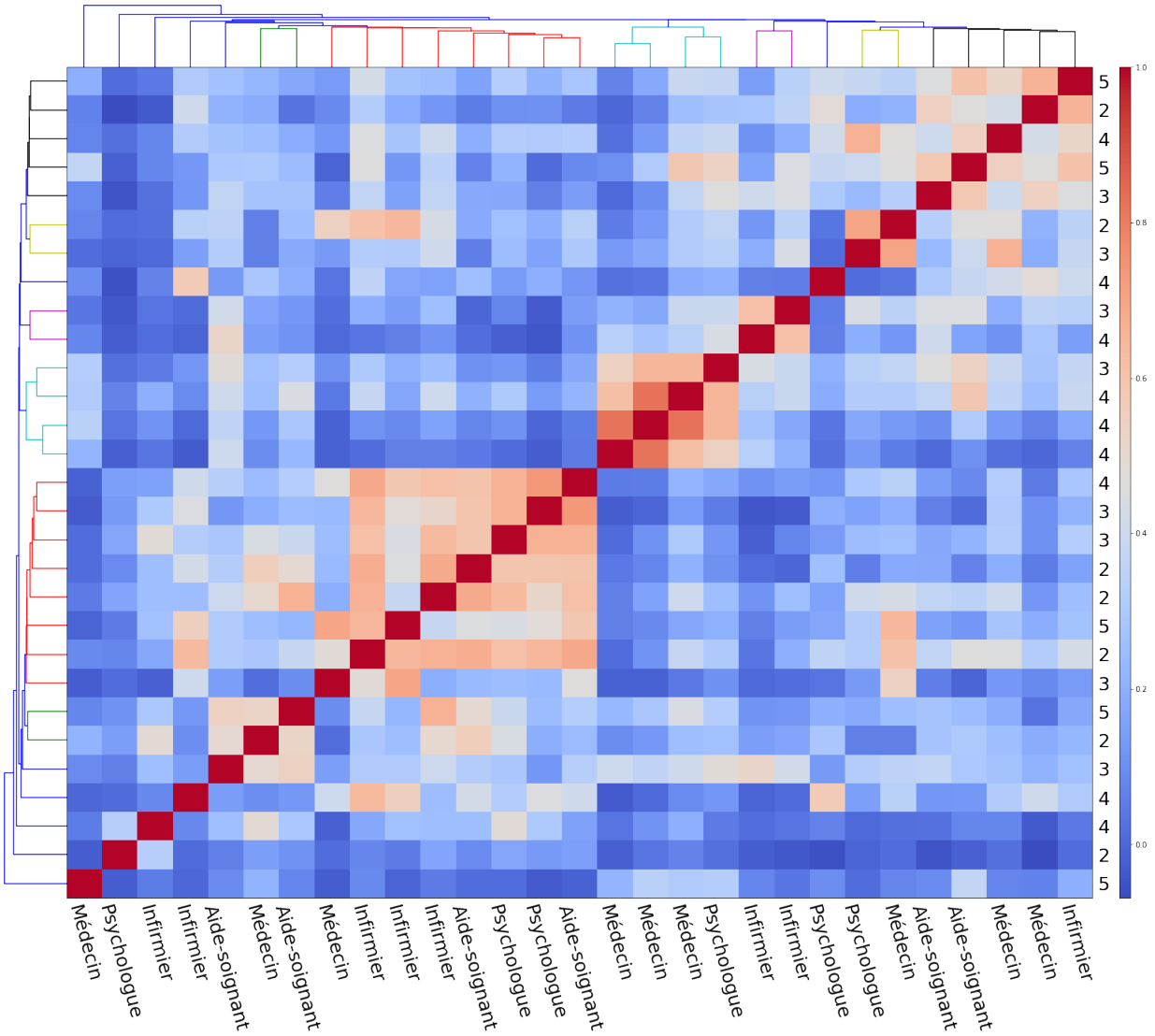

Figure 3.7. Matrice de similarité de notre corpus, sur le dessus et le côté de laquelle se trouvent les dendogrammes de l'algorithme de clustering hiérarchique. Les différentes couleurs représentent les différents clusters trouvés par cet algorithme. Les labels x représentent la profession et les labels y représentent la catégorie d'âge des participants.

Tableau 3.6. Étude en contexte des interactions grâce à la récurrence. On voit ici que les catégories interpersonnelles n'influent pas sur un comportement global. Du moins du point de vue de l'analyse par quantification de récurrence.

Catégorie	RR	TT	L	Ent
Cluster bleue	0.178(0.073)	25.654(3.77)	13.783(1.887)	3.394(0.167)
Cluster rose	0.253(0.116)	41.864(14.139)	22.125(7.211)	3.655(0.325)
Cluster jaune	0.221(0.043)	26.529(8.781)	14.164(4.377)	3.355(0.272)
Cluster rouge	0.135(0.035)	29.276(6.491)	15.689(3.254)	3.423(0.218)
Cluster noir	0.145(0.036)	21.781(5.762)	11.86(2.934)	3.2(0.25)
r(10)	0.167(0.001)	26.537(6.497)	14.291(1.663)	3.332(0.007)
r(5)	0.168(0.001)	26.822(17.33)	14.436(4.438)	3.338(0.018)

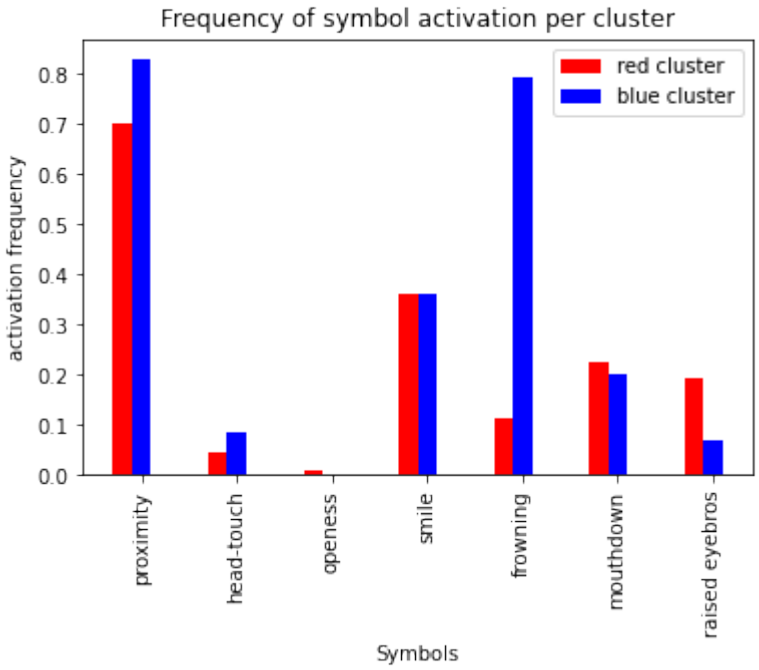

Figure 3.8. Fréquence d'activation des symboles parmi les deux clusters de la figure 3.7

Nous analysons maintenant ces clusters au moyen de l'analyse par récurrence comme cela a été fait dans la section 3.2. Les résultats sont reproduits dans le tableau 3.6 pour l'analyse par récurrence et dans le tableau 3.7 pour l'analyse par récurrence croisée.
Au niveau de l'analyse par récurrence croisée nous voyons que les valeurs de synchronie les plus importantes sont dans les clusters roses et jaunes. Cependant, ces clusters ne comportent que deux interactions chacun, ainsi il s'agit peut-être d'un artéfact dû au hasard ou d'interaction trop pauvre en terme de comportement. On voit notamment dans le cluster rose un Trapping Time, un L, un Lmax et une entropie élevés qui semblent en être le signe.
Les trois clusters restants sont alors le noir, le rouge et le bleu. Pour ces trois clusters, nous voyons des comportements assez similaires en terme de récurrence et, à l'inverse, des résultats de la section 3.2 indiquant des comportempents riches et similaires au sein de ces groupes d'interactions et donc confirmant le résultat du clustering.

Tableau 3.7. Étude en contexte des interactions grâce à la récurrence croisée. On voit ici que les catégories interpersonnelles n'influent pas sur un comportement global. Du moins du point de vue de l'analyse par quantification de récurrence.

Catégorie	**RR**	**DET**	**Lam**	**TT**	**Lmax**	**L**	**Ent**
Cluster bleue	0.131(0.044)	0.989(0.004)	0.995(0.003)	23.077(4.784)	257.0(93.718)	13.437(1.78)	3.38(0.153)
Cluster rose	0.16(N/A)	0.988(N/A)	0.994(N/A)	42.956(N/A)	480.0(N/A)	19.775(N/A)	3.643(N/A)
Cluster jaune	0.201(N/A)	0.983(N/A)	0.994(N/A)	18.581(N/A)	174.0(N/A)	12.93(N/A)	3.363(N/A)
Cluster noir	0.083(0.015)	0.987(0.003)	0.993(0.001)	23.815(3.995)	254.5(43.371)	13.733(1.838)	3.338(0.133)
Cluster rouge	0.109(0.022)	0.982(0.005)	0.995(0.003)	21.996(6.722)	169.19(59.806)	10.877(1.868)	3.155(0.181)
r(5)	0.066(0.009)	0.976(0.004)	0.99(0.002)	21.41(2.281)	160.774(22.214)	10.836(0.906)	3.067(0.085)
r(10)	0.066(0.016)	0.976(0.007)	0.99(0.004)	21.315(3.723)	160.598(39.938)	10.798(1.529)	3.063(0.144)

Conclusion

Nous avons lors de cette section présenté l'analyse d'une première collecte de données avec la mise en place de notre dispositif expérimental.
Ces données sont très riches et nous avons justifié le choix que nous avons fait de nous intéresser seulement au comportement du soignant lors des phases principales de l'interaction avec le Patient-Virtuel.
Notre analyse nous a permis d'identifier un groupe le participant le moins expérimenté et de voir qu'ils se comporte significativement tous de manière différente.
Nous avons aussi identifié grâce à des méthodes de clustering hiérarchique deux clusters de participant nous permettant d'avancer dans la direction d'une classification du comportement des soignants en types de profils différents.
Ces résultats ont été obtenus avec notre base de donnée de 29 participants et sont donc à confirmer avec d'autres passations.

Chapitre 4

Modélisation non-supervisée du comportement non-verbal

Sommaire

1 Introduction

Le chapitre précédent repose sur la modélisation du comportement non-verbal en mots. L'étude de la fréquence de ces mots permet de construire des métriques de similarités, et d'appliquer des outils d'analyses pertinents avec assez peu de données.
La détection de nouveauté comble alors un des problèmes majeurs de cette modélisation du comportement non-verbal. En effet, les symboles qui composent les mots sont prédéfinis et spécifiques, donc la modélisation est focalisée localement, sur les parties du corps concernées et leurs dynamiques.
C'est pourquoi nous avons opté pour donner un retour sur la dimension temporelle en plus de la dimension spatiale. Nous nous intéressons donc aux méthodes de **partition** du comportement non-verbal, c'est à dire de segmentation temporelle puis de classification des différents segments. La segmentation permet de découper le comportement en différentes actions et la classifications de ces segments permet de compter et de localiser les actions différentes. Cette partition permet alors de créer des mots de manière non-supervisée.
Nous pensons que cette approche est complémentaire par rapport à celle présentée dans la section 3 dans le sens où elle ne focalise pas l'attention du soignant sur une zone du corps en particulier mais sur des segments temporels.
Cette méthode est un prototype. Le but de ce chapitre est de prouver qu'elle peut être appliquée dans un cadre plus simple que la partition de comportements non-verbaux qui est celle de la partition d'actions. Pour ce faire, nous employons des réseaux de neurones siamois et les comparons avec des méthodes classiques de segmentation du comportement qui ne sont pas applicables dans notre modélisation, mais qui permettent tout de même la partition d'action lorsque des labels sont disponibles.

