L'intelligence artificielle au service de la formation professionnelle basée sur la simulation

Matei Mancas, Service d'Information, Signal et Intelligence artificielle, UMONS

François Rocca, Service d'Information, Signal et Intelligence artificielle, UMONS

Laurie-Anna Dubois, Service de Psychologie du travail, UMONS

Antoine Derobertmasure, Service de Méthodologie et formation, UMONS

1. Introduction

Depuis une dizaine d'années, l'arrivée de l'Intelligence Artificielle (IA) basée sur l'apprentissage profond (deep learning) dans le domaine de l'observation et de la mesure centrée sur l'humain bouleverse l'état de l'art technologique. En effet, les possibilités de détection et de suivi des personnes ainsi que de leurs interactions sont désormais plus matures et permettent d'obtenir des informations précises en temps réel. Analyser l'activité des apprenants dans des situations à visée de formation professionnelle devient donc une application possible de l'IA. Dans un contexte de formation professionnelle, la simulation est un outil utilisé par les formateurs pour former de (futurs) professionnels issus de différents secteurs d'activités (par exemple : les forces de l'ordre, la sécurité civile, les soins de santé, l'enseignement…) à tel ou tel type de tâches ou situations de travail (Béguin & Weill-Fassina, 1997). De nos jours, il existe de nombreux dispositifs permettant l'enregistrement audio-vidéo de l'activité des apprenants dans des formations professionnelles basées sur la simulation. Mais force est de constater que ces dispositifs, bien que pertinents pour l'observation en temps réel, ne réalisent aucune analyse automatique de cette activité. Ils ne « déchargent cognitivement » donc en rien le formateur qui,

en simulation, remplit de nombreuses missions (missions d'observateur, de modérateur, de médiateur, d'acteur…).

Face à un tel constat, des laboratoires de recherche se mettent à coopérer. Ces laboratoires de recherche sont pour une part spécialisés dans le domaine de la formation professionnelle et pour une autre part davantage familiarisés aux nouvelles technologies. L'objectif poursuivi, dans le cadre de cette collaboration, est de mutualiser les connaissances et, sur cette base, développer des outils technologiques contribuant à rendre l'activité du formateur plus efficace en simulation en permettant par exemple une labélisation semi-automatique des vidéos, un résumé automatique des moments-clés, etc.

Ce chapitre est le fruit d'une coopération (encore à ses débuts) de tels laboratoires de recherche. Il est structuré en trois grandes parties : une première partie (cf. point 2) qui vise à clarifier les objectifs pédagogiques poursuivis dans le cadre d'une formation professionnelle par simulation et à caractériser l'activité du formateur dans pareil dispositif. La deuxième partie (cf. point 3) a pour objectif de souligner l'intérêt qu'il peut y avoir pour un formateur de mobiliser les nouvelles technologies dans le cadre de sa pratique professionnelle en simulation. Elle vise également à dresser un rapide état de l'art relatif aux possibilités techniques déjà disponibles (ou en passe de le devenir) et à donner un aperçu des pistes à creuser en matière de technologies IA centrées sur l'humain susceptibles de soutenir l'activité du formateur dans le cadre d'une formation par simulation. Enfin, la troisième partie (cf. point 4) porte sur les conditions à satisfaire pour viser une adaptation réciproque et optimale entre formateur et nouvelles technologies dans le cadre d'une formation par simulation. Le pari est fait que les technologies recourant à l'IA, dans un futur très proche, se verront de plus en plus développées et utilisées. S'il est pratiquement certain que ces avancées vont largement transformer la recherche sur la formation, rien ne nous garantit, en effet, que la rencontre de ces deux mondes s'inscrive dans une logique de continuité.

2. Les simulations à visée de formation professionnelle : quels objectifs pédagogiques poursuivis ? Quelle activité du formateur ?

2.1 Les objectifs pédagogiques des formations basées sur la simulation

La simulation est un outil qui, de nos jours, est de plus en plus utilisé pour former de (futurs) professionnels issus de différents secteurs d'activités tels que les forces de l'ordre, la sécurité civile, les soins de santé, l'enseignement, etc. Dans le cadre particulier de la formation des enseignants, les simulations visent à amener chaque futur enseignant (l'apprenant) à présenter une leçon de 40 minutes (sur la base d'une leçon qu'il a préparée) devant ses collègues endossant le rôle d'observateur ou celui d'élève du niveau visé. Ces simulations se déroulent sous le regard attentif d'un formateur, chargé de l'organisation et de la gestion des simulations auxquelles prend part l'enseignant qui est ici l'apprenant (Dubois, Bocquillon, Romanus, & Derobertmasure, 2019).

Une formation par simulation se caractérise habituellement par trois phases (Samurçay, 2009) : le briefing, la séance de simulation et le débriefing. Le briefing est une phase qui permet de préparer la séance de simulation. La séance de simulation correspond au moment où l'apprenant est confronté à la situation simulée et où il construit (ou met en œuvre) des compétences opérationnelles. Enfin, le débriefing est une étape au cours de laquelle l'apprenant doit, en étant guidé par le formateur, porter un regard réflexif sur son activité en séance. Cette étape contribue à une exploitation plus poussée de la simulation : il s'agit de construire son savoir professionnel par la réflexion sur l'action et non uniquement par sa reproduction (Pastré, 2009).

En formation, les simulations peuvent poursuivre deux grandes catégories d'objectifs pédagogiques (Béguin & Weill-Fassina, 1997) : viser la réussite de l'action en situation en agissant sur la performance des apprenants et/ou viser le développement de compétences permettant de « réussir » dans d'autres situations.

2.1.1 Agir sur la performance des apprenants et viser la réussite de l'action

Les simulations peuvent permettre aux apprenants d'apprendre à faire. De telles simulations peuvent prendre la forme d'exercices de « drill » visant la mise en pratique (cf. l'application) de techniques ainsi que l'acquisition de gestes professionnels qui y sont liés (par exemple : la mise en pratique de manœuvres d'accouchement dans le cadre d'une formation par simulation pour sages-femmes). Le but de l'enseignement par simulation est dans ce cas-ci la réussite de l'action par l'apprenant. Le formateur vise en simulation à agir sur la performance des apprenants.

2.1.2 Agir sur le développement des compétences et viser la réussite dans d'autres situations

Les simulations peuvent aussi avoir pour objectif d'apprendre à savoir faire en situation. Dans ce cas, le formateur cherche à agir sur les compétences des apprenants. Il ne s'agit pas seulement pour les apprenants de mettre en pratique ce qui leur a été préalablement enseigné de manière théorique (c'est-à-dire d'appliquer des règles prescrites ou de reproduire un geste technique…), il s'agit aussi de mettre en place des réponses adaptées aux problèmes posés, et ce, dans des situations complexes, ce qui doit amener les apprenants à adapter, dans certains cas, ce qui leur a été enseigné aux situations auxquelles ils sont confrontés en simulation. Et pour cause, les cas d'école ne se rencontrent que rarement sur le terrain (Caens-Martin, 2009).

Dans une formation par simulation, le formateur peut donc chercher à agir sur la performance des apprenants ou chercher à agir sur leurs compétences (Béguin & Weill-Fassina, 1997). Lorsque le formateur cherche à agir sur les compétences des apprenants, on peut dire qu'on dépasse le niveau de la réussite (immédiate) de l'action et qu'on vise l'acquisition de compétences permettant ultérieurement et dans d'autres situations de réussir. Partant de ces constats, on conçoit aisément l'intérêt des formations professionnelles basées sur la simulation.

On ne peut également ignorer le rôle non négligeable exercé par les formateurs sur les compétences construites par les apprenants dans pareils dispositifs.

2.2 L'activité du formateur dans le cadre d'une formation par simulation

2.2.1 Une activité professionnelle

Tout dispositif de formation professionnelle basé sur la simulation implique (au moins) deux catégories d'acteurs : le formateur et l'apprenant. Tous deux développent en formation une activité.

L'activité du formateur en situation de formation (par simulation) peut être envisagée comme une activité professionnelle (Rogalski, 2003, 2007, 2012 ; Vidal-Gomel & Rogalski, 2009). Comme pour tout professionnel, des tâches sont prescrites au formateur : il doit atteindre des buts sous certaines conditions (Leplat & Hoc, 1983). L'activité qu'il réalise renvoie non seulement à ce qu'il fait (ce qui est de l'ordre de l'observable tel que ses déplacements, ses gestes…), mais aussi à ses diagnostics, ses anticipations, ses représentations (à savoir, ce qui n'est pas directement observable : son activité mentale) et à ce qu'il s'empêche éventuellement de faire, ou ce qu'il souhaiterait faire mais ne peut pas faire (en raison de certaines contraintes institutionnelles auxquelles il est soumis par exemple) (Rogalski, 2007). Rogalski (2003, 2007, 2012) souligne que l'activité du formateur est déterminée par ses propres caractéristiques (les déterminants « intrinsèques » tels que ses compétences vis-à-vis de la matière qu'il est chargé d'enseigner) mais aussi par la situation de travail dans laquelle son activité se déploie (les déterminants « extrinsèques » tels que les caractéristiques des apprenants ou encore le contexte de la situation de formation). Toujours selon cette auteure, l'activité telle que déployée par le formateur génère un double système d'effets : des effets sur le formateur lui-même et des effets sur la situation, en particulier, sur l'objet de l'action du formateur. L'objet de l'action du formateur porte sur le rapport entre les apprenants et le contenu enseigné. Le formateur compare

l'effet de son action au but à atteindre (but qu'il s'est donné ou qui lui a été prescrit). Le résultat de cette comparaison est à l'origine du processus de régulation de l'action (Rogalski & Colin, 2018) : si le formateur estime que l'effet de son action est trop éloigné de l'état-cible, il procède alors à des ajustements de son action, soit dans le moment même de l'action, soit à plus long terme.

