



# Contribution à l'apprentissage humain de gestes à l'aide de techniques de clustering pour l'analyse de mouvements capturés

Quentin Couland

## ► To cite this version:

Quentin Couland. Contribution à l'apprentissage humain de gestes à l'aide de techniques de clustering pour l'analyse de mouvements capturés. Environnements Informatiques pour l'Apprentissage Humain. Le Mans Université, 2020. Français. NNT : 2020LEMA1011 . tel-03052482

**HAL Id: tel-03052482**

**<https://theses.hal.science/tel-03052482>**

Submitted on 10 Dec 2020

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

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

# THESE DE DOCTORAT DE

LE MANS UNIVERSITE

ECOLE DOCTORALE N° 601

*Mathématiques et Sciences et Technologies  
de l'Information et de la Communication*

Spécialité : Informatique (CNU 27)

Par

**Quentin COULAND**

**Contribution à l'apprentissage humain de gestes à l'aide de techniques  
de clustering pour l'analyse de mouvements capturés**

Thèse présentée et soutenue à Laval, le 27/05/2020

Unité de recherche : LIUM

Thèse N° : 2020LEMA1011

## Rapporteurs avant soutenance :

Franck HETROY-WHEELER  
Vanda LUENGO

Professeur des Universités  
Professeur des Universités

Université de Strasbourg  
Sorbonne Université

## Composition du Jury :

Examineurs : Jules FRANÇOISE

Franck HETROY-WHEELER  
Richard KULPA  
Vanda LUENGO

Chargé de Recherche CNRS

Professeur des Universités  
Maître de Conférences, HDR  
Professeur des Universités

Université Paris-Sud - Université  
Paris-Saclay  
Université de Strasbourg  
Université Rennes 2  
Sorbonne Université

Directeur de thèse : Sébastien GEORGE Professeur des universités

Le Mans Université

Co-encadrant. de thèse : Ludovic HAMON Maître de Conférences

Le Mans Université



---

## REMERCIEMENTS

---

Mes remerciements à **Sébastien George** et **Ludovic Hamon**, pour m'avoir accepté en thèse, pour l'encadrement de cette thèse, la bonne humeur, les conseils toujours pertinents, la patience, le soutien tout au long des étapes de cette thèse, ainsi que pour m'avoir apporté la rigueur qui me faisait tant défaut avant d'arriver en thèse.

Je souhaite remercier également l'équipe du LITIS, qui m'a initié au monde de la recherche, et plus particulièrement messieurs **Sébastien Adam**, **Pierre Héroux** et **Laurent Heutte**, pour m'avoir chacun à leur manière fait découvrir la recherche et donné une forte envie et motivation pour continuer dans ce domaine.

Merci à mes collègues doctorants, ex-doctorants, docteurs ou post-doctorants : **Guillaume**, **Inès**, **Damien**, **Vincent**, **Aïcha**, **Aous**, **Esteban**, **Zeyneb**, **Jean**, **Oussema**, **Claire**, **Dalal**, **Ibtissem**, **Pierre-Yves** et tous les autres que j'oublie peut-être (pardon), pour avoir fait du temps passé au CERIUM<sup>2</sup> un agréable moment. Je souhaite remercier particulièrement **Inès Dabbebi** pour m'avoir supporté pendant 2 longues années, moi, ma musique, mes thés, ma nourriture dans le bureau, mes gros mots, ainsi que **Guillaume Loup**, pour sa gentillesse, son aide et son soutien autant technique que moral, et de manière générale pour avoir été mon parrain de thèse pendant ces quelques années passées ici. On notera également la patience légendaire dont il a dû faire preuve pour me supporter pendant son temps ici, car il n'a jamais pu se débarrasser de moi, même quand il a déménagé en info, où je l'ai suivi (pour le meilleur, et pas vraiment de pire au final). Les loooooo(ooongues conversations que nous avons (souvent) eu tous les trois resteront un souvenir très agréable pour moi.

Merci à tous les membres de l'équipe **IEIAH** et plus largement du **LIUM** pour leur accueil. Je remercie également tous mes collègues de l'**IUT de Laval**, en particulier ceux du **département Informatique**. Un grand merci au **service administratif de l'IUT de Laval** également, grâce à qui toutes mes démarches administratives se sont toujours très bien passées, ainsi qu'à **Elisabeth** et **Marie**, nos deux secrétaires du département informatique.

Merci à **Yann Walkowiak** et **Pierre Laforcade** pour les conversations intéressantes et les blagues d'un genre douteux comme je les aime, **Clément Laborie** pour avoir été mon bae ici (et je pense qu'il peut aussi me remercier pour avoir été son exutoire contre la SNCF), et enfin à **Nathalie Vieillard** pour avoir été ma mère adoptive de l'IUT de Laval.

Depuis petit, j'ai toujours voulu enseigner. Lors de ma thèse, j'ai eu la chance de pouvoir donner des cours, et je tenais à remercier l'ensemble des promotions auxquelles j'ai enseigné, car ça n'a été que du plaisir (même si j'ai parfois *légèrement* haussé la voix). Plus particulièrement, j'aimerais remercier **Hugo Morali**, **Florian Pellegrin**, **Antoine Lambert**, **Clément Fievez**, **Alexis Nolat**, **Maëli Coué**, **Justin Martin**, **Bibite**, **Quentin Pineau**, **Evan Delaunay**, **Euntoine**, **Florian Plaut** et une bonne partie des sous-doués de la promo 2018-2020 (ça fait un peu trop de monde,

pas envie de tout écrire).

Sans **Valval**, **Princess Mayva**, **Léo the tasty virgin**, **Queen Juju** et bien évidemment **Marie Sproutch Bouin**, la 3ème année de thèse n'aurait pas été aussi bien que ce qu'elle a été. Un grand merci à eux pour les nombreuses visites et excellents moments passés (ainsi que les innombrables thés), autant dans le bureau qu'en dehors de l'IUT.

Merci également à tous mes amis restés sur Rouen, pour m'avoir permis de passer des moments agréables quand je rentrais : **PH**, **Fourré**, **Dodu**, **Dandelinos**, **Kiki**, **Guigui**, et tous les autres. Même s'ils ne sont pas à Rouen, je tiens à remercier plusieurs autres personnes, à commencer par **Bastien Resse**, pour être l'homme de classe et de culture qu'il est, ayant même réussi à m'intéresser à la politique sous un angle neutre. Également le tenant de ma maison sur Paris, **Freddy " DiFFtY "**, qui me supporte depuis beaucoup trop longtemps pour quelqu'un qui serait sain d'esprit, ainsi que **Chacha la dame du CDI**, deux personnes avec qui j'ai toujours passé d'excellents moments (et dieu sait que je n'aime pourtant pas Paris). Je remercie également (sous la pression) **Lussie Leroy**, bien que ce soit plutôt à elle de me remercier au final, vu les efforts que je doit déployer pour la supporter au quotidien, elle et ses excentricités.

Je pense que sans **PH** et **Dandelinos** en vocal presque tous les soirs depuis que je suis arrivé à Laval, je n'aurais pas tenu aussi longtemps sans devenir fou dans cette ville. Aussi, il me semble naturel de les remercier tout particulièrement pour les discussions distinguées, intelligentes et matures (150 de QI à nous trois), les games de LoL, le coaching de qualité sur le ladder SC2 et enfin pour les games de CS:GO où je carry toute l'équipe.

Je remercie ma **famille** pour ses encouragements, et en particulier mes **parents** pour m'avoir toujours soutenu lors de mes études et avoir été fier de moi, même quand j'ai parfois pu faire des choix à première vue un peu douteux. Sans leur soutien moral ainsi que financier (car oui, ça coûte des sous de louer un appartement, s'inscrire à la fac, *etc.*), rien de tout ça n'aurait été possible.

Enfin, je voudrais terminer par une citation qui m'a inspiré, et qui devrait tous nous inspirer, à toujours aller de l'avant quelles que soient les difficultés rencontrées :

*« Qu'est-ce que tu crois qu'il aurait fait Toto à ta place, tu crois qu'il aurait lâché l'affaire, qu'il aurait abandonné comme ça Toto, hein ? Non Toto il aurait continué, ouais Toto il aurait continué ouais. »*

*Cette thèse est dédiée à une personne exceptionnelle  
que j'ai eu la chance de rencontrer ici.  
Sa seule rencontre fait que cette thèse  
valait la peine d'être faite.  
Quoique l'avenir nous réserve,  
Je chérirai toujours les souvenirs que je garde  
de son passage dans ma vie,  
et les moments que nous avons passés ensemble ici.*



*Merci pour tout.*



---

## TABLE DES MATIÈRES

---

### I INTRODUCTION

1	CADRE ET CONTEXTE	15
1.1	Environnements Informatiques pour l'Apprentissage Humain (EIAH)	16
1.2	Le mouvement humain . . . . .	16
1.2.1	Définition . . . . .	16
1.2.2	Méthodes d'apprentissage du mouvement . . . . .	17
1.2.3	Apprentissage de gestes avec des d'EIAH . . . . .	17
1.2.4	Analyse automatique par méthodes d'apprentissage supervisé / non-supervisé . . . . .	18
1.2.5	Limites des systèmes existants . . . . .	19
1.3	Problématiques de recherche . . . . .	20
1.3.1	Verrous scientifiques . . . . .	20
1.3.2	Questions de recherche . . . . .	21
1.3.3	Hypothèses de recherche . . . . .	21
1.4	Aperçu des contributions . . . . .	22
1.4.1	Motion learning Analytics (MLA) . . . . .	22
1.4.2	Traitement et segmentation des mouvements . . . . .	23
1.4.3	Séparation des données de mouvement et sémantique associée	23
1.4.4	Système de retour visuel . . . . .	23
1.5	Expérimentations . . . . .	23
1.6	Plan du manuscrit . . . . .	23

### II ÉTAT DE L'ART

2	LES EIAH POUR L'APPRENTISSAGE DE GESTES	27
2.1	Représentation du mouvement . . . . .	27
2.2	Matériel de capture et données obtenues . . . . .	29
2.2.1	Extraction de mouvement par vidéo . . . . .	30
2.2.2	Extraction de mouvement par vidéo et capteur de profondeur	30
2.2.3	Extraction de mouvement par caméra infrarouge . . . . .	31
2.2.4	Extraction de mouvement par capteurs inertiels portables . . .	32
2.2.5	Matériel de captation pour la manipulation d'objets . . . . .	33
2.2.6	Bilan sur les dispositifs de captation de mouvement . . . . .	35
2.3	Environnements informatiques pour l'apprentissage de gestes . . . . .	36
2.3.1	Quelques exemples significatifs d'EIAH pour apprendre des gestes . . . . .	36
2.3.2	Tableau de synthèse . . . . .	43
2.4	Discussion . . . . .	46
3	DESCRIPTEURS DU MOUVEMENT	49
3.1	Définition et principes . . . . .	49
3.2	Revue des descripteurs existants, classification et modèle de représentation . . . . .	50
3.2.1	Descripteurs pour l'analyse de la performance . . . . .	51
3.2.2	Descripteurs pour l'indexation et la recherche de mouvements	55
3.2.3	Descripteurs pour la reconnaissance d'actions . . . . .	57
3.2.4	Discussion . . . . .	58
3.3	Classification de descripteurs élémentaires . . . . .	60



3.4	Discussion . . . . .	64
4	ANALYSE DU MOUVEMENT CAPTURÉ . . . . .	65
4.1	Analyse du geste par observation humaine . . . . .	65
4.2	Analyse par observation d'indicateurs calculables du geste . . . . .	68
4.3	Analyse par détection d'une séquence ordonnée d'actions . . . . .	72
4.4	Analyse fondée sur des techniques d'apprentissage automatique . . . . .	74
4.5	Bilan des systèmes existants et positionnement . . . . .	75
 <b>III CONTRIBUTIONS</b>		
5	TRAITEMENT DES DONNÉES DE MOUVEMENT . . . . .	83
5.1	Motion Learning Analytics (MLA) . . . . .	83
5.1.1	Données de mouvement au format BVH . . . . .	84
5.1.2	Pré-traitements des données . . . . .	85
5.1.3	Analyse automatique . . . . .	87
5.1.4	Feedback . . . . .	88
5.2	Capture de mouvement (Perception Neuron) . . . . .	88
5.3	Pré-traitements des données . . . . .	90
5.4	Traitements des données de mouvement . . . . .	90
5.4.1	Variation de la taille des membres . . . . .	90
5.4.2	Incohérences de position des articulations d'une frame à une autre . . . . .	91
5.4.3	Extraction du moment d'intérêt du mouvement . . . . .	93
5.4.4	Aide à l'expert et retour à l'apprenant . . . . .	98
5.5	Résumé des contributions . . . . .	101
6	EXPÉRIMENTATIONS . . . . .	105
6.1	Lancer de balles dans une corbeille . . . . .	105
6.1.1	Motivations et objectifs . . . . .	105
6.1.2	Protocole . . . . .	106
6.1.3	Données capturées . . . . .	109
6.1.4	Données et métriques utilisées . . . . .	110
6.1.5	Résultats . . . . .	110
6.2	Bottle Flip Challenge . . . . .	112
6.2.1	Motivations et objectifs . . . . .	113
6.2.2	Protocole . . . . .	113
6.2.3	Données capturées . . . . .	114
6.2.4	Données et métriques utilisées . . . . .	115
6.2.5	Résultats . . . . .	115
6.2.6	Discussion . . . . .	118
6.3	Expérimentation : lancer de fléchettes . . . . .	118
6.3.1	Motivations et objectifs . . . . .	119
6.3.2	Fléchettes : principes et description . . . . .	119
6.3.3	Protocole . . . . .	124
6.3.4	Résultats . . . . .	129
6.4	Synthèse des contributions . . . . .	156
 <b>IV CONCLUSION</b>		
7	BILAN ET PERSPECTIVES . . . . .	161
7.1	Synthèse des contributions . . . . .	161
7.2	Limites des travaux et du système MLA . . . . .	162
7.3	Perspectives sur le projet MLA . . . . .	163
7.3.1	Interface graphique . . . . .	163

7.3.2	Explication des données non-alignées de l'apprenant . . . . .	163
7.3.3	Création de nouveaux descripteurs à l'aide d'un système fondé sur des patrons de conception . . . . .	165
7.3.4	Expérimentation à plus grande échelle . . . . .	166
7.3.5	Expérimentation d'apprentissage en autonomie . . . . .	166
7.3.6	Extension des expérimentations à d'autres gestes . . . . .	166
7.3.7	Clustering récursif . . . . .	167
A	ANNEXES . . . . .	179
A.1	Annexe 1 : pré-questionnaire et post-questionnaire de l'expérimenta- tion du lancer de fléchettes . . . . .	179
A.2	Annexe 2 : Tableaux d'analyses complets des résultats de l'expérimen- tation des fléchettes . . . . .	183



---

## TABLE DES FIGURES

---

FIGURE 1	Représentation informatique d'un squelette humain en 3D . .	28
FIGURE 2	Données d'un squelette 3D . . . . .	29
FIGURE 3	Squelette reconstitué à partir de vidéo stéréo (LIU, HAN et S. LEE 2016) . . . . .	30
FIGURE 4	Squelette et articulations captés par la Kinect V2 (AHMED, POLASH PAUL et GAVRILOVA 2015) . . . . .	31
FIGURE 5	utilisation de multiples caméras RGB et RGB-B pour éviter l'occlusion (REGAZZONI, VECCHI et RIZZI 2014) . . . . .	31
FIGURE 6	Système de captation par caméra infrarouge (MOREL 2017) . .	32
FIGURE 7	Bras haptique simulant l'insertion d'une sonde naso-gastrique (CHOI <i>et al.</i> 2015) . . . . .	34
FIGURE 8	Marqueurs réfléchissants sur une personne et un objet (NG 2008) . . . . .	35
FIGURE 9	EIAH pour la rééducation (BALDOMINOS, SAEZ et POZO 2015) .	37
FIGURE 10	EIAH pour la chirurgie orthopédique (TOUSSAINT 2015) . . . .	38
FIGURE 11	EIAH pour l'apprentissage de la langue des signes (GAMEIRO, CARDOSO et RYBARCZYK 2014) . . . . .	39
FIGURE 12	EIAH permettant de corriger des parties spécifiques du geste (XU <i>et al.</i> 2019) . . . . .	40
FIGURE 13	EIAH pour l'évaluation de gestes sportifs (MOREL 2017) . . .	42
FIGURE 14	EVAH superposant les gestes de l'apprenant à ceux de l'expert (LE NAOUR, HAMON et BRESCIANI 2019) . . . . .	43
FIGURE 15	Le système de partition de Laban . . . . .	53
FIGURE 16	Corrélation entre le geste de l'apprenant et de l'expert . . . .	54
FIGURE 17	Aires extraites du polygone du squelette (SAKURAI <i>et al.</i> 2014) .	56
FIGURE 18	Descripteurs basés sur les positions relatives des articulations entre elles (XIAO <i>et al.</i> 2015) . . . . .	56
FIGURE 19	Mouvement retournés à partir de postures dessinées (SAKURAI <i>et al.</i> 2014) . . . . .	57
FIGURE 20	Enveloppe convexe . . . . .	61
FIGURE 21	Descripteurs pour l'analyse de la démarche (S. CHEN <i>et al.</i> 2016) .	66
FIGURE 22	Environnement virtuel pour l'apprentissage du karaté (BURNS, ANNE-MARIE <i>et al.</i> 2011) . . . . .	67
FIGURE 23	Correction du mouvement de l'apprenant (YOSHINAGA et SOGA 2015) . . . . .	68
FIGURE 24	Système d'apprentissage du lancer de disque (YAMAOKA <i>et al.</i> 2013b) . . . . .	70
FIGURE 25	Système d'apprentissage de mouvement utilisant un CAVE pour la captation et la comparaison (KYAN <i>et al.</i> 2015) . . . . .	72
FIGURE 26	EVAH permettant de générer des parcours à suivre pour un objet ou l'apprenant (DJADJA, HAMON et GEORGE 2019) . . . .	73
FIGURE 27	Étapes de normalisation et d'analyse automatique du mouvement (PATRONA <i>et al.</i> 2018) . . . . .	75
FIGURE 28	Architecture de la plateforme <i>Motion Learning Analytics</i> (MLA). .	84
FIGURE 29	Le workflow général de la plateforme MLA. . . . .	85