1.1 État de la recherche sur l'analyse de nouveautés

Comme nous l'avons présenté dans le chapitre 2, notre comportement non-verbal peut être considéré comme l'émanation d'un langage (Guerra Filho and Aloimonos, 2007). De cette approche, beaucoup de travaux, parmi lesquels ceux de (Cheng et al., 2015), (Cong et al., 2011) et de (Zhao and Li, 2011), ont utilisé des algorithmes de détection de nouveautés pour modéliser le comportement humain. Ce que nous appelons détection de nouveautés peut être trouvé dans la littérature comme détection d'anomalies, de données aberrantes, de bruit, d'exceptions. La détection de nouveautés peut être définie comme l'identification d'un pattern non-observé dans les données (Chalapathy and Chawla, 2019). Elle est notamment appliquée pour :

- Détecter les fraudes.
- Détecter les problèmes de santé.
- Détecter les erreurs dans un texte.
- Détecter les attaques dans un serveur informatique.
- Détection d'intrusion.
- Filtrer le spam.
- Contrôler la qualité de produits.
- Maintenance préventive.

De plus, la détection de nouveautés est une étape déjà empruntée dans la modélisation du comportement humain. On peut aussi se réferer à (Ionescu et al., 2019),(Kiran et al., 2018) et (Duman and Erdem, 2019) où il s'agit de détecter des comportements inhabituels dans des dispositifs de vidéosurveillance ou encore à (Kadri et al., 2008), (Fergani et al., 2008) et (Hogg et al., 2019) qui détectent des tours de parole dans des conversations audio.
D'autres travaux ont été faits sur la reconnaissance de comportement humain dans des vidéos (Dwibedi et al., 2019), (Revaud et al., 2013), mais ils requièrent des bases de données massives et ces algorithmes sont trop spécifiques à la base de donnée d'apprentissage.

Tous les travaux cités s'intéressent, plus ou moins directement, à la segmentation temporelle de comportement humain. Ils adressent aussi une problématique commune, à savoir : qu'est-ce qui définit une action ?
Par opposition à la reconnaissance/classification d'actions où une série temporelle correspond à une action, dans la partition/segmentation une série temporelle correspond à d actions avec $d > 2$.
Des modèles spécifiques existent aussi pour segmenter des actions, par exemple (Cheema et al., 2019) et (Hosseini et al., 2019) construisent des réseaux de neurones supervisés sur des données de motion capture. Cependant, ces méthodes requièrent des bases de données importantes pour l'apprentissage et dans ces bases de données, chaque point de chaque série temporelle doit être labellisé, ce qui n'est pas notre cas. Le travail le plus proche du nôtre est celui de (Zhou et al., 2008), (Zhou et al., 2013) : il s'agit de segmentation non-supervisée grâce à des algorithmes de clustering.
Ces derniers travaux apparaissent souvent au cours de ce chapitre. C'est de là que proviennent les algorithmes auxquels nous comparons les réseaux de neurones siamois sur les tâches de partition.

1.2 Les réseaux de neurones siamois

L'entraînement de réseaux de neurones profonds a la réputation d'être une tâche difficile et gourmande en données. Les déviations des tâches supervisées classiques et les nouvelles architectures qui en découlent, telles que Word2Vec, Bert et Gan, ont rendu cet apprentissage encore plus difficile. En effet, le but principal de ces réseaux n'est pas de différencier certains attributs des données mais d'apprendre à représenter les données, à les résumer ou même à les recréer. Par conséquent, ces tâches nécessitent une meilleure connaissance des données. Une autre application populaire de réseaux de neurones profonds capables de généraliser et d'apprendre une représentation de haut niveau des données est celle des réseaux de neurones siamois (Bromley et al., 1993).

Les réseaux de neurones siamois sont souvent associés à des tâches plus complexes que la classification. Par exemple, ils peuvent être appliqués à des tâches telles que la reconnaissance faciale (Schroff et al., 2015), surmonter le déséquilibrage des données (Bedi et al., 2020) ou d'autres applications semi-supervisées (Sahito et al., 2019). En raison de cette capacité de généralisation, les réseaux de neurones siamois sont notoirement connus pour être difficiles à entraîner, et de nombreuses méthodes d'optimisation ont été développées. Certaines se concentrent sur la création/modification des pertes, le *Metrique Learning* ou le choix a posteriori des données sur lesquelles s'entraîner. Dans cette section, nous proposons un point de vue différent et complémentaire sur ce que nous pensons être la principale raison pour laquelle les réseaux de neurones

siamois sont difficiles à entraîner. Le principal problème qui ressort de notre analyse est que, dans leur apprentissage, les réseaux de neurones siamois ont un problème de déséquilibre de classe sur les paires d'entraînement. Après avoir théoriquement analysé ce problème, nous présentons une manière très simple et efficace d'améliorer les réseaux de neurones siamois en utilisant le *Curriculum Learning* (Bengio et al., 2009). Avec notre modèle amélioré, nous effectuons une analyse basée sur deux tâches connexes :

- Classification, faite sur la base de données NTU-rgbd (Liu et al., 2019).
- Partition d'actions, faite sur la base de données IXmas (Weinland et al., 2006) avec les modèles issues de la tâche de classification précédente.

Nous nous basons sur ces deux tâches et différentes métriques afin de confirmer l'amélioration du modèle à l'aide du *Curriculum Learning*. Ceci est de la plus haute importance puisque ce que nous voulons voir, c'est l'équilibre de notre apprentissage avec notre modèle amélioré. Notre travail dans cette section consiste à :

(i) Identifier et répondre à la difficulté d'apprentissage au sein du réseaux de neurones siamois.

(ii) Mettre en pratique le pouvoir de généralisation d'une telle architecture avec une application sur la partition de séries temporelles.

(iii) Vérifier que nous pouvons concurrencer les algorithmes de l'état de l'art.

Les deux premières tâches sont dépendantes et nous observons par la suite que l'identification et la résolution des difficultés d'apprentissage dans le réseau de neurones siamois conduisent à de meilleurs résultats dans la tâche de segmentation des séries temporelles. Ces tâches sont une étape nécessaire au bon apprentissage d'un réseau de neurones siamois et pour s'assurer de sa capacité à généraliser.
La dernière tâche est, elle, liée à la création de mots de manière non-supervisée.

Présentation des différentes parties
Ce chapitre est organisé comme suit : la section 3 formalise le déséquilibre au sein du réseaux de neurones siamois et présente les différentes méthodes d'optimisation afin de le résoudre. Ensuite, nous présentons la procédure expérimentale dans la section 4. Enfin, nous présentons les résultats de notre algorithme sur les deux tâches liées à la compréhension du comportement humain dans la section 5. Mais d'abord, nous présentons une vue d'ensemble du réseaux de neurones siamois dans la section 2.

2 Contexte

2.1 Présentation des réseaux de neurones siamois

Les réseaux de neurones siamois sont des réseaux de neurones qui sont entraînés à identifier si une paire de points de données possède ou non le même label. Un exemple de réseau de neurones siamois est représenté dans la figure 4.1.

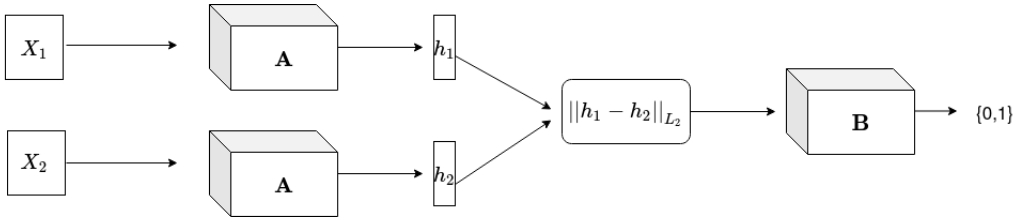

Figure 4.1. Architecture d'un réseaux de neurones siamois. Il prend en entrée deux matrices X_1 et X_2 représentant deux points de données et prédit si elles ont le même label ou non.

Où X_1 et X_2 sont deux points de notre jeu de données, $\mathcal{A}$ est un réseau de neurones qui leur est appliqué, et, h_1 et h_2 sont les sorties respectives du réseau de neurones. Une distance entre ces deux sorties est ensuite calculée et envoyée dans un autre réseau de neurones pour prédire si nos deux points ont le même label. Il est à noter que certaines différences peuvent apparaître dans la littérature, notamment au niveau de l'architecture des réseaux, des pertes et du nombre d'exemples qui sont en entrée du réseau.
Notre but est donc d'apprendre la fonction suivante :

$$F : \mathbb{R}^{d_f} \mathrm{x} \mathbb{R}^{d_f} \longrightarrow \{0, 1\}$$
$$(X_i, X_j) \longmapsto F(X_i, X_j) = \begin{cases} 0 & \text{if } C(X_i) = C(X_j) \\ 1 & \text{if } C(X_i) \neq C(X_j) \end{cases}$$

Où d_f est la dimension de l'espace de nos points de données, $C()$ est la fonction classe, c'est-à-dire la fonction qui fait correspondre un point de données à son label. Si X_i et X_j représentent deux points partageant le même label, alors $F(X_i, X_j) = 0$, et inversement pour des labels dissimilaires.

2.2 Les réseaux siamois dans la littérature

Les réseaux de neurones siamois (Bromley et al., 1993) offrent un compromis entre la puissance d'apprentissage d'un réseau de neurones et la capacité de généralisation des tâches de détection d'anomalies. De ce fait, il est courant d'utiliser des bases de données supplémentaires pour mesurer la qualité de l'apprentissage de ces réseaux (Berlemont et al., 2015). Elles permettent de le tester sur des données et des labels jamais vus, donnant ainsi une métrique sur la capacité de généralisation qu'a notre modèle. On appelle alors ces données des données "*out of distribution (OoD)*".
Les applications des réseaux de neurones siamois sont nombreuses et vont de la reconnaissance

faciale (Hermans et al., 2017) à la détection de changement dans des images satellites (Faiz Ur Rahman et al., 2018), ou encore, à la détection de cerveaux en mauvaise santé grâce aux données IRM de cerveaux en bonne santé (Alaverdyan et al., 2020). Ces travaux utilisent des réseaux de neurones siamois dans l'optique de résoudre un problème de détection de nouveauté. Récemment, on assiste aussi à une augmentation de l'utilisation de réseaux siamois sur des tâches de *Transfer Learning* (Chalapathy and Chawla, 2019), (Chen et al., 2020).
Dans un contexte plus proche de notre problématique, les réseaux de neurones siamois ont été appliqués dans le but de construire une représentation non-supervisée d'actions humaines (Berlemont et al., 2015). Cette étude utilise K-means (Bishop and Nasrabadi, 2007) sur les représentations apprises de ces réseaux de neurones siamois dans le but de classifier les actions. Grâce à cette méthode, ils peuvent aussi introduire un niveau de rejet pour dire si une action est trop différente de celles précédemment apprises.
Bien que le problème d'équilibrage de classe dans les réseaux de neurones siamois ne soit pas nouveau (Berlemont et al., 2017), à notre connaissance, il n'a jamais été formalisé. Ainsi, la plupart des améliorations faites le sont dans la direction de l'architecture du réseau (LSTM, transformers...), la création d'une nouvelle perte (Masana et al., 2018) et/ou en modifiant la taille des entrées du réseau (triplets (Hoffer and Ailon, 2015), quadruplets (Chen et al., 2017), ou même un ensemble représentant toutes les classes possibles (Berlemont et al., 2017)).
Au vu de la multiplicité des architectures, acheminement des données et pertes [1] possibles, identifier clairement le problème de déséquilibre de classe peut être noyé dans le nombre de variables que nous devons prendre en compte. Ainsi, nous comparerons principalement les améliorations sur quelques variables présélectionnées. Enfin, nous mettons de côté les pertes contrastives et triples dues aux faibles performances de ces dernières lors de leur apprentissage dans les mêmes conditions que nos réseaux de neurones. En effet, ces réseaux n'ont jamais réussi à concurrencer l'aléa lors de la phase d'apprentissage ou de test et ce dans les mêmes conditions expérimentales décrites lors de la section 4.4.1.
Pour l'architecture générale du réseau de neurones, nous utilisons des LSTM bi-directionnels (Schuster and Paliwal, 1997) avec un processus d'attention (Vaswani et al., 2017).