2.2.2 Une activité de gestion de deux environnements dynamiques emboîtés

Dans le cadre d'une formation par simulation, le formateur met en œuvre une activité qui peut être appréhendée comme un cas particulier de gestion d'un environnement dynamique (Rogalski, 2003, 2007, 2012), à savoir un environnement qui « *a comme caractéristique d'évoluer même en l'absence d'intervention d'un acteur* » (Rogalski, 2012, p.11). Comme mentionné dans le point précédent, l'objet de l'action du formateur porte sur le rapport entre les apprenants et le contenu enseigné. Plus concrètement, le formateur cherche à modifier ce rapport de façon à atteindre des objectifs de compétence (Rogalski, 2007). Or, Rogalski (2007, p.9) précise que « *le rapport entre apprenant et contenu enseigné est évolutif* ». En effet, le développement de compétences chez les apprenants ne dépend pas uniquement des actions du formateur. En fait, *« le résultat de l'action du formateur sur les apprenants dépend à la fois de ses actions mais aussi de la dynamique propre du travail et de l'apprentissage des apprenants* » (Rogalski, 2007, p.6). Lors de la séance de simulation, le formateur est amené à gérer cet environnement puisqu'il doit agir sur la dynamique de développement des compétences chez les apprenants. Pour ce faire, le diagnostic de l'état des compétences des apprenants à un moment donné et le pronostic de ses évolutions à court et à plus long terme (après la formation) constitue une activité centrale du formateur. Ils contribuent à déterminer le choix des situations de simulation qui seront proposées aux apprenants (Vidal-Gomel, Boccara, Rogalski, & Delhomme, 2008 ; Vidal-Gomel & Rogalski, 2009).

Il convient de préciser qu'un autre processus dynamique au sein duquel la construction de compétences chez les apprenants est emboîtée doit aussi être géré au cours de la séance de simulation. Il s'agit de la situation de simulation elle-même qui est généralement caractérisée par une évolution propre (c'est-à-dire par une évolution en partie indépendante des actions des participants). Le formateur se doit donc aussi d'agir en temps réel dans le but de maintenir la situation de simulation dans la zone proche du développement de l'apprenant[1] (Vygotsky, 1934/1997). Plus concrètement, il doit analyser la situation de simulation, vérifier que les actions entreprises par l'apprenant sont et vont être pertinentes pour gérer les éventuels risques que comporte cette situation, guider l'apprenant pour l'aider à faire face aux éventuels problèmes rencontrés, voire prendre en charge (lorsque cela est nécessaire) une partie de l'activité de l'apprenant (Boccara, Vidal-Gomel, & Rogalski, 2013 ; Vidal-Gomel, Boccara, Rogalski, & Delhomme, 2008). De ce fait, le formateur est amené à gérer deux environnements dynamiques emboîtés : la dynamique propre au développement des compétences de l'apprenant et la dynamique de la situation de formation. Pour ce faire, il doit donc constamment prélever des informations sur l'activité des apprenants mais aussi sur la situation en cours (Vidal-Gomel & Rogalski, 2009).

2.2.3 Une activité de médiation

Comme le souligne le point précédent, les fonctions et missions du formateur dans le cadre d'une formation par simulation ne se résument pas à concevoir des formations ou encore à prescrire des tâches. Le formateur joue également un rôle de médiateur entre les apprenants et les compétences que ces derniers doivent construire. L'activité de médiation du formateur peut à la fois s'opérer au travers des tâches données à l'apprenant et de manière plus directe « *par*

[1] *Cette zone se situe entre la zone d'autonomie et la zone de rupture. La zone d'autonomie renvoie à la zone où l'apprenant est capable de faire la tâche de manière autonome (sans aide) tandis que la zone de rupture correspond à la zone où l'apprenant arrivera difficilement à faire la tâche même avec beaucoup d'aide. Ainsi la zone proximale de développement se définit comme la zone où l'apprenant est capable de réaliser la tâche moyennant l'aide d'autrui (Rivière, 1990).*

des actions portant sur l'activité de l'apprenant lors de la réalisation des tâches » (Rogalski, 2007, p.9). En outre, il convient de noter que cette activité se déploie lors de la conduite des trois phases de la simulation : le briefing, la séance de simulation et le débriefing.

L'activité du formateur lors du briefing

Lors du briefing, le formateur amène les apprenants à préparer et à planifier l'action qu'ils déploieront lors de la séance de simulation. Le briefing constitue également le moment propice pour négocier le contrat didactique. Rogalski (1997) souligne l'importance du contrat didactique qui renvoie aux attentes mutuelles du formateur et des apprenants ainsi qu'aux objectifs de la formation. Dans le cadre d'une formation basée sur la simulation, un déterminant commun de l'activité du formateur et de l'apprenant est le dispositif de formation lui-même (Rogalski & Colin, 2018) : l'activité du formateur tout comme celle de l'apprenant doivent poursuivre en formation un même but : acquérir (du point de vue de l'apprenant) ou faire acquérir (du point de vue du formateur) des compétences « cibles », à savoir les compétences attendues des tâches qui sont la cible de la formation. Selon ces mêmes auteurs (2018, p.8), « *un élément central dans l'articulation de l'activité de l'apprenant et du formateur en formation est donc la convergence des buts* » évoqués ci-avant. Toutefois cette convergence n'est pas donnée a priori : elle est à constituer. L'apprenant est certes un objet de l'action du formateur mais il est aussi le sujet de son activité avec des motivations et des préoccupations qui ne sont pas forcément tournées vers l'apprentissage. De ce fait, il est préconisé pour le formateur d'agir à plusieurs niveaux (Rogalski, 2007, p.14) : il doit certes « *préparer et gérer la « route didactique » conçue pour faire agir les apprenants sur des tâches visant leur apprentissage* ». Mais il lui faut également « *enrôler les apprenants dans le procédé didactique retenu* ». En effet, prescrire des tâches ne suffit pas à engager les apprenants dans l'activité voulue. Il s'avère nécessaire d'établir un contrat didactique. Cependant, plusieurs facteurs sont susceptibles de favoriser ou d'entraver cet enrôlement. En formation initiale, Rogalski et Colin

(2018) soulignent que l'organisation de la situation de simulation représente un composant de cet enrôlement et que les caractéristiques personnelles du formateur (par exemple : son expérience d'opérationnel) en est un autre. Il convient également de noter que les démarches visant l'enrôlement des apprenants ne doivent pas uniquement être entreprises au moment de l'entrée dans les tâches. Elles doivent également viser à maintenir les apprenants sur la route didactique que le formateur veut leur faire suivre (Rogalski, 2007). En cours de séance, cela peut se traduire par un rappel des conventions de l'exercice.

L'activité du formateur lors de la séance de simulation

Comme déjà évoqué dans le point 2.2.2, l'activité du formateur en séance peut être analysée en termes de gestion d'un environnement dynamique (Samurçay & Rogalski, 1998). Plus concrètement, le formateur est amené à gérer à la fois la dynamique et le tempo de la simulation et ceux de l'activité des apprenants sous des contraintes de temps liées à sa propre activité (telles que la durée fixée pour la séance). Le formateur doit élaborer un diagnostic sur la nature des problèmes rencontrés par les apprenants dans la réalisation des tâches et doit choisir d'intervenir en temps réel ou de manière différée sur l'activité de ceux-ci, mais aussi sur des paramètres de la situation simulée (c'est notamment le cas lorsqu'il prend la décision de « geler » l'évolution de la situation ou encore d'arrêter prématurément la séance). Les interventions du formateur en séance peuvent porter sur les étapes qui précèdent la réalisation d'une action : le formateur peut intervenir pour transmettre des informations ou aider au repérage d'un problème (alerte). Il peut aussi intervenir dans le cadre de l'identification d'un but. De plus, le formateur peut intervenir pendant l'action : il peut aider à exécuter cette action. Enfin, le formateur peut également intervenir lors de l'étape relative au contrôle des effets de l'action. Si les effets de l'action se révèlent trop éloignés du but visé, les interventions du formateur peuvent avoir pour but d'aider à orienter l'ajustement de l'action ou à réaliser les ajustements nécessaires (Rogalski & Colin, 2018). Globalement, Samurçay et Rogalski (1998) soulignent que l'activité du formateur en

séance peut être repartie entre trois catégories : la gestion didactique de la séance (apport de connaissances, contrôle des acquis et guidage des apprenants), la gestion de la simulation elle-même (modifications des paramètres de la situation) et la gestion de l'activité propre (gestion de la temporalité des séances, du contrat institutionnel...).

L'activité du formateur lors du débriefing

La simulation se conclut généralement sur un débriefing, durant lequel l'activité de médiation des formateurs doit concerner la compréhension de l'action mise en œuvre ainsi que des résultats de celle-ci (Olry & Vidal- Gomel, 2011). Pour le formateur, un piège fréquent doit être évité s'il veut assurer une bonne conduite du débriefing. Ce piège consiste à s'engouffrer et à persister dans un débat portant uniquement sur les aspects techniques et les prescriptions (Vidal-Gomel, Fauquet-Alekhine, & Guibert, 2011). Par ailleurs, pour un formateur, il n'est pas suffisant de pouvoir uniquement statuer, globalement, sur la réussite ou non de l'activité des apprenants. Le débriefing doit être vu comme une discussion du métier qui se base sur ce qui a été vécu en cours de simulation. La qualité du contenu du débriefing est donc dépendante de ce qui s'est passé en séance de simulation. Elle dépend également de la qualité et de la pertinence des informations recueillies par le formateur concernant l'activité des apprenants en séance. Or, la récolte de ces informations ne se révèle pas toujours aisée à réaliser compte tenu des nombreuses missions décrites ci-avant que doit remplir simultanément le formateur en simulation. Partant de ce constat, il peut être jugé pertinent de soutenir l'activité du formateur par des aides techniques lui permettant de décoder et d'analyser finement l'activité des apprenants en simulation afin de mieux paramétrer la gestion des séances et/ou des débriefings.

3. Technologies d'analyse automatique pour la simulation à visée de formation professionnelle

Les techniques de captation liées au comportement humain ont beaucoup évolué ces dernières années. On peut désormais extraire des informations très diverses liées au corps (postures, comportements sociaux liés aux espaces interpersonnels ou à la disposition relative des corps), liées au visage (expressions, âge, sexe, ethnie…), liées à la direction du visage et des yeux. Enfin, le domaine vocal permet aussi d'extraire des informations précieuses sur l'état de la personne (notamment sur son état émotionnel) sans entrer dans le domaine de la parole et donc d'une langue en particulier. Toutes ces informations peuvent être rendues nominatives puisqu'il est possible de suivre une personne pendant de bien plus longues périodes qu'auparavant grâce aux technologies utilisant de l'intelligence artificielle. Il est donc possible d'extraire en temps réel un flux d'informations de plus en plus important d'une personne ou d'un groupe de personnes dans des situations écologiques, ce qui ouvre des portes dans de nombreux domaines dont celui de la formation et, en particulier, de la formation par simulation.