FIGURE 30	Exemple de squelette dans un fichier BVH . . . . .	85
FIGURE 31	Exemple de données de mouvement dans un fichier BVH . . .	85
FIGURE 32	Architecture des pré-traitements et de l’affichage du mouve- ment capturé de <i>MLA</i> . . . . .	86
FIGURE 33	Diagramme de classes du mouvement . . . . .	86
FIGURE 34	Diagramme de classes de la structure de stockage des données	87
FIGURE 35	Exemple de descripteurs au format JSON . . . . .	87
FIGURE 36	Configurations possibles pour le Perception Neuron . . . . .	89
FIGURE 37	Vitesse linéaire avec artéfacts . . . . .	91
FIGURE 38	Comparaison de l’influence des paramètres d’un filtre Savitsky- Golay . . . . .	94
FIGURE 39	Vitesses filtrées par un filtre de Savitsky-Golay . . . . .	94
FIGURE 40	Vitesse linéaire filtrée et segmentée . . . . .	97
FIGURE 41	Signal filtré décomposé en trois étapes . . . . .	97
FIGURE 42	Clusters formés par les données de l’expert . . . . .	99
FIGURE 43	Feedback visuel proposé à l’expert . . . . .	102
FIGURE 44	Schéma de l’expérimentation du lancer de balle. . . . .	106
FIGURE 45	Exemple du mouvement du Bottle Flip Challenge . . . . .	113
FIGURE 46	L’installation du matériel lors de l’expérimentation du Bottle Flip Challenge (vue du dessus). . . . .	114
FIGURE 47	Cible de fléchettes . . . . .	120
FIGURE 48	Zones d’une cible de fléchettes . . . . .	120
FIGURE 49	Fléchette . . . . .	120
FIGURE 50	Défaut du « leaning » (se pencher en avant) . . . . .	122
FIGURE 51	Défaut du « elbow move » (déplacement du coude) . . . . .	122
FIGURE 52	Défaut du « javelin » (lancer en javelot) . . . . .	123
FIGURE 53	Défaut du « align arm » (alignement du bras) . . . . .	124
FIGURE 54	Installation de l’expérimentation. . . . .	125
FIGURE 55	Position latérale imposée aux participants. . . . .	126
FIGURE 56	Distance relevée entre les fléchettes et le centre de la cible . .	127
FIGURE 57	Exemple de visualisation à la fin des séries de lancer . . . . .	128
FIGURE 58	Moyennes des distances des fléchettes par rapport au centre en intra-groupe . . . . .	133
FIGURE 59	Moyennes des distances des fléchettes par rapport au centre en inter-groupe . . . . .	134
FIGURE 60	Moyenne des distances du défaut <i>align arm</i> en intra-groupe .	138
FIGURE 61	Distance du défaut <i>align arm</i> en inter-groupe . . . . .	138
FIGURE 62	Moyenne des distances du défaut <i>elbow move</i> en intra-groupe	141
FIGURE 63	Moyenne des distances du défaut <i>elbow move</i> en inter-groupe	141
FIGURE 64	Moyenne des distances du défaut <i>javelin</i> en intra-groupe . . .	144
FIGURE 65	Moyenne des distances du défaut <i>javelin</i> en inter-groupe . . .	144
FIGURE 66	Moyenne des distances du défaut <i>leaning</i> en intra-groupe . . .	147
FIGURE 67	Moyenne des distances du défaut <i>leaning</i> en inter-groupe . . .	147
FIGURE 68	Moyenne des distances de tous les défauts en intra-groupe . .	150
FIGURE 69	Moyenne des distances de tous les défauts en inter-groupe . .	150
FIGURE 70	Résultats des questionnaires . . . . .	154
FIGURE 71	Une première itération de l’interface graphique des modules d’analyse de <i>MLA</i> . . . . .	164
FIGURE 72	Un exemple de non-alignement des données de l’apprenant par rapport à celles de l’expert . . . . .	165

---

## LISTE DES TABLEAUX

---

TABLE 1	Tableau résumant les principales caractéristiques des EIAH vu dans le chapitre. . . . .	45
TABLE 2	Différentes familles de descripteurs . . . . .	62
TABLE 3	Classification des descripteurs élémentaires . . . . .	63
TABLE 4	Résumé des hypothèses sur les incohérences de positions . . .	92
TABLE 5	Métriques calculées sur les résultats du clustering pour le lancer de balle. . . . .	111
TABLE 6	Métriques calculées sur les résultats du clustering pour le Bottle Flip Challenge. . . . .	116
TABLE 7	Distance relative entre les centroïdes des clusters, pour la main droite, avec les vecteurs vitesse en X, Y, Z pour $k = 2$ . .	117
TABLE 8	Valeurs de l’Average Silhouette Score et de l’Adjusted Rand Index pour les données de l’expert sur les quatre défauts. . .	129
TABLE 9	Moyennes et écarts-types des données des apprenants (distance au centre de la cible). . . . .	131
TABLE 10	Résultats des tests statistiques en intergroupe (distance par rapport au centre). . . . .	132
TABLE 11	Résultats des tests statistiques en intragroupe (distance par rapport au centre). . . . .	135
TABLE 12	Fréquences des conseils donnés en fonction des jeux. . . . .	136
TABLE 13	Fréquences des conseils donnés pour chaque groupe, en fonction des jeux. . . . .	136
TABLE 14	Moyennes et écarts-types des données des apprenants (alignement du bras). . . . .	136
TABLE 15	Résultats des tests statistiques en intergroupe (alignement du bras). . . . .	137
TABLE 16	Moyennes et écarts-types des données des apprenants (mouvement du coude). . . . .	137
TABLE 17	Résultats des tests statistiques en intragroupe (alignement du bras). . . . .	139
TABLE 18	Résultats des tests statistiques en intergroupe (mouvement du coude). . . . .	140
TABLE 19	Moyennes et écarts-types des données des apprenants (mouvement du coude). . . . .	140
TABLE 20	Résultats des tests statistique en intragroupe (mouvement du coude). . . . .	142
TABLE 21	Résultats des tests statistiques en intergroupe (lancer de type javelot). . . . .	143
TABLE 22	Moyennes et écarts-types des données des apprenants (corps penché). . . . .	143
TABLE 23	Résultats des tests statistiques en intragroupe (lancer de type javelot). . . . .	145
TABLE 24	Résultats des tests statistiques en intergroupe (corps penché). .	146
TABLE 26	Résultats des tests statistiques en intragroupe (corps penché). .	148
TABLE 27	Résultats des tests statistiques en intergroupe (tous les défauts). .	149

TABLE 25	Moyennes et écarts-types des données des apprenants (tous les défauts). . . . .	149
TABLE 28	Résultats des tests statistiques en intragroupe (tous les défauts).	151
TABLE 29	Valeurs et p-values (entre parenthèses) des tau de Kendall entre les défauts du Groupe 1. . . . .	152
TABLE 30	Valeurs et p-values (entre parenthèses) des tau de Kendall entre les défauts du Groupe 2. . . . .	152
TABLE 31	Valeurs et p-values (entre parenthèses) des tau de Kendall entre les défauts du Groupe 3. . . . .	152
TABLE 32	Moyenne et écart-type des réponses aux questions en fonction de chaque groupe. . . . .	153
TABLE 33	Indices de corrélations entre les moyennes des réponses aux questionnaires pour chaque groupe. . . . .	154
TABLE 34	Résultats complets des tests statistiques en intergroupe (distance par rapport au centre). . . . .	183
TABLE 35	Résultats complets des tests statistiques en intergroupe (alignement du bras). . . . .	183
TABLE 37	Résultats complets des tests statistiques en intragroupe (distance par rapport au centre). . . . .	184
TABLE 38	Résultats complets des tests statistiques en intragroupe (alignement du bras). . . . .	185
TABLE 39	Résultats complets des tests statistiques en intragroupe (mouvement du coude). . . . .	186
TABLE 36	Résultats complets des tests statistiques en intergroupe (mouvement du coude). . . . .	187
TABLE 40	Résultats complets des tests statistiques en intergroupe (lancer de type javelot). . . . .	187
TABLE 41	Résultats complets des tests statistiques en intergroupe (corps penché). . . . .	187
TABLE 42	Résultats complets des tests statistiques en intergroupe (tous les défauts). . . . .	188
TABLE 43	Résultats complets des tests statistiques en intragroupe (lancer de type javelot). . . . .	189
TABLE 44	Résultats complets des tests statistiques en intragroupe (corps penché). . . . .	190
TABLE 45	Résultats complets des tests statistiques en intragroupe (tous les défauts). . . . .	191

Première partie

## INTRODUCTION





---

## CADRE ET CONTEXTE

---

Ce chapitre introduit le cadre de recherche de cette thèse, à la croisée des domaines de l'apprentissage humain du mouvement et des Environnements Informatiques pour l'Apprentissage Humain (EIAH). À l'issue d'une mise en contexte, les problématiques de recherche sont soulevées, et la contribution apportée est présentée.

L'apprentissage est un élément fondamental de l'humain. Notre capacité à observer, analyser, retenir et reproduire diverses choses nous permet de progresser tout au long de notre vie et de nous perfectionner. Cet apprentissage peut prendre plusieurs formes : formel / informel, par transmission orale, par imitation, mettant en jeu des connaissances tacites/explicites, *etc.* (ERAUT 2000). Les récentes avancées technologiques font que l'informatique est maintenant profondément ancrée dans notre société. L'utilisation conjointe des possibilités offertes par l'informatique et du savoir et de l'expertise humaine dans les processus d'apprentissage permet la création d'environnements informatiques pour soutenir l'enseignement et l'apprentissage humain. En particulier, ces environnements peuvent s'appliquer à la formation aux gestes et aux mouvements.

Le mouvement humain peut être vu comme une succession de postures d'un corps dans l'espace, variant au cours du temps. Son apprentissage peut se faire de différentes manières : par imitation, c'est-à-dire la reproduction de chaque posture d'une personne experte démontrant le mouvement (p. ex. apprentissage de la danse), par acquisition des propriétés biomécaniques clés connues au préalable du geste cible (p. ex. coude en bas pour un lancer de pétanque, flexion puis extension des jambes vers l'avant et le haut permettant d'initier le service au tennis) ou par recherche des gestes permettant d'atteindre un état ciblé des objets manipulés par l'intermédiaire des mouvements à apprendre (opération chirurgicale nécessitant la manipulation d'un outil tel qu'un scalpel). L'apprentissage d'un geste peut poser plusieurs contraintes, en fonction du mouvement à effectuer : contraintes matérielles, d'espace, de mobilité, d'environnement, *etc.* l'objectif de cet apprentissage peut être multiple : caractère moteur (le mouvement) mais aussi un caractère fonctionnel (liée à l'objectif de l'action) ou caractère structurant (liées aux connaissances du sujet et du contexte d'apprentissage).

La capture des mouvements peut se faire à l'aide de nombreux dispositifs, tels que : les caméras RGB-D (*Red Green Blue Depth*), popularisées grâce à l'apparition de systèmes grand public abordables comme la Kinect de Microsoft, utilisable dans des contextes de recherche où les interactions gestuelles sont " simples " (c'est-à-dire quelques mouvements brefs sur un ensemble restreint d'articulations). Des combinaisons constituées de capteurs inertiels, telles que le Perception Neuron de NOITOM, proposent une capture intégrale du corps humain, tout en ayant un encombrement réduit. Pour des captures de haute précision, l'industrie du cinéma a souvent recours à des dispositifs de captation utilisant des caméras infrarouges, disposées autour de la scène, à l'aide d'une combinaison de marqueurs réfléchissants, portée par les

acteurs. En fonction des besoins et du coût, et grâce au large choix des moyens de captation, il est possible d'obtenir des données de mouvements exploitables informatiquement et de construire des EIAH dédiés à leurs apprentissages.

La conception d'un environnement pour apprendre des mouvements nécessite de relever des défis propres aux domaines de l'étude des mouvements et de l'ingénierie des EIAH. Les prochaines sections sont dédiées à la définition des différents éléments constituant le domaine de recherche concerné par ce manuscrit, ainsi qu'à la problématique de la thèse et l'approche proposée.

## 1.1 ENVIRONNEMENTS INFORMATIQUES POUR L'APPRENTISSAGE HUMAIN (EIAH)

Les Environnements Informatiques pour l'Apprentissage Humain (EIAH) sont des environnements ayant pour but de favoriser l'apprentissage, en aidant, guidant et évaluant les apprenants d'une part et en assistant les enseignants d'autre part : que ce soit en présentiel, à distance, ou en situation mixte (TCHOUNIKINE 2009). Une situation d'apprentissage impliquant un EIAH est en conséquence constituée d'au moins un artefact informatique, d'un enseignant (bien que la connaissance experte puisse être injectée dans l'EIAH), d'un apprenant, du matériel à manipuler (dans le contexte d'un apprentissage de gestes sportifs ou de gestes chirurgicaux par exemple), de l'environnement spécifique où a lieu l'apprentissage (p. ex. étude géologique sur le terrain), etc. Cependant, le point fondamental d'une situation d'apprentissage impliquant l'utilisation d'un EIAH est son aspect pédagogique. Ainsi, un EIAH doit être développé afin de servir et s'adapter à une ou plusieurs situations d'apprentissage. Autour des EIAH s'articulent plusieurs domaines de recherche : informatique, sciences de l'éducation, science de l'information et de la communication, didactique, psychologie, etc. Ces domaines sont complémentaires pour aboutir à la création d'un EIAH supportant de manière efficace le processus d'apprentissage. De plus, la fonction visée de l'EIAH (apprentissage théorique, pratique, par imitation, sur une ou plusieurs séances, etc.) influe sur les compétences à mettre en œuvre pour son développement. La recherche en informatique sur les EIAH s'intéresse d'une part à la conception d'un tel système, mais également à son implémentation ainsi qu'à l'évaluation des impacts de son utilisation. Nous nous intéressons ici à l'accompagnement de l'expert, lors de sa tâche de transmission de savoir, et de l'apprenant dans sa tâche d'apprentissage. Les travaux présentés dans ce manuscrit n'ont pas vocation à remplacer l'expert lors de l'apprentissage, mais à l'assister dans la tâche d'observation, d'analyse et de retours donnés à l'apprenant.

## 1.2 LE MOUVEMENT HUMAIN

### 1.2.1 Définition

Le mouvement entendu comme le déplacement d'un corps dans l'espace à un moment donné, par rapport à un référentiel, est composé de variations à la fois spatiales et temporelles. Ce sont les différents paramètres de ces variations qui vont permettre de différencier un mouvement d'un autre, et leur donner une sémantique. En effet, le mouvement humain est porteur d'informations : il est possible d'extraire des informations bas-niveau, liées à la cinématique et la dynamique (NUNES et MOREIRA 2016). Le geste est également porteur de sens dans le contexte de la

communication verbale (HUANG *et al.* 2015), ou non-verbale (CHANG *et al.* 2013a). En plus de cela, il est également possible d'inférer des informations de haut-niveau, par rapport à l'émotion (KOBAYASHI 2007), l'intention (YU et M. LEE 2015a), et l'action (KAPSOURAS et NIKOLAIDIS 2014a).

### 1.2.2 Méthodes d'apprentissage du mouvement

Plusieurs méthodes d'apprentissage du mouvement existent. Une de ces méthodes est l'apprentissage par la reproduction du mouvement de l'expert. Dans ce cas, l'expert fait le mouvement cible, et l'apprenant s'applique à reproduire la succession de postures composant le mouvement. La décomposition du mouvement original en postures clés permet de fragmenter l'apprentissage du mouvement en plusieurs étapes successives, facilitant ainsi la reproduction du geste par l'apprenant (MAES, AMELYNCK et LEMAN 2012).

Il est également possible d'apprendre le geste en acquérant les propriétés biomécaniques clés, connues au préalable, du geste cible. Cela nécessite d'être capable d'identifier, de qualifier voire de quantifier les différents paramètres (rotation, déplacement, durée, vitesse, etc.) des articulations ou des membres concernées par le mouvement. Dans ce cas, il faut être en mesure de pouvoir mesurer les propriétés du mouvement de l'apprenant, afin de les comparer avec les valeurs de références, soit spécifiées par l'expert *a priori*, soit directement extraites du mouvement de ce dernier.

Enfin, lorsque l'objectif du geste est la manipulation d'un objet, l'apprentissage concerne la recherche des mouvements spécifiques (et leur optimisation) permettant d'atteindre l'état ciblé de l'objet. Un exemple d'un tel apprentissage est la manipulation d'outils chirurgicaux afin d'effectuer une opération. Dans ce cas, l'objectif est d'être capable d'effectuer le geste qui permet à l'outil d'arriver à la position et l'orientation désirée, ou d'être manipulé de la façon désirée, tout en prenant en compte les contraintes de manipulation propres au corps humain.

### 1.2.3 Apprentissage de gestes avec des d'EIAH

Les EIAH peuvent être conçus pour apprendre des gestes à des humains. Ces systèmes proposent différentes méthodes pour faciliter l'apprentissage du mouvement : un simple affichage du geste à effectuer et à reproduire, la manipulation d'un ou de plusieurs objets, l'immersion de l'apprenant dans un environnement en réalité augmentée, virtuelle ou mixte, etc.

L'affichage du mouvement à reproduire est une méthode analogue à l'apprentissage par reproduction du mouvement. Le support étant virtuel, il est possible de fournir un affichage superposé de l'apprenant avec celui du mouvement à effectuer (KORA, SOGA et TAKI 2015), de proposer une comparaison avec un ou plusieurs experts (YOSHINAGA et SOGA 2015), ou même de permettre à l'apprenant de visualiser quelles sont les différences entre son geste et celui de l'expert (CHAN *et al.* 2011). Il est possible de fournir une indication de performance quant au degré de similitude du geste de l'apprenant par rapport à celui de l'expert, soit à l'aide d'un score (MAES, AMELYNCK et LEMAN 2012), soit à l'aide d'indications de positions directement sur le geste de l'apprenant (YAMAOKA *et al.* 2013a).

Lors de la manipulation d'un objet, les systèmes peuvent contrôler non seulement la position de l'objet, mais également différents paramètres propres à la situation d'apprentissage (force appliquée, position du regard lors du mouvement, durée

d'une partie spécifique du mouvement, etc.) (CHELLALI *et al.* 2016). Un retour visuel peut être donné à l'apprenant, en affichant la progression de la position et l'orientation de l'objet dans une scène en temps réel (CHOI *et al.* 2015), mais il est également possible de donner des retours physiques, lors de l'utilisation de dispositifs tels que les bras haptiques (GILLESPIE 2018). La mise en situation de l'apprenant dans un environnement virtuel pour la manipulation d'un objet permet de s'affranchir des contraintes imposées par les apprentissages classiques nécessitant un environnement spécifique.

Enfin, les systèmes utilisant les environnements virtuels permettent de renforcer l'engagement des apprenants dans la tâche à accomplir. Ces environnements peuvent par exemple être utilisés dans la rééducation post-attaque cardiaque, afin de faire réaliser les mouvements répétitifs au patient tout en proposant un contexte attrayant, réaliste (BALDOMINOS, SAEZ et POZO 2015) ou ludique (ALANKUS *et al.* 2010).

La finalité d'un EIAH dédié à l'apprentissage de gestes étant la bonne réussite de ceux-ci, il est nécessaire de pouvoir analyser les données de mouvements de l'apprenant. L'analyse des données peut se faire de manière empirique par l'expert, ou de manière automatique par le système.

Lorsqu'il n'est pas possible de formaliser les caractéristiques qui font que le geste est considéré comme étant correct, l'analyse est faite de manière empirique. Dans le sport, les différences morphologiques font qu'il n'est pas toujours possible de proposer une comparaison pertinente avec les données de l'expert (BURNS, ANNE-MARIE *et al.* 2011), bien que des travaux pallient ces problèmes en utilisant plusieurs experts (YOSHINAGA et SOGA 2015). Dans le domaine médical, l'analyse empirique de données de mouvements reste la méthode prédominante (S. CHEN *et al.* 2016 ; R. WANG *et al.* 2013 ; ALANKUS *et al.* 2010), car l'expertise du médecin peut être difficile à formaliser sans être contextualisée en fonction du patient et de ses antécédents, ainsi que de la progression de la rééducation.