3 Déséquilibre dans les réseaux de neurones siamois

3.1 D'un point de vue théorique

Les réseaux de neurones siamois diffèrent des méthodes supervisées par de nombreux points. L'un en particulier, réside dans le déséquilibre de classe, ou plutôt de paires.
De manière générale, dans la plupart des tâches supervisées, nous extrayons des batchs de données pour que notre modèle apprenne dessus ; si notre base de données est équilibrée, alors les batchs seront équilibrés (notamment grâce à la loi des grands nombres (Bernoulli, 1713)).
Avec les réseaux de neurones siamois, ceci n'est plus vrai car ils apprennent avec des paires de données similaires et dissimilaires (voir section 2.1). Et, quelle que soit la construction de paires réalisée, la proportion est forcément très inégale (si le nombre de classes de la base de données est strictement supérieure à deux).
Formellement, soit B un batch de taille N contenant des variables aléatoires appartenant à d classes, i.e. $\forall X \in B \;\; \exists i \in [\![1, d]\!] \; s.t. \; X \sim \mathcal{L}^{(i)}$ où $\mathcal{L}^{(i)}$ est une loi de probabilité.

1. *loss* en anglais

Sans perte de généralité, nous considérons qu'un batch est parfaitement équilibré entre d classes, ceci permettant une analyse rigoureuse et plus lisible. De plus, un batch équilibré est une représentation fidèle d'un tirage moyen de données à partir d'une base de données équilibrée (la loi des grands nombres, encore une fois (Bernoulli, 1713))
Nous devons maintenant construire des paires d'exemples de données similaires et dissimilaires, laissant la question suivante :

Combien de couples similaires et dissimilaires sont constructibles dans un batch à partir de ces d *classes ?*

Soit #P, #S^i, #S, #D, représentant respectivement le nombre total de paires, le nombre de paires similaires de classe $i \in [\![1, d]\!]$, le nombre total de paires similaires, le nombre total de paires dissimilaires.
Comme notre batch est parfaitement équilibré, nous pouvons d'ores et déjà déduire que #S$= d$#S^i pour tout i.
Nous voyons aussi que #D=#P$-$#S car une paire est soit similaire soit dissimilaire.
Nous pouvons maintenant en déduire les valeurs suivantes :

$$\begin{aligned}
\#P &= \binom{N}{2} = \frac{N(N-1)}{2} \\
\#\mathrm{S}^i &= \binom{\frac{N}{d}}{2} = \frac{(\frac{N^2}{d} - N)}{2d} \\
\#\mathrm{S} &= \frac{(N^2 - Nd)}{2d} = \frac{N^2}{2d} - \frac{N}{2} \\
\#D &= \frac{N(N-1)}{2} - \frac{N(N-d)}{2d} = \frac{N^2 d - Nd - N^2 + Nd}{2d} = \frac{N^2(d-1)}{2d}
\end{aligned}$$

Nous remarquons ensuite que, quelle que soit la valeur de d, nous devons avoir $N > d$ puisque nous considérons B comme parfaitement équilibré. Ceci implique aussi que :

$$\exists n \in \mathbb{N}^* \text{ s.t. } N = nd \tag{4.1}$$

De plus, pour les paires dissimilaires, le fait que $\#D = \frac{N^2(d-1)}{2d}$ nous informe que si d est assez grand, il n'a pas d'impact sur le nombre de paires dissimilaires possiblement constructibles car $\frac{d-1}{d} \underset{d\to\infty}{\sim} 1$. Et, ce nombre est environ $\frac{N^2}{2}$.
Tout de suite, nous pouvons voir que quand le nombre de classes est assez grand, la différence entre le nombre de paires similaires et dissimilaires possiblement constructibles est quadratique. Pour voir cela plus clairement, nous utilisons le fait que B est un batch parfaitement équilibré, et notamment l'équation 4.1. Ce qui nous donne :

$$\begin{aligned}
\#\mathrm{S} &= \frac{(nd)^2}{2d} - \frac{nd}{2} < \frac{n^2 d}{2} \\
\#D &= \frac{n^2 d(d-1)}{2} \approx \frac{n^2 d^2}{2}
\end{aligned}$$

Nous observons nettement que $\frac{n^2 d}{2} << \frac{n^2 d^2}{2}$ quand nous prenons comme variable n, d et donc #S $<<$ #D.

Pour illustrer cette différence, nous prenons un exemple concret, qui correspond notamment au données décrites dans la section 4.1.1.
Si nous considérons un batch de taille 250 et que nous avons 50 classes différentes, nous pouvons alors, à partir de ce batch, construire 500 paires similaires et 30625 paires non-similaires.

De manière générale, il existe typiquement deux des stratégies pour faciliter l'entraînement (More, 2016), soit nous utilisons toutes les données disponibles et nous mettons des poids sur les pertes quand nous optimisons notre réseau, soit nous mettons une partie des données de côté [2]. Nous utilisons cette dernière méthode couplée avec du *Curriculum Learning* dans le but d'améliorer l'apprentissage des réseaux de neurones siamois.
Nous avons présenté une analyse formelle de la construction de paires à partir de batch pour des réseaux siamois. Cette analyse reste valide pour les triplets utilisés avec la *triplet loss* également. Ceci est dû au fait que, dans la *triplet loss*, il doit y avoir forcément un exemple positif et un exemple négatif.

3.2 D'un point de vue pratique

Nous présentons dans cette section, comment la *triplet loss* et notre méthode utilisant du *Curriculum Learning* fonctionnent. Nous mettons aussi en avant quelques points sur lesquels nous pensons que notre méthode est plus avantageuse que la *triplet loss*.

3.2.1 *Triplet loss*

La *triplet loss* fonctionne comme décrit dans (Schroff et al., 2015) : on sélectionne un exemple appelé "ancre" parmi nos données, auquel on associe un exemple dit positif (qui a le même label) et un exemple négatif (qui a un label différent). Le but de cette perte est alors de rapprocher l'ancre de l'exemple positif, et, en même temps, d'éloigner l'ancre de l'exemple négatif.
Il existe alors deux stratégies d'apprentissage : On peut entraîner notre réseau sur tous les triplets possiblement constructibles ou sur les triplets qui ont le plus mal performé.
Cette dernière stratégie apporte son lot d'avantages, cependant l'optimisation est faite a posteriori, c'est-à-dire une fois que les pertes sont calculées, ce qui nécessite que les données soient évaluées par le réseau. En pratique, ceci veut dire que cette stratégie d'optimisation est réalisée sur le GPU où la mémoire est très rare, et se traduit par des batchs de petite taille et un entraînement difficile.

3.2.2 *Curriculum Learning*

Une création de batch typique pour les réseaux de neurones siamois avec une perte classique de type softmax utilise de l'*undersampling*. Il s'agit de sélectionner aléatoirement autant de paires dissimilaires que l'on peut construire des paires similaires.
Du problème d'équilibrage de paires identifié dans la section 3.1, notre première intuition est de choisir correctement les paires dissimilaires sur lesquelles notre réseau de neurones siamois

2. *Undersampling*

s'entraînera. Cette procédure se rapproche de ce que l'on appele le *Curriculum Learning* (Bengio et al., 2009).
Cette stratégie s'applique à l'étape *sampling M couples* de la figure 4.2. En effet, une fois que notre réseau a appris une bonne représentation de nos données, c'est-à-dire à partir d'une certaine *epoch* de l'apprentissage, nous utilisons la base de données de validation pour créer des poids qui aideront ensuite à la sélection de couples lors de l'entraînement.

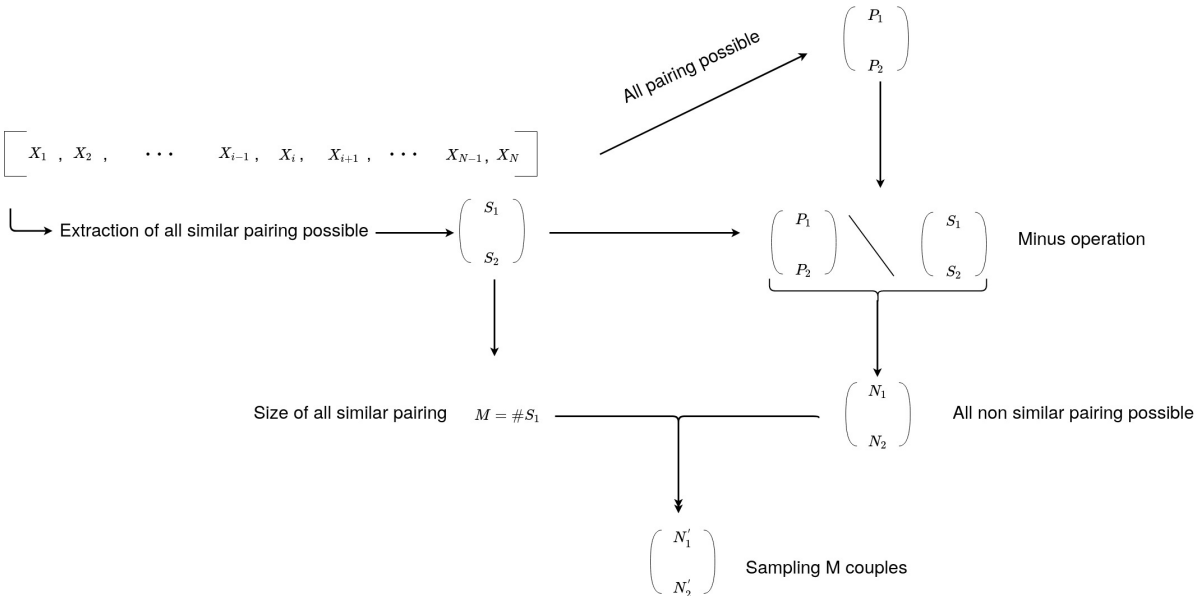

Figure 4.2. Construction de paires similaires et dissimilaires à partir d'un batch de données. À chaque points de données X_i dans le batch correspond un label. Ainsi, nous pouvons créer toutes les paires similaires et dissimilaires possibles. De l'ensemble des paires dissimilaires, nous sélectionnons aléatoirement M paires, où M est le nombre total de paires similaires. Nous concaténons et mélangeons ensuite les deux ensembles de paires pour obtenir un batch d'apprentissage. La stratégie de *Curriculum Learning* s'inscrit alors dans l'étape : "*sampling M couples*", le tirage aléatoire n'est alors plus uniforme mais prend en compte les difficultés de prédiction sur la base de données de validation.

La base de données de validation est alors traitée comme décrit dans l'algorithme 2 et comme suit :
Les labels des paires dissimilaires évaluées sont enregistrés en même temps que les résultats de la prédiction sur ces paires : prédiction correcte ou incorrecte pour chaque paire. De ces valeurs, nous construisons pour chaque paires dissimilaires une distribution de probabilité, basée sur leur fréquence de prédiction incorrecte. Si une paire de label a une très faible *accuracy* durant l'*epoch* de validation, alors elle aura une forte probabilité d'être sélectionnée durant l'*epoch* d'entraînement. L'utilisation de cette distribution durant l'*epoch* d'entraînement est représentée dans l'algorithme 3.
Ce faisant, notre méthode de *curriculum leanring* peut être réalisée sur le CPU. De plus, l'itération et l'énumération de tous les couples possibles sont faites sur les labels et les index des points de données. Ceci permet une utilisation très faible de la mémoire et n'a pas d'impact sur l'apprentissage qui est fait côté GPU.