3.1 Intérêt des nouvelles technologies pour l'activité du formateur en simulation

Les présupposés théoriques sous-jacents à un plaidoyer pour une assistance des formateurs reposent sur plusieurs arguments majeurs. Tous d'abord, plusieurs recherches (Labrucherie, 2011 ; Rogalski, Plat, & Antolin-Glenn, 2002 ; Salas & Cannon-Bowers, 2000) tendent à montrer que les formateurs (mêmes expérimentés) éprouvent des difficultés à observer, analyser et guider l'activité des apprenants en simulation. Autrement dit, il ne s'avère pas aisé pour un formateur de mener à bien les différentes missions qui lui sont assignées en simulation. Ensuite, dans les situations simulées sans simulateur technologique (à savoir, des formations par simulation se déroulant à partir de mises en situation), il existe peu d'aides aux formateurs leur permettant d'analyser l'activité des apprenants en cours de simulation et de mieux paramétrer la gestion des séances ou le débriefing. Les outils existants, comme The Observer ou Vosaic

Connect[2], s'ils outillent le formateur pour l'observation en live, ne réalisent cependant aucune analyse automatique des comportements et ne « déchargent cognitivement » donc en rien le formateur. Or, pour un formateur, pouvoir uniquement statuer globalement sur la réussite ou non de l'activité des apprenants est insuffisant : il est nécessaire de lui permettre de décoder et d'analyser finement cette activité afin de guider la séance de simulation et/ou de réinjecter ces données lors des débriefings. A ce niveau, les techniques de captation liées au comportement humain telles que décrites dans le point 3 de ce présent chapitre détiennent un potentiel non négligeable en matière de recueil d'informations riches et très diversifiées sur l'activité des apprenants. Certains de ces outils technologiques (par exemple : outils de détection et de suivi des mouvements du corps) peuvent contribuer à renseigner le formateur sur la part observable de l'activité des apprenants (par exemple : leurs déplacements, leurs gestes, leurs postures en séance) tandis que d'autres (par exemple : le suivi oculaire ou eye-tracking) peuvent contribuer à capter la part inobservable de cette même activité (par exemple : la prise d'informations des apprenants en séance). Enfin, ce besoin d'outils est d'autant plus crucial que, d'une part, le formateur n'est pas forcément compétent pour analyser l'activité des apprenants en cours de simulation et que, d'autre part, pour des raisons économiques et chronologiques, les séances de simulation sont le plus souvent peu nombreuses, ce qui met en exergue la nécessité d'être efficace d'emblée. En outre, le débriefing suit rapidement la séance de simulation proprement dite, laissant peu de temps à une préparation poussée.

[2] *Pour plus d'informations sur l'outil, voir : https://vosaic.com/products/vosaic-connect*

3.2 Avancées des technologies d'analyse utiles pour la simulation à visée de formation professionnelle

3.2.1 Détection et suivi des mouvements du corps

Dans le but de pouvoir analyser le comportement qu'exprime une personne, il faut d'abord être capable de détecter et de suivre ses mouvements d'une manière suffisamment précise. Afin de représenter un individu et de pouvoir facilement suivre ses mouvements tout en gardant une fidélité, son corps est modélisé par un squelette numérique. Ce squelette se compose de « points caractéristiques » reliés par des « bâtonnets » (Figures 1 et 2) et dont leur suivi 3D au cours du temps permet une représentation du mouvement du corps de la personne. Ces points et bâtonnets peuvent être assimilés aux articulations et aux os d'un squelette humain simplifié.

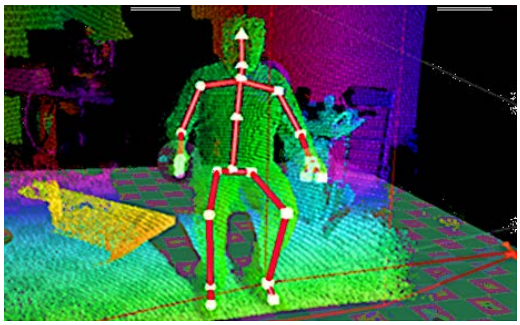
Aujourd'hui, il existe deux familles de méthodes de capture de mouvements (MOCAP) : la capture de mouvements sans marqueur et celle avec marqueurs. La capture de mouvements avec marqueurs nécessite des équipements couteux et encombrants ainsi que le port de costumes équipés : 1) de marqueurs pouvant par exemple réfléchir une lumière infrarouge ou 2) de capteurs actifs permettant de fournir des informations de position (gyroscope, accéléromètre, magnétomètre…). Invasive mais qualitative, cette approche est toujours utilisée dans des applications nécessitant une grande précision, comme pour l'industrie du cinéma ou du jeu vidéo. En ce qui concerne la capture de mouvements sans marqueur, le suivi du squelette requiert une ou plusieurs caméras idéalement munies de capteurs de profondeur qui permettent d'obtenir directement des informations de profondeur d'un objet par rapport à la caméra et donc des données en 3 dimensions pour suivre le mouvement d'un être humain.

A partir de 2010, la commercialisation par Microsoft du capteur Kinect ouvre la capture de mouvements sans marqueur au grand public. Par la suite, plusieurs constructeurs ont commercialisé des capteurs et logiciels de suivi de mouvements permettant aux chercheurs et

développeurs de créer leurs propres projets et applications de MOCAP à moindre coût (ASUS Xtion, Intel® RealSense™…). Les caméras Kinect possèdent des capteurs de profondeur. Ces caméras différenciaient le corps humain de l'arrière-plan et utilisaient des forêts d'arbres de décision pour identifier les parties du corps. Les positions étaient quant à elles identifiées à l'aide d'un certain nombre de points caractéristiques ou d'articulations appelées « joints » (telles que les épaules, les genoux, les coudes et les mains) pour former un squelette (voir Figure 1) (Shotton, et al., 2013).

Figure 1

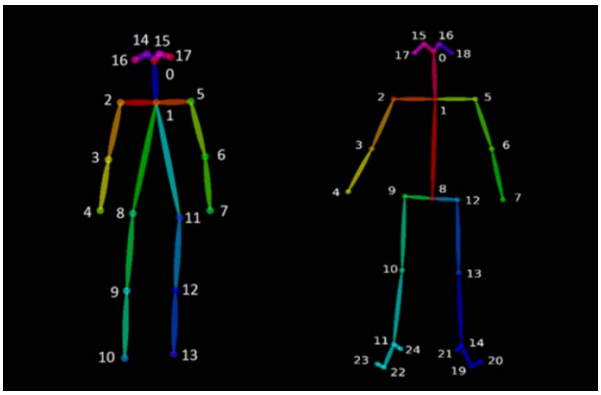
Superposition du squelette sur les données 3D capturées par la Kinect



L'arrivée des réseaux de neurones profonds ou « deep neural networks » (DNNs) a révolutionné le domaine en permettant de modéliser des squelettes plus complexes et plus stables même à partir de simples caméras 2D sans avoir besoin de capteurs spécifiques de profondeur. OpenPose est un des algorithmes pionniers dans le domaine qui a connu un énorme succès. Ce système propose un squelette de 18 points puis de 25 points plus rapide à calculer qui fonctionne directement sur des images 2D à partir de simples caméras couleurs ou noir et blanc comme celui d'images infra-rouges (Figure 2). Le calcul des squelettes implique cependant une machine puissante pour avoir des résultats en temps réel et le paiement d'une licence au coût non négligeable en cas d'utilisation commerciale (Cao, Hidalgo, Simon, Wei, & Sheikh, 2019).

Figure 2

Squelettes de OpenPose (18 points puis 25 points)



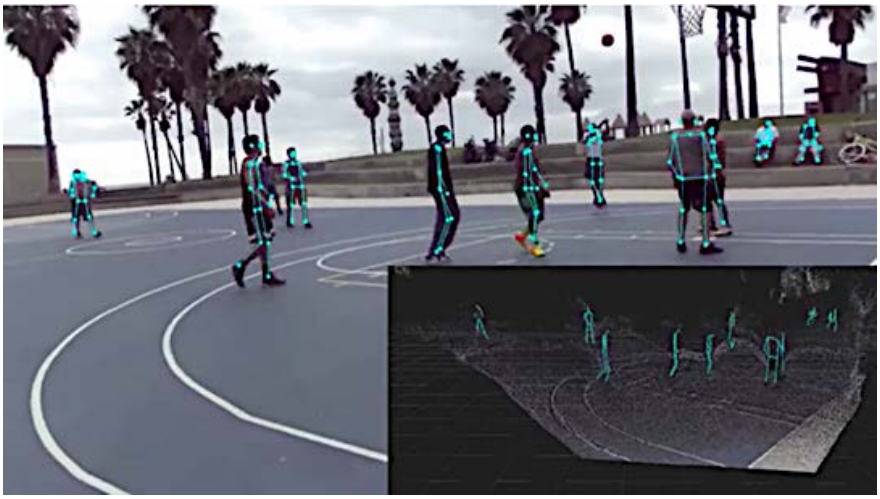
D'autres méthodes de détection de pose (à savoir, des techniques de vision par ordinateur pour suivre les mouvements d'une personne ou d'un objet), dont certaines bien plus légères en termes de calcul existent directement dans des plateformes comme TensorFlow (Abadi et al., 2015) et peuvent être implémentées même sur des téléphones portables. Pour avoir une vue plus globale sur des algorithmes de détection de pose, l'outil MMPose (MMPose Contributors, 2020) en propose, à l'heure où nous écrivons, 18 modèles.

Une fois le squelette détecté, il est alors suivi au cours du temps en se basant sur la proximité des différents points du squelette entre les différentes images de la vidéo. Cette approche fonctionne assez bien en 2D, mais performe moins bien lorsqu'une personne passe derrière une autre (occlusion). Les coordonnées 2D ne sont alors pas suffisantes. Dans ce cas, l'utilisation de caméras qui ont la possibilité de calculer la profondeur permet d'obtenir des coordonnées 3D qui permettent une résistance beaucoup plus grande aux occlusions. En outre, des caméras comme la OAK-D (OpenCV (D), 2021) (OpenCV (D-PoE), 2021) ou la ZED2 de Stereolabs (Stereolabs, 2021) (ZED2i (Stereolabs (i), 2021)) permettent d'extraire de l'information de profondeur dans des situations écologiques. Les caméras ZED2 permettent par exemple de faire un suivi plus robuste des personnes et donc une modélisation plus fidèle de leur mouvement

avec des méthodes de détection de squelette qui ne nécessitent pas des licences supplémentaires en cas d'utilisation commerciale (Figure 3).