L'analyse automatique va permettre de proposer des retours sans l'intervention systématique d'un expert. Lorsque la connaissance experte est formalisable et quantifiable, il est possible de retourner à l'apprenant non seulement le degré de réussite de son geste (MAES, AMELYNCK et LEMAN 2012), mais également des précisions quant aux différentes parties du mouvement (YAMAOKA *et al.* 2013a). L'analyse des différences entre les gestes experts et ceux de l'apprenant permet également de mettre en lumière les caractéristiques déterminantes lors de la réalisation du geste en question (MAKIO, TANAKA et UEHARA 2007). Il est même possible de découvrir quels sont les ensembles de règles implicites utilisées par des experts pour déterminer le degré de réussite d'un geste (PIRSIAVASH, VONDRICK et TORRALBA 2014).

#### 1.2.4 Analyse automatique par méthodes d'apprentissage supervisé / non-supervisé

Il existe d'ores et déjà des EIAH qui analysent automatiquement le mouvement à l'aide de techniques de *Machine Learning*. Les algorithmes sous-jacents ont pour objectif d'apprendre à séparer des données, selon des critères variables. Ce domaine comprend plusieurs familles d'algorithmes : l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage semi-supervisé et l'apprentissage par renforcement.

L'apprentissage supervisé est un terme qui regroupe un ensemble d'algorithmes permettant d'apprendre à une machine à classer des données sous la forme d'une fonction de séparation et de prédiction construite à partir d'une base de données annotées.

L'apprentissage non-supervisé cherche à découvrir une ou des structure(s) au sein des données non-étiquetées. Les regroupements obtenus n'ont pas de sémantique associée par l'algorithme, à l'inverse de l'apprentissage supervisé. Ces regroupements se font en fonction de propriétés communes entre les différentes données fournies en entrée de l'algorithme. Les critères d'évaluations de l'efficacité des regroupements obtenus se basent sur des critères de proximité des points au sein d'un même groupe et entre les groupes (distance intra-cluster vs distance inter-cluster) (KASSAB et LAMIREL 2008).

L'apprentissage semi-supervisé fait usage d'un petit nombre de données annotées. Les avantages sont multiples : l'utilisation d'un corpus de données annotées restreint limite le coût en temps d'annotation, il est possible de comparer les groupes obtenus avec les données étiquetées, permettant de juger de la répartition des données au sein des groupes, et il est possible de donner du sens aux groupes ainsi obtenus.

L'approche supervisée pour l'analyse de mouvement requiert un corpus suffisamment grand de données étiquetées pour chaque domaine et tâche ciblés. En pratique, le temps de capture et d'annotation n'est pas adapté à une utilisation *in vivo*. Dans ce contexte, l'approche semi-supervisée ou non-supervisée est la plus adaptée. En effet, avec peu de données, l'utilisation de techniques de *clustering* permet de :

- regrouper les gestes en fonction des propriétés communes qui peuvent être spécifiées au préalable par l'enseignant,
- identifier les groupes de gestes acceptables en fonction de ces propriétés,
- positionner le geste de l'apprenant en termes de distance par rapport à ces groupes,
- suivre graphiquement l'évolution de ces propriétés individuellement ou conjointement.

Ces avantages combinés font que leur utilisation est pertinente dans le cadre d'une aide à l'apprentissage *in vivo*. Cependant, il n'y a que peu de travaux qui font usage de tels algorithmes pour l'apprentissage de gestes. Ainsi, l'utilisation d'algorithmes de *clustering* pour faciliter l'analyse et l'apprentissage de gestes est une piste explorée dans ce travail.

#### 1.2.5 Limites des systèmes existants

Les EIAH permettant d'assister l'apprentissage du geste et d'analyser les données de mouvement sont souvent conçus de façon *ad-hoc*, c'est-à-dire créés pour une situation d'apprentissage précise, dans un contexte fixé. Il est ainsi difficile, voire impossible, de réutiliser ces EIAH dans d'autres contextes sans un travail de réingénierie conséquent. Cela mène au développement de multiples EIAH dans des domaines pourtant analogues. De plus, l'intégration de la connaissance experte dans le système est souvent réalisée de manière fortement couplée, ce qui rend la genericité difficile. L'impossibilité d'introduire de nouveaux ensembles de règles pour une autre situation d'apprentissage fait que leur portée est limitée au domaine spécifique pour lequel ils ont été développés.

L'utilisation de techniques d'apprentissage supervisé pour l'analyse des gestes pose également le problème de la constitution d'un corpus spécifique pour la tâche concernée. Il n'est pas aisé de capturer, de segmenter et d'annoter un nombre de données suffisant pour qu'un tel algorithme puisse apprendre la fonction de séparation de manière correcte. En pratique, un système réutilisable dans différents contextes devra être en mesure de proposer des algorithmes d'apprentissage automatique

pouvant travailler sur un nombre restreint de données, tout en permettant d'apporter une aide suffisante pour l'expert dans sa tâche d'analyse du geste, capturables par l'expert en amont de sa tâche d'enseignement : l'utilisation des algorithmes d'apprentissage supervisé n'est donc pas adapté à un EIAH dédié à l'apprentissage de gestes dans une situation *in vivo*.

Enfin, dans le cas de l'apprentissage de mouvements, la présence d'un expert permet à l'apprenant de formuler des demandes précises en terme d'observation du geste à reproduire. Dans le cas d'un apprentissage à l'aide d'un EIAH dédié, les rôles de l'apprenant et l'enseignant ne sont pas toujours les mêmes en comparaison avec une situation d'apprentissage classique. Par exemple, certains systèmes écartent totalement l'expert du processus d'apprentissage. Bien que permettant un enseignement à distance, ou diffusé de manière plus large, il n'est pas possible pour l'apprenant de questionner l'expert sur des parties spécifiques du mouvement à effectuer. Il faut également que l'EIAH soit en mesure de prendre en compte des paramètres autres que le geste (morphologie, analyse fine des mouvements de l'apprenant, observations d'autres parties du corps pouvant mener à un geste incorrect, etc.), et donc que ces paramètres soient intégrés *a priori* dans le système. En pratique, les données spécifiques « annexes » (c'est-à-dire non-spécifiques au geste à réaliser) à intégrer dépendent du contexte, mais également de l'appréciation de l'expert par rapport à l'apprenant et le geste à réaliser. Ainsi, les EIAH écartant l'expert du processus peuvent mener à un apprentissage moins efficace pour l'apprenant.

### 1.3 PROBLÉMATIQUES DE RECHERCHE

#### 1.3.1 *Verrous scientifiques*

Dans le cadre de l'utilisation d'EIAH pour l'apprentissage de gestes, notre objectif est d'améliorer à la fois la situation d'apprentissage, tant pour l'enseignant que pour l'apprenant, mais également de faciliter la réutilisation d'un tel EIAH dans des contextes variés d'apprentissage de gestes.

Ainsi, un premier verrou réside dans l'analyse « fine » des mouvements, en vue de leur reproduction ou de la détermination de leurs propriétés acceptables. Cette analyse constitue une tâche souvent fastidieuse, contextuelle à la tâche d'apprentissage et nécessitant des connaissances scientifiques solides en géométrie, physique et biomécanique. Le plus souvent, l'apprenant et l'enseignant ne possèdent pas de telles connaissances. Les méthodes automatiques d'analyse du mouvement utilisant l'approche supervisée ne sont pas utilisables dans une situation d'apprentissage *in vivo*, car ils nécessitent un nombre de données étiquetées conséquent (de l'ordre du millier de données) et spécifique à la situation d'apprentissage considérée. De plus, les différences morphologiques entre les différents acteurs de l'apprentissage doivent être prises en compte si une analyse par superposition partielle ou totale des mouvements de l'expert et de l'apprenant est envisagée notamment.

Un moyen de résoudre ce problème est d'utiliser des descripteurs cinématiques du mouvement, ces derniers étant invariants à la morphologie. Une autre méthode pour pallier ce problème réside dans l'apprentissage par observation et reproduction du mouvement, à l'aide du EIAH dédié permettant de montrer le du geste de l'expert. Bien que l'étudiant puisse s'appuyer sur une telle représentation, les propriétés acceptables du geste ne seront pas nécessairement explicitement perçues ou identifiées. Cela constitue un deuxième verrou lié à la formalisation de l'expertise de l'enseignant qui doit être intégré au système formatif sous forme de règles ou de contraintes spatiales et temporelles. Il peut être difficile, voire impossible, pour un

expert d'être en mesure d'expliciter de manière quantitative les observations qu'il réalise afin de juger le geste de l'apprenant.

Un troisième verrou est lié à la représentation intelligible d'indicateurs de progression contextuels et adaptés à l'apprentissage, dépendant de l'expertise, mais aussi des capacités de perception des acteurs. En effet, bien qu'il soit possible d'extraire de nombreuses informations à partir de données de mouvements capturés brutes, un système d'aide à l'apprentissage de mouvement doit être en mesure de proposer des indicateurs pertinents en fonction de la situation d'apprentissage considérée et des objectifs pédagogiques visés. La visualisation de ces indicateurs doit être adaptée non seulement à leur type (discret ou continu), leur sémantique mais également aux besoins d'observation et d'analyse de l'enseignant.

Le quatrième verrou est lié aux retours donnés à la suite de la réalisation du geste par l'apprenant. C'est cette étape qui permet à l'apprenant d'améliorer son geste. Si ce retour est fait par un EIAH, il faut être en mesure de déterminer en amont les propriétés qui font qu'un retour est pertinent pour l'apprentissage et la progression du geste de l'apprenant. Cette connaissance est parfois tacite pour l'expert. Ces paramètres peuvent donc varier en fonction de la situation d'apprentissage considérée, mais également de l'apprenant lui-même. De plus, un système réutilisable et adaptable dans d'autres contextes faisant intervenir le mouvement ne doit pas nécessiter un processus de réingénierie lourd.

### 1.3.2 Questions de recherche

Le développement d'EIAH support à l'enseignement de gestes, extensibles au-delà de la tâche pour laquelle elles ont été conçues et ayant un coût minimal en termes de réingénierie, représente un défi qui soulève plusieurs questions de recherche :

- **Q1** : Comment développer un système permettant de caractériser le geste à l'aide de l'intégration de l'expertise d'un enseignant ?
- **Q2** : Comment évaluer et comparer le geste (ou ses propriétés) de l'apprenant avec celui de l'enseignant afin d'évaluer la progression de l'apprentissage ?
- **Q3** : Comment, dans une situation d'apprentissage de gestes donnée, proposer des retours pertinents et compréhensibles aux acteurs de l'apprentissage non spécialistes en analyse du mouvement ?

### 1.3.3 Hypothèses de recherche

À partir des éléments précédemment expliqués, ainsi que des questions de recherches formulées, six hypothèses seront vérifiées dans ces travaux :

- **H1** : Il est possible de regrouper les gestes selon leurs propriétés cinématiques communes. (**Q2**)
- **H2** : Il est possible de séparer les gestes des apprenants en deux groupes correspondant à une dichotomie geste réussi / geste raté afin de déterminer, pour une situation d'apprentissage donnée, les propriétés d'un ensemble fini de gestes réussis. (**Q1**)
- **H3** : Il est possible de séparer les gestes en fonction de propriétés attendues et identifiées au préalable par l'expert. (**Q1**)
- **H4** : Il est possible de corriger chaque défaut du geste de l'apprenant, en lui indiquant les défauts majeurs à corriger en premier. Un défaut majeur



est identifié par la plus grande distance séparant le mouvement courant de l'apprenant, du groupe de gestes acceptables ayant éliminé ce défaut. (Q2)

- H5 : L'utilisation du système MLA basée sur l'hypothèse 4 permet d'améliorer l'apprentissage du geste par rapport à une situation sans le système MLA. (Q3)
- H6 : L'utilisation du système MLA en tant qu'assistant à l'enseignant permet d'améliorer l'apprentissage du geste par rapport à une situation sans MLA, et à une situation avec MLA et sans enseignant. (Q3)

La validation ou non de ces hypothèses s'effectuera à travers le système d'aide à l'apprentissage de mouvements développé au cours de cette thèse.

#### 1.4 APERÇU DES CONTRIBUTIONS

##### 1.4.1 *Motion learning Analytics (MLA)*

Afin de répondre à ces questions, notre approche s'est axée autour du système *Motion Learning Analytics* (MLA), développé pour les besoins de cette thèse. Il s'agit d'un système d'analyse de mouvement, combinant la captation, le filtrage, le traitement et l'analyse (à l'aide de méthodes d'apprentissage non-supervisé ou semi-supervisé) du mouvement. L'approche consiste à traiter des données de mouvement brutes, afin de les rendre : (i) exploitables, en éliminant les défauts inhérents au système de capture utilisé et (ii), compréhensibles, en extrayant des descripteurs de mouvement à partir des données filtrées. Une fois ces étapes réalisées, l'objectif du système est alors double : (i) proposer automatiquement une évaluation et des conseils pour l'amélioration du geste de l'apprenant à partir de la comparaison de propriétés du mouvement de l'apprenant avec celles attendues et calculées à partir des mouvements de l'expert, et (ii) proposer un affichage graphique de la progression de l'apprenant, en fonction des défauts considérés, sur lequel peut s'appuyer l'expert.

Cette plateforme a permis de valider notre approche, et d'apporter ainsi une preuve de concept de notre chaîne allant de l'acquisition du mouvement jusqu'à l'analyse de ces derniers :

- L'acquisition du mouvement se fait à l'aide du Perception Neuron, une combinaison constituée de multiples capteurs inertiels, permettant de capturer le mouvement en temps réel.
- Les données exportées au format BVH sont passées dans une chaîne de traitements, afin d'améliorer le signal. Les données sont ensuite utilisées afin de calculer des descripteurs, correspondant à des aspects spécifiques du mouvement (vitesse, accélération, etc.).
- Les descripteurs sont utilisés dans un processus de *clustering*, afin d'étudier la séparation possible des gestes, selon leur degré de réussite, leurs défauts ou les propriétés d'acceptabilité identifiées en amont par un expert.
- Les données de l'apprenant sont comparées à celles de l'expert, afin de déterminer quels sont les conseils à donner à l'apprenant afin qu'il améliore son geste.

#### 1.4.2 *Traitement et segmentation des mouvements*

La chaîne de traitement des données permet non seulement de filtrer les artéfacts propres aux données obtenues à partir du Perception Neuron, mais également de n'en garder que la partie jugée utile à l'aide d'un algorithme de segmentation et d'en extraire des descripteurs jugés pertinents dans la situation d'apprentissage concernée.

#### 1.4.3 *Séparation des données de mouvement et sémantique associée*

L'utilisation d'algorithmes de *clustering* permet non seulement de travailler avec un nombre restreint de données, mais également de proposer un regroupement des données en fonction de propriétés définies au préalable par l'expert. L'ajout d'un petit nombre de données étiquetées permet de donner un sens aux groupements obtenus, permettant une comparaison des données de l'apprenant par rapport aux gestes jugés bons ou mauvais.

#### 1.4.4 *Système de retour visuel*

Enfin, un système de retour visuel, permettant à l'utilisateur de juger de la distance des données de l'apprenant par rapport à celles de l'expert, permet de consolider l'avis de l'expert ou de donner des pistes quant aux défauts les plus significatifs du geste de l'apprenant, dans le cas où le jugement est difficile.

### 1.5 EXPÉRIMENTATIONS

Le système a été testé au travers de trois expérimentations, permettant chacune de tester les hypothèses présentées dans cette introduction. Ces trois expérimentations consistent en la réalisation de gestes dont deux exigent une certaine technicité, puis à l'utilisation de ces données pour tester les hypothèses. La première expérimentation consistait à lancer une balle dans deux corbeilles éloignées l'une de l'autre. L'objectif ici était de voir s'il était possible d'obtenir une séparation acceptable du geste en fonction d'une ou plusieurs propriétés définies au préalable, correspondant à différentes stratégies de lancers. Une séparation correspondant au type de lancer a été obtenue. Le geste de la deuxième expérimentation était celui du *Bottle Flip Challenge*, où il faut lancer une bouteille tout en lui imprimant une rotation sur son axe horizontal, puis la faire retomber debout sur une surface plane. Les données de cette expérimentation ont été utilisées afin de vérifier s'il était possible de séparer les gestes en deux groupes correspondant aux gestes réussis et ceux ratés. Il a été possible d'obtenir une bonne séparation mais pas une séparation en fonction des groupes considérés. Enfin, la troisième expérimentation s'est appuyée sur le lancer de fléchettes comme domaine applicatif, afin de tester la pertinence du système MLA en tant qu'outil de regroupements des gestes suivant des propriétés identifiées en amont par un expert, et en tant qu'outil de retours à l'apprenant et d'assistance à l'enseignant.

### 1.6 PLAN DU MANUSCRIT

Le chapitre suivant porte sur un état de l'art des EIAH dédiés à l'apprentissage du geste. Y sont introduit les différentes modalités utilisées au sein de ces EIAH :

objectif du système, données utilisées et leur représentation informatique, méthodes de captation, analyses effectuées, retours fournis à l'apprenant et place de l'expert dans le processus.

Le chapitre trois propose une étude bibliographique des informations extractibles du mouvement, appelées descripteurs du mouvement. Ces descripteurs permettent de caractériser le mouvement afin de pouvoir l'analyser en fonction des besoins d'observations ou des propriétés cinématique du mouvement. Une analyse bibliographique des différents descripteurs, ainsi que de leur utilisation dans des contextes de recherche sur le mouvement est réalisée. Un tableau présentant des descripteurs couvrant un ensemble de propriétés observables est proposé à la fin de ce chapitre.

Le chapitre quatre se concentre sur les méthodes d'analyses appliquées à des mouvements capturés. À partir des descripteurs calculés, il est possible de réaliser des analyses précises sur le mouvement, soit par un expert, soit par un processus automatique. Ce chapitre présente ces analyses sous deux angles majeurs : l'analyse empirique qualitative et l'analyse automatique quantitative.

Le chapitre cinq présente la contribution apportée au travers du système MLA, développé pour l'aide à l'évaluation de gestes humains. Y sont présentées les différentes étapes, du traitement des données des mouvements brutes jusqu'à l'aide proposée à l'expert et aux apprenants.

Le chapitre six est dédié aux trois expérimentations qui ont permis de tester les hypothèses de recherche.

Pour terminer, le dernier chapitre fournit un résumé des apports de cette thèse, et ouvre plusieurs pistes et perspectives de recherche.

Deuxième partie

ÉTAT DE L'ART



---

## LES EIAH POUR L'APPRENTISSAGE DE GESTES

---

Ce chapitre présente un état de l'art sur les différents éléments nécessaires à la conception d'EIAH pour soutenir l'apprentissage humain de gestes. La première partie rappelle et décrit l'une des représentations informatiques les plus connues pour modéliser le mouvement humain i.e. la représentation par une succession temporelle de squelettes. La seconde partie fait état des différents moyens pour capturer le mouvement. La troisième partie se concentre sur les EIAH existants, le matériel utilisé pour la capture, leurs principales caractéristiques (*e.g.* domaines applicatifs, types de données utilisées, etc.) et les retours faits à l'apprenant. De plus, elle étudie ces systèmes sous l'angle de la réingénierie et la place/le rôle des usagers au sein de ces systèmes.