Algorithm 2 Construction du *Curriculum Learning* sur la base de données de validation.

global $\mathcal{D}$; `// 𝒟 is the tab where the couples are stored`
for $(i,j) \in [\![1,d]\!]^2$ ***if*** $i \neq j$ **do**
  $\mathcal{D}[(i,j)]$[occurrences] $\leftarrow 0$
  $\mathcal{D}[(i,j)]$[missed] $\leftarrow 0$
  $\mathcal{D}[(i,j)]$[frequency] $\leftarrow 0$
end
while *Validation_data* **is not** *empty* **do**
  $\mathcal{B}, \mathcal{Y} \leftarrow gen\_batch(Validation\_data)$
  $\bar{\mathcal{Y}} \leftarrow$ Siamese($\mathcal{B}$)
  $W \leftarrow \mathcal{Y} == \bar{\mathcal{Y}}$; `// right/wrong predictions`
  for $k \in [\![0, len(W)]\!]$ **do**
    $i, j =$ labels($\mathcal{B}[k]$) ; `// labels() returéseaux de neurones siamois the couple labels`
    if $i \neq j$ **then**
      $\mathcal{D}[(i,j)]$[occurences] $\leftarrow \mathcal{D}[(i,j)]$[occurences] $+ 1$
      if $W[k] = 0$ **then**
        $\mathcal{D}[(i,j)]$[missed] $\leftarrow \mathcal{D}[(i,j)]$[missed] $+ 1$
      end
    end
  end
end
for $(i,j) \in [\![1,d]\!]^2$ ***if*** $i \neq j$ **do**
  $\mathcal{D}[(i,j)]$[frequency] $\leftarrow \frac{\mathcal{D}[(i,j)][\text{missed}]}{\mathcal{D}[(i,j)][\text{occurrences}]}$
end

Algorithm 3 Utilisation du *Curriculum Learning* lors de l'apprentissage.

global $\mathcal{D}, i$; `// 𝒟 is loaded after validation epoch of Algorithm 2`
$B \leftarrow$ Train_data$[i * \text{bs} : (i+1) * \text{bs}]$
$i = i + 1$
$S, \mathcal{Y}_S \leftarrow$ get_similar_couples(B)
$M \leftarrow len(S)$
$D, \mathcal{Y}_D \leftarrow$ get_disimilar_couples(B)
$W \leftarrow [\,]$
for $k \in [\![1, len(D)]\!]$ **do**
  $i, j =$ labels($D[k]$)
  $W[i] = \mathcal{D}[(i,j)]$[frequency]
end
$\mathbf{p} =$ Softmax(W)
$D_M, \mathcal{Y}_{D_M} \leftarrow Choice(D, M, \mathbf{p})$
$\mathcal{B}, \mathcal{Y} =$ concatenate_and_shuffle($(S, D_M), (\mathcal{Y}_S, \mathcal{Y}_{D_M})$)
return $\mathcal{B}, \mathcal{Y}$

4 Bases de données et métriques

Cette section présente l'environnement dans lequel nos algorithmes sont testés. Nous utilisons deux tâches pour souligner la pertinence de notre méthode de *Curriculum Learning* : la classification supervisée et la partition d'actions non-supervisées. Nous présentons, pour chaque tâche, la base de données utilisée, les métriques qui permettent de mesurer la performance des algorithmes et enfin, les baselines utilisées.

4.1 Bases de données

4.1.1 Données pour la tâche supervisée

Nous utilisons ici la base de données NTU RGB-D (Liu et al., 2019). Elle consiste en 114480 vidéos réparties de manière équitable en 120 classes d'action qui sont représentées dans la figure 4.3. Les actions qui nous intéressent sont seulement celles qui impliquent une seule personne, c'est-à-dire celles qui ne sont pas coopératives. Nous restreignons donc cette base de données aux 94 actions qui correspondent à des actions solitaires.

1.1 Daily Actions (82)

A1: drink water	A2: eat meal	A3: brush teeth	A4: brush hair
A5: drop	A6: pick up	A7: throw	A8: sit down
A9: stand up	A10: clapping	A11: reading	A12: writing
A13: tear up paper	A14: put on jacket	A15: take off jacket	A16: put on a shoe
A17: take off a shoe	A18: put on glasses	A19: take off glasses	A20: put on a hat/cap
A21: take off a hat/cap	A22: cheer up	A23: hand waving	A24: kicking something
A25: reach into pocket	A26: hopping	A27: jump up	A28: phone call
A29: play with phone/tablet	A30: type on a keyboard	A31: point to something	A32: taking a selfie
A33: check time (from watch)	A34: rub two hands	A35: nod head/bow	A36: shake head
A37: wipe face	A38: salute	A39: put palms together	A40: cross hands in front
A61: put on headphone	A62: take off headphone	A63: shoot at basket	A64: bounce ball
A65: tennis bat swing	A66: juggle table tennis ball	A67: hush	A68: flick hair
A69: thumb up	A70: thumb down	A71: make OK sign	A72: make victory sign
A73: staple book	A74: counting money	A75: cutting nails	A76: cutting paper
A77: snap fingers	A78: open bottle	A79: sniff/smell	A80: squat down
A81: toss a coin	A82: fold paper	A83: ball up paper	A84: play magic cube
A85: apply cream on face	A86: apply cream on hand	A87: put on bag	A88: take off bag
A89: put object into bag	A90: take object out of bag	A91: open a box	A92: move heavy objects
A93: shake fist	A94: throw up cap/hat	A95: capitulate	A96: cross arms
A97: arm circles	A98: arm swings	A99: run on the spot	A100: butt kicks
A101: cross toe touch	A102: side kick	-	-

1.2 Medical Conditions (12)

A41: sneeze/cough	A42: staggering	A43: falling down	A44: headache
A45: chest pain	A46: back pain	A47: neck pain	A48: nausea/vomiting
A49: fan self	A103: yawn	A104: stretch oneself	A105: blow nose

Figure 4.3. Liste des actions présentes dans la base de données NTU-RGB-D. Tableau issue du cite officiel de la base de données [3].

2. https://rose1.ntu.edu.sg/dataset/actionRecognition/

Nous nous intéressons aussi à quantifier la performance de nos algorithmes sur des données jamais vues et des labels jamais vus. Nous créons donc une base de données OoD qui incorpore la moitié de la base de données. C'est-à-dire en choisissant aléatoirement 47 actions des 94 possibles et en mettant de côté toutes les données qui correspondent à ces actions. Nous nous retrouvons alors avec les jeux de données suivants : apprentissage, validation, test, test OoD.

4.1.2 Données pour la tâche non-supervisée

Nous utilisons ici la base de données IXMAS (Weinland et al., 2006), où chaque vidéo correspond à 16 actions. Un exemple des labels présents dans une vidéo est montré dans la figure 4.4 et l'ensemble des labels est décrit dans le tableau 4.1. À chaque frame de chaque vidéo est associé un label correspondant à l'action réalisée.
Chacun des 12 participants est filmé par 5 caméras et réalise l'expérience 3 fois pour un total de 180 vidéos.
Nous pensons que cette base de données possède assez de vidéos pour servir de benchmark à nos algorithmes de partition. De plus, les algorithmes considérés ne sont pas impactés par la répétition des caméras et des participants car ils sont non-supervisés.

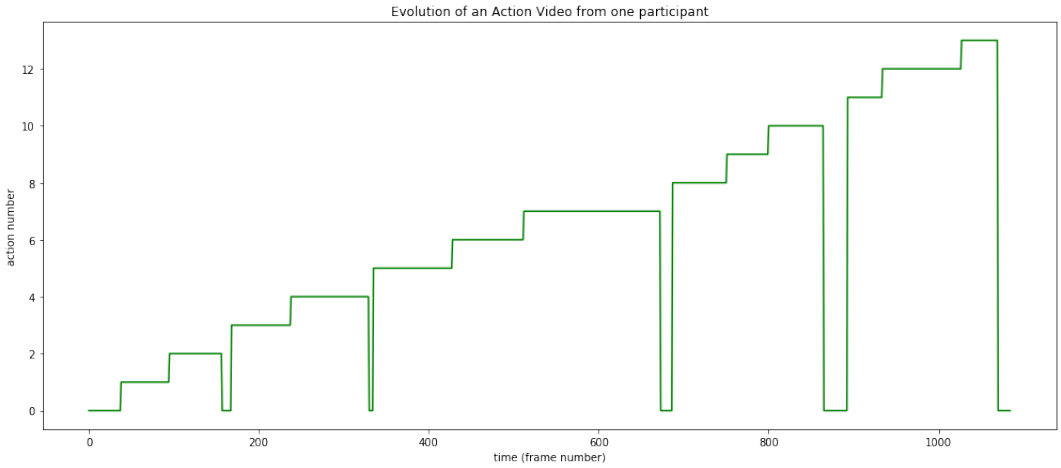

Figure 4.4. Évolution des actions d'un participant frame par frame.

Index	Label	Index	Label
0	Neutre	8	Signe de la main
1	Regarder sa montre	9	coup de poing
2	croiser les bras	10	coup de pied
3	se gratter la tête	11	pointer du doigt
4	s'asseoir	12	ramasser un objet
5	se relever	13	lancer l'objet 1
6	marcher en rond	14	lancer l'objet 2
7	marcher	15	autre

Tableau 4.1. Liste des actions présentes dans la base de données IXMAS.

4.2 Extraction de *features*

L'extraction de *features* est réalisée de la même manière pour les deux bases de données. Nous extrayons les données du corps et des mains de chaque participant grâce au logiciel OpenPose (voir section 3.1.1). Ceci nous évite de travailler sur les pixels des vidéos directement mais sur des donnés liées au corps et donc permet un apprentissage plus rapide. De plus, nous ne travaillons pas avec des données liées au visage des participants car elles ne sont pas assez informatives lorsqu'il s'agit de reconnaissance d'actions.
Comme nous travaillons avec des vidéos, la taille des données peut être un problème pour l'apprentissage. Par exemple, une vidéo de 10 secondes séquencée à 25 fps et avec une résolution 224 x 224 quand elle est extraite dans un tenseur pèse approximativement 300 MB. En comparaison, les données extraites par OpenPose sur cette même vidéo pèsent environ $0,3$ MB.
Les données extraites sont celles du corps (25 coordonnées x, y) et celles des mains (2 x 21 coordonnées x, y) résultant en un vecteur de taille 134 par frame. Les données manquantes sont remplacées par des zéros.
Nous construisons alors une matrice de *features* X_i représentant la $i^{ème}$ vidéo en empilant les vecteurs de *features* extraites. Cependant, en utiliant OpenPose, nous perdons beaucoup d'informations, surtout quand l'action porte sur une manipulation d'objet. Enfin, OpenPose possède aussi un taux d'erreur qui aura des conséquences sur l'apprentissage.

4.3 Quantifier l'apprentissage de nos modèles

Choisir une bonne métrique est cruciale pour avoir une bonne perspective sur l'apprentissage de notre algorithme. Pour la tâche supervisée, par exemple, si nous considérions juste l'*accuracy* sur la base de données de test, nous arriverions à des conclusions erronées et nous ne sélectionnerions pas le modèle ayant apprit la meilleure représentation.

4.3.1 Métriques sur la tâche supervisée

Notre réseau de neurones siamois est entraîné comme une tâche de classification, donc les métriques de performances habituelles s'appliquent : *accuracy*, *recall*, *precision*. Cependant, et comme nous l'avons démontré dans la section 3, nous sommes dans un cadre déséquilibré. Nous nous intéressons donc aussi à une vision symétrique de métriques de *precision* et de *recall*, ce qui est pertinent dans ce contexte (More, 2016).
Nous présentons cela en décrivant ces métriques :

	Classe 0	*Classe 1*
Prédiction de la classe 0	True Positive (TP)	False Positive (FP)
Prédiction de la classe 1	False Negative (FN)	True Negative (TN)

Tableau 4.2. Table de classification binaire. Les colonnes représentent les classes 0 et 1 ; les lignes représentent les prédictions correctes et incorrectes faites sur ces classes.

Dans le cadre de la classification binaire, un label prédit depuis un point de données peut appartenir à quatre états comme représenté dans le tableau 4.2. Avec cette catégorisation, nous

pouvons construire des métriques qui aident à comprendre l'apprentissage de nos algorithmes. Ces métriques sont le *recall* (R) et la *precision* (P), qui, pour la classe 0, sont définies comme suit :

$$R = \frac{TP}{TP + FN}$$
$$P = \frac{TP}{TP + FP}$$

Ces métriques sont pertinentes car elles répondent aux questions suivantes :

Precision : Parmi les prédictions d'une classe donnée, à quel point est-on précis ?
Recall : Parmi les vrais labels d'une classe donnée, à quel point est-on précis ?