Figure 3

Suivi des squelettes de personnes en 3D avec la caméra ZED2



Le suivi de personnes (ou « tracking ») peut être effectué en utilisant des squelettes ou bien des boites englobantes (qui vont fournir moins de données qu'un squelette). La figue 4 montre un exemple d'application de suivi de personnes dans un contexte de simulation de micro-enseignement. Les différents élèves sont détectés et chacun possède une boite englobante d'une couleur différente qui restera la même durant l'ensemble du cours. L'enseignant (ici l'apprenant en simulation) possède aussi une boite englobante qui va suivre ses mouvements. Cette simulation est relativement simple du point de vue de la détection et du suivi de personnes : en effet, la seule personne à pouvoir changer de position librement est l'enseignant, les élèves étant relativement fixes. Les résultats d'un suivi de personnes 3D sont généralement satisfaisants dans ce type d'environnement.

Figure 4

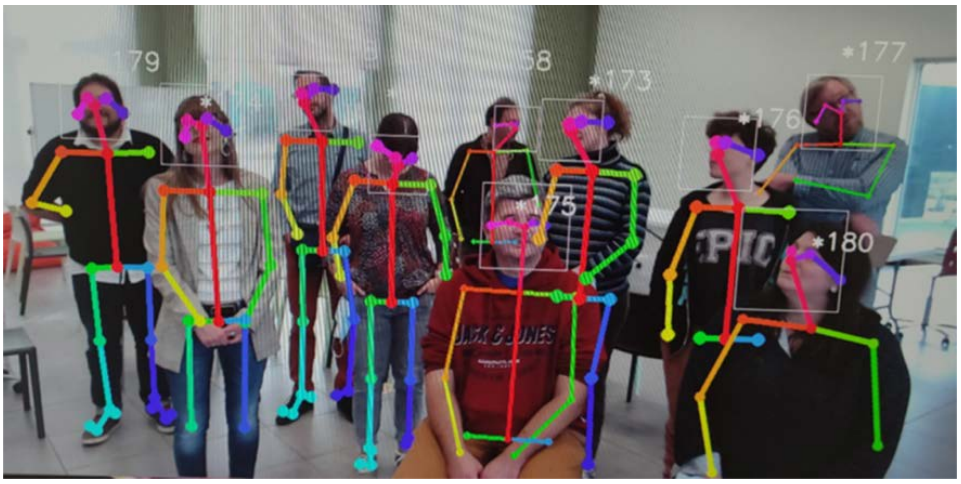
Un identifiant est attribué à chacun suivi long terme plus robuste. Suivi basé sur des boites englobantes dans un environnement de micro-enseignement.

Bien que le suivi des squelettes ou de boites englobantes se soit grandement amélioré en une dizaine d'années, des erreurs de détection et de suivi restent possibles dans des situations de mouvements plus complexes. Deux possibilités sont alors envisageables pour un suivi long-terme des squelettes. Pour éviter un maximum les occlusions, il peut s'avérer pertinent soit de placer les caméras en hauteur, soit d'utiliser plusieurs caméras. Cette démarche peut être entreprise à l'aide de caméras 2D ou directement avec des caméras 3D. Dans les deux cas, les caméras doivent avoir un moyen de synchronisation (software ou hardware) suffisamment efficace pour être en mesure de fusionner les bonnes données au bon moment. En cas de perte du suivi malgré l'approche multi-caméra, qui arrivera tôt ou tard si l'environnement à suivre est complexe, il reste une deuxième possibilité : la réidentification (REID). L'idée ici est de pouvoir reconnaître « qui est qui » au moment où les méthodes de suivi donnent une confiance de suivi faible. Dans le cas où la probabilité de perte du suivi est forte, il faut réidentifier chaque personne afin de repartir sur un suivi avec la même identité (ID) qu'avant l'évènement problématique (perte d'ancien ID et création de nouvel ID, inversion d'IDs…). La réidentification peut se baser sur l'apparence de la personne et ses mouvements (Zhou & Xiang, 2019). Pour faciliter la réidentification dans un environnement écologique, deux pistes sont à envisager. La première est l'utilisation de marqueurs visibles sur les personnes comme des QRCodes par exemple. Chaque QRCode aura un ID unique qu'il suffira de relire correctement

lorsque l'algorithme de suivi de personne est en difficulté. Toutefois, il n'est pas toujours possible d'appliquer un QRCode sur les personnes dans un environnement écologique. L'autre solution consiste alors à utiliser la reconnaissance de visages, en utilisant FaceNet (Schroff, Kalenichenko, & Philbin, 2015 ; Siv, Mancas, Sreng, Chhun, & Gosselin, 2020) (Figure 5). Cette approche présente tout de même quelques inconvénients : 1) gérer de grandes bases de données de personnes peut se révéler fastidieux et générer des difficultés en termes de gestions de données d'un point éthique, 2) la taille du visage doit être suffisamment grande sur l'image et le visage doit être suffisamment visible pour que l'algorithme fonctionne, 3) l'utilisation de masque sur le visage (période de crise sanitaire) risque d'entraver le processus d'identification et de reconnaissance des visages, 4) la personne doit rester statique pendant quelques secondes afin de permettre l'identification et la reconnaissance du visage par l'algorithme. La figure 5 montre le suivi d'un groupe dans des situations très complexes où les occlusions potentielles sont courantes. Dans ces cas, l'utilisation de la reconnaissance de visage est la seule technique sans marqueur qui fonctionne suffisamment bien pour reconnaître une personne à l'heure où nous écrivons. Les méthodes basées sur l'apparence (vêtements…) ne sont en effet pas encore suffisamment précises.

Figure 5

Un identifiant est attribué à chacun suivi long terme plus robuste. Suivi basé sur des squelettes dans un environnement de groupe complexe (en termes d'occlusions) utilisant la REID de visage



3.2.2 Analyse des mouvements du corps

Une fois le corps suivi grâce aux techniques décrites précédemment, il est possible d'en tirer des informations de haut niveau. En effet, inconsciemment, les gens créent des zones autour d'eux définissant les interactions qu'ils peuvent avoir avec leur environnement et avec d'autres personnes. La distance qui sépare deux individus, appelée distance interpersonnelle, peut donner des informations quant à leur relation sociale. Dans les années 1960, le psychologue et comportementaliste Edward T. Hall fut l'un des premiers scientifiques à proposer un modèle de relations entre les individus en fonction des distances interpersonnelles qu'il appellera « proxémie ». Les études de la proxémie menées par Edward T. Hall (1963, 1966) ont permis de décrire la manière dont l'homme organise ses distances interpersonnelles en fonction de différentes données sensorielles perçues. Ce modèle est théorique car il dépend évidemment du contexte comme la place disponible et des cultures. Il permet cependant de tirer des informations dans le cadre d'une formation par simulation. Concernant le classement des distances, 4 grandes zones sont mises en évidence :

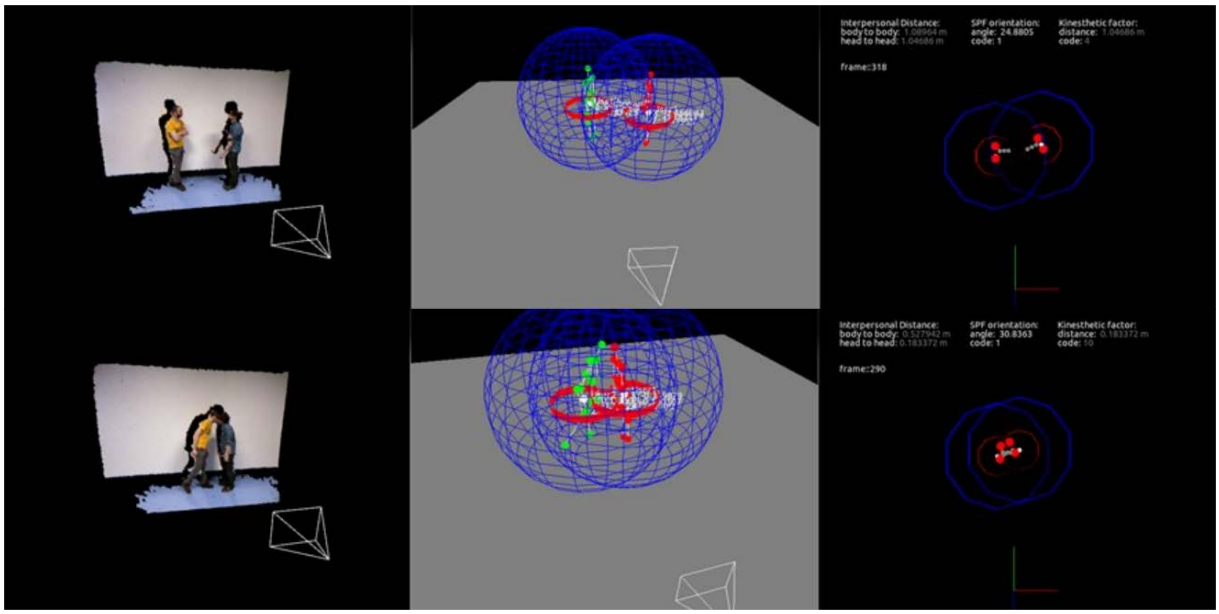
- La distance intime (de 0 à 45 cm) correspond à la distance pour toucher, murmurer ou embrasser quelqu'un. À distance intime, la présence d'une personne étrangère est inconfortable/intolérable en raison de l'apport sensoriel intensifié au travers d'élément comme l'olfaction, la chaleur du corps de l'autre, le son, l'odeur et la sensation de la respiration.
- La distance personnelle (45 cm à 1,2 m) renvoie à la distance pour interagir avec des proches. La pénétration non sollicitée de cet espace provoquera des postures défensives ou d'évitement.
- La distance sociale (1,2 m à 3,5 m) est la distance naturelle en cas de rencontre d'un étranger pour établir un processus de communication avec lui. Elle correspond à des interactions plus formelles ou impersonnelles.

- La distance publique (3,5 m à l'infini) se rapport à la distance perçue comme adéquate dans le cadre d'une réunion de groupe, salle de conférence ou interactions avec des personnalités importantes. La vision centrale permet d'englober plusieurs visages et la vision périphérique permet de voir plusieurs personnes.