Dans la suite de ce manuscrit, nous utiliserons le terme de « geste cible », pour faire référence au geste à apprendre pour atteindre l'objectif de la tâche à réaliser. En effet, il existe parfois, pour une même tâche d'apprentissage, plusieurs gestes possibles jugés corrects par un ou plusieurs experts. Ces gestes constituent l'objectif d'apprentissage. Si l'on prend l'exemple du service au tennis, et que l'on définit la tâche comme étant le fait de frapper la balle de manière à ce qu'elle arrive dans la zone légale de service adverse, plusieurs mouvements existent : le lancer " droit ", le lancer " arrière " et le lancer " devant ". De plus, il est possible d'imprimer un effet à la balle lors de la frappe (*lifté* ou *slicé*), ce qui induit une variation du geste de frappe (ZAPPALA *et al.* 2017). Au bowling, il existe à minima trois types de lancers différents (*Stroker*, *Cranker*, *Spinner*) pour atteindre l'objectif, qui est de renverser un maximum de quilles en un minimum de lancers (B. TAN, AZIZ et CHUAN 2000). Ces deux exemples parmi tant d'autres montrent qu'il ne serait possible de définir un seul geste comme étant le geste parfait à atteindre pour une situation d'apprentissage donnée. En conséquence, nous introduisons le terme de " geste cible ", qui est le geste dont les caractéristiques sont : (i) connues par l'expert et (ii) doivent être transmises à, et maîtrisées par, l'apprenant. Le geste cible peut être différent d'un expert à l'autre pour un mouvement donné, en fonction de l'objectif recherché et des spécificités morphologiques.

### 2.1 REPRÉSENTATION DU MOUVEMENT

Afin de pouvoir exploiter les données de mouvement au sein d'un EIAH, il faut pouvoir formaliser, discrétiser puis opérationnaliser les mouvements du corps. Une des techniques les plus utilisées consiste à utiliser des graphes, et plus particulièrement des arbres, afin de modéliser les mouvements du corps humain sous la forme d'une succession de squelettes 3D, c'est-à-dire des postures espacées selon un pas de temps fixe ou variable (LEWIS, CORDNER et FONG 2000).

Le corps humain peut ainsi être vu comme un ensemble structuré d'articulations sous la forme d'une arborescence (Fig. 1). Dans ce cas, les nœuds de l'arbre

représentent les articulations, et les arcs représentent les différents membres du corps.

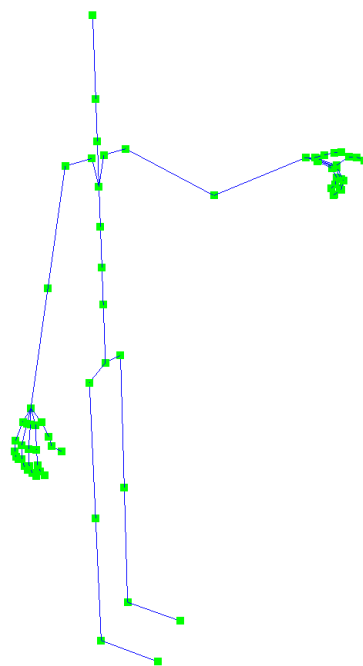


FIGURE 1 – Un exemple de squelette humain 3D représenté sous la forme d'un arbre. Les articulations sont représentées à l'aide de nœuds (en vert), et les membres du corps sont les arcs reliant ces nœuds entre eux (en bleu).

Chaque nœud contient généralement plusieurs informations : le nom, la position et/ou l'orientation de l'articulation, ainsi que des références vers les articulations parentes et filles. L'articulation racine généralement choisie dans le cadre d'un squelette est celle des hanches, car située près du centre de masse du corps humain. On définit également souvent sa position comme étant l'origine  $(0,0,0)$  du repère tridimensionnel cartésien (Fig. 2). Les systèmes de captures usuels ne peuvent capter qu'une fraction des éléments constituant le corps humain : un corps humain est composé d'environ 360 articulations, 206 os et 640 muscles. L'ensemble de ces éléments ne sont pas nécessairement modélisés pour étudier les mouvements humains suivant le contexte. Ainsi, pour caractériser le mouvement d'une personne s'asseyant, la variation sur l'axe vertical du centre de masse, ou encore la position de la projection du centre de masse par rapport au support du corps (les pieds) sera suffisante (SALAMAH, ZHANG et BRUNETT 2015).

Cette structure est particulièrement adaptée pour modéliser tout ou une partie du corps. La représentation du corps humain sous forme de chaîne articulaire 3D, afin de représenter un mouvement, sera celle utilisée dans la suite de ce manuscrit.

Afin d'obtenir un mouvement représenté sous forme d'un arbre, et ainsi exploitable à l'aide d'outils informatiques, il faut être en mesure de le capturer. La capture de mouvements peut être réalisée à partir de plusieurs dispositifs, permettant soit d'obtenir des données 2D (*i.e.* images ou vidéos), ou directement des données 3D (squelette de la personne, nuage de points, modèles surfaciques, *etc.*). En fonction du matériel de capture utilisé, le processus d'acquisition, ainsi que les contraintes associées, la qualité des données obtenues ne sera pas la même. Le choix d'un dispositif de capture adapté à un contexte d'apprentissage de geste doit se faire en fonction des besoins d'observation et d'analyse du geste, tout en prenant en compte

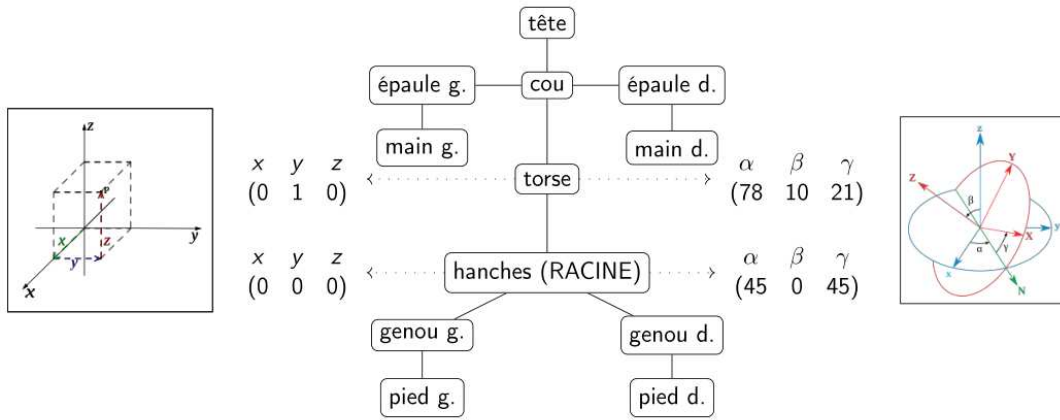


FIGURE 2 – Un exemple de squelette humain représenté sous forme informatique avec, pour chaque articulation, des informations de position et d'orientation relatives à l'articulation parente.

les contraintes imposées par la réalisation du geste : amplitude, parties du corps impliquées, manipulation d'objet, environnement spécifique, *etc.*.

## 2.2 MATÉRIEL DE CAPTURE ET DONNÉES OBTENUES

Il est possible de séparer l'apprentissage de gestes en deux grandes familles non exclusives : les cas où le geste est la finalité de l'apprentissage, et le cas où la manipulation d'un objet est l'objectif final. Dans le premier cas, le corps humain (une partie ou dans son entièreté) va être au cœur de l'analyse, alors que dans le deuxième cas, la position de l'objet, et parfois le geste permettant d'y arriver seront les points d'intérêt principaux. Les dispositifs utilisés sont ainsi choisis en fonction de l'objet considéré pour la captation (*i.e.* le corps ou un objet), mais aussi du rapport coût/caractéristiques souhaitées de l'équipement, déduits de la situation d'apprentissage et des besoins d'observation. En termes de caractéristiques, l'encombrement, l'espace de travail, la chaîne de traitements des données et leur précision, ainsi que les retours sensoriels souhaités du système (visuels et/ou haptiques) sont souvent considérés pour ce choix (CHAMARET *et al.* 2010). Cette section s'intéresse d'abord aux dispositifs dédiés à la capture d'un corps humain, puis aux dispositifs dédiés à la simulation ou la capture de position et d'orientation d'objets. Les coûts de ces dispositifs peuvent être très importants, mais il est maintenant possible d'obtenir des données de mouvements capturés d'une qualité suffisante pour effectuer des analyses de geste d'apprenant à partir de systèmes relativement peu onéreux. Cet aspect *low/middle-cost* de la captation de mouvement est intéressant pour des captures prenant place dans des situations variées (par opposition à la captation dans un environnement dédié, par exemple en laboratoire), ce qui en fait le choix retenu pour cette thèse.

Dans des domaines tels que le sport (YAMAOKA *et al.* 2013a) (lancer de disque) (YOSHINAGA et SOGA 2015) (archerie) (CHAN *et al.* 2011) (danse), le corps entier de l'apprenant va être capturé et utilisé. Dans ces cas-ci, l'analyse va concerner le mouvement d'une partie ou du corps entier. Il est possible d'obtenir ces données à partir de vidéos, de caméras RGB-D (*Red Green Blue Depth*) (YOSHINAGA et SOGA 2015) (KORA, SOGA et TAKI 2015) (YAMAOKA *et al.* 2013a), d'un ensemble de capteurs réduit placés à des endroits spécifiques du corps (PORCIUNCULA *et al.* 2018) ou d'une combinaison de capteurs (CHANG *et al.* 2013a).



### 2.2.1 Extraction de mouvement par vidéo

Il est possible d'extraire des données de mouvement à l'aide de vidéos. Ces méthodes reposent sur l'inférence du mouvement à partir d'un ensemble de pixels (Z. Lv *et al.* 2015). Il existe une multitude de techniques, et certaines d'entre elles permettent d'obtenir un squelette 3D. Une revue des principales méthodes utilisées est proposée dans (SARAFIANOS *et al.* 2016) : projection dans un espace 2D de données de mouvement 3D, afin d'obtenir une correspondance avec l'image / la vidéo, utilisation de contraintes rigides sur plusieurs images successives afin de retrouver le squelette le plus probable, *etc.* Il est également possible d'utiliser des vidéos stéréos, permettant ainsi d'obtenir des points de repère dans l'espace et de faciliter le suivi de points spécifiques du corps, afin de reconstituer un squelette 3D correspondant au mouvement (Fig. 3) (LIU, HAN et S. LEE 2016).

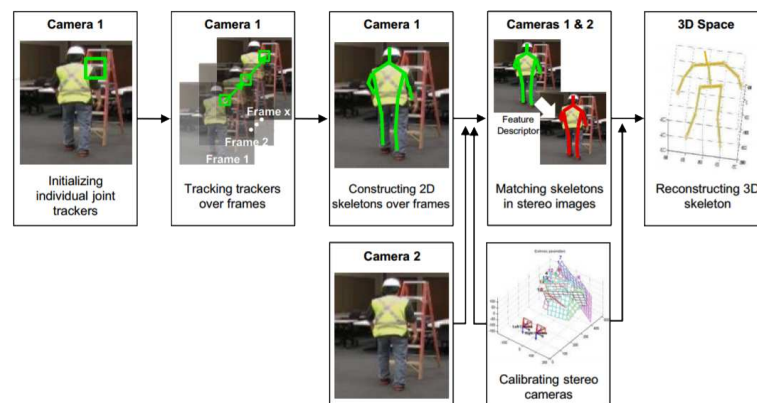


FIGURE 3 – Le squelette 3D peut être reconstitué à partir de vidéo stéréo (LIU, HAN et S. LEE 2016).

### 2.2.2 Extraction de mouvement par vidéo et capteur de profondeur

Les caméras RGB-D sont des dispositifs de capture se basant sur la combinaison d'une capture vidéo 2D et d'un mapping de profondeur sur chaque pixel, généralement obtenu à l'aide d'un capteur infrarouge. L'ajout de l'information de profondeur à la vidéo 2D permet d'obtenir une information supplémentaire notamment sur le déplacement des acteurs dans la scène filmée. Leur encombrement relativement réduit, combiné à la non nécessité d'avoir des marqueurs équipés sur le corps font que ces dispositifs ont été largement utilisés dans des contextes de recherche tels que : la capture en temps réel de multiples personnes (BASSO *et al.* 2013), la reconstruction de scènes interactives en 3D (IZADI *et al.* 2011) ou encore la capture du mouvement (YOSHINAGA et SOGA 2015). En effet, il est possible d'obtenir le squelette d'une personne à l'aide des outils fournis par les entreprises développant ces caméras (Fig. 4) : jusqu'à 6 squelettes de 20 articulations chacun peuvent être capturés simultanément (25 articulations par squelette pour la Kinect v2, voir Fig. 4). La précision de la Kinect permet de capter les pouces de chaque main, mais pas les doigts pris individuellement.

Les caméras RGB-D possèdent cependant des inconvénients. En effet, le problème de l'occlusion est fréquent. Il apparaît dès lors qu'une partie du corps est masquée par une autre, par rapport à l'objectif de la caméra. Il est possible de pallier ce problème grâce à l'utilisation de plusieurs caméras en simultanées (Fig. 5), et dont

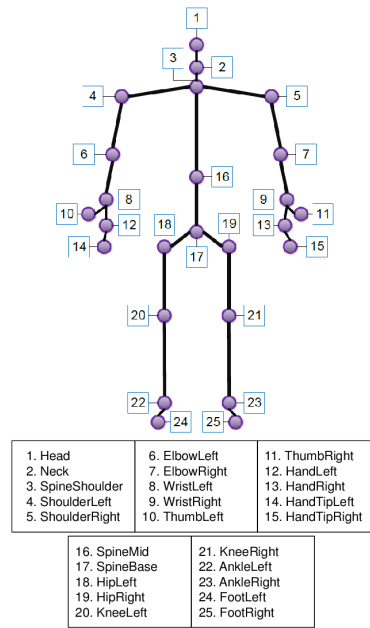


FIGURE 4 – Squelette et articulations captés par la Kinect V2 (AHMED, POLASH PAUL et GAVRILOVA 2015). Le nombre d’articulation est suffisant pour pouvoir représenter les changement de position et d’orientation de tous les membres du corps humain.

les différentes images sont utilisées pour reconstruire le squelette final (REGAZZONI, VECCHI et RIZZI 2014). Cette méthode nécessite un calibrage très précis, ainsi qu’un environnement dédié à la captation des mouvements.

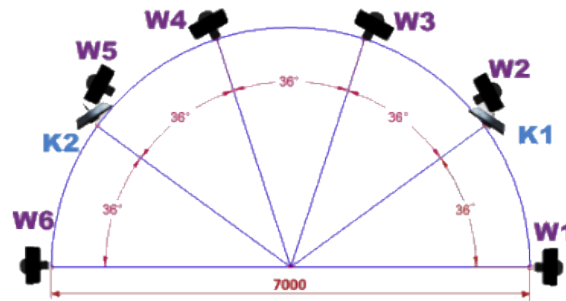


FIGURE 5 – L’utilisation de plusieurs caméras RGB, ainsi que RGB-D permet d’éviter les problèmes d’occlusion (REGAZZONI, VECCHI et RIZZI 2014).

### 2.2.3 Extraction de mouvement par caméra infrarouge

La méthode la plus précise, en termes de qualité des données, pour capturer un corps entier à l’heure actuelle est celle basée sur les caméras infrarouges utilisées conjointement avec des marqueurs réfléchissants (Fig. 6). Ces systèmes sont les plus précis pour l’acquisition de mouvements d’une ou plusieurs personnes en simultanée (PFISTER *et al.* 2014) (C. YANG *et al.* 2016). Ils sont très utilisés dans l’industrie du cinéma et du jeu vidéo, afin de proposer des mouvements les plus fidèles possibles. Les inconvénients majeurs de ces systèmes sont d’une part le coût (chiffré en plusieurs centaines de milliers d’euros pour un système basé sur 8 caméras), la non portabilité du matériel de capture (nécessité d’avoir un emplacement dédié et fixe) et la chaîne lourde de traitements pour reconstituer le squelette 3D (*e.g.* capture

des points, filtrage pour la réduction des phénomènes occlusions, reconstitution de la chaîne articulaire et des orientations de chaque articulation). Bien que des entreprises proposent des solutions tout-en-un pour obtenir le squelette 3D d'un ou plusieurs acteurs (*e.g.* Qualysis, Vicon, Optitrak), ils nécessitent les compétences d'un informaticien qualifié pour être utilisé. En effet, le placement des capteurs doit se faire de manière optimale (soit en suivant les recommandations fournies par les fabricants des systèmes utilisés, soit en fonction des besoins de captation) afin de s'assurer que les marqueurs réfléchissants soient visibles par au moins une caméra afin de limiter au maximum les phénomènes d'occultation, et les données doivent ensuite être filtrées à l'aide des logiciels fournis par les fabricants ou en utilisant des logiciels externes. Ces filtres peuvent avoir plusieurs objectifs : ré-interpolation du mouvement à une fréquence plus ou moins rapide, lissage des données de mouvement et d'orientation par moyennage des postures proches, effacement des pics de valeurs en fonction des valeurs voisines, *etc.*



FIGURE 6 – Un système de captation à base de caméras infrarouges et de marqueurs disposés sur le corps (eu haut). Les données sont obtenues sont la forme d'un nuage de point (en bas à gauche), à partir desquels les squelettes sont reconstitués (en bas à droite) (MOREL 2017).

#### 2.2.4 Extraction de mouvement par capteurs inertiels portables

L'utilisation de capteurs portés par le corps représente une autre méthode pour enregistrer les mouvements. Ces capteurs sont généralement reliés entre eux, et permettent de transmettre les données par une liaison filaire ou sans fil, se libérant ainsi des contraintes de l'espace fixe (PORCIUNCULA *et al.* 2018). Un capteur renferme un ou plusieurs dispositifs de mesure tels que : un accéléromètre, permettant de mesurer l'accélération linéaire, un gyromètre, permettant de mesurer la vitesse

angulaire, et un magnétomètre, qui permet de mesurer la force et la direction d'un champ magnétique. La combinaison de ces trois informations permet de déduire les positions et orientations d'un corps en mouvement dans l'espace. Ainsi, il est possible d'obtenir des données de mouvements dans des contextes plus variés que les systèmes nécessitant une installation fixe. Afin de permettre l'acquisition d'un squelette humain, il est possible de combiner plusieurs de ces capteurs, posés sur plusieurs parties du corps (au niveau des articulations).

Cependant, les données obtenues peuvent être moins précises qu'avec des systèmes fixes (*e.g.* perturbations liées à l'environnement, qualité de transmission des données variable, *etc.*). Un des problèmes souvent rencontrés est celui de la dérive. Il s'agit de la variation des données renvoyées par un capteur, sans que cette variation ne soit le résultat d'un changement de position ou d'orientation réel. Bien que des méthodes existent pour corriger ces défauts, cela implique généralement d'utiliser d'autres types de données pour compenser les erreurs propres au matériel (données GPS, par exemple (BEVLY 2004)). Cependant, en fonction du degré de précision nécessaire, ces méthodes ne sont pas toujours adaptées et réutilisables dans tous les contextes. Ainsi, il est parfois nécessaire de disposer d'une chaîne de traitements adaptée afin d'éliminer ces erreurs. Ainsi, une technique de filtrage des données proposée par Roetenberg *et al.* est celle basée sur un filtre de Kalman, fusionnant plusieurs informations provenant de capteurs différents (gyroscopes, accéléromètres, magnétomètres) afin d'estimer l'erreur et ainsi la corriger à la volée (ROETENBERG *et al.* 2005). Les casques de réalité virtuelle récents font également usage de capteurs inertiels (disposés sur le casque lui-même) afin de proposer une plus grande immersion dans les jeux, en rajoutant le mouvement comme interaction. L'ajout de ces capteurs inertiels permet également de réduire les problèmes de nausées propres à ces casques (HTC 2016).