4.3.2 Métriques sur la tâche non-supervisée

La métrique appliquée à la tâche non-supervisée vient de l'aglorithme Hungarian (Kuhn and Yaw, 1955), qui est un algorithme de programmation dynamique cherchant une couverture optimale maximisant, dans notre cas, l'*accuracy*. L'*accuracy* étant définie comme suit :
Soit X une vidéo composée de n frames, i.e. $X = (x_1, x_2, \ldots, x_n)$. À chaque frame est associé un label $y_i \in [\![1, d]\!]$ pour $i \in [\![1, n]\!]$
Soit $\hat{y}_i$ le label prédit à la frame i, alors l'*accuracy* est définie comme :

$$A_X = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\hat{y}_i = y_i}$$

4.4 Conditions expérimentales

4.4.1 Recherche quadrillée pour la tâche supervisée

Pour la tâche supervisée, comme dit précédemment, nous ne considérons pas le score de la *triplet loss* car cette méthode n'a pas été capable d'apprendre dans notre cadre d'étude.
Nous adoptons alors une procédure de recherche quadrillée pour choisir les paramètres des réseaux de neurones siamois et pour tester la procédure de *Curriculum Learning*.
Nous avons testé quatre types de stratégies pour l'entraînement des réseaux de neurones siamois : avec/sans *Curriculum Learning*, avec/sans pré-entraînement. Le pré-entraînment consiste en l'entraînement du réseau de neurones $\mathcal{A}$ de la figure 4.1 sur une tâche de reconnaissance d'actions. La tâche de reconnaissance d'action consiste à prédire les actions de la base de données NTU-RGB-D. Pour cela, nous découpons cette base de données de la même manière que pour l'apprentissage des réseaux de neurones siamois (voir section 4.1.1).
Ces réseaux de neurones siamois ont des paramètres qui sont : la proportion de label que nous mettons de côté pour la construiction du jeu de données OoD {0.5,0.3}, la taille du

batch {151,250}, le *learning rate* {0.01,0.001}, le nombre de neurones dans le réseau Bi-LSTM {191,250,300} et le *dropout rate* {0.1,0.2}. Nous avons aussi sélectionné le modèle pré-entraîné avec les mêmes paramètres et sur une recherche quadrillée. Nous avons ensuite sélectionné le modèle avec la plus haute accuracy.
Au total, 160 modèles ont été entraînés, répartis équitablement entre modèles : avec/sans *Curriculum Learning*, avec/sans pré-entraînement.
De tous ces modèles, nous avons sélectionné pour chacune des quatre stratégies le modèles avec la plus haute *accuracy* OoD. C'est-à-dire :

- SiamNOCL : sans *Curriculum Learning*, sans pré-entraînement.
- SiamCL : avec *Curriculum Learning*, sans pré-entraînement.
- PretrainSiamNOCL : sans *Curriculum Learning*, avec pré-entraînement.
- PretrainSiamCL : avec *Curriculum Learning*, avec pré-entraînement.

4.4.2 Baseline pour la tâche non-supervisée

Pour la tâche non-supervisée, nous choisissons le travail de (Zhou et al., 2013) comme point de comparaison. Leur travail consiste en itérations de l'algorithme Hierarchical Aligned Cluster Analysis (HACA). Nous comparons aussi nos résultats avec un de leurs précédents algorithmes, Aligned Cluster Analysis (ACA) (Zhou et al., 2008) et avec des Modèles de Mixture Gaussien (GMM) (Bishop and Nasrabadi, 2007). Tous ces algorithmes étant publiquement disponibles sur leur github [4].

5 Résultats

5.1 Résultats sur la tâche supervisée

Nous notons tout d'abord que lorsque le paramètre de la taille du batch est trop bas (e.g 72), peu importe la stratégie employée, le réseau de neurones siamois n'apprend pas. Nous pensons que ceci est dû au fait que le nombre de paires constructibles décroît au fur et à mesure que la taille du batch décroît.
De plus, sur les 160 réseaux de neurones siamois entraînés, seulement 32 ont appris une représentation des données correcte (mesurée par une *accuracy* OoD > 0.7). Ce faible nombre rappelle la difficulté de la tâche et nous empêche de recourir à des tests statistiques. Nous allons plutôt nous intéresser au meilleur modèle de chaque stratégie, comme il est traditionnellement fait en deep learning.
Les résultats sur la base de données test OoD sont affichés dans le tableau 4.3. Les résultats sur la base de données test sont affichés dans le tableau 4.4.

Nous pouvons alors voir que notre meilleur modèle est PretrainSiamCL, et qu'il est parfaitement équilibré. Il possède la même *accuracy* OoD que PretrainSiamNOCL, mais ce dernier est plus

4. https://github.com/zhfe99/aca

déséquilibré. Si nous regardons le tableau 4.4, nous voyons le même phénomène.
Si nous nous concentrons sur les modèles qui n'ont pas été pré-entraînés nous voyons, sur la base de données OoD, une meilleure *accuracy* de SiamCL par rapport à SiamNOCL. Ceci implique donc une meilleure représentation des données car ce modèle est plus efficace sur des labels jamais vus provenant de données jamais vues. Il est intéressant de noter que ce n'est pas le cas sur la base de données de test classique.
De plus, si nous nous étions seulement intéressé à la base de données test, nous aurions pu croire que SiamNOCL était le meilleur algorithme car il possède la meilleure accuracy. Nous voyons alors ici tout l'intérêt d'établir une base de données OoD et de prendre toutes ces métriques en compte. Ces résultats sont aussi confirmés par les résultats de la tâche de partition du comportement humain.

Model	***accuracy*-OoD**	**P-Dissim**	**R-Sim**	**P-Sim**	**R-Dissim**
SiamNOCL	0.71	0.7	0.7	0.72	0.72
SiamCL	0.72	0.74	0.73	0.70	0.71
PretrainSiamNOCL	0.73	0.75	0.74	0.71	0.72
PretrainSiamCL	0.73	0.73	0.73	0.73	0.73

Tableau 4.3. Résultats sur la base de données test OoD. Ce tableau présente aussi l'équilibrage de l'entraînement grâce à des métriques de *precision* et *recall* pour les deux classes.

Model	***accuracy***	**P-Dissim**	**R-Sim**	**P-Sim**	**R-Dissim**
SiamNOCL	0.79	0.75	0.77	0.83	0.81
SiamCL	0.77	0.75	0.76	0.79	0.78
PretrainSiamNOCL	0.78	0.83	0.81	0.73	0.75
PretrainSiamCL	0.78	0.83	0.81	0.74	0.76

Tableau 4.4. Résultats sur la base de données test. Ce tableau présente aussi l'équilibrage de l'entraînement grâce à des métriques de *precision* et *recall* pour les deux classes.

5.2 Partition d'actions

5.2.1 Motivations

Les résultats de la section précédente nous motivent pour tester nos quatre modèles sur la tâche de partition non-supervisée du comportement humain.
En effet, notre but principal ici est d'identifier la meilleure stratégie d'entraînement de réseaux de neurones siamois. Comme nous l'avons vu, l'utilisation de la base de donnée test OoD nous a déjà permis de ne pas nous tromper sur le modèle. Nous voudrions donc confirmer ces résultats. Pour ne pas biaiser l'interprétation des résultats sur cette nouvelle tâche, nous n'ajouterons pas d'entraînement supplémentaire. Nous convertissons de la manière la plus simple possible les réseaux de neurones siamois pour qu'ils puissent être capables de partitionner une vidéo comprenant des actions. Ceci est explicité dans la section 5.2.2.
Ainsi, nous sommes capable de voir comment nos modèles performent sur cette tâche sans ajouter d'aléa. Des paramètres comme la taille de la fenêtre temporelle ou le palier existent tout de

même et sont partagés par l'ensemble des modèles.

5.2.2 Convertir un réseau de neurones siamois

Nous utilisons deux fenêtres temporelles adjacentes de taille $\mathcal{W}$ pour diviser la vidéo en segments cohérents. Ainsi, nous listons tous les points de dissimilarité, c'est-à-dire tous les points où notre modèle prédit que les deux segments ne représentent pas la même action. Les points de dissimilarité sont ensuite nettoyés : les points qui sont trop proches les uns des autres sont regroupés. Cette procédure est représentée dans l'algorithme 4 et le résultat de la segmentation et du nettoyage est présenté dans la figure 4.5.
Maintenant, avec l'ensemble des points de dissimilarité, nous pouvons lister tous les segments d'actions et possiblement les regrouper. Nous construisons alors une matrice de similarité grâce aux réseaux de neurones siamois. Cette matrice est alors traitée pour extraire les segments similaires.

Algorithm 4 Partition d'action humaine grâce au réseaux de neurone siamois.

```
breaking_points = [ ]
for i in range(W, len(X) − W) do
    P = S(X[i − W : i], X[i : i + W])
    if P > θ then
        breaking_points.append(i)
    end
end
breaking_points = clean(breaking_points)
couples =list(zip(breaking_points,breaking_points))
Mat = [ ]
for i, j in enumerate(breaking_points) do
    Line = [ ]
    for k in breaking_points do
        P = 1 − S(X[j[0] : j[1]], X[k[0] : k[1]])
        Line.append(P)
    end
    Mat.append(Line)
end
R = Construct_Sim(Mat)
return R
```

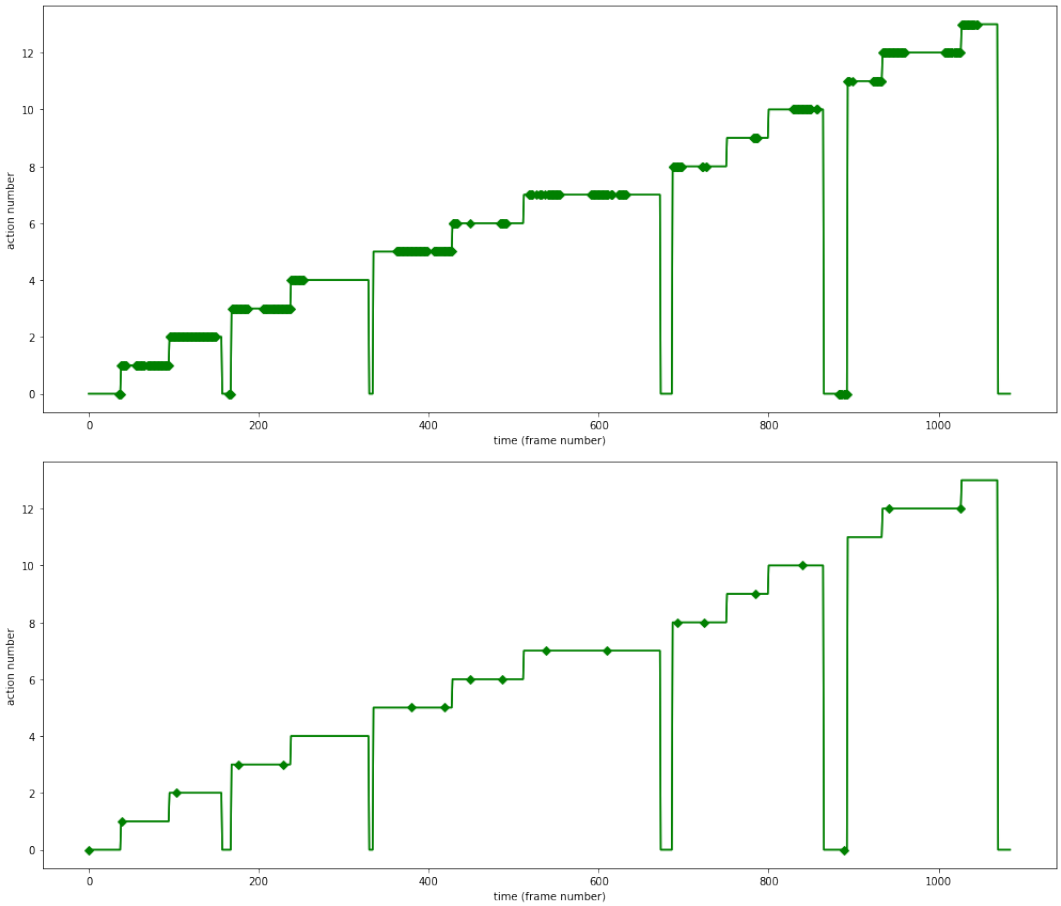

Figure 4.5. Représentation de points de dissimilarité et de leurs nettoyage sur un point de données.