Une fois la détection et le suivi de personnes effectués correctement, de nombreuses mesures de signal social peuvent être extraites (Dingler, Funk, & Alt, 2015 ; Leroy, Mancas, & Gosselin, 2011 ; Mancas, Riche, Leroy, Gosselin, & Dutoit, 2011 ; Mead & Mataric, 2016). Comme le montre la figure 6, il s'avère possible de visualiser l'intersection des zones de proxémie de deux individus représentées par les sphères colorées entourant les squelettes modélisant les deux individus. Dans un contexte de formation par simulation, la mesure et l'étude de la proxémie permet d'apporter de nombreuses informations quant aux contacts, aux relations et échanges interpersonnels entre les apprenants mais également vis-à-vis du formateur.

Figure 6

Données 3D (gauche), squelettes et espaces intimes (cylindre rouge) et personnel (sphère bleue) au centre, facteurs kinesthésiques et orientation mutuelle des personnes (droite). Résultats basés sur un capteur Kinect



3.2.3 Comportement de la tête et du visage

L'analyse de certaines caractéristiques extraites du suivi de la tête et du visage permet d'estimer l'âge, le sexe, d'analyser les expressions ou encore de déterminer la direction de la tête afin d'en déduire la direction du regard. Comme pour le corps, il y a deux familles de méthodes : celle basée sur des marqueurs, et celle sans marqueur. Les méthodes avec marqueurs exigent d'équiper chaque personne pour détecter et effectuer le suivi de leur tête. L'approche sans marqueur rend quant à elle la détection et le suivi de tête moins intrusif pour la personne observée. Dans le cadre d'une formation par simulation (cf. une situation de micro-enseignement), ce type de démarche peut permettre au formateur de récolter des informations précises sur l'état émotionnel des élèves mais aussi de l'enseignant (ici l'apprenant en simulation) grâce à l'analyse des expressions du visage. Il peut aussi contribuer à repérer parmi les élèves, ceux qui sont distraits (les élèves dont la tête est orientée en direction de la fenêtre alors que la situation ne le requiert pas) et d'établir des statistiques de participation par exemple en fonction des personnes.

3.2.3.1 Détection du visage et ses caractéristiques générales (âge, sexe, ethnie …)

Avant d'extraire des informations du visage, la première étape est d'être capable de détecter automatiquement les visages. A ce sujet, l'arrivée des réseaux de neurones profonds (DNNs) a permis le développement de modèles précis comme Dlib (King, 2009), MTCNN (Zhang, Zhang, Li, & Qiao, 2016) ou encore RetinaFace (Figure 7) (Deng, et al., 2019).

Figure 7

Détection de visages basée sur RetinaFace capable de détecter des visages quelle que soit leur taille ou orientation

Aujourd'hui, grâce à des algorithmes tels que RetinaFace, il s'avère possible d'extraire de nombreux attributs des visages observés tels que le sexe, l'âge ou encore l'ethnie. Pour ce faire, ces réseaux de neurones profonds procèdent à une analyse des visages et comparent les informations récoltées à ceux d'une base de données. Pour rendre ce travail d'analyse et de comparaison performant, il s'avère nécessaire d'intégrer dans la base de données des informations représentatives de la population en termes de sexe, d'ethnie ou de style (longueur des cheveux, barbe, lunettes...). Alors que la détection du visage est à la base des sections qui suivent, les données statistiques liées au sexe, à l'âge ou à l'ethnicité peuvent aussi présenter un intérêt pour le formateur. Y a-t-il plus de réponses et de participation de la part des élèves hommes de certaines ethnies dans un cours de sciences par exemple ? Est-ce que l'enseignant (ici l'apprenant en simulation) passe plus de temps à interagir avec certains élèves plus que d'autres ? Est-ce qu'il y a quelque chose à changer au niveau des interactions pour un enseignement plus inclusif ?

3.2.3.2 Les expressions du visage

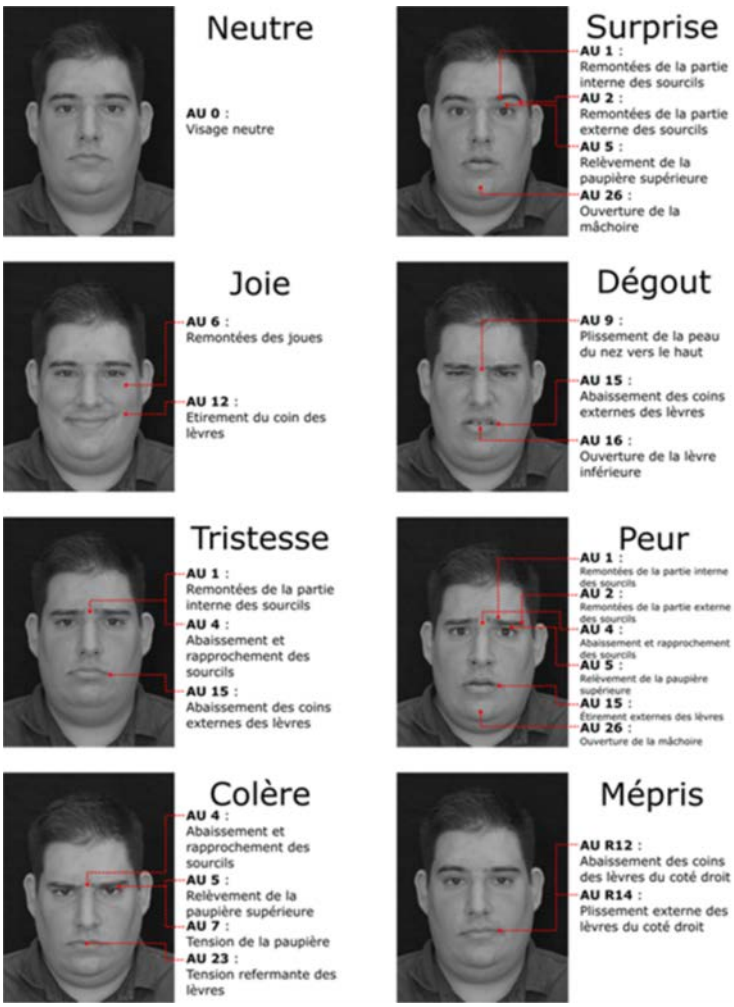
Si l'on se concentre sur l'analyse des mouvements au niveau du visage, il est possible d'estimer l'état émotionnel d'une personne. Les expressions du visage ne sont qu'un moyen parmi d'autres d'exprimer des émotions. En effet, les émotions peuvent aussi se manifester au travers de la voix (voir point 3.2.5), des gestes, des postures, etc. Les expressions faciales font pleinement partie de la communication non-verbale et peuvent être involontaires ou volontaires (clin d'œil, expression actée/simulée, etc.). Dans le cadre de la formation des enseignants (micro-enseignement), l'étude et l'analyse de l'expressions des émotions permet d'apporter des informations non verbales sur l'état émotionnel des élèves (peur, joie, etc.) ou lors des interactions entre les élèves et l'enseignant (ici l'apprenant en simulation), ainsi que lors des échanges entre les élèves.

Jusqu'à la moitié du XX^{ème} siècle, peu de travaux ont été réalisés sur l'expression des émotions chez l'homme privilégiant le fait que l'expressivité est uniquement culturelle. Ce n'est que vers 1960 que Paul Ekman entreprit des recherches détaillées sur l'expression des émotions en étudiant les contractions des muscles du visage en lien avec les émotions. Il en tira six expressions universelles : la peur, le dégoût, la colère, le bonheur, la tristesse et la surprise, sur lesquelles tous les individus s'accordent, quelle que soit leur culture (Ekman, 1971). A ces 6 expressions d'émotions, on y ajoute parfois le « mépris » considéré comme un mélange de colère et de dégout, mais non considéré comme une expression d'émotion de base (Figure 8).

Les recherche sur l'expression des émotions ont permis à Paul Ekman et à Wallace Friesen, de créer un guide de codification appelé « FACS » pour « Facial Action Coding System », publié en 1978 et révisé en 2002 (Ekman & Friesen, 1978). Le FACS est un index des expressions faciales, dont le but est de lister des unités d'action (AU) qui sont les actions fondamentales des muscles ou des groupes de muscles individuels (Figure 8).

Figure 8

Codification des expressions de base selon le FACS



La détection d'unités d'action (AUs) sur des visages présents dans des vidéos au moyen de modèles d'apprentissage automatique est actuellement envisageable. L'existence de plusieurs collections de vidéos annotées (Zhang, et al., 2014), (Mavadati, Mahoor, Bartlett, Trinh, & Cohn, 2013) permettent de concevoir et d'entraîner de tels systèmes.

L'index des expressions faciales qu'est le FACS comporte presqu'une centaine d'unités d'action. Certaines permettent de décrire les expressions principales, mais d'autres portent plutôt sur le comportement du visage, des yeux ou même des mouvements de la tête. L'analyse de ces autres mouvements caractéristiques du visage et de la tête permet par exemple d'identifier des mouvements de la mâchoire, des lèvres, des joues que l'on peut lier à la parole

(et donc au temps de parole), à la mastication, à un bâillement (et donc à la fatigue), ou encore à des grimaces. L'analyse des yeux permet aussi de coder des clignements d'œil et la fermeture des yeux (et donc la vigilance).

De nombreux modèles de détection des 6 émotions (en plus de l'émotion dite « neutre ») ont été mis au point et on retrouve les émotions aussi dans de nombreux APIs comme ceux d'IBM, de Microsoft, de Google, … (Faceplusplus, 2021). En général, ils fonctionnent assez bien sur des grands visages de face. Il convient toutefois de souligner que les résultats peuvent être très variables selon les émotions. Alors que « joie » versus « non-joie » est relativement facile à reconnaître, il en est tout autre pour les autres émotions. Des recherches sont en cours pour améliorer les bases de données existantes (Wang, Peng, Yang, Lu, & Qiao, 2020).