### 2.2.5 Matériel de captation pour la manipulation d'objets

Dans le cas où la manipulation de l'objet est au cœur de l'apprentissage, bien que le geste soit parfois important, l'état de l'objet (position, orientation, intégrité, *etc.*) est l'objectif de l'apprentissage du geste, et donc de la captation. Cette différence se retrouve notamment au niveau des données capturées et utilisées : dans les domaines tels que la chirurgie (TOUSSAINT 2015) (CHOI *et al.* 2015) ou la musique (NG 2008), la position de l'objet est souvent la donnée de mouvement principale récupérée pendant la tâche (possiblement utilisée conjointement avec d'autres informations, telles que la position du regard (TOUSSAINT 2015)). Ces données peuvent être capturées à l'aide de dispositifs adaptés : les interfaces haptiques en sont un exemple (Fig. 7).

Ces interfaces permettent non seulement une manipulation plus aisée d'un objet virtuel, mais également un retour sensitif au niveau du toucher lors de la manipulation (SREELAKSHMI et SUBASH 2017). Ainsi, il est possible d'obtenir des retours de force ou tactile permettant d'émuler respectivement la forme ou la texture d'un objet. En combinant ces retours à des environnements de réalité virtuelle, il est possible de proposer une sensation de saisie, de collision et de toucher, en conjonction avec ce qui est visuellement présenté à l'utilisateur (WHITMIRE *et al.* 2018). Ce dispositif est très utilisé dans le cadre de l'apprentissage de manipulation d'objets médicaux, tels que pour l'insertion d'aiguilles sous la peau (CORRÊA *et al.* 2019), la manipulation de sondes (CHOI *et al.* 2015), le cathétérisme (PEPLEY *et al.* 2017), ou encore l'entraînement du personnel à la planification pré-opératoire par télé-chirurgie (HALABI et HALWANI 2018).

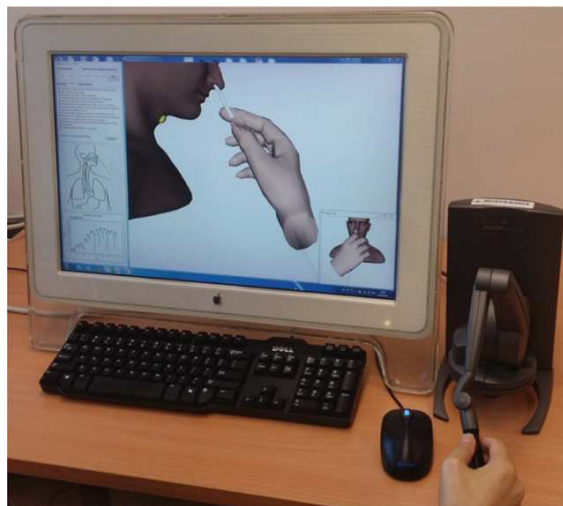


FIGURE 7 – Exemple de bras haptique servant à simuler l'insertion d'une sonde nasogastrique, avec un modèle 3D mis à jour en temps réel à côté (CHOI *et al.* 2015).

Par exemple, dans le cadre de la simulation de l'insertion d'une sonde nasogastrique, Choi *et al.* ont utilisé un modèle physique 3D de tête humaine conjointement avec un bras haptique 7 (CHOI *et al.* 2015). Ce système permet de générer la sensation d'une insertion et de butée si jamais la sonde touche une partie de la gorge, et ainsi de pouvoir ajuster le geste en temps réel.

Dans le domaine industriel, Chamaret *et al.* ont capturé des mouvements à l'aide d'un dispositif appelé « spidar » (CHAMARET *et al.* 2010). Le but était d'étudier et d'évaluer différentes procédures d'assemblage et de démontage d'un phare de voiture, dans un environnement virtuel. Les modèles 3D, ainsi que le retour de force, étaient adaptés à la morphologie de l'utilisateur. L'objectif était d'évaluer empiriquement la capacité de l'utilisateur à accomplir la tâche demandée, en accord avec le nombre de collisions.

Les dispositifs haptiques disposent souvent d'un espace de travail limité (bras à retours d'efforts), contraignant l'utilisateur dans ses mouvements à cet espace (bras à retour d'effort à socle fixe) ou par le poids de ces dispositifs (veste tactile, bras à retours d'efforts portatifs, gants à retour d'efforts). Cependant, les situations d'apprentissage du geste ne se limitent pas toujours à un espace de travail restreint. Dans le contexte de l'apprentissage de gestes sportifs ou artistiques par exemple, le corps ainsi que les objets manipulés peuvent évoluer dans un espace 3D avec un volume/espace de travail conséquent.

Dans ces cas, plusieurs méthodes de suivi de l'objet existent. L'une d'elles utilise des marqueurs réfléchissants placés sur l'objet à suivre, afin d'en obtenir la trajectoire dans l'espace et le temps, au même titre que les méthodes vues précédemment servant à capter le corps d'un humain. Le positionnement de ces capteurs sur l'objet en question (Fig. 8) doit faire l'objet d'une attention particulière, afin de limiter les problèmes d'occlusion. Shapira *et al.* utilisent un ensemble de caméra infrarouges combinées à des casques de réalité virtuelle et des objets (des cubes mous et des jouets divers) équipés de de sphère réfléchissantes, afin d'être captées et utilisées au sein de l'environnement virtuel (SHAPIRA, AMORES et BENAVIDES 2016). Un autre test en utilisant des bâtons réfléchissants à la place des sphères a été réalisé, cependant, les objets étant manipulés par des enfants, ces marqueurs étaient le plus souvent occultés par ces derniers. Ainsi, les sphères réfléchissantes positionnées sur les





FIGURE 8 – Exemple de marqueurs réfléchissants positionnés sur une personne et un objet (Ng 2008).

arrêtes des objets à manipuler se sont révélées plus efficaces pour réussir à suivre les déplacements et rotations des objets dans l'espace.

#### 2.2.6 Bilan sur les dispositifs de captation de mouvement

La qualité des données produites par les différentes méthodes d'acquisition peut varier grandement, notamment à cause des problèmes spécifiques aux matériels utilisés (problèmes d'occlusion pour les caméras RGB-D, problème de dérive pour les capteurs inertiels portés). Cependant, les matériels de captation fournissant les données plus précises possèdent un coût d'achat parfois prohibitif, ainsi qu'une mise en place plus contraignante. De plus, ils nécessitent parfois l'utilisation d'une chaîne de traitement (intégrée aux logiciels fournis avec système de capture ou non) afin d'obtenir des données exploitables. Concernant les systèmes moins coûteux, tels que la captation à l'aide de caméras *RGB-D* ou de combinaison de capteurs inertiels, la mise en place du matériel est moins contraignante (en terme de temps et d'espace), et nécessite moins de traitements afin d'obtenir des données utilisables. Quelle que soit l'utilisation faite du mouvement, il faut être en mesure de disposer de données non-bruitées par les systèmes de capture. Les systèmes au sein desquels sont utilisées des données de mouvement doivent souvent intégrer une chaîne de traitements du mouvement, soit pour le filtrage, soit pour l'extraction d'informations sur une partie précise du mouvement global. Ce besoin est crucial dans les systèmes dédiés à l'apprentissage de mouvements, indépendamment du contexte applicatif.

Dans le contexte de la thèse, la captation par combinaison de capteurs inertiels a été retenue. Plus précisément, il s'agit de la combinaison Perception NEURON. L'installation et la calibration d'une telle combinaison est rapide, et elle n'encombre que marginalement la personne la portant. La possibilité d'utiliser de tels dispositifs sans être contraint à un espace de travail fixe permet de mieux s'adapter au contexte du geste à effectuer (contraintes d'environnement, d'espace nécessaire, *etc.*). Le logiciel fourni, servant à récupérer les informations transmises par la combinaison, intègre une série de pré-traitements, afin de filtrer les données de mouvement obtenues. Les retours proposés par le système se basant sur le mouvement seul, et étant donnés par un expert ou le système lui-même, il n'y a pas besoin d'utiliser un système de retours haptiques. Enfin, le coût faible d'une telle combinaison par rapport aux systèmes de captation plus perfectionnés est des atouts dans le contexte d'un EIAH permettant d'aider à l'apprentissage de geste qui se veut utilisables par tous.

### 2.3 ENVIRONNEMENTS INFORMATIQUES POUR L'APPRENTISSAGE DE GESTES

Il existe de nombreux EIAH dédiés au geste. La finalité de l'apprentissage des gestes au sein de ces systèmes est variée : apprentissage actif impliquant d'avantage l'étudiant (XU *et al.* 2019), rééducation (ALANKUS *et al.* 2010) (BALDOMINOS, SAEZ et POZO 2015), amélioration de la performance sportive (MAES, AMELYNCK et LEMAN 2012), *etc.* Ces catégories ne sont pas exclusives, et les systèmes développés se retrouvent souvent à l'intersection de plusieurs d'entre elles. Les domaines d'application sont également variés : médical (*e.g.* analyse de postures (AMINIAN et NAJAFI 2004), chirurgie (TOUSSAINT 2015), *etc.*), sport (*e.g.* danse (MAES, AMELYNCK et LEMAN 2012), archerie japonaise (YOSHINAGA et SOGA 2015), lancer de disque (YAMAOKA *et al.* 2013b)), et les techniques utilisées peuvent également faire appel à la réalité virtuelle, par exemple (KYAN *et al.* 2015). Pour chaque EIAH, il est possible d'identifier des caractéristiques qui lui sont propres : domaine d'application, objectif visé de l'apprentissage, objectif visé de la ou des tâches à apprendre, prise en compte des besoin d'observation et fonctionnels de l'enseignant, utilisabilité vis-à-vis de l'utilisateur afin de créer et d'évaluer une nouvelle tâche, type de données utilisées pour l'analyse du geste, méthodes d'analyse, retours faits à l'apprenant, *etc.* En effet, un EIAH sera efficace et démocratisable si l'utilisateur a été placé au cœur de sa conception afin de : (i) répondre à ses besoins d'observation et fonctionnels, (ii) prouver son utilisabilité et (iii), des modèles génériques de conception ont été extraits afin de minimiser la réingénierie en cas de modification de la tâche, des observations et du domaine d'application. Cette section étudie les principaux EIAHs basés « geste » selon ces caractéristiques.

#### 2.3.1 Quelques exemples significatifs d'EIAH pour apprendre des gestes

Dans le domaine médical, la rééducation nécessite par exemple la répétition de gestes permettant de renforcer une partie du corps, ou de réapprendre à utiliser un membre. Il est possible de faciliter la répétition de certains de ces gestes à l'aide d'interfaces dédiées. Baldominos *et al.* ont développé un EIAH permettant de faire répéter les gestes requis par le médecin pour la rééducation de parties spécifiques du corps (BALDOMINOS, SAEZ et POZO 2015). Dans cette étude, les exercices sont focalisés sur un travail des adducteurs, ainsi que les abducteurs. La mise en situation proposée est celle d'un gardien de but (Fig. 9). L'expert peut intervenir ou non dans la séance de rééducation. Si l'expert est présent, il peut choisir le moment du tir du ballon, la vitesse ainsi que de la hauteur de la balle. Dans le cas d'une session d'apprentissage sans expert, le système commence par effectuer des tirs bas, espacés dans le temps, puis augmente ou diminue l'intensité des deux paramètres (hauteur et temps entre les tirs) en fonction du score du patient. Dans les deux cas, une augmentation progressive de la difficulté de la tâche à effectuer permet d'augmenter l'amplitude du mouvement nécessaire, et ainsi de faire progresser le patient. De plus, l'utilisation de dispositifs tels qu'un casque de réalité virtuelle masque les parties du corps, ce qui demande une concentration accrue afin de synchroniser ses mouvements et d'être capable de se situer dans l'espace.

Les travaux de Alankus *et al.* proposent un système permettant d'engager plus fortement les victimes d'attaque cardiaque lors de la rééducation à domicile (ALANKUS *et al.* 2010). Après une attaque cardiaque, il y a souvent une perte de mobilité de la part du patient. Dans le cadre de la rééducation post-attaque cardiaque, il est souvent nécessaire pour le patient de répéter des gestes, afin de retrouver une certaine mobilité. Ces gestes sont définis par l'expert, en fonction de la gravité des



FIGURE 9 – L’environnement de réalité virtuelle proposé par (BALDOMINOS, SAEZ et POZO 2015). La mise en situation dans un cadre crédible par rapport aux exercices à effectuer permet d’augmenter l’implication du patient (BALDOMINOS, SAEZ et POZO 2015).

lésions cérébrales et de l’avancée de la rééducation. Bien que les séances avec les spécialistes soient dédiées à ces gestes, la durée, ainsi que la fréquence de ces séances est insuffisante ; c’est pourquoi le patient doit également réaliser des gestes répétés chez lui. Dans ce contexte, les mouvements sont capturés à l’aide de dispositifs communs et abordables : webcam et Wiimote™. Plusieurs jeux ont été développés : déplacement de personnages dans l’espace, jeu de course automobile, lancer et rattrapage de balle de baseball, jardinage (élimination des mauvaises herbes en laissant les fleurs intactes), *etc.* Chacun de ces jeux se concentre sur un mouvement précis à réaliser, allant de mouvement amples (balayage du bras) à des mouvements fins (utilisation de contrôleurs analogiques pour déplacer un personnage). Ils ont été développés en collaboration avec des médecins, afin de solliciter les capacités cognitives du patient, ainsi que les parties du corps concernées par la rééducation. Ils permettent d’activer des parties du cerveau différentes, en stimulant soit la mémoire, le contrôle d’une partie du corps spécifique, la coordination des membres du corps ou les facultés d’anticipation. Le système peut lui-même modifier la difficulté des jeux (et donc l’intensité des exercices), et il est également possible pour l’expert de personnaliser les jeux, afin de les adapter aux différents patients. Par exemple, il peut définir un ensemble de valeurs correspondant à différents niveaux de difficultés, afin que cette difficulté s’adapte au patient pour que l’exercice propose toujours un défi suffisamment élevé afin d’améliorer l’efficacité de la rééducation. Dans le cas du jeu Pong (un des jeux proposés par le système), la taille de raquette et de la balle, ainsi que la vitesse de la balle peuvent être changées.

Lorsque l’objectif est de réussir un geste menant à la manipulation d’un objet, comme dans le domaine de la chirurgie, le système peut combiner plusieurs types de données hétérogènes, afin d’évaluer plusieurs modalités. Par exemple, le système utilisé par Toussaint permet de combiner les données de mouvement d’un bras haptique (utilisé pour simuler la manipulation d’un trocard) ainsi que le retour de force associé (simulant les variations de densités du corps humain et des vertèbres), et d’un oculomètre, afin de suivre la trajectoire du regard au cours d’une opération (Fig. 10) (TOUSSAINT 2015). Les connaissances mises en jeu sont appelées « perceptivo-gestuelles », car elles combinent la vision et le geste. Un algorithme d’extraction de *patterns*, *PhARules*, inspiré de *CMRules*, permet d’extraire des séquences d’actions



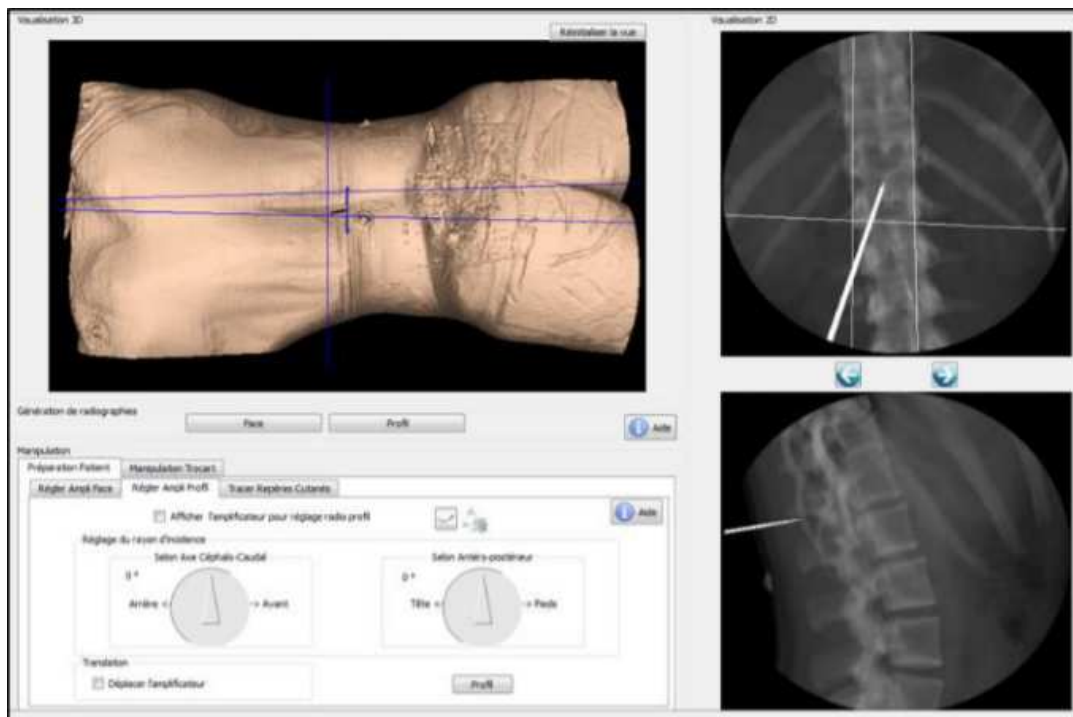


FIGURE 10 – Le système utilisé par (TOUSSAINT 2015) permet de combiner les données d'un oculomètre ainsi que d'un bras haptique afin de donner un retour couvrant tous les aspects importants d'une procédure chirurgicale orthopédique (TOUSSAINT 2015).

en fonction des phases dans lesquelles elles sont réalisées. Les détails du parcours de l'apprenant, ainsi que les interactions pour chaque séquence de l'opération sont fournis. Ces interactions peuvent être de type « contrôle *a priori* » (e.g. une observation des signes vitaux du patient avant l'opération), « contrôle *a posteriori* » (e.g. une observation des signes vitaux du patient après être intervenu sur le patient) ou « mixte » (e.g. prise d'information et manipulation du patient en simultané). Selon le degré de réussite de la procédure, le système propose un retour à l'apprenant : réalisation d'un exercice complémentaire sur un nouveau cas clinique simulé afin de renforcer l'acquisition des connaissances sur un point précis, proposition de révision des cours théoriques, approfondissement de la connaissance du cas clinique qui vient d'être traité.