5.2.3 Résultats sur la tâche non-supervisée

Les résultats sont présentés dans le tableau 4.5, et ils confirment que le modèle qui a le mieux apprit est PretrainSiamCL. C'est donc le modèle avec la plus haute *accuracy* OoD et l'apprentissage le plus équilibré.
En second vient alors SiamCL et PretrainSiamNOCL en troisième. Ceci suggère que notre stratégie de *Curriculum Learning* a un impact positif sur l'apprentissage et de même pour le pré-entraînement. Ces trois modèles battent aussi la baseline et ont des avantages non négligeables : pas d'entraînement, pas besoin de connaître le nombre total de labels ni leur emplacement précis, algorithmes plus rapide et aucun aléa.

Model	*accuracy*
Hierarchical Aligned Cluster Analysis	54.10
Aligned Cluster Analysis	53.59
Gaussian Mixture Models	26.84
SiamNOCL	48.20
SiamCL	54.53
PretrainSiamNOCL	54.44
PretrainSiamCL	**54.73**

Tableau 4.5. Métrique d'*accuracy* sur la tâche de partition non-supervisée. Tous les réseaux de neurones siamois partagent les mêmes paramètres pour éviter un quelconque biais.

Conclusion

Dans cette section, nous avons rigoureusement identifié un des problèmes majeurs des réseaux de neurones siamois : le déséquilibre des classes. Nous avons prouvé qu'il y avait une différence quadratique entre le nombre de paires similaires et dissimilaires qui sont constructibles à partir d'un batch équilibré.
À partir de cette analyse, nous avons créé une stratégie simple et efficace reposant sur une combinaison de sous-sampling et de *Curriculum Learning*. Cette méthode est alors testée sur deux tâches liées à la compréhension du comportement humain où elle performe mieux. Notre travail souligne aussi l'importance du jeu de données test OoD et des différentes métriques employées sans lesquels nous aurions des interprétations erronées.
Transférer des réseaux de neurones siamois sur une tâche de partition non-supervisée du comportement humain était, si ce n'est trivial, relativement facile.
Nous avons donc prouvé qu'il était possible de construire de manière non-supervisée des mots à partir de comportement non-verbal humain. Le comportement ici présent étant des actions, nous ne pouvons donc logiquement pas l'appliquer aux données VirtuAlZ. Cependant, c'est une bonne piste, et la suite pourrait être une collecte de données similaires aux bases de données employées mais sur des expressions faciales, par exemple.

Conclusions et perspectives

Conclusion

Lors de cette thèse nous avons apporté des éléments de réponse à la problématique $\mathcal{Q}$. Nous avons souligné l'impossibilité, en l'état actuel des choses, de quantifier automatiquement la qualité des comportements non-verbaux des soignants en interaction avec un Patient-Virtuel. Cette notion étant trop complexe et spécifique à chaque patient. Nous avons donc choisit de répondre aux deux sous-objectifs $\mathcal{O}1$ et $\mathcal{O}2$.
Pour ce faire, nous avons, en collaboration avec les autres membres du projet VirtuAlZ, élaboré un protocole expérimental dans lequel un soignant interagit avec un Patient-Virtuel sur une tâche spécifique. Le but du soignant étant de donner un médicament au Patient-Virtuel.
Nous avons alors collecté et analysé les interactions de 29 soignants. Les principales analyse réalisées dans le but de répondre aux objectifs $\mathcal{O}1$ et $\mathcal{O}2$ concluent que certains groupes de soignants ont des patterns de comportements non-verbaux similaires mais, qu'a priori, aucune caractéristiques personnelles n'a de rôle là-dedans. On observe, cependant, que le groupe des soignants les moins expérimentés ont tous un comportement non-verbal significativement différent les uns des autres. Nous arrivons à ces conclusions grâce aux outils de quantification de récurrence et de clusterisation employés. Ce résultat reste quand même à confirmer au vu de la faible taille de notre base de données.
Les outils d'analyse déployés requièrent au préalable une modélisation du comportement non-verbal des soignants. Cette modélisation est présentée rigoureusement dans cette thèse et s'inspire du traitement du langage naturel. Les signaux de comportement non-verbal des soignants sont détectés grâce aux logiciels OpenPose et OpenFace. De ces signaux, nous extrayons des informations temporelles symboliques qui nous permettent de construire des *mots* de comportement non-verbal. Enfin, grâce à ces mots, nous pouvons utiliser d'autres modélisation et analyses issues du traitement du langage naturel qui sont adaptées à la faible taille et grande richesse de notre base de données.
Finalement, et au vu de l'importance de la création des mots de comportement non-verbal, nous

choisissons aussi de présenter une preuve de concept sur l'identification de symboles de manière non-supervisée. Pour ce faire, nous développons :

- Des réseaux de neurones siamois, Bi-LSTM avec un processus d'attention.
- Une stratégie d'entraînement des réseaux de neurones siamois qui utilise une combinaison d'undersampling et de *curriculum learning*.
- Un algorithme déterministe qui permet de porter les réseaux de neurones siamois sur des tâches de segmentation temporelle et de classification de ces segments.

Ces trois contributions permettent de concurrencer les algorithmes de l'état de l'art.

Perspectives

Plusieurs améliorations et développements restent à poursuivre.
Dans le cadre du projet VirtuAlZ, il serait nécessaire de collecter plus de données. Une bonne stratégie serait de faire passer des personnes qui n'ont aucune expérience avec le milieu médical et d'étudier les différences de comportement avec les données que nous avons déjà. Cependant pour qu'une analyse comparative soit possible, cette collecte de données doit être réalisée dans les mêmes conditions que la précédente et notamment avec le même Magicien d'Oz. Avoir des données de comportement annotées par des professionnels pourrait aussi permettre de faire des analyses plus poussées.
Au niveau de la modélisation du comportement non-verbal employée, certaines améliorations pourraient aussi être développées. Nous pourrions extraire des symboles différents, utiliser des normes différentes, notamment dans la construction des matrices de similarités.
Sur la partie des réseaux de neurones siamois de nombreuses améliorations sont aussi envisagée. Premièrement sur les données extraites par OpenPose. Nous perdons beaucoup d'information et, avec la puissance de calcul nécessaire, il est possible de travailler directement à partir des images. Nous pourrions aussi ré-entraîné l'algorithme d'OpenPose et l'intégrer au réseaux de neurones siamois. Une possibilité serait alors d'utiliser directement le réseaux de neurones pré-entraîné d'OpenPose et de le brancher intelligemment à notre réseau de neurones Siamois.
Au niveau de l'architecture du réseau de neurones beaucoup de variantes existe et nous pourrions par exemple utiliser des transformers. Ce type d'architecture est très performant sur des données séquentielles et est notamment utilisé en traitement du langage naturel, en particulier sur des tâches de traduction. Les transformers sont la suite direct des réseaux de neurones que nous avons développés car nous utilisons déjà des mécanismes d'attention sur des données séquentielles.
Enfin, la transformation des réseaux de neurones siamois pour être capable de partitionner des actions est faite de la manière la plus simple possible. Ajouter des algorithmes de clustering est envisageable de même que reprendre l'entraînement du réseaux de neurones siamois de manière non-supervisée. En effet, tout le processus de partition que nous avons développé est différentiable et nous pourrions donc envisager un clustering plus fin qui intégrerait un réapprentissage d'une petite partie de notre réseau de neurones Siamois.

Finalement, et pour lier les deux parties principales de cette thèse, il est aussi possible de

poursuivre notre travail vers une application de la détection non-supervisée de mots à l'analyse du corpus collecté.
Il faut tout d'abord entraîner notre réseau de neurones siamois sur des comportements Humains qu'il serait pertinent de détecter dans une interaction soignant Patient-Virtuel. Pour ce faire, nous pourrions envisager, par exemple, d'entraîner notre réseau sur une base de données de vidéos d'expressions faciales, où une vidéo correspond à une seule expression. Ceci permettrait alors de partitionner les vidéos des soignants de notre corpus en fonction de changements dans leurs expressions faciales. Cependant, il nous faudrait vérifier la capacité de généralisation de notre réseau entraîné. Soit grâce à une base de données vidéos où une vidéo correspond à plusieurs expressions faciales, toutes différentes de celle de la base de données d'entraînement. Soit en recrutant des annotateurs qui jugeront de la pertinence des mots trouvés.
Avoir un modèle capable de découper les comportements des soignants en mots permettrait alors de réaliser les analyses suivantes :

- Étude comparative avec les mots et les analyses déjà faites. Étudier les similarités et dissimilarités entre les deux types de modélisations possibles.
- Faire une étude des signaux faibles sur le comportement d'un soignant ou alors d'un groupe de soignant.

Bibliographie

Agrawal, N., Kumar, S., Balasubramaniam, S., Bhargava, S., Sinha, P., Bakshi, B., and Sood, B. (2015). Effectiveness of virtual classroom training in improving the knowledge and key maternal neonatal health skills of general midwifery students in bihar, india : A pre – and post- intervention study. *Nurse education today*, 2. *2 citations pages 11 et 16*

Aisopos, F., Papadakis, G., and Varvarigou, T. (2011). Sentiment analysis of social media content using n-gram graphs. In *Proceedings of the 3rd ACM SIGMM International Workshop on Social Media*, WSM '11, page 9–14, New York, NY, USA. Association for Computing Machinery. *Cité page 35*

Alaverdyan, Z., Jung, J., Bouet, R., and Lartizien, C. (2020). Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging : Application to epilepsy lesion screening. *Medical Image Analysis*, 60 :101618. *Cité page 64*

Alexander, L., Sheen, J., Rinehart, N., Hay, M., and Boyd, L. (2018). Mental health simulation with student nurses : A qualitative review. *Clinical Simulation in Nursing*, 14 :8–14. DNP and PhD projects. *2 citations pages 12 et 17*

Baba, K., Nakatoh, T., and Minami, T. (2017). Plagiarism detection using document similarity based on distributed representation. *Procedia Computer Science*, 111 :382–387. The 8th International Conference on Advances in Information Technology. *Cité page 42*

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. (2018). Openface 2.0 : Facial behavior analysis toolkit. pages 59–66. *Cité page 29*

Bartneck, C., Kulic, D., Croft, E., and Zoghbi, S. (2008). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1 :71–81. *Cité page 25*

Bates, J. (1995). The role of emotion in believable agents. *Communications of the ACM*, 37. *Cité page 2*

Beal, R., Afrin, T., Farheen, A., and Adjeroh, D. (2016). A new algorithm for “the lcs problem” with application in compressing genome resequencing data. *BMC Genomics*, 17. *Cité page 42*

Becerril Ortega, R., Lucie, P., and Vanderstichel, H. (2019). Élaboration d’un outil de simulation pour la formation de soignant.e.s en gériatrie. Expérimenter pour apprendre ou questionner ses pratiques. In *5° colloque international de la didactique professionnelle*, Montréal, Canada. *Cité page 18*

Becerril-Ortega, R., Vanderstichel, H., Petit, L., Urbiolagallegos, M. J., Schoch, J., Benamara, S. D. A., Ravenet, B., Zagdoun, J., and Chaby, L. (2020). Design process for a virtual simulation environment for training healthcare professionals in geriatrics. *2 citations pages 2 et 18*

Bedi, P., Gupta, N., and Jindal, V. (2020). Siam-ids : Handling class imbalance problem in intrusion detection systems using siamese neural network. *Procedia Computer Science*, 171 :780–789. Third International Conference on Computing and Network Communications (CoCoNet’19). *Cité page 61*

Belzil, G. and Vézina, J. (2015). Impact of caregivers’ behaviors on resistiveness to care and collaboration in persons with dementia in the context of hygienic care : An interactional perspective. *International psychogeriatrics/IPA*, pages 1–13. *Cité page 18*

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). Curriculum learning. volume 60, page 6. *2 citations pages 62 et 67*

Berlemont, S., Lefebvre, G., Duffner, S., and Garcia, C. (2015). Siamese neural network based similarity metric for inertial gesture classification and rejection. *2 citations pages 63 et 64*