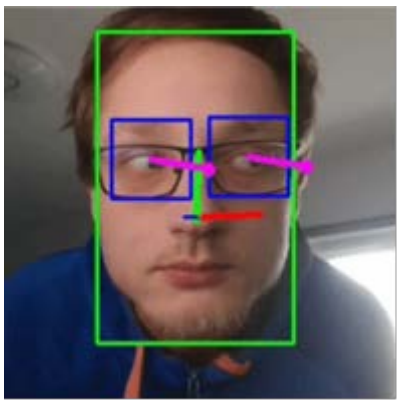
3.2.3.3 Les mouvements de la tête et des yeux

Connaître la direction du visage d'un individu nous donne des informations quant à son comportement : regarde-t-il une personne ou un objet particulier ? Sur quoi le sujet porte-t-il son attention ? Combien de temps dure la fixation de cette personne ou cet objet ?

Si le visage peut être détecté, sa direction peut l'être aussi sur des caméras 2D et encore plus grâce à des caméras RGB-D qui donnent de l'information 3D. A titre d'exemple, la caméra OAK-D light (OpenCV (Lite), 2021) permet d'extraire la direction du visage mais aussi d'obtenir une approximation de la direction du regard si le visage est suffisamment grand et que les yeux sont visibles (Figure 9).

Figure 9

Direction du visage et des yeux (OAK – D light)



Si une direction du regard performante reste très difficile à obtenir sur des images classiques et impossible lorsque les yeux deviennent petits, la direction du visage permet, dans certaines circonstances, de résoudre partiellement le problème. Des études (Langton, Honeyman, & Tessler, 2004 ; Rocca, Mancas, & Gosselin, 2014) ont montré que le regard provient d'une combinaison de la direction des yeux et de la direction de la tête et qu'à défaut de pouvoir faire le suivi oculaire (eye tracking), l'orientation de la tête donnait une indication fiable sur l'attention ou la concentration. Cette corrélation est forte dans le cas précis où l'action est proche et couvre un large espace qui implique des mouvements de tête. La distance entre le visage et le capteur est évidement fondamentale : plus la personne est éloignée du capteur qui cherche à l'analyser, plus les erreurs sur les mesures augmentent. A très courte distance (<1 mètre), le suivi oculaire est ce qui apporte les résultats les plus précis pour savoir sur quoi l'utilisateur porte son attention (Seeing Machines, 2010 ; Tobii, 2021). Au-delà de cette distance, la direction du regard peut être substituée par l'estimation de l'orientation de la tête dans certains scénarios d'utilisation comme dans le cas de l'attention visuelle face à un écran de télévision (Rocca F., et al., 2015).

3.2.4 Comportement oculaire

Le regard de l'Homme ne se pose pas sur son environnement de manière linéaire, mais au contraire, il se concentre sur des zones spécifiques de son environnement dans une exploration très dynamique (Mancas, Ferrera, Riche, & Taylor, 2016). Cette approche permet de prioriser les informations entrantes dans le cerveau en fonction de : 1) tâches/volontés précises (attention top-down) ou de 2) la difficulté de comprendre/compresser l'information, une information difficile à compresser étant vécue par une personne comme « surprenante » et donc « intéressante » (attention bottom-up). Dans ce sens, analyser le mouvement oculaire d'une personne peut livrer de nombreuses informations sur cette personne telles que son état de fatigue, son niveau de concentration ainsi que ce qui attire son attention à tel et tel moment. Il s'agit d'une information très intéressante à obtenir pour un formateur (en simulation) dans le sens où le regard est lié aux tâches qui sont effectuées (attention top-down) et qu'il peut témoigner du degré d'assimilation des connaissances en lien avec ces tâches. En outre, l'analyse du mouvement oculaire peut aussi renseigner le formateur (en simulation) sur les facteurs qui facilitent la réalisation de ces tâches ou, au contraire l'entravent comme la présence de distracteurs (attention bottom-up).

En ce qui concerne la technologie visant à capter la direction du regard, il existe trois grandes approches permettant d'extraire des données utilisables dans des situations écologiques. La première nécessite de maintenir le sujet à une distance constante de caméras spéciales qui travaillent dans l'infra-rouge. Des constructeurs tels que Tobii fournissent des systèmes de ce type sous la forme de barrettes mobiles aisément transportables et qui peuvent être branchées en USB (Tobii, 2021). Ce type d'approche implémentable sur PC et sur des tablettes vise à faire du suivi sur un écran même si l'utilisation de caméras de scène reste possible (mais plus complexes à mettre en œuvre) pour observer l'interaction de l'utilisateur avec son environnement (Figure 10, en haut).

Figure 10

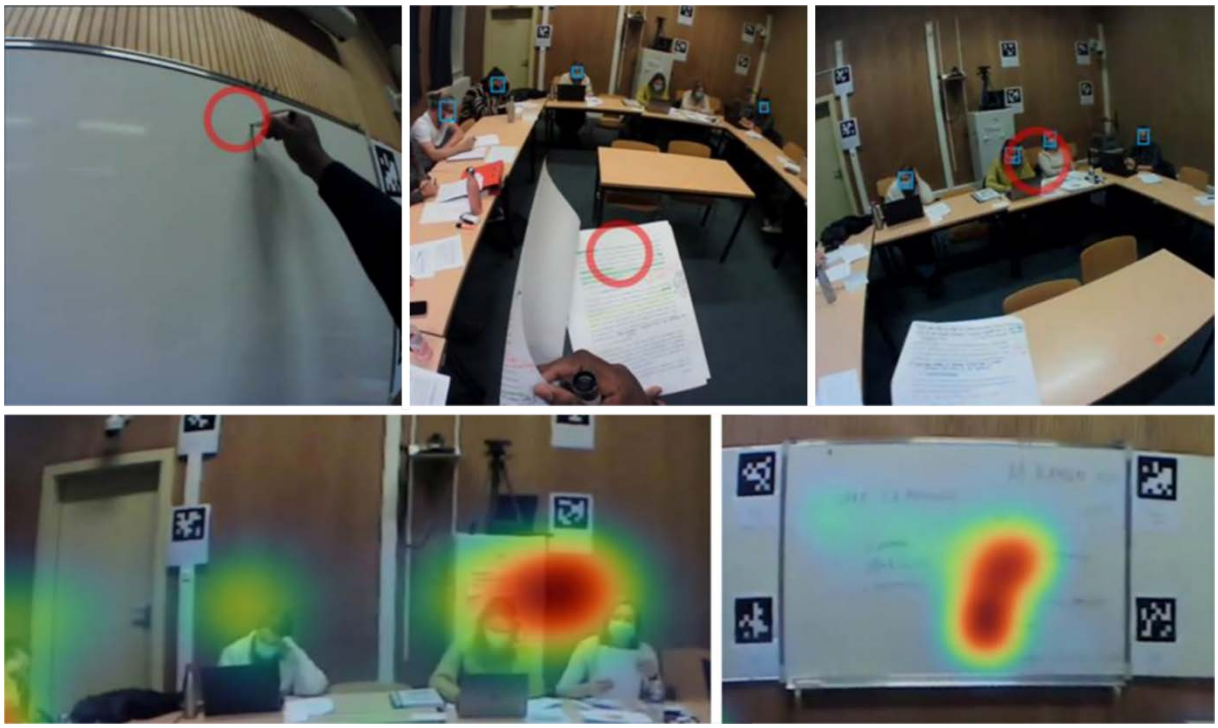
Haut : barrettes de suivi du regard (ici Tobii PCEye), Milieu : Lunettes de suivi du regard (Pupil invisible à gauche et Tobii glasses à droite), Bas (dispositifs AR à gauche et dispositifs VR à droite)

La deuxième approche consiste à placer les caméras proche de l'œil du sujet pour lequel on souhaite obtenir des informations sur le mouvement oculaire. Parmi les technologies inhérentes à ce type d'approche, on retrouve les lunettes de suivi du regard (Figure 10, au milieu). Plus discret, ce type de dispositif permet de recueillir des données relatives aux interactions de plusieurs utilisateurs avec des objets dans leur environnement réel. Le potentiel de cette technologie ne se limite donc pas au suivi du mouvement oculaire d'un sujet statique ou sur un écran. De plus, les lunettes de suivi du regard permettent d'enregistrer les données sur des petits appareils de type téléphones portables facilement transportables par les utilisateurs. Ceci laisse les mains des utilisateurs libres pour accomplir des gestes en lien avec leur activité principale. La Figure 11 montre les possibilités de retour à l'apprenant (ici le futur enseignant) qui porte des lunettes de suivi du regard. En effet, le formateur pourra, lors de la phase de débriefing, donner à l'apprenant un feedback « instantané » avec la position du regard à un moment précis

(Figure 11, en haut) ou un retour avec une carte de chaleur qui agrège toutes les données oculaires sur la période (ou une partie) du micro-enseignement (Figure 11, en bas).

Figure 11

Résultats de suivi du regard sur un futur enseignant (ici l'apprenant) dans un environnement de micro-enseignement en utilisant des lunettes Pupil Invisible. Haut : position du regard (cercle rouge), Bas : agrégation du regard depuis le début du cours sous forme de carte de chaleur



Une troisième approche, qui prend de plus en plus d'importance, est celle de l'intégration du suivi du regard dans des casques de réalité augmentée (AR) (comme les Microsoft Hololens 2) (Microsoft HoloLens, 2021) ou des casques de réalité virtuelle (VR) (tels que le HTC Vive Pro Eye)(HTC Corporation, 2021) ou le Pico Neo 3 Pro Eye (Pico Interactive, 2021)) (Figure 10, en bas). Ces technologies permettent d'immerger des apprenants dans des environnements de travail proche de la réalité, de les confronter à des situations à risques, d'urgence… pour lesquelles il s'avère nécessaire de développer des compétences professionnelles et de recueillir des informations, via l'analyse du mouvement oculaire, sur les interactions des apprenants avec les objets de leur environnement.

3.2.5 Comportement vocal

L'extraction d'informations de la voix a été traditionnellement effectuée à l'aide de valeurs appelées descripteurs ou caractéristiques. Ces descripteurs ont été pensés et conçus à partir d'équations visant à représenter ou modéliser certains phénomènes acoustiques ou même morphologiques dans la production de parole. Ces équations ont été notamment établies manuellement en étudiant le signal lui-même. Le pitch (terme utilisé dans la littérature pour désigner « la fréquence fondamentale ») ou les Mel-Frequency Cepstral Coefficients (MFCC) font partie des descripteurs traditionnels les plus connus pour la représentation de la répartition et de la dynamique fréquentielle de la voix.