L'apprentissage de la langue des signes est difficile, et nécessite une pratique régulière pour les non-natifs (COOPER, HOLT et BOWDEN 2011). La captation et l'analyse des gestes de la langue des signes nécessitent de prendre en compte : (i) toutes les parties du corps situées au-dessus du bassin, en tant qu'éléments signifiant du discours et (ii) la précision et la rapidité des mouvements effectués par certaines parties du corps telles que les mains et les doigts (GIBET *et al.* 2016). En utilisant des données enregistrées avec un dispositif low-cost (Microsoft Kinect™), le système développé par Gameiro *et al.* permet de faciliter l'apprentissage de la langue des signes grâce à deux modes d'utilisation : le mode « classe », et le mode « compétition » (Fig. 11) (GAMEIRO, CARDOSO et RYBARCZYK 2014). Le mode classe met l'apprenant face à une situation d'apprentissage où il doit répéter les signes présents à l'écran, jusqu'à ce que le signe soit correctement fait. Le mode compétition est gamifié, et propose une situation de type quizz télévisé, où le but est de répondre en signant à des questions permettant de tester les connaissances de la personne. L'autre défi du mode compétition consiste en une évaluation des capacités de l'apprenant à

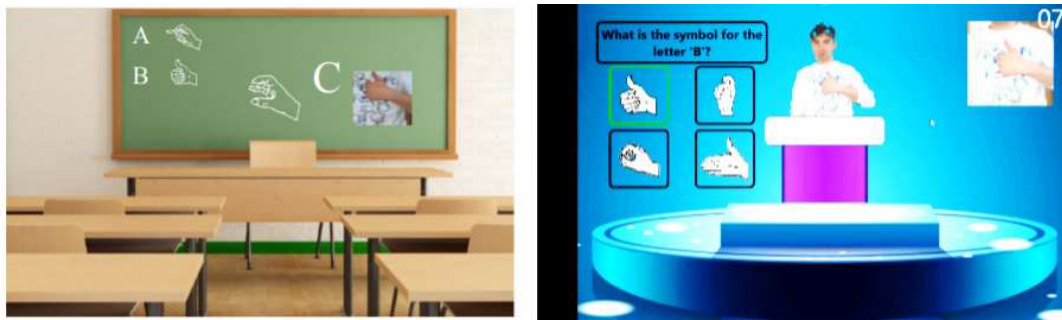


FIGURE 11 – Le système proposé par (GAMEIRO, CARDOSO et RYBARCZYK 2014) met d'une part l'apprenant dans une situation classique d'apprentissage (à gauche), puis permet de tester les connaissances à l'aide d'une gamification type "quizz télévisé" (à droite) (GAMEIRO, CARDOSO et RYBARCZYK 2014).

épeler et signer correctement un mot, en composant les signes des lettres successives. Dans ce cas, l'évaluation est binaire : soit le geste correspond au signe à réaliser, soit il ne correspond pas. L'évaluation se fait à l'aide d'un masquage des données de l'apprenant par les signes correspondant aux lettres à signer présents dans la base de données. Bien que présentant des résultats positifs avec des classes d'étudiants en langage des signes, les auteurs notent que la reconnaissance automatique des signes de l'apprenant est meilleure lorsque la base de données de signes de référence est construite à partir des mouvements de l'apprenant. Cela s'explique par les différences morphologiques entre les différentes personnes ayant utilisé le système, à cause de l'absence de normalisation ou de mise à l'échelle des données capturées par la caméra.

L'EIAH développé par Xu *et al.* permet de faire apprendre à des enfants certains gestes, complexes ou non (Xu *et al.* 2019). Dans cette étude, ces gestes étaient liés à la culture Chinoise : manipuler un métier à tisser, tirer à l'arc, chevaucher un cheval, manipuler des panneaux, *etc.* Les gestes sont acquis à l'aide d'une caméra RGB-D. Le mouvement est segmenté dans un premier temps, à l'aide d'un modèle de Markov caché, comparant le mouvement complet de l'apprenant à plusieurs combinaisons de mouvements issus de la base de données de mouvements à effectuer. La combinaison de mouvements ayant la plus grande similitude avec les gestes de l'apprenant est choisie comme modèle du geste réalisé par l'apprenant, et le découpage du mouvement de l'apprenant s'effectue selon cette combinaison. La comparaison du geste de l'apprenant à celui à effectuer intervient à la suite de la segmentation, et est également réalisée par un modèle de Markov caché, à l'aide plusieurs modules permettant chacun de reconnaître un geste. Ainsi, en décomposant le mouvement en plusieurs parties, il est possible de comparer chaque étape d'un geste continu aux gestes cible présents dans la base de données de gestes, et donc de déterminer si le geste est réussi ou non. S'il ne l'est pas, le système va proposer un exercice permettant de se focaliser sur la ou les parties du geste ratées, à l'aide d'un autre modèle de Markov caché (Fig. 12). Ce réseau se base sur les résultats de l'analyse des gestes du précédent : ainsi, en chaque point où le geste de l'apprenant n'est pas conforme au geste attendu, le système va être en mesure de proposer plusieurs ressources différentes (explications textuelles, vidéo de démonstration, *etc.*) pour expliquer ou montrer à l'apprenant quel est le geste attendu. Par exemple, dans le cas de la manipulation du métier à tisser, le système peut détecter si l'apprenant ne regarde pas le métier à tisser pendant la tâche, ou si la distance entre son corps et le métier à tisser est trop grande, et lui faire visualiser une vidéo lui montrant la posture à adopter pour réaliser cette tâche. L'apprentissage du geste est ainsi

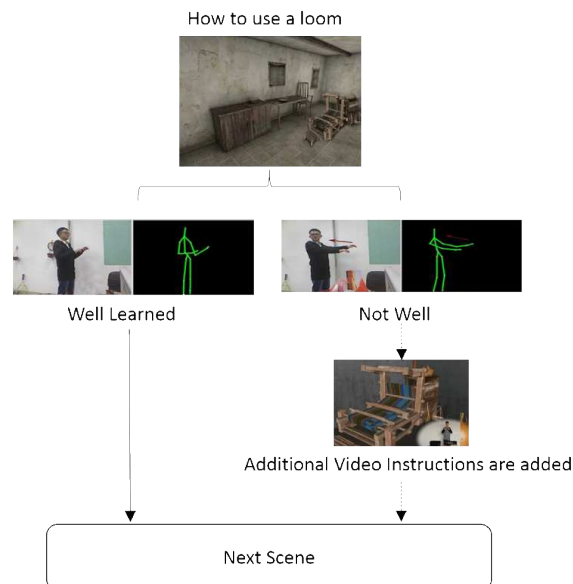


FIGURE 12 – L'adaptation du parcours d'apprentissage proposé par Xu *et al.* permet de corriger des parties spécifiques du mouvement de l'apprenant (Xu *et al.* 2019).

amélioré en faisant répéter des parties spécifiques à l'apprenant, avant de passer à la suite du parcours. De cette manière, il est possible de mettre l'accent sur les mouvements que l'apprenant a du mal à réaliser.

Le domaine artistique et sportif fait usage des EIAH pour l'apprentissage de gestes. Chan *et al.* ont développé un système permettant d'apprendre des mouvements de danse, en se basant sur des données de mouvement d'expert enregistrées au préalable à l'aide d'un système basé sur une combinaison de capteurs réfléchissants et des caméras infrarouge, afin de l'afficher sur un écran (CHAN *et al.* 2011). En observant les gestes de l'expert, l'apprenant le reproduit ensuite, et ses données sont capturées, afin d'être comparées à celles de l'expert. Les données sont stockées sous forme de squelettes 3D, à partir desquels plusieurs analyses sont faites. Dans ce système, les différences morphologiques entre l'apprenant et l'expert sont atténuées à l'aide d'une normalisation des positions des articulations par la somme des membres du corps. Plusieurs retours sont offerts à l'apprenant. En premier lieu, une visualisation en temps réel du mouvement de l'expert et de celui de l'apprenant est proposée, en colorant les différentes parties du corps de l'apprenant en fonction des différences avec le geste de l'expert (lorsqu'un certain seuil de distance est franchi entre les positions des articulations de l'expert et de l'apprenant). Il est également possible de rejouer le mouvement au ralenti, afin de mieux voir les différences entre les gestes. Enfin, un score de performance global, correspondant à la somme des distances des articulations de l'apprenant par rapport à celles de l'expert, est fourni à l'apprenant.

Le système d'aide à l'apprentissage de la danse développé par Maes *et al.* propose une approche différente pour la visualisation des mouvements de l'expert (MAES, AMELYNCK et LEMAN 2012). Trois étapes sont proposées : la visualisation du mouvement réalisé par l'expert, l'apprentissage en réalisant le geste de manière décomposée afin d'apprendre chaque étape, et l'évaluation à l'aide d'un score indiquant la performance de l'apprenant. La décomposition du mouvement est réalisée de manière automatique, en analysant d'une part le rythme et le tempo de la musique, et d'autre part le mouvement de l'expert réalisant le geste. Les points d'inflexions dans ces valeurs synchronisées avec les tempos sont utilisés comme séparateurs du mouvement. L'expert peut ensuite valider cette séparation avant d'intégrer le geste au système. Le score se base sur le coefficient de corrélation croisée

de Pearson entre les données de mouvement de l'expert et celles de l'apprenant, un score de 1 indiquant une similitude parfaite entre les deux gestes, alors qu'un score de 0 indique une dissimilitude complète. L'utilisation de systèmes intégrant un tuteur intelligent (souvent appelés « coachs ») permet de s'affranchir de la présence d'un expert, et donc de l'obligation de se rendre dans un endroit spécifique pour pratiquer, et permet également d'entraîner des personnes en parallèle. Dans ces systèmes, l'expert n'intervient que dans l'enregistrement et l'intégration des données au système, le retour à l'apprenant est fait par le système à l'aide d'un score ou de conseils par exemple.

Dans le domaine du sport, il est également possible de développer de tels systèmes pour la découverte automatique de la différence entre le geste d'un expert et d'un débutant, comme le système proposé par Makio *et al.*, servant à séparer les gestes de service au tennis des professionnels de ceux des amateurs (MAKIO, TANAKA et UEHARA 2007). Dans ce cas, le système permet d'extraire dans un premier temps des règles d'associations entre les différentes parties du corps, afin de découvrir les mouvements effectués par l'expert. Dans un premier temps, le mouvement est segmenté à partir des points où la vitesse (sur un des trois axes) d'une articulation spécifiée est égale à zéro. Les parties du mouvement ainsi segmentées sont ensuite comparées entre elles à l'aide du Dynamic Time Warping (DTW), afin d'obtenir un score de similitude. Les mouvements sont ensuite regroupés à l'aide d'un algorithme de clustering, afin de réunir au sein d'un même groupe plusieurs sous-parties du mouvement correspondant au même geste. L'utilisation de deux fenêtres glissantes, l'une active (correspondant aux mouvements ayant une influence sur d'autres mouvements) et l'autre passive (correspondant aux mouvements étant influencés par d'autres mouvements) permet ensuite de détecter les dépendances récurrentes entre les parties du corps lors de la réalisation d'un geste, correspondant aux règles d'association. Une règle ainsi obtenue peut s'écrire

$$if \begin{pmatrix} \text{left ankle } 166 \\ \text{right wrist } 48 \end{pmatrix} then(\text{left ankle } 167)$$

ce qui correspond à « Si le centre de gravité se situe sur le pied gauche alors que la main droite se déplace vers le bas, alors le pied gauche se déplacera vers le haut ». Dans un deuxième temps, des motifs fréquents de déplacement sont extraits au sein des mouvements, afin d'obtenir une décomposition de ce mouvement. Cette extraction se base sur une réduction de la dimensionnalité des données 3D dans un espace 1D à l'aide d'une analyse en composantes principales (PCA), afin de réduire les temps de calcul. Les séquences sont ensuite réduites à un ensemble symbolique à l'aide du principe de longueur de description minimale (RISSANEN 1998), puis comparées entre elles à l'aide d'une distance euclidienne. Ici, la connaissance experte est directement intégrée au système lors de sa création : dans le cas présent, il s'agit de la trajectoire des poignets dans le contexte d'un service au tennis. Ces données peuvent ensuite être utilisées afin de comparer les données de l'apprenant à celles de l'expert. Ce système permet d'obtenir une visualisation graphique de la trajectoire de la raquette, et permet de mettre en évidence les différences de trajectoires entre les débutants et les experts.

Les travaux de Morel sur l'évaluation de gestes sportifs à l'aide de séries temporelles ont conduit à la création d'un EIAH permettant d'apprendre un geste (MOREL 2017). Ce système est développé de manière à être indépendant du sport choisi. Le geste à atteindre est modélisé par le squelette de l'expert effectuant ce geste, et les données sont capturées à l'aide d'un système de marqueurs réfléchissants et caméras

infrarouges. Les squelettes de l'apprenant et de l'expert sont recalés spatialement (à l'aide d'une transformation linéaire) et temporellement (à l'aide de l'algorithme du *Subsequence DTW*). Les erreurs spatiales et temporelles sont calculées, à l'aide d'un respectivement, permettant ainsi de donner des indications simples par rapport au geste : « trop haut », « trop bas », « trop à gauche », « trop à droite », « trop en avant » , « trop en arrière » pour les variations spatiales, et « en retard » ou « en avance » pour les variations temporelles. La plus grande erreur de l'apprenant est affichée sous la forme d'une animation de son mouvement, superposé à celui de l'expert, avec la partie du corps en faute colorée en rouge (Fig. 13). Le processus d'intégration de la connaissance experte au système n'est pas précisé dans l'étude, mais aucune interface de sélection de caractéristiques à observer n'est montée, ce qui suggère que cette intégration se fait au moment de la conception du système, limitant ainsi la facilité de réutilisation du système dans d'autres contextes.

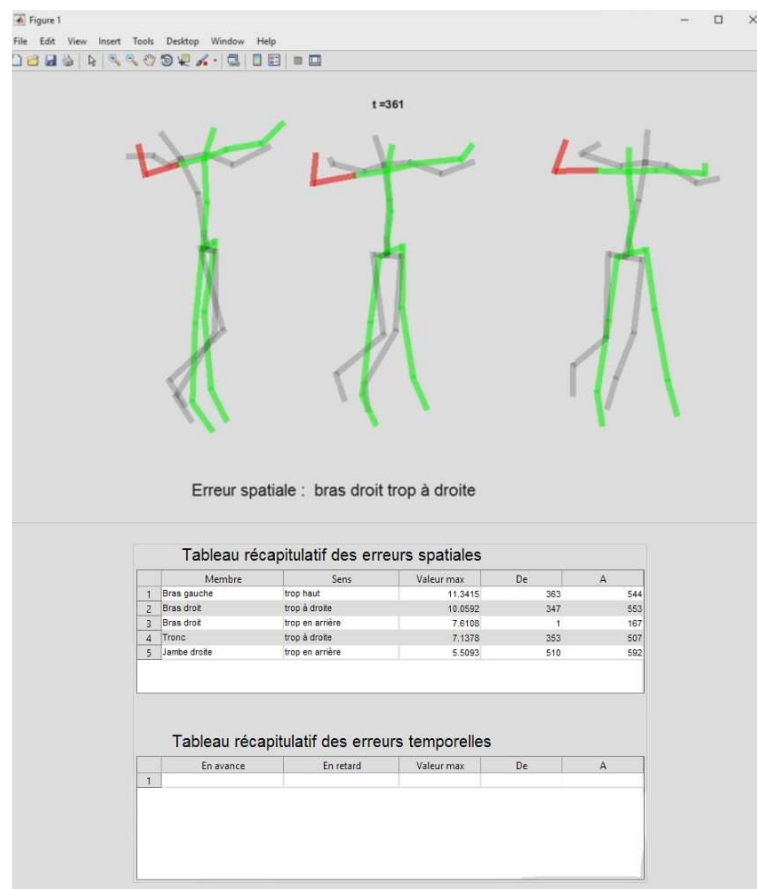


FIGURE 13 – Le système développé par Morel permet d'afficher les plus grandes erreurs spatiales et temporelles du mouvement de l'apprenant, par rapport à celui de l'expert (MOREL 2017).

Enfin, le système proposé par Le Naour *et al.* a pour but de fournir une aide à l'apprentissage de mouvement en proposant une superposition du modèle de l'apprenant à celui de l'expert, dans un environnement virtuel (LE NAOUR, HAMON et BRESCIANI 2019) (Fig. 14). Les mouvements sont capturés à l'aide d'un système de marqueurs réfléchissants et caméras infrarouges. L'expérimentation présentée dans ces travaux porte sur le lancer au football américain. En premier lieu, une démonstration est effectuée par un expert, et des conseils sont donnés par ce dernier sur les caractéristiques propres à un bon geste (*e.g.* position et rotation des articulations impliquées dans le mouvement, prise de la balle, *etc.*). Après dix lancers d'entraînement, les participants devaient effectuer cinquante lancers, tout en



suivant le plus possible le mouvement de l'expert. Plusieurs groupes ont été formés : un groupe sans aucun retour sur la performance, un groupe où le mouvement de l'expert était montré après chaque lancer, un autre où le mouvement de l'expert était montré au participant pendant leur mouvement, suivi de la visualisation de la superposition du mouvement du participant à celui de l'expert, un autre groupe où le mouvement de l'expert était superposé à celui du participant pendant le geste et où le geste de l'expert était visualisé après le lancer, et un dernier où le modèle expert superposé à celui de l'apprenant était visualisé pendant et après le geste. Ont été relevés l'écart-type du calcul de la DTW entre le mouvement de l'apprenant et de l'expert, et la distance par rapport au centre de la cible (le résultat du mouvement). Les résultats montrent que les groupes ayant reçu les retours sous forme de superposition des modèles expert et apprenant ont montrés une amélioration de leur geste. Il est intéressant de noter que l'amélioration du geste n'améliore pas systématiquement la distance par rapport au centre de la cible, et inversement. Ainsi, lors de l'analyse du résultat d'une telle expérimentation, il faut être en mesure de positionner l'objectif final de l'apprentissage : amélioration du geste ou du résultat du geste.

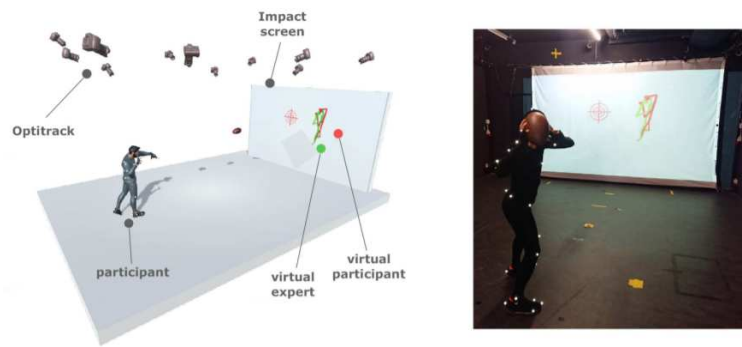


FIGURE 14 – Le système développé par LeNaour *et al.* propose une superposition en temps réel des mouvements de l'apprenant et de l'expert dans un environnement virtuel (LE NAOUR, HAMON et BRESCIANI 2019).