Berlemont, S., Lefebvre, G., Duffner, S., and Garcia, C. (2017). Class-balanced siamese neural networks. *Neurocomputing*. *Cité page 64*

Bernoulli, J. (1713). Ars conjectandi. *2 citations pages 64 et 65*

Biancardi, B. (2019). *Les premières secondes comptent ; Gérer les premières impressions pour un agent virtuel plus engageant*. PhD thesis, Sorbonne-Université, UPMC. *Cité page 30*

Bishop, C. M. and Nasrabadi, N. (2007). Pattern recognition and machine learning. *J. Electronic Imaging*, 16 :049901. *3 citations pages 53, 64 et 73*

Borg Sapiano, A., Sammut, R., and Trapani, J. (2017). The effectiveness of virtual simulation in improving student nurses’ knowledge and performance during patient deterioration : A pre and post test design. *Nurse Education Today*, 62. *2 citations pages 12 et 16*

Bracq, M.-S., Michinov, E., and Jannin, P. (2019). Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals. *Simulation in Healthcare*, 14(3) :188–194. *Cité page 8*

Bromley, J., Bentz, J., Bottou, L., Guyon, I., Lecun, Y., Moore, C., Sackinger, E., and Shah, R. (1993). Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7 :25. *2 citations pages 61 et 63*

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2018). OpenPose : realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv :1812.08008*. *Cité page 27*

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*. *Cité page 27*

Chalapathy, R. and Chawla, S. (2019). Deep learning for anomaly detection : A survey. *2 citations pages 60 et 64*

Cheema, N., Hosseini, S., Sprenger, J., Herrmann, E., Du, H., Fischer, K., and Slusallek, P. (2019). Fine-grained semantic segmentation of motion capture data using dilated temporal fully-convolutional networks. *Cité page 61*

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. E. (2020). A simple framework for contrastive learning of visual representations. *CoRR*, abs/2002.05709. *Cité page 64*

Chen, W., Chen, X., Zhang, J., and Huang, K. (2017). Beyond triplet loss : a deep quadruplet network for person re-identification. *Cité page 64*

Cheng, K., Chen, Y., and Fang, W. (2015). Video anomaly detection and localization using hierarchical feature representation and gaussian process regression. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2909–2917. *2 citations pages 35 et 60*

Chowdhury, D. (2021). A new dna sequencing alignment methodology using the longest common subsequence technique. *American Journal of Applied Mathematics and Computing*, Vol 1 :5–9. *Cité page 42*

Collins, L., Schrimmer, A., Diamond, J., and Burke, J. (2011). Evaluating verbal and non-verbal communication skills, in an ethnogeriatric osce. *Patient education and counseling*, 83 :158–62. *2 citations pages 3 et 30*

Cong, Y., Yuan, J., and Liu, J. (2011). Sparse reconstruction cost for abnormal event detection. In *CVPR 2011*, pages 3449–3456. *2 citations pages 35 et 60*

Coyne, E., Calleja, P., Forster, E., and Lin, F. (2021). A review of virtual-simulation for assessing healthcare students' clinical competency. *Nurse Education Today*, 96 :104623. *4 citations pages 8, 9, 17 et 18*

Damashek, M. (1995a). Gauging similarity with n-grams : Language-independent categorization of text. *Science*, 267(5199) :843–848. *Cité page 35*

Damashek, M. (1995b). Gauging similarity with n-grams : Language-independent categorization of text. *Science*, 267(5199) :843–848. *Cité page 41*

De la Torre, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., and Cohn, J. (2015). Intraface. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 1, pages 1–8. *2 citations pages iii et 29*

Delaherche, E. and Chetouani, M. (2010). Multimodal coordination : exploring relevant features and measures. In *SSPW '10*. *Cité page 47*

Doolen, J., Giddings, M., Johnson, M., Guizado de Nathan, G., and Abadia, L. (2014). An evaluation of mental health simulation with standardized patients. *International journal of nursing education scholarship*, 11. *2 citations pages 11 et 16*

Dubuisson Duplessis, G., Langlet, C., Clavel, C., and Landragin, F. (2021). Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, 55. *Cité page 41*

Duman, E. and Erdem, O. A. (2019). Anomaly detection in videos using optical flow and convolutional autoencoder. *IEEE Access*, 7 :183914 – 183923. *Cité page 61*

Dwibedi, D., Aytar, Y., Tompson, J., Sermanet, P., and Zisserman, A. (2019). Temporal cycle-consistency learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *Cité page 61*

Elhadi, M. and Al-Tobi, A. (2009). Duplicate detection in documents and webpages using improved longest common subsequence and documents syntactical structures. pages 679 – 684. *Cité page 42*

Esogbue, A. O. (1979). Simulation in surgery, anesthesia and medical interviewing. *SIGSIM Simul. Dig.*, 10(4) :19–24. *Cité page 8*

Faiz Ur Rahman, F., Vasu, B., Cor, J., Kerekes, J., and Savakis, A. (2018). Siamese network with multi-level features for patch-based change detection in satellite imagery. *Cité page 64*

Fergani, B., Davy, M., and Houacine, A. (2008). Speaker diarization using one-class support vector machines. *Speech Communication*, 50(5) :355 – 365. *Cité page 61*

Flin, R., O'Connor, P., and Crichton, M. (2017). *Safety at the Sharp End : A Guide to Non-Technical Skills*. *Cité page 8*

Foronda, C., Gattamorta, K., Snowden, K., and Bauman, E. (2013). Use of virtual clinical simulation to improve communication skills of baccalaureate nursing students : A pilot study. *Nurse education today*, 34. *Cité page 17*

Fossen, P. and Stoeckel, P. (2016). Nursing students' perceptions of a hearing voices simulation and role-play : Preparation for mental health clinical practice. *The Journal of nursing education*, 55 :203–208. *2 citations pages 11 et 16*

Fusaroli, R., Konvalinka, I., and Wallot, S. (2014). Analyzing social interactions : The promises and challenges of using cross recurrence quantification analysis. *Springer Proceedings in Mathematics & Statistics*, 103 :137–155. *Cité page 47*

Gauffman, E. (1974). *Frame Analysis : An Essay on the Organization of Experience*. Harvard University Press. *Cité page 18*

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., and Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Frontiers in Human Neuroscience*, 8 :807. *2 citations pages 9 et 15*

Giwa, O. and Davel, M. (2013). N-gram based language identification of individual words. *Cité page 35*

GOH, Y.-S., Selvarajan, S., Chng, M.-L., Tan, C.-S., and Yobas, P. (2016). Using standardized patients in enhancing undergraduate students' learning experience in mental health nursing. *Nurse Education Today*, 45 :167–172. *2 citations pages 11 et 16*

Grier, R., Bangor, A., Kortum, P., and Peres, S. (2013). The system usability scale. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57 :187–191. *Cité page 25*

Grover, S. C., Garg, A., Scaffidi, M. A., Yu, J. J., Plener, I. S., Yong, E., Cino, M., Grantcharov, T. P., and Walsh, C. M. (2015). Impact of a simulation training curriculum on technical and nontechnical skills in colonoscopy : a randomized trial. *Gastrointestinal Endoscopy*, 82(6) :1072–1079. *Cité page 14*

Guerra Filho, G. and Aloimonos, Y. (2007). A language for human action. *Computer*, 40 :42 – 51. *2 citations pages 35 et 60*

Ha, J. F. and Longnecker, N. (2010). Doctor-patient communication : a review. *Ochsner Journal*, 10(1) :38–43. *Cité page 7*

Hall, J. A., Horgan, T. G., and Murphy, N. A. (2019). Nonverbal communication. *Annual Review of Psychology*, 70 :271–294. *Cité page 7*

Harrigan, J. A. (1985). Self-touching as an indicator of underlying affect and language processes. *Social Science & Medicine*, 20(11) :1161–1168. *Cité page 32*

Harwood, R. H., O'Brien, R., Goldberg, S. E., Allwood, R., Pilnick, A., Beeke, S., Thomson, L., Murray, M., Parry, R., Kearney, F., Baxendale, B., Sartain, K., and Schneider, J. (2018). A staff training intervention to improve communication between people living with dementia and health-care professionals in hospital : the voice mixed-methods development and evaluation study. *Health Services and Delivery Research*, 6 :1–134. *Cité page 18*

H.A.S. (2009.). Maladie d'alzheimer et maladies apparentées : Prise en charge des troubles du comportement perturbateurs. page 11. *Cité page 18*

Henderson, S., Dalton, M., and Cartmel, J. (2016). Using interprofessional learning for continuing education : Development and evaluation of the graduate certificate program in health professional education for clinicians. *Journal of Continuing Education in the Health Professions*, 36 :211–217. *Cité page 8*

Hermans, A., Beyer, L., and Leibe, B. (2017). In defense of the triplet loss for person re-identification. *CoRR*, abs/1703.07737. *Cité page 64*

Heyselaar, E., Hagoort, P., and Segaert, K. (2015). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior research methods*, 49. *Cité page 2*

Hoffer, E. and Ailon, N. (2015). Deep metric learning using triplet network. *Lecture Notes in Computer Science*, page 84–92. *Cité page 64*

Hogg, A., Evers, C., and Naylor, P. (2019). Speaker change detection using fundamental frequency with application to multi-talker segmentation. *Cité page 61*

Hosseini, B., Montagne, R., and Hammer, B. (2019). Deep-aligned convolutional neural network for skeleton-based action recognition and segmentation. *2019 IEEE International Conference on Data Mining (ICDM)*. *Cité page 61*

Howe, L. and Leibowitz, K. (2019). When your doctor gets it and gets you : The critical role of competence and warmth in the patient provider interaction. *Frontiers in Psychiatry*, 10. *Cité page 7*

Ionescu, R. T., Khan, F., Georgescu, M., and Shao, L. (2019). Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. pages 7834–7843. *Cité page 61*

Jack, D., Gerolamo, A. M., Frederick, D., Szajna, A., and Muccitelli, J. (2014). Using a trained actor to model mental health nursing care. *Clinical Simulation in Nursing*, 10(10) :515–520. *2 citations pages 11 et 16*

Jacobs, A. and Venter, I. (2017). Standardised patient-simulated practice learning : A rich pedagogical environment for psychiatric nursing education. *African Journal of Health Professions Education*, 9 :107. *2 citations pages 12 et 16*

Jia, S., Wang, S., Hu, C., Webster, P. J., and Li, X. (2021). Detection of genuine and posed facial expressions of emotion : Databases and methods. *Frontiers in Psychology*, 11. *2 citations pages iii et 29*

Kadri, H., Davy, M., Rabaoui, A., Lachiri, Z., and Ellouze, N. (2008). Robust audio speaker segmentation using one class svms. In *2008 16th European Signal Processing Conference*, pages 1–5. *Cité page 61*

Khan, R., Scaffidi, M. A., Walsh, C. M., Lin, P., Al-Mazroui, A., Chana, B., Kalaichandran, R., Lee, W., Grantcharov, T. P., and Grover, S. C. (2017). Simulation-based training of non-technical skills in colonoscopy : Protocol for a randomized controlled trial. *JMIR Res Protoc*, 6(8) :e153. *Cité page 14*

Khullar, D. (2019). Building trust in health care—why, where, and how. *Jama*, 322(6) :507–509. *Cité page 7*

Kipp, M. (2001). Anvil - a generic annotation tool for multimodal dialogue. pages 1367–1370. *Cité page 25*

Kiran, B., Thomas, D., and Parakkal, R. (2018). An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging*, 4. *Cité page 61*

Kneebone, R., Weldon, S.-M., and Bello, F. (2016). Engaging patients and clinicians through simulation : rebalancing the dynamics of care. *Advances in simulation (London, England)*, 1 :19. *Cité page 16*

Kostakis, O. and Papapetrou, P. (2015). Finding the longest common sub-pattern in sequences of temporal intervals. *Data Mining and Knowledge Discovery*, 29. *Cité page 46*