Ces descripteurs ont contribué au développement de nombreuses applications technologiques liées à la parole comme la reconnaissance de locuteur, la transcription de voix en texte, la génération de voix à partir de texte, etc. Récemment une grande attention s'est portée vers les émotions et l'expressivité. Des groupes de descripteurs, dont une grande partie dérive directement du pitch et des MFCC, ont même été établis uniquement pour cette tache-là, comme les descripteurs eGeMAPs (Eyben, et al., 2015).

Avec l'avancée de l'apprentissage machine et des réseaux neuronaux profonds (DNNs) en particulier, est arrivée une nouvelle forme de représentation des descripteurs plus performante : les embeddings. De nos jours, ces embeddings sont très explorés pour l'extraction d'informations de la voix et des émotions que l'on peut y découvrir (Salamon & Bello, 2017 ; Stowell, Giannoulis, Benetos, Lagrange, & Plumbley, 2015). Parmi les systèmes les plus connus grâce à leur robustesse et potentiel de généralisation, on trouve Soundnet (Aytar, Vondrick, & Torralba, 2016) et VGG-ish (Hershey et al., 2017), système crée par Google.

Plusieurs travaux ont prouvé leur efficacité par rapport aux descripteurs traditionnels dans différents domaines comme la reconnaissance d'émotions (Nandan & Vepa, 2020 ; Tits,

Haddad, & Dutoit, 2018). L'intérêt de ces méthodes est d'utiliser la voix en dehors de toute considération liée à la langue pour y trouver des informations liées à l'état d'esprit des apprenants qui parlent plutôt qu'au contenu de leur parole.

4. L'usage d'outils technologiques dans le cadre de formations professionnelles basées sur la simulation : sous quelles conditions ?

Comme décrit dans le point 3 du chapitre, de nombreuses technologies peuvent contribuer à renseigner le formateur sur l'activité des apprenants dans le cadre d'une formation, en particulier une formation par simulation. Elles permettent ainsi de « décharger cognitivement » le formateur et de le soutenir dans son activité de médiation en séance et lors du débriefing. Toutefois l'usage d'outils technologiques dans le cadre de formations professionnelles par simulation ne peut s'opérer que sous certaines conditions.

Le premier écueil à éviter, tout comme pour la réalisation de la recherche, serait de considérer que le recours à la technologie (et ici les technologies IA centrées sur l'humain) constitue a priori la panacée ou même le gage d'une quelconque efficacité ou plus-value dans les difficultés susceptibles d'être rencontrées par le formateur en formation par simulation. Elles peuvent certes constituer une aide en matière de prise et d'analyse d'informations en séance (concernant l'activité des apprenants), et ainsi contribuer à soutenir l'activité du formateur lors de la conduite des différentes phases d'une simulation, mais elles nécessiteront toujours l'activité réflexive du formateur... qu'il faut d'ailleurs veiller à ne pas « étouffer » du fait du recours à une multitude d'informations (risque de surcharge cognitive).

L'une des pistes à privilégier pour maximiser l'impact de ces technologies reste donc bien une centration sur le formateur et sa formation, tant technique que (voire surtout) « pédagogique » à la mise en œuvre du dispositif/du scénario pédagogique recourant à la technologie. Ses compétences doivent à la fois être valorisées et perçues comme contributives à sa performance

en simulation mais également « soutenues » dans la perspective d'un développement professionnel et la mise en œuvre d'une véritable formation continuée du formateur (notamment en l'amenant à consulter les résultats de la recherche sur les dispositifs recourant aux modalités pédagogiques qu'il mobilise dans sa pratique).

Toujours dans une perspective de développement professionnel du formateur, nous soutenons l'idée que ces outils technologiques devraient également avoir pour but de recueillir des informations sur l'activité du formateur en simulation. Ces informations ainsi recueillies pourraient être exploitées ultérieurement par le formateur afin d'améliorer sa pratique professionnelle.

Partant de ces constats, l'implémentation de technologies (de la technologie) dans le cadre d'une formation par simulation doit être pensée en amont en traitant les questions suivantes : quelle acceptation par les formateurs et les apprenants (et réflexion sur le caractère intrusif ou non de la solution retenue, le type de traitement réservé aux données) ?, quel niveau de complexité en regard des compétences des formateurs et/ou des perspectives de formation ?, quel objectif et quelle efficacité visée (approprié, perçu comme tel par les formateurs, atteignable, justifiant le déploiement technologique) ?

5. Conclusion : Où en est-on et que reste-t-il encore à faire ?

Les points 3 et 4 de ce chapitre soulignent d'une part la possibilité d'extraire une multitude d'informations en temps-réel au sujet d'humains en formation et, de l'autre, la nécessité de garder le formateur au centre de la boucle d'interactions de la formation, que cela soit avant (briefing), durant la formation proprement dite et après (débriefing). Cette approche va dans le sens plus large du fait de garder l'humain dans la boucle de l'IA (Human-in-the-AI-loop) au lieu de le remplacer complétement. Quand on regarde les applications créatives de l'IA, à chaque fois que l'IA est laissée « seule » le résultat est pour le moins étrange et pas très qualitatif

(pensons aux IA qui écrivent seules des livres par exemple). L'IA est là pour lui fournir des informations objectives sur les apprenants et des propositions de modification du processus de formation en temps réel d'une part et un résumé de l'expérience de formation pour le débriefing offline.

Pour les propositions de changement en temps réel dans le scénario de formation, il s'agit aussi d'expliquer pourquoi la proposition de changement de scénario est effectuée et laisser la possibilité au formateur de valider ou non la proposition. En cas de validation, il revient également au formateur de choisir le moment où ce changement dans le scénario peut être le plus efficace.

Pour les propositions de débriefing, l'expérience de formation est automatiquement annotée avec des informations compréhensibles par le formateur et liées à l'activité des apprenants en séance de simulation.

Dans les deux cas, le système doit recueillir, en plus des informations relatives à l'activité des apprenants, la réaction du formateur par rapport aux propositions de l'IA ou par rapport aux annotations qu'elle produit afin d'améliorer le système de recommandation et d'annotation automatique dans le temps. Le système doit également donner un feedback au formateur lequel va peut-être, en fonction, adapter sa façon de faire. Il s'agit donc bien de chercher une symbiose entre le formateur et l'IA pour servir les objectifs de la formation professionnelle.

En ce qui concerne les problèmes éthiques liés à l'extraction de données sur les apprenants et éventuellement le formateur, il s'agit d'un réel point d'attention. Les systèmes mis en place doivent être pensés éthiquement dès le départ afin qu'ils soient acceptables et acceptés par les utilisateurs dans des sociétés démocratiques où il est possible de refuser l'utilisation de systèmes technologiques. Il faut donc penser à utiliser des données globales agrégées plutôt que des données que l'on peut affecter à une personne précise. Le fait de ne pas enregistrer des

données compréhensibles par de humains (vidéos, visages, etc.) mais seulement des informations de haut niveau est très important. Pour cela il faut être capable de faire un pré-traitement des données au plus près des capteurs en utilisant des algorithmes suffisamment légers et rapides pour cela. L'éthique est donc un facteur de créativité important dans les technologies liées à l'IA car il pose des défis technologiques très intéressants et il est très important de considérer le facteur éthique comme un allié du système plutôt qu'un obstacle.

Alors, l'IA dans la formation : rupture radicale ou continuité ? Tout semble indiquer qu'il s'agit bien d'une continuité où les deux mondes, celui de la formation et celui de l'IA doivent apprendre l'un de l'autre afin de pouvoir vivre dans une relative symbiose. Dans ce sens, le défi technologique le plus important dans la formation n'est paradoxalement pas la quantité d'informations qu'il est possible d'extraire des humains et de leurs interactions, mais bien la manière de simplifier ces informations. Parallèlement, la conception de l'interface qui permettra aux formateurs d'utiliser toute cette information sans rajouter une charge cognitive et qui permettra une réelle plus-value dans la compréhension de la formation par le formateur constituera un maillon important du dispositif. Idéalement, l'enregistrement de l'utilisation de cette interface permettra d'étudier l'activité du formateur.

A court terme, en ce qui concerne les développements, c'est l'étape permettant au formateur de réaliser le débriefing offline qui apparait la plus plausible. L'enrichissement de l'interface d'annotations et la proposition des zones problématiques qui nécessitent que le formateur se penche dessus lors du débriefing est la tâche la plus réaliste d'un point de vue technologique dans un horizon à court terme. Par zones problématiques, il s'agit de zones « atypiques » par rapport au développement normal de l'activité... cette détection pouvant potentiellement être réalisée automatiquement par rapport à l'étude d'un grand nombre de scènes similaires à celles reproduites en simulation. A moyen et plus long terme, l'aide en temps réel du formateur durant la formation est la prochaine étape dans le mouvement de transformation de la formation.

Références bibliographiques

Abadi, M. a., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., . . . Devin, M. (2015). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. Retrieved from https://www.tensorflow.org/

Aytar, Y., Vondrick, C., & Torralba, A. (2016). Soundnet: Learning sound representations from unlabeled video. *Advances in neural information processing systems*, 892-900.

Béguin, P., & Weill-Fassina, A. (1997). De la simulation des situations de travail à la situation de simulation. Dans P. Béguin & Weill-Fassina (dirs.), *La simulation en ergonomie : connaître, agir et interagir* (pp. 5-28). Octarès.

Boccara, V., Vidal-Gomel, C., & Rogalski, J. (2013, 5-7 juin). *Analyse multiniveaux de l'activité de médiation des formateurs* [Communication]. Colloque international : Les question vives en éducation et formation : regards croisés France-Canada, Nantes (France). https://www.researchgate.net/publication/296831739_Analyse_multiniveaux_de_l%27activite_de_mediation_des_formateurs

Caens-Martin, S. (2009). Concevoir un simulateur pour apprendre à gérer un système vivant à des fins de production : la taille de la vigne. Dans P. Pastré & P. Rabardel (dirs.), *Apprendre par la simulation. De l'analyse du travail aux apprentissages professionnels* (pp. 81-106). Octarès.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence, 43*(1), 172-186.

Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., & Zafeiriou, S. (2019). Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641*.

Dingler, T., Funk, M., & Alt, F. (2015). Interaction proxemics: Combining physical spaces for seamless gesture interaction. *Proceedings of the 4th International Symposium on Pervasive Displays*, 107-114.