### 2.3.2 Tableau de synthèse

Tous ces systèmes ont pour point commun d'aider à l'apprentissage de geste, que ce soit en proposant une aide à la décision à l'enseignant, en intégrant les connaissances expertes à un système permettant de s'affranchir de la présence de l'expert, en proposant une évaluation du geste en fonction de critères pré-définis, ou en analysant automatiquement les différences entre les gestes, avec un *a priori* ou non.

Les systèmes où la connaissance experte est intégrée dès la conception permettent de focaliser la conception de l'EIAH sur un problème précis, et ainsi de l'adapter du mieux possible au geste cible (données relevées et utilisées, environnement d'apprentissage du geste, retours à faire à l'apprenant, etc.) (BALDOMINOS, SAEZ et POZO 2015; ALANKUS *et al.* 2010; GAMEIRO, CARDOSO et RYBARCZYK 2014; TOUSSAINT 2015; LE NAOUR, HAMON et BRESCIANI 2019). Dans ces cas, il est possible de faciliter l'apprentissage du geste, mais également de proposer des évaluations sous forme de score pertinent dans le contexte applicatif choisi ou de retours donnés par l'expert après visualisation des indicateurs de la performance. Cependant, comme le montre le tableau 1, ces systèmes ne sont pas réutilisables dans d'autres contextes sans nécessiter un processus de réingénierie conséquent : il faut ainsi re-formaliser et

réintégrer une autre connaissance experte dans une nouvelle phase de conception, possiblement ré-adapter le système au nouvel environnement dans lequel le geste est réalisé, acquérir de nouvelles données puis les intégrer au processus de traitement, et enfin ré-adapter les retours donnés en fonction des objectifs d'apprentissage.

L'utilisation d'algorithmes d'apprentissage automatique est une méthode qui permet de séparer les gestes de l'apprenant de ceux de l'expert de manière automatique selon des caractéristiques prédéfinies (NG 2008). En outre, elle permet également d'observer dans quelle mesure les caractéristiques déterminées au préalable du geste observé varient entre ces mouvements (TOUSSAINT 2015). Ces méthodes nécessitent de disposer d'une base de données d'apprentissage conséquente (*i.e.* de l'ordre du millier de données de mouvement) afin d'être capable de construire une bonne fonction de séparation. En pratique, il est fastidieux de devoir constituer un corpus de données annotées spécifique à un type de mouvement, bien que des bases de données existent (e.g. *Graphics Lab Motion Capture Database* (UNIVERSITY 2016), *Motion Capture Database HDMo5* (MÜLLER *et al.* 2007), *etc.*). Cependant, ces bases sont rarement spécialisées sur une tâche précise, et des variations et spécificités peuvent exister sur le geste à apprendre pour une même tâche et un même enseignant. De plus, l'utilisation d'algorithmes d'apprentissage supervisé suppose d'avoir la connaissance des différentes classes en sortie de l'algorithme, ce qui limite les types de restitution par rapport aux connaissances d'un expert humain, comme montré dans le tableau 1. Bien que l'expert soit en mesure de proposer des caractéristiques, il est possible que les variations de ces caractéristiques puissent mener à la formation de groupes possédant une sémantique plus fine que « réussi / raté » par exemple, permettant de proposer une séparation prenant en compte plusieurs niveaux de réussite du geste, ou menant à la découverte de caractéristiques spécifiques à un type de difficulté ou d'apprenant.

Les systèmes où la découverte des différences significatives entre des groupes de gestes se fait de manière totalement automatique s'affranchissent de l'intégration de la connaissance experte à l'aide du vocabulaire métier, où l'intègrent sous la forme de descripteurs cinématiques, géométriques ou biomécaniques (MAKIO, TANAKA et UEHARA 2007) (PIRSIAVASH, VONDRICK et TORRALBA 2014). De nouveaux critères de séparation des gestes acceptables ou non peuvent être découverts. Ils manquent cependant de sémantique dans leur retour : ainsi, il n'est pas possible d'identifier par exemple des groupes ou ensembles de gestes acceptables et non acceptables typiques afin d'étudier les caractéristiques des meilleures stratégies et des erreurs les plus fréquentes. De plus, il peut être difficile de fournir des retours pertinents à des personnes ne possédant pas des compétences poussées en physique, mécanique, *etc.*, car l'interprétation de données cinématiques, géométriques, *etc.* exige d'avoir ces connaissances. Ainsi, les conseils apportés par un tel système doivent non seulement être pertinent dans le contexte applicatif retenu, mais également compréhensibles et utilisables par des personnes sans connaissances scientifiques avancées, ce qui n'est pas toujours réalisable.

Auteurs	Domaine	Objectif visé	Acquisition des données	Modélisation des connaissances	Analyse	Acteur de la restitution	Information restituée	Visualisation du geste ou de la restitution	Généricité
Yoshinaga et Soga	Sport Archerie Japonaise	Amélioration du geste par rapport aux modèles experts	Caméra RGB-D (Kinect)	Corpus de squelettes 3D d'experts (enregistrement ad-hoc)	Empirique : analyse visuelle de l'expert du squelette 3D de l'apprenant effectuant le mouvement	Expert	Différences de positions entre les mouvements de l'expert et l'apprenant	Superposition de l'apprenant et des modèles de coaches	Générique, nécessite de réintégrer des modèles experts dans le système
Kora et al.	Sport Golf	Amélioration du swing	Caméra RGB-D, caméra sur casque de réalité virtuelle	Un modèle d'expert (squelette 3D)	Empirique (par l'étudiant) : visualisation des différences entre le modèle expert et son geste	Aucune restitution	Aucune information restituée	Superposition de l'apprenant et du modèle expert en temps réel	Dépendance totale du système au matériel et au domaine applicatif
Yamaoka et al.	Sport Lancer de disque	Position des articulations lors du lancer	Caméra RGB-D (Kinect)	Position relative des articulations entre elles dans les différentes étapes du mouvement (défini par un expert)	Automatique : comparaison de valeurs à des seuils	système (textuel)	Positions des membres à corriger	Retours générés automatiquement à partir de l'analyse	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception)
Ng	Musique Violon	Détermination du style de jeu	Capteurs réfléchissants (sur l'archet et le violon)	Représentation symbolique du son (système SMR), descripteurs du mouvement (amplitude, vitesse, durée)	Automatique (clustering sur les descripteurs extraits)	Enseignant et système (visualisation de la similitude des mouvements entre eux)	Mode de jeu, durée, position de l'archet et de l'instrument, vitesse de jeu, amplitude du mouvement	Visualisation 3D du mouvement du violon et de l'archet, regroupement des mouvements en fonction de leurs caractéristiques communes	Dépendance totale du système au matériel et au domaine applicatif
Choi et al.	Médical Insertion de sonde nasogastrique	Apprentissage de la procédure d'insertion	Bras haptique	Modèle 3D de patient vérifiant les collisions frictions, etc. Procédure complète et étapes clés modélisées sous la forme d'un arbre de décision	Empirique par l'expert (à partir de métriques : durée, min et max de force appliquées, nombre d'insertion et nombre de déformations du tube)	Expert (retours empiriques) et le système (retours visuels et haptiques)	Visualisation de la position de la sonde dans le modèle 3D, retour de force de l'instrument, durée, min et max de force appliquées, nombre d'insertion et nombre de déformations du tube	Visualisation 3D en temps réel de la position de la sonde dans un modèle 3D d'humain	Dépendance totale du système au matériel et au domaine applicatif
Baldominos et al.	Médical Réadaptation physique	Rééducation des patients	Caméra RGB-D (Intelisense) et casque de réalité virtuelle (Oculus Rift)	Position du patient comparée à des références fournies par les médecins pour les parties du corps concernées	Empirique : par l'expert médical	Expert médical	Rééducation suffisante ou non	Environnement de réalité virtuelle dans lequel le patient effectue les gestes désirés	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception), re-développement de scène virtuelle, dépendance du système au matériel utilisé
Toussaint et al.	Médical Chirurgie orthopédique percutanée	Apprentissage de la manipulation de l'objet pendant la procédure	Bras haptique et oculomètre	Séquences d'actions à effectuer par l'apprenant au cours de l'opération	Automatique : analyse de l'ordre des séquences d'action	Système	Propose de revoir une partie du cours, consulter un cas clinique ou résoudre un autre problème sur le simulateur	Visualisation en temps réel de la manipulation de l'objet	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception)
He et al.	Sport Chinese Taichi	Apprentissage du Taichi	Caméra RGB-D (Kinect)	Squelette 3D de référence pour le geste	Automatique : comparaison avec modèle expert au niveau de la position spatio-temporelle selon 3 métriques	Système	Score prenant en compte les 3 métriques calculées sur le mouvement de l'apprenant par rapport au mouvement de l'expert	Environnement de réalité virtuelle et CAVE	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception)
Chan et al.	Danse	Apprentissage de la danse	Caméra infrarouge + marqueurs	Un modèle d'expert (avatar reproduit à l'aide du squelette 3D)	Automatique : comparaison de la position des articulations entre l'apprenant et le modèle expert	Système	Visualisation du mouvement de l'apprenant et de celui de l'expert, coloration des parties du corps non synchronisées avec celles de l'expert	Environnement 3D (OpenGL) montrant le squelette de l'apprenant et le modèle de l'expert en train de réaliser les mouvements de danse	Générique par rapport au domaine applicatif, nécessite de réintégrer des modèles experts dans le système
Morel (thèse)	Sport	Amélioration du geste sportif	Caméra infrarouge + marqueurs	Un modèle d'expert (squelette 3D)	Automatique : décalage spatial et temporel à l'aide de la SDTW	Système	Visuelle et textuelle (affichage des 5 erreurs les plus significatives, ainsi que le temps auquel elles apparaissent)	Superposition du mouvement de l'apprenant par-dessus celui de l'expert, mise en évidence des erreurs les plus significatives	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception)
Gaimero et al.	Langue des signes	Amélioration du signage de l'apprenant	Caméra RGB-D (Kinect)	Images des signes fait par un ou des expert(s)	Automatique : masquage des données de l'apprenant par rapport au bon signe présent dans la base de données	Système	Réussite ou non du geste (restitution binaire)	Affichage de l'image du geste capturé	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception)
Xu et al.	Apprentissage de gestes (pas de contexte particulier)	Apprentissage de gestes spécifiques	Caméra RGB-D	Corpus de mouvements unitaires	Automatique : modèle de Markov caché	Système	Proposition d'exercices adaptés au renforcement de l'apprentissage des gestes non-maîtrisés	Sur le système, par une visualisation du geste correct à effectuer et des erreurs	Générique par rapport au domaine applicatif, nécessite de réintégrer des modèles experts dans le système
Maes et al.	Danse	Apprentissage de la danse	Caméra infrarouge + marqueurs	Descripteurs du mouvement (déplacement du centre du corps par rapport au tempo, rotation du corps autour de l'axe vertical)	Automatique : calcul de corrélation entre les descripteur du mouvement de l'apprenant et de l'expert	Système	Score final, correspondant à la similitude entre le geste de l'apprenant et celui de l'expert	Affichage du score	Dépendance totale du système au domaine applicatif
Makio et al.	Sport - générique	Séparation des gestes expert / apprenant	Non précisé	Descripteurs du mouvement	Automatique	Système	Différence au niveau des descripteurs du mouvement calculés	Affichage graphique de la différence entre les données de l'apprenant et de l'expert	Générique par rapport au domaine applicatif, nécessite de réintégrer des modèles experts dans le système
Le Naour et al.	Sport lancer au football américain	Amélioration du geste sportif	Caméra infrarouge + marqueurs	Descripteurs du mouvement et résultat du lancer	Automatique	Système	Différence visuelle entre le mouvement de l'apprenant et de l'expert	Visualisation du mouvement de l'apprenant superposé ou non à celui de l'expert (en fonction du groupe)	Nécessite de re-formaliser la connaissance experte au sein du système (re-conception)

TABLE 1 – Tableau résumant les principales caractéristiques des EIAH vu dans le chapitre.



## 2.4 DISCUSSION

L'apprentissage du geste peut prendre place dans de nombreux contextes. Dans ce chapitre, nous avons vu les différents matériels et EIAHs actuellement utilisés pour l'apprentissage du geste à l'aide d'un environnement informatique. Bien que ces systèmes présentent de nombreuses différences tant au niveau du matériel, que des données utilisées, des retours fait à l'apprenant, des méthodes d'intégration des besoins d'observation et des connaissances de l'expert et de la place de ce dernier dans l'usage, ils ont pour point commun de mettre à la disposition des enseignants et des apprenants un ensemble d'outils permettant de faciliter l'apprentissage, l'analyse ou l'évaluation des gestes. Ces systèmes proposent généralement une analyse du geste, sous une ou plusieurs formes : sous la forme de comparaison visuelle entre les données de l'expert et celles de l'apprenant, en proposant une évaluation sous forme de score, en permettant de visualiser le parcours de l'apprenant dans le cas de multiples gestes qui se suivent, *etc.* L'intégration de l'analyse automatique permet de proposer une aide à l'évaluation, voire de remplacer l'expert dans cette tâche. Bien que l'apprentissage supervisé soit utilisé pour l'analyse de gestes, afin de séparer automatiquement les gestes en fonction de leur sémantique ou de leur degré de réussite, peu de travaux explorent l'utilisation d'algorithmes d'apprentissage non-supervisé. Cela peut s'expliquer par le manque de sémantique associée aux groupes obtenus en sortie. Il est cependant possible d'utiliser un petit nombre de données étiquetées, dans le contexte de l'apprentissage semi-supervisé, afin d'être en mesure de donner un sens aux groupes ainsi obtenus.

Ce chapitre a mis en relief le manque de systèmes assistant l'enseignant qui soient réutilisables dans plusieurs contextes sans une réingénierie conséquente, malgré la variété existante d'EIAH dédiés à l'apprentissage du geste. Une des hypothèses de ce travail de thèse est qu'il est possible de proposer un EIAH réutilisable dans d'autres contextes avec un coût de réingénierie minimisé, accompagnant l'expert dans sa tâche d'enseignement des gestes, en combinant plusieurs des techniques utilisées au sein des systèmes déjà présentés. L'intégration de la connaissance experte en amont de la séance d'apprentissage, puis l'utilisation d'algorithmes d'apprentissage semi-supervisés permettraient de proposer une aide à la décision, en fonction de caractéristiques observées à partir des captures des gestes de l'apprenant. L'avantage d'un tel système serait de permettre à l'enseignant d'analyser toutes les facettes voulues du mouvement sur un même geste, ce qui n'est pas toujours aisé ou possible dans le cadre de mouvements faisant appel à plusieurs parties du corps en même temps (par exemple, l'analyse du jeu d'un étudiant à la guitare).

Que le processus d'analyse soit automatique ou non, il peut être dur d'inférer des informations quant à la bonne réussite du geste ou non à partir de données de mouvement seules. Les EIAH présentés dans cette partie n'utilisent parfois que les données de mouvements capturés. Ces systèmes se basent sur une comparaison du mouvement à une référence, qui peut être l'expert, ou simplement une comparaison des paramètres clés du mouvement avec ceux de référence. Il est ainsi possible ainsi de visualiser les moments de décalages (autant temporels que spatiaux) entre le mouvement de l'apprenant et celui du mouvement attendu. Ces retours donnés à l'apprenant sont cependant souvent " bas-niveau ", c'est-à-dire qu'ils ne sont donnés qu'en terme de valeurs cinématiques, dynamiques, géométriques et nécessitent des connaissances scientifiques afin d'être interprétable par un apprenant. L'adjonction de l'analyse de l'expert n'est pas systématiquement possible, car certains de ces systèmes sont conçus pour pouvoir être utilisés en autonomie.

Ainsi, il faut être en mesure de fournir une analyse fine à l'aide d'indications précises quant aux différentes caractéristiques du geste tout en utilisant le vocabulaire métier du domaine applicatif. Dans ce contexte, il est important d'être capable d'extraire des valeurs plus facilement interprétables à partir de données de mouvements capturés, ou plus spécifiquement adaptées au contexte. Le prochain chapitre présente les différentes informations qu'il est possible d'extraire à partir de mouvements capturés.



---

## DESCRIPTEURS DU MOUVEMENT

---

Le mouvement humain est porteur d'informations, que ce soit lors d'une communication verbale (JIE HUANG *et al.* 2015) ou non-verbale (CHANG *et al.* 2013b). Ces informations sont d'ordres différents et peuvent caractériser l'état psychologique de l'Homme (détection de l'émotion (YUICHI KOBAYASHI 2007) et de l'intention (YU et M. LEE 2015b) ou décrire les interactions dans son environnement (reconnaissance d'actions (KAPSOURAS et NIKOLAIDIS 2014b)). Les gestes de la vie quotidienne (prendre un objet, s'asseoir, s'allonger, etc.) peuvent être détectés par des systèmes automatiques, tout comme leurs perturbations liées à un contexte émotionnel ou affectif (perturbation du geste lié à l'énervement, la compassion, etc.).

Si l'on se place dans un contexte d'analyse du geste, il faut être en mesure d'observer certaines caractéristiques du mouvement, relatives à, par exemple, la vitesse des articulations lors d'un lancer ou le déplacement du corps lors d'une danse, afin de pouvoir en extraire des informations pertinentes dans le contexte de l'analyse. Larboulette et Gibet définissent le geste comme une combinaison d'éléments multiples, associant le sens, le style et l'expressivité (LARBOULETTE et GIBET 2015). Le sens est décrit par un ensemble d'actions réalisées au sein du geste, permettant ainsi de définir l'objectif du mouvement. Le style est constitué des éléments propres à la personne effectuant le geste : sa morphologie, son genre, sa personnalité, ainsi que la manière d'effectuer le geste (se déplacer de manière gracieuse ou pesante, par exemple). L'expressivité englobe les nuances émotionnelles qui vont caractériser l'état d'esprit de la personne réalisant le geste.

Ces informations peuvent être déduites par analyse du mouvement brut capturé, en observant par exemple, les variations du même mouvement effectué dans le but d'accomplir la même tâche par différentes personnes. Si on considère une capture de « qualité » du corps humain (c'est-à-dire suffisamment de points de capture sur les parties du corps ciblées, à une fréquence suffisamment élevée, les valeurs acceptables de ces deux caractéristiques étant dépendantes de l'objectif d'apprentissage), il est alors possible de calculer, à partir des données brutes, des indicateurs exprimant des informations utiles pour l'apprentissage, en fonction des besoins d'analyse de la situation pédagogique. Ces indicateurs extraits sont nommés « descripteurs » dans la suite de ce manuscrit.

### 3.1 DÉFINITION ET PRINCIPES

Un descripteur est un indicateur calculé à partir du mouvement brut. Il est utilisé pour représenter des informations d'une partie ou de l'ensemble du corps, calculé à des temps  $t$  précis ou selon un intervalle de temps pouvant aller jusqu'à la durée totale du mouvement. Un parallèle peut être fait avec la définition des indicateurs proposés par Choquet et Iksal : « des observables signifiant sur le plan pédagogique » calculés à partir de traces (tous types de données, générées à partir des interactions de l'étudiant avec le système) ou d'autres indicateurs (CHOQUET et IKHAL 2007). Dans

notre contexte, un descripteur permet de quantifier une ou plusieurs caractéristiques du mouvement, facilitant ainsi l'analyse manuelle ou automatique en diminuant, notamment, la quantité de données numériques à analyser. En effet, si l'on considère une représentation d'un squelette constitué de 20 articulations lors d'un mouvement d'une durée de 15 secondes, enregistré à 60 postures par seconde, avec 3 valeurs pour la position (espace tridimensionnel) et 3 valeurs pour l'orientation (représentation d'Euler) de chaque articulation, on obtient ainsi le nombre suivant de données à

analyser  $\overbrace{15}^{\text{temps}} * \overbrace{60}^{\text{nb frames}} * \overbrace{20}^{\text{nb articulations}} * \overbrace{3*3}^{\text{position et orientation}} = 162000$ . De plus, les données de rotation sont souvent représentées sous forme de quaternion, de manière à pallier le problème du blocage de cadran (perte d'un degré de liberté résultant de la coplanarité de deux axes) propre à la représentation d'Euler (KUIPERS 2000), ce qui augmente le nombre de données obtenues. En conséquence, il est difficilement possible pour un humain de pouvoir observer, puis inférer des informations sur le mouvement à partir de l'affichage de ces valeurs numériques brutes. De plus, les informations sous forme d'un ensemble de positions et d'orientations nécessitent des connaissances scientifiques (au moins en géométrie) afin d'être interprétées. Le calcul de descripteurs à partir de données brutes a notamment pour objectif de réduire le nombre de valeurs à analyser, et de donner une nouvelle sémantique plus proche des connaissances de l'utilisateur et de son vocabulaire métier, par utilisation des modèles sémantiques hiérarchiques ou de réification (BAUDOUIN *et al.* 2008). Par exemple, à partir de la position, il est possible de calculer la distance d'écart entre deux membres, puis de qualifier l'amplitude du mouvement en fonction de valeurs seuils de la distance. De même, à partir de la position, il est possible de calculer des vitesses et des accélérations, puis de définir des valeurs seuil pour qualifier le mouvement comme un mouvement « régulier », « haché », « fluide » *etc.* selon le vocabulaire métier.

Enfin, notons qu'il est possible d'extraire des descripteurs de mouvement à partir de deux types de données brutes principalement : (i) les squelettes (*i.e.* ensemble de points capturés dans l'espace tridimensionnel, dont la chaîne articulaire a été reconstruite) capturés à l'aide d'une combinaison de capteurs inertiels, de caméra RGB-D ou de système de marqueurs avec des caméras infrarouges, et (ii), les images, c'est-à-dire des ensembles de pixels RGB ou nuage de points RGB-D provenant d'enregistrements vidéo.

### 3.2 REVUES DES DESCRIPTEURS EXISTANTS, CLASSIFICATION ET MODÈLE DE REPRÉSENTATION

Plusieurs travaux ont proposé de classer des ensembles de descripteurs du mouvement. Ces classifications ont souvent pour but de faciliter le choix d'un ensemble de descripteurs potentiels à partir des besoins d'analyse du geste : c'est le cas d'études telles que celles de Larboulette et Gibet (LARBOULETTE et GIBET 2015) ou Morel (MOREL 2017). D'autres travaux s'intéressent au regroupement d'ensembles de descripteurs pertinents dans un contexte applicatif précis, par exemple la danse (BOUCHARD et BADLER 2007a), l'écriture manuscrite (DELAYE et ANQUETIL 2013), *etc.* Une présentation de différentes classifications de descripteurs, ainsi que de différents ensembles de descripteurs sont proposés dans cette partie.

### 3.2.1 Descripteurs pour l'analyse de la performance

Les travaux de Larboulette et Gibet (LARBOULETTE et GIBET 2015) proposent de classer les descripteurs communément utilisés pour décrire le geste dans plusieurs autres études, selon plusieurs niveaux de regroupement : bas-niveau / haut-niveau, cinématique / géométrique, corps / environnement. La frontière bas-niveau / haut-niveau est définie par l'information donnée par le descripteur : un descripteur concernant des quantités cinématiques, dynamiques ou géométriques sera classé comme un descripteur bas-niveau, alors qu'un descripteur permettant d'exprimer des quantités exploitables dans une évaluation visuelle du mouvement ou par un expert sera considéré comme haut-niveau. La classification des descripteurs haut-niveau comme étant « exploitables » ne prend cependant pas en compte les connaissances de l'utilisateur final. La deuxième caractéristique propre aux descripteurs haut-niveau est qu'ils sont systématiquement calculés à partir de descripteurs bas-niveau, au même titre que les indicateurs de Choquet et Iksal sont calculés à partir des traces. Cette classification fournit également un formalisme mathématique homogène en posant une notation unifiée pour décrire le mouvement, les positions et les orientations à partir desquelles sont calculés les descripteurs. Ces travaux ne traitent cependant pas de l'utilisation possible de ces descripteurs, tant au niveau de l'analyse qu'il est possible d'en tirer (empirique ou non) que des cadres scientifiques ou applicatifs dans lesquels ils sont pertinents ou habituellement considérés. De plus, ces descripteurs nécessitent pour la majorité d'entre eux une connaissance scientifique poussée (biomécanique, par exemple) pour en tirer des informations utiles dans le cadre de l'apprentissage d'un geste.

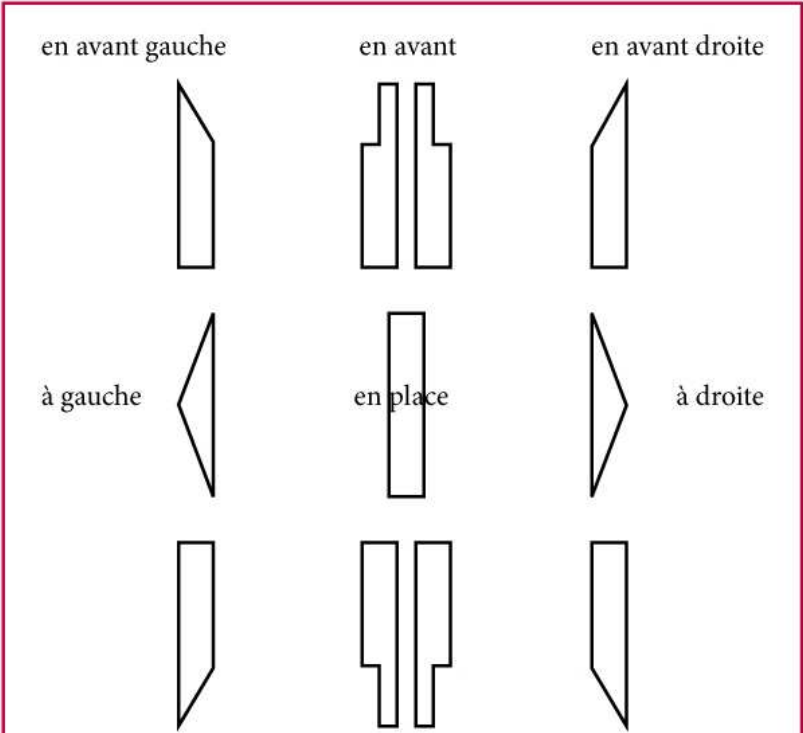
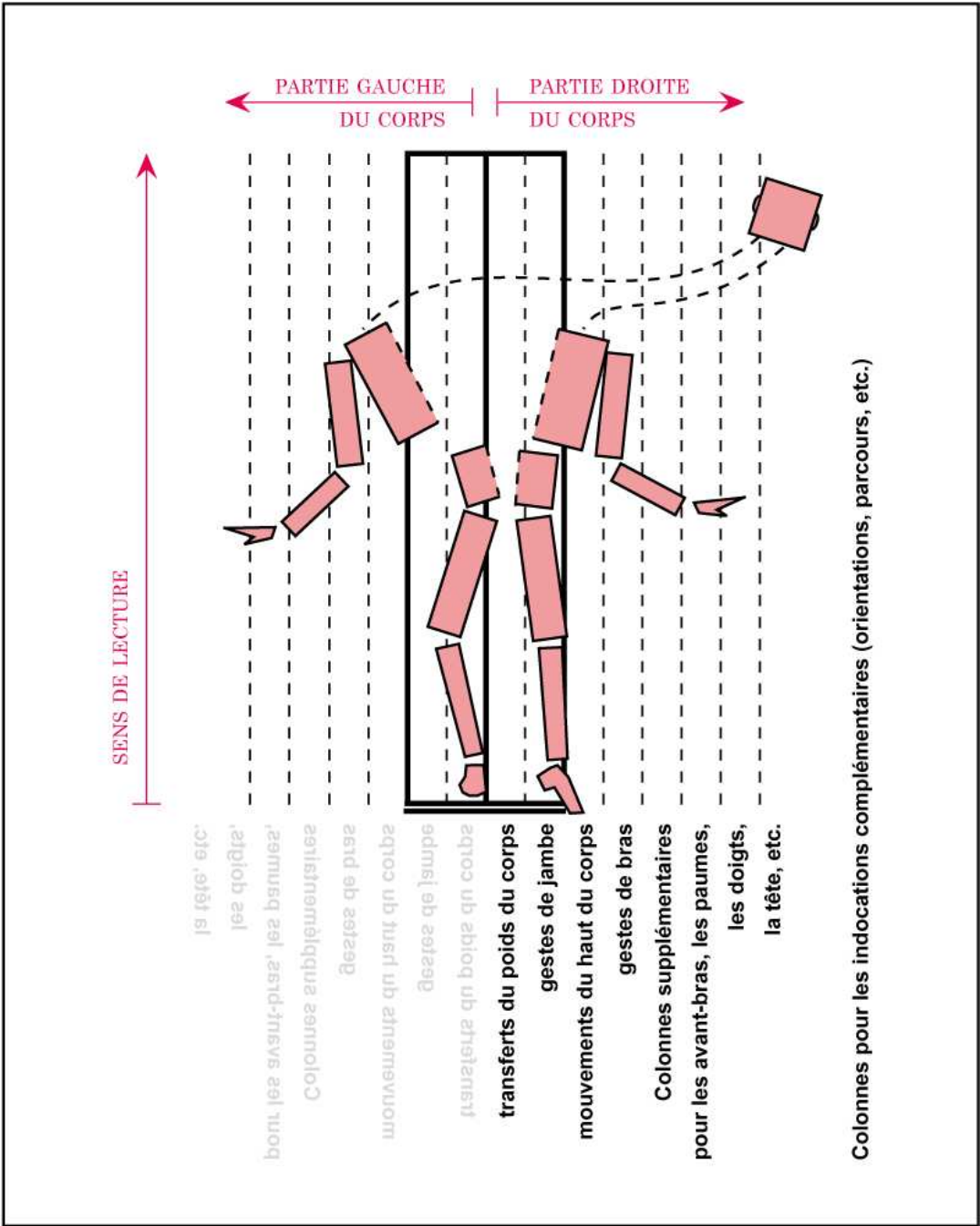
Dans le cadre de son travail de thèse, Morel a également présenté une classification des descripteurs (MOREL 2017), en trois familles : les descripteurs reposant sur un modèle du corps humain, les descripteurs holistiques, qui utilisent la dynamique globale de l'objet considéré (corps humain ou autre), et les descripteurs locaux, caractérisant les mouvements à partir de points d'intérêt isolés. Pour chacune des familles, les descripteurs peuvent être calculés de manière à former une chaîne temporelle (« codage temporel », c'est-à-dire que le signal est codé à chaque instant) ou de manière globale (« codage global », un vecteur de caractéristiques représente le geste dans son entièreté). La famille des descripteurs reposant sur un modèle de corps humain est constituée des descripteurs de trajectoire, vitesse, accélération (linéaire ou angulaire), courbure, *etc.* du mouvement. Sont également inclus les descripteurs géométriques, tels que les formes englobantes (cubes, ellipsoïdes, *etc.*), et les descripteurs d'efforts, correspondant à l'effort déployé par une partie du corps sur l'intégralité du mouvement. La deuxième famille, celle des descripteurs holistiques, contient les descripteurs relatifs au mouvement de l'objet considéré dans sa globalité, sans s'intéresser aux spécificités dudit objet. Ainsi, vont être inclus les descripteurs calculés à partir de vidéos : zone en mouvement au sein de la vidéo, calcul d'une fonction d'énergie à partir des différences entre les images successives d'une vidéo, variation du gradient des images, *etc.* Enfin, la troisième famille de descripteurs locaux se focalisent sur les points d'intérêts d'un objet, déterminés de manière automatique ou non. En observant le voisinage spatial ou temporel de ces points, il est ainsi possible d'obtenir des informations par rapport au mouvement de l'objet considéré. Des traitements visant à réduire les ensembles de descripteurs à un sous-ensemble sont également étudiés, tels que l'analyse en composantes principales (PCA) ou la décomposition en valeurs singulières. Cette réduction conserve le même pouvoir discriminant pour la séparation des gestes par rapport à leurs caractéristiques initiales prédominantes. Bien qu'une étude analysant

la pertinences de certains descripteurs en fonction de différents contextes applicatifs se trouve dans ce manuscrit, celle-ci ne concerne pas directement les familles de descripteurs proposées, ce qui fait qu'il n'est pas possible de directement mettre en contexte les descripteurs ou les familles et de proposer des cas applicatifs précis.

Il est intéressant de noter que le concept de bas-niveau / haut-niveau proposé par Larboulette et Gibet n'est ici pas un critère de séparation des descripteurs ; en effet, on retrouve des descripteurs bas-niveaux (*i.e.* vitesse) et des descripteurs haut-niveau (*i.e.* forme englobante) au sein d'une même catégorie. De plus, la classification de Morel mélange également des descripteurs issus de squelettes 3D et des descripteurs extraits de vidéos.

À l'origine de la description et la formalisation de mouvement humains, le système LBA (Laban Movement Analysis) est un système qui a beaucoup été utilisé, autant pour l'analyse que pour la séparation de gestes (BOUCHARD et BADLER 2007a). L'objectif de ce système descriptif de gestes est de proposer une annotation hiérarchique dans l'objectif de caractériser originellement les mouvements de danse. Cette notation quantifie quatre informations sur le mouvement : sa direction, sa durée, la partie du corps concernée et sa dynamique. Les trois premières informations sont représentées par des signes (voir Fig. 15, en bas) et le placement de ces signes sur une portée définit la dynamique du mouvement (Fig. 15, en haut) : la succession des mouvements se lit sur l'axe vertical, de bas en haut, et les parties du corps concernées sont représentées sur l'axe horizontal. Ces travaux ont été les premiers à proposer un ensemble de descripteurs unifiés afin de décrire un geste dans sa continuité. Cependant, dans le contexte d'une opérationnalisation du système LBA, il est nécessaire de choisir et d'implémenter au préalable un ou plusieurs descripteurs de bas ou de haut niveau (au sens de la classification de Larboulette et Gibet) pour chaque dimension considérée du modèle, en fonction du mouvement à analyser.

Aristidou *et al.* proposent un ensemble de descripteurs caractérisant la danse folklorique selon quatre composantes du système de Laban, à savoir le corps, l'effort, la forme et l'espace (Andreas ARISTIDOU *et al.* 2015). Les positions et les orientations, la hauteur du pelvis, la distance des jambes par rapport au sol et la taille des pas (distance entre les deux pieds) sont utilisés pour caractériser le corps. L'effort est divisé en quatre sous-catégories, possédant chacune deux polarités possibles : (i) l'espace occupé direct (précis et spécifique) ou indirect (non concentré sur un seul point), décrit à l'aide de l'orientation de la tête par rapport à la trajectoire du corps, (ii) le poids appliqué, fort ou léger, décrit à l'aide de la décélération du mouvement de la hanche, (iii) le temps, soudain (sentiment d'urgence, mouvement inattendu ou isolé) ou continu (réalisation du mouvement s'étalant dans le temps), décrit à l'aide de la vitesse et l'accélération du mouvement sur une période de temps prédéfinie, et (iv) le flux du mouvement, restreint (contrôlé, attentionné) ou libre (relâché, épanché, fluide), décrit à l'aide de la saccade du mouvement. La forme est décrite à l'aide du volume occupé par le corps de la personne, pris sur les cinq articulations extrêmes du corps (la tête, les mains et les pieds), la hauteur du torse, et la position des mains par rapport au corps. Enfin, l'espace est décrit à l'aide de la distance couverte par les hanches au cours du temps, ainsi que l'aire correspondant à ce déplacement. L'organisation en quatre familles des descripteurs possède quelques similitudes avec la classification bas-niveau / haut-niveau qui est proposée dans les travaux de Larboulette et Gibet. Après extraction, les descripteurs de l'apprenant sont comparés avec ceux de l'expert, afin d'obtenir un score de corrélation correspondant à la similitude entre les deux gestes (Fig. 16). Ce score, correspondant à la valeur absolue du coefficient de corrélation linéaire de Pearson, met en évidence les moments où le geste de l'apprenant est éloigné de celui de l'expert, mais ne propose pas une





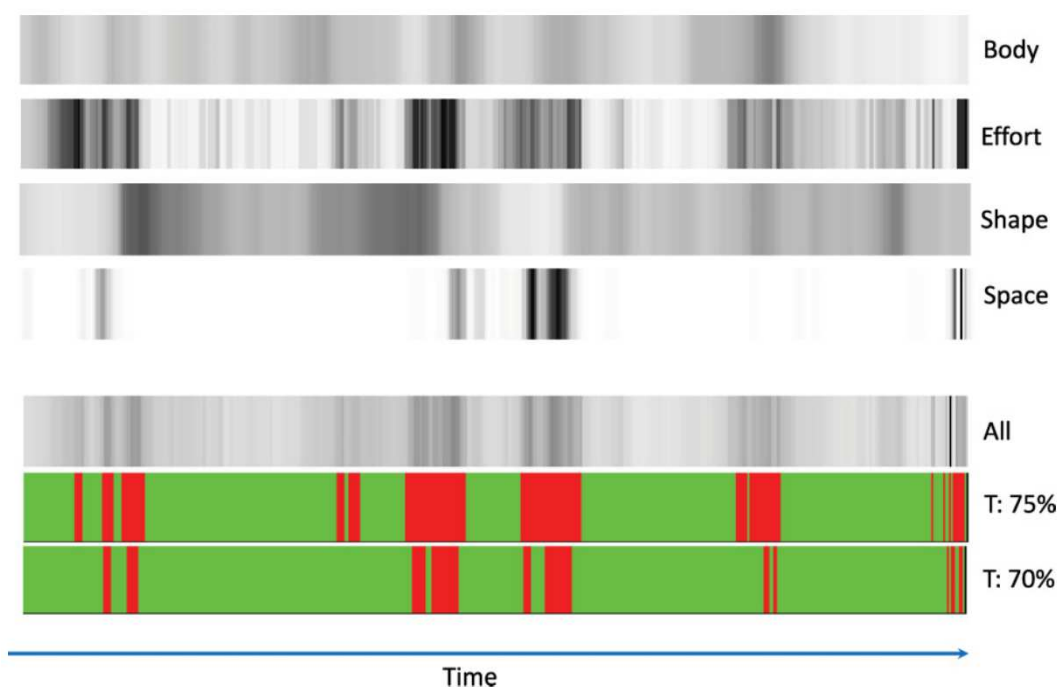


FIGURE 16 – La corrélation entre le geste de l'apprenant et de l'expert est calculée selon les quatre composantes du système de Laban considérées, puis moyennées. Le seuil  $T$  choisi permet de changer la corrélation