Kron, F., Fetters, M., Scerbo, M., White, C., Lypson, M., Padilla, M., Gliva-McConvey, G., Belfore, L., West, T., Wallace, A., Guetterman, T., Schleicher, L., Kennedy, R., Mangrulkar, R., Cleary, J., Marsella, S., and Becker, D. (2016). Using a computer simulation for teaching communication skills : A blinded multisite mixed methods randomized controlled trial. *Patient Education and Counseling*, 100. *4 citations pages 11, 14, 16 et 18*

Kuhn, H. W. and Yaw, B. (1955). The hungarian method for the assignment problem. *Naval Res. Logist. Quart*, pages 83–97. *Cité page 72*

Lehr, S. T. and Kaplan, B. (2013). A mental health simulation experience for baccalaureate student nurses. *Clinical Simulation in Nursing*, 9(10) :e425–e431. *2 citations pages 10 et 17*

Lester, J. and Stone, B. (1997). Increasing believability in animated pedagogical agents. pages 16–21. *Cité page 2*

Liaw, S. Y., Ooi, S. W., Rusli, K. D. B., Lau, T. C., Tam, W. W. S., and Chua, W. L. (2020a). Nurse-physician communication team training in virtual reality versus live simulations : Randomized controlled trial on team communication and teamwork attitudes. *J Med Internet Res*, 22(4) :e17279. *Cité page 17*

Liaw, S. Y. and Wu, L. T. (2019). Getting everyone on the same page : Interprofessional team training to develop shared mental models on interprofessional rounds. *2 citations pages 13 et 16*

Liaw, S. Y., Wu, L. T., Soh, S. L. H., Ringsted, C., Lau, T. C., and Lim, W. S. (2020b). Virtual reality simulation in interprofessional round training for health care students : A qualitative evaluation study. *Clinical Simulation in Nursing*, 45 :42–46. Patient Safety. *2 citations pages 13 et 16*

Lira, D., González-Rosales, K., Castillo Guevara, R., Spencer, R., and Fresno, A. (2018). Categorical cross-recurrence quantification analysis applied to communicative interaction during ainsworth's strange situation. *Complexity*, 2018 :1–15. *Cité page 41*

Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L.-Y., and Kot, A. C. (2019). Ntu rgb+d 120 : A large-scale benchmark for 3d human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. *2 citations pages 62 et 69*

Louise, H., Sonal, A., Rajesh, A., Ara, D., Charles, V., and Nick, S. (2011). The impact of nontechnical skills on technical performance in surgery : A systematic review. *Journal of the American College of Surgeons*, 214 :214–30. *Cité page 8*

Marano, C., Murianni, L., and Sticchi, L. (2005). To err is human. building a safer health system. *Italian Journal of Public Health*, 2. *Cité page 8*

Masana, M., Ruiz, I., Serrat, J., van de Weijer, J., and Lopez, A. M. (2018). Metric learning for novelty and anomaly detection. In *British Machine Vision Conference (BMVC)*. *Cité page 64*

Mast, M. (2007). On the importance of nonverbal communication in the physician-patient interaction. *Patient education and counseling*, 67 :315–318. *3 citations pages 3, 7 et 31*

May, W., Park, J. H., and Lee, J. P. (2009). A ten-year review of the literature on the use of standardized patients in teaching and learning : 1996–2005. *Medical Teacher*, 31(6) :487–492. *Cité page 17*

Mccamley, J., Denton, W., Lyden, E., and Yentes, J. (2017). Measuring coupling of rhythmical time series using cross sample entropy and cross recurrence quantification analysis. *Computational and Mathematical Methods in Medicine*, 2017 :1–11. *Cité page 47*

Menzel, N., Willson, L., and Doolen, J. (2014a). Effectiveness of a poverty simulation in second life® : Changing nursing student attitudes toward poor people. *International journal of nursing education scholarship*, 11. *Cité page 11*

Menzel, N., Willson, L., and Doolen, J. (2014b). Effectiveness of a poverty simulation in second life® : Changing nursing student attitudes toward poor people. *International journal of nursing education scholarship*, 11. *Cité page 17*

More, A. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. *2 citations pages 66 et 71*

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7 :33–35. *Cité page 2*

Nguyen, V. H., Nguyen, H. T., Duong, H. N., and Snasel, V. (2016). N-gram-based text compression. *Intell. Neuroscience*, 2016 :9. *Cité page 35*

Okuda, Y., Bryson, E. O., DeMaria Jr, S., Jacobson, L., Quinones, J., Shen, B., and Levine, A. I. (2009). The utility of simulation in medical education : what is the evidence ? *Mount Sinai Journal of Medicine : A Journal of Translational and Personalized Medicine : A Journal of Translational and Personalized Medicine*, 76(4) :330–343. *Cité page 9*

Oudshoorn, A. and Sinclair, B. (2015). Using unfolding simulations to teach mental health concepts in undergraduate nursing education. *Clinical Simulation in Nursing*, 11(9) :396–401. *2 citations pages 11 et 16*

Padilha, M., Machado, P., Ribeiro, A., and Ramos, J. (2018). Clinical virtual simulation in nursing education. *Clinical Simulation in Nursing*, 15 :13–18. *2 citations pages 12 et 17*

Piéron, H. (1931). *La Psychologie comme science biologique ducomportement des organismes.* *Cité page 7*

Przyrembel, M., Smallwood, J., Pauen, M., and Singer, T. (2012). Illuminating the dark matter of social neuroscience : Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in Human Neuroscience*, 6 :190. *Cité page 9*

Ramseyer, F. and Tschacher, W. (2006). Synchrony - a core concept for a constructivist approach to psychotherapy. *Cité page 47*

Ren, H., Liu, W., Olsen, S., Escalera, S., and Moeslund, T. (2015). Unsupervised behavior-specific dictionary learning for abnormal event detection. *Cité page 35*

Revaud, J., Douze, M., Schmid, C., and Jégou, H. (2013). Event retrieval in large video collections with circulant temporal encoding. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2459–2466. *Cité page 61*

Roberts, S. D. and England, W. L. (1980). Simulation and health care delivery. Technical report, Institute of Electrical and Electronics Engineers (IEEE). *Cité page 8*

Rosenthal-von der Pütten, A. M., Krämer, N., and Gratch, J. (2010). How our personality shapes our interactions with embodied conversational agents - implications for research and development. *Cité page 18*

Saha, D., Islam, T., and Hasan, M. (2018). Longest common substring in dna sequence. *International Journal of Scientific and Engineering Research*, 9. *Cité page 42*

Sahito, A., Frank, E., and Pfahringer, B. (2019). Semi-supervised learning using siamese networks. In *Australasian Conference on Artificial Intelligence*, pages 586–597. *Cité page 61*

Saint-Georges, C., Mahdhaoui, A., Chetouani, M., Cassel, R. S., Laznik, M.-C., Apicella, F., Muratori, P., Maestro, S., Muratori, F., and Cohen, D. (2011). Do parents recognize autistic deviant behavior long before diagnosis ? taking into account interaction using computational methods. *PLOS ONE*, 6(7) :1–13. *Cité page 35*

Sankaranarayanan, G., Wooley, L., Hogg, D., Dorozhkin, D., Olasky, J., Chauhan, S., Fleshman, J., De, S., Scott, D., and Jones, D. (2018). Immersive virtual reality-based training improves response in a simulated operating room fire scenario. *Surgical Endoscopy*, 32. *2 citations pages 12 et 17*

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet : A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832. *2 citations pages 61 et 66*

Schuster, M. and Paliwal, K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11) :2673–2681. *Cité page 64*

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27 :379–423. *Cité page 35*

Simon, T., Joo, H., Matthews, I., and Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*. *Cité page 27*

Smidt, A., Balandin, S., Sigafoos, J., and Reed, V. (2009). The kirkpatrick model : A useful tool for evaluating training outcomes. *Journal of intellectual & developmental disability*, 34 :266–74. *Cité page 25*

Spencer, F. and Eiseman, B. (1962). Technique of carotid endarterectomy. *Surgery, gynecology &amp ; obstetrics*, 115 :115—117. *Cité page 8*

Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., and Iloreta, A. (2017). Immersive virtual reality as a teaching tool for neuroanatomy : Immersive vr as a neuroanatomy teaching tool. *International Forum of Allergy & Rhinology*, 7. *2 citations pages 12 et 17*

Takano, W., Imagawa, H., and Nakamura, Y. (2011). Prediction of human behaviors in the future through symbolic inference. *2 citations pages 35 et 53*

Tariman, J. D., Berry, D. L., Halpenny, B., Wolpin, S., and Schepp, K. (2011). Validation and testing of the acceptability e-scale for web-based patient-reported outcomes in cancer care. *Applied nursing research : ANR*, 24(1) :53—58. *Cité page 25*

Thomas, P., Chandès, G., Hazif-Thomas, C., and Fontanille, J. (2017). Analyse du sens d'un trouble du comportement dans la démence. *Annales Médico-psychologiques, revue psychiatrique*, 175(10) :906 – 913. *Cité page 7*

Thurau, C. and Hlaváč, V. (2007). N-grams of action primitives for recognizing human behavior. In *Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns*, CAIP'07, page 93–100, Berlin, Heidelberg. Springer-Verlag. *2 citations pages 35 et 53*

Timmermann, C., Uhrenfeldt, L., and Birkelund, R. (2017). Ethics in the communicative encounter : seriously ill patients' experiences of health professionals' nonverbal communication. *Scandinavian journal of caring sciences*, 31(1) :63–71. *Cité page 7*

Tomović, A., Janičić, P., and Kešelj, V. (2006). n-gram-based classification and unsupervised hierarchical clustering of genome sequences. *Computer Methods and Programs in Biomedicine*, 81(2) :137–153. *Cité page 35*

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc. *Cité page 64*

Verkuyl, M., Lapum, J., St-Amant, O., Hughes, M., Romaniuk, D., and McCulloch, T. (2020). Exploring debriefing combinations after a virtual simulation. *Clinical Simulation in Nursing*, 40 :36–42. *Cité page 13*

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301) :236–244. *2 citations pages 35 et 53*

Weaver Moore, L., Moore, L. W., and Proffitt, C. (1993). Communicating effectively with elderly surgical patients. *AORN Journal*, 58(2) :345–355. *Cité page 7*

Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y. (2016). Convolutional pose machines. In *CVPR*. *Cité page 27*

Weinland, D., Ronfard, R., and Boyer, E. (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, 104 :249–257. *2 citations pages 62 et 70*

Witt, M. A., McGaughan, K., and Smaldone, A. (2018). Standardized patient simulation experiences improves mental health assessment and communication. *Clinical Simulation in Nursing*, 23 :16–20. *2 citations pages 13 et 16*

Wooldridge, M. and Veloso, M. (1999). *Artificial Intelligence Today : Recent Trends and Developments*. *Cité page 2*

Wright, R., Tinnon, E., and Newton, R. (2017). Evaluation of vsim for nursing in an adult health nursing course : A multisite pilot study. *CIN : Computers, Informatics, Nursing*, 36 :1. *Cité page 12*

Zagdoun, J., Chaby, L., Benamara, A., Urbiolla Gallegos, M. J., and Chetouani, M. (2021). Non-verbal behaviors analysis of healthcare professionals engaged with a virtual-patient. *Socially-Informed AI for Healthcare, ICMI'21, Montreal, Canada*. *Cité page 38*

Zhang, F., Jhi, Y.-C., Wu, D., Liu, P., and Zhu, S. (2012). A first step towards algorithm plagiarism detection. ISSTA 2012, page 111–121, New York, NY, USA. Association for Computing Machinery. *Cité page 42*

Zhao, B. and Li, F. F. (2011). Online detection of unusual events in videos via dynamic sparse coding. pages 3313–3320. *2 citations pages 35 et 60*

Zhou, F., De la Torre, F., and Hodgins, J. K. (2008). Aligned cluster analysis for temporal segmentation of human motion. In *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–7. *2 citations pages 61 et 73*

Zhou, F., De la Torre, F., and Hodgins, J. K. (2013). Hierarchical aligned cluster analysis for temporal clustering of human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3) :582–596. *2 citations pages 61 et 73*

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort : An Introduction to Human Ecology*. Addison-Wesley, Cambridge, Massachusetts. *Cité page 41*