Dubois, L-A., Bocquillon, M., Romanus, C., Derobertmasure, A. (2019). Usage d'un modèle commun de la réflexivité pour l'analyse de débriefings post-simulation : le cas de futurs policiers, sages-femmes et enseignants. *Le travail humain, 82*(3), 213-251.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotions. *Nebraska Symposium on Motivation*, 207–283.

Ekman, P., & Friesen., W. (1978). *Facial Action Coding System (FACS): Manual.* Consulting Psychologists Press.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre}, E., Busso, C., . . . Narayanan, S. (2015). F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," in IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190-202, 1 April-. *IEEE transactions on affective computing*, 190-202.

Faceplusplus. (2021). Retrieved from Face ++: https://www.faceplusplus.com

Hall, E. T. (1963). A system for the notation of proxemic behavior. *American anthropologist, 65*(5), 1003-1026.

Hall, E. T. (1966). *The Hidden Dimension* (Vol. 6). Doubleday.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J., Jansen, A., Moore, C., . . . Wilson, K. (2017). CNN architectures for large-scale audio classification. *IEEE ICASSP, New Orleans, LA, USA*, 131-135.

HTC Corporation. (2021, 10). *Vive Pro Eye*. Retrieved from Vive: https://www.vive.com/fr/product/vive-pro-eye/overview/

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 1755-1758.

Labrucherie, M. (2011). Le pilotage des avions de ligne. Dans Ph. Fauquet-Alekhine & N. Pehuet (dirs.), *Améliorer la pratique professionnelle par la simulation* (pp. 9-36). Octarès.

Langton, S. R., Honeyman, H., & Tessler, E. (2004). The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & psychophysics, 66*(5), 752-771.

Leplat J., & Hoc J-M. (1983). Tâche et activité dans l'analyse psychologique des situations. *Cahiers de Psychologie cognitive, 3*(1), 49-63.

Leroy, J., Mancas, M., & Gosselin, B. (2011). Personal space augmented reality tool. *First joint WIC/IEEE SP Symposium on Information Theory and Signal Processing in the Benelux*.

Mancas, M., Ferrera, V. P., Riche, N., & Taylor, J. G. (2016). From Human Attention to Computational Attention. *Springer, 2*.

Mancas, M., Riche, N., Leroy, J., Gosselin, B., & Dutoit, T. (2011). Toward a social attentive machine. *2011 AAAI Fall Symposium Series*.

Mavadati, S., Mahoor, M., Bartlett, K., Trinh, P., & Cohn, J. (2013). DISFA: A spontaneous facial action intensity database. *Affective Computing, IEEE Transactions*, 151-160.

Mead, R., & Mataric, M. J. (2016). Perceptual models of human-robot proxemics. *Experimental robotics. Springer*, 261-276.

Microsoft. (2021). *Kinect pour Windows*. Retrieved from https://developer.microsoft.com/fr-fr/windows/kinect/

Microsoft HoloLens. (2021, 10). *HoloLens 2*. Retrieved from https://www.microsoft.com/fr-fr/hololens/buy

MMPose Contributors. (2020, 8). *OpenMMLab Pose Estimation Toolbox and Benchmark*. Retrieved from https://github.com/open-mmlab/mmpose

Nandan, A., & Vepa, J. (2020). Language agnostic speech embeddings for emotion classification.

Olry, P., & Vidal-Gomel, C. (2011). Conception de formation professionnelle continue : tensions croisées et apports de l'ergonomie, de la didactique professionnelle et des pratiques d'ingénierie. *Activités, 8*(2), 115-149. https://doi.org/10.4000/activites.2604

OpenCV (D). (2021). *OpenCV AI Kit: OAK—D*. Retrieved from OpenCV Store: https://store.opencv.ai/products/oak-d

OpenCV (D-PoE). (2021). *OpenCV AI Kit: OAK—D-PoE*. Retrieved from OpenCV Store: https://store.opencv.ai/products/oak-d-poe

OpenCV (Lite). (2021). *OpenCV AI Kit - Lite (and Tiny)*. Retrieved from Kickstarter: https://www.kickstarter.com/projects/opencv/opencv-ai-kit-oak-depth-camera-4k-cv-edge-object-detection/posts

Pastré, P. (2009). Apprendre par la résolution de problèmes : le rôle de la simulation. Dans P. Pastré & P. Rabardel (dirs.), *Apprendre par la simulation. De l'analyse du travail aux apprentissages professionnels* (pp. 17-40). Toulouse : Octarès.

Pico Interactive. (2021, 10). *Neo 3 pro - Neo3 pro eye*. Retrieved from https://www.pico-interactive.com/us/neo3.html

Rivière, A. (1990). Les relations entre apprentissage et développement. La zone proximale de développement. Dans A. Rivière (dir.), *La psychologie de Vygotsky* (pp. 89-95). Mardaga.

Rocca, F., De Deken, P., Grisard, F., Mancas, M., & Gosselin, B. (2015b). Real-time marker-less implicit behavior tracking for user profiling in a TV context. *28th International Conference on Computer Animation and Social Agents (CASA 2015)*.

Rocca, F., Mancas, M., & Gosselin, B. (2014). Head pose estimation by perspective-n-point solution based on 2d markerless face tracking. *International Conference on Intelligent Technologies for Interactive Entertainment*, 67-76.

Rocca, F., Mancas, M., Grisard, F., Leroy, J., Ravet, T., & Gosselin, B. (2015a). Head pose estimation \& TV Context: current technology. *Rocca, Francois and Mancas, Matei and Grisard, Fabien and Leroy, Julien and Ravet, Thierry and Gosselin, Bernard, 2*(3).

Rogalski, J. (1997). Simulations : fonctionnalités ? validité ? Dans P. Béguin & A. Weill-Fassina (dirs.), *La simulation en ergonomie : connaître, agir et interagir* (pp. 55-76). Octarès.

Rogalski, J. (2003*).* Y a-t-il un pilote dans la classe ? Une analyse de l'activité de l'enseignant comme gestion d'un environnement dynamique ouvert. *Recherches en Didactique des Mathématiques, 23*(3)*, 343-388.*

Rogalski, J. (2007, 11-15 juin). *Approche de psychologie ergonomique de l'activité de l'enseignant* [Communication]. Séminaire international : La professionnalisation des enseignants de l'éducation de base : les recrutements sans formation initiale, Sèvres. https://www.archives.philippeclauzard.com/Rogalski_ApprocheErgoActiviteEnseignant.pdf

Rogalski, J. (2012). Théorie de l'activité et didactique, pour l'analyse conjointe des activités de l'enseignant et de l'élève. *International Journal for Studies in Mathematics Education, 5*(1), 1-37.

Rogalski, J., & Colin, B. (2018). Le rôle du formateur dans l'articulation des compétences acquises sur simulateur et des compétences cibles ("terrain"). Le cas du moniteur dans la formation de pilotes militaires d'hélicoptères - armée de Terre. *Activités, 15*(2), 1-25. https://doi.org/10.4000/activites.3333

Rogalski, J., Plat, M., & Antolin-Glenn, P. (2002). Training for collective competence in rare and unpredictable situations. Dans N. Boreham, R. Samurçay & M. Fischer (dirs.), *Work process knowledge* (pp. 134-147). Routledge.

Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*, 279-283.

Salas, E., & Cannon-Bowers, J. A. (2000). The anatomy of team training. Dans S. Tobias & J.D. Fletcher (dirs.), *Training and retraining: A handbook for business, industry, government, and the military* (pp. 312– 335). Macmillan Reference.

Samurçay, R. (2009). Concevoir des situations didactiques pour la formation professionnelle : une approche didactique. Dans P. Rabardel & P. Pastré (dirs.), *Modèles du sujet pour la conception* (pp. 53-72). Octarès.

Samurçay, R., & Rogalski, J. (1998). Exploitation didactique des situations de simulation. *Le travail humain, 61*(4), 333-359.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815-823.

Seeing Machines. (2010). FaceLAB.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., . . . Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM, 56*(1), 116-124.

Siv, R., Mancas, M., Sreng, S., Chhun, S., & Gosselin, B. (2020). People Tracking and Re-Identifying in Distributed Contexts: PoseTReID Framework and Dataset. *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 323-328.

Stereolabs (i). (2021). *ZED 2i - Industrial AI Stereo Camera | Stereolabs*. Retrieved from Stereolabs: https://www.stereolabs.com/zed-2i/

Stereolabs. (2021). *ZED 2 - AI Stereo Camera | Stereolabs*. Retrieved from Stereolabs: https://www.stereolabs.com/zed-2/

Stowell, D., Giannoulis, D., Benetos, E., Lagrange, M., & Plumbley, M. D. (2015). Detection and classification of acoustic scenes and events}. *IEEE Transactions on Multimedia*, 1733-1746.

Tits, N., Haddad, K. E., & Dutoit, T. (2018). Asr-based features for emotion recognition: A transfer learning approach. *arXiv preprint arXiv:1805.09197*.

Tobii. (2021). *Tobii - Hardware, software, and services*. Retrieved from Tobii Tech: https://tech.tobii.com/products/

Vidal-Gomel, C., Boccara, V., Rogalski, J., & Delhomme, P. (2008). Les activités de guidage des formateurs au cours d'un audit destiné à des conducteurs expérimentés et âgés. *Travail et Apprentissage, 2*, 46-64.

Vidal-Gomel, C., Fauquet-Alekhine, P., & Guibert, S. (2011). Réflexions et apports théoriques sur la pratique des formateurs et de la simulation. Dans Ph. Fauquet-Alekhine & N. Pehuet (dirs.), *Améliorer la pratique professionnelle par la simulation* (pp. 115-141). Octarès.

Vidal-Gomel, C., & Rogalski, J. (2009). Analyser l'activité des formateurs en conduite automobile : une étude exploratoire des aspects collectifs du travail. *Savoirs, 20*(2), 85–118. https://www.cairn.info/revue-savoirs-2009-2-page-85.htm

Vygotski, L. (1934/1997). *Pensée et Langage*. La dispute.

Wang, K., Peng, X., Yang, J., Lu, S., & Qiao, Y. (2020). Suppressing uncertainties for large-scale facial expression recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6897-6906.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 1499-1503.

Zhang, X., Yin, L., Cohn, J., Canavan, S., Reale, M., Horowitz, A., . . . Girard, J. (2014). BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image and Vision Computing*, 692-706.

Zhou, K., & Xiang, T. (2019). orchreid: A library for deep learning person re-identification in pytorch. *arXiv preprint arXiv:1910.10093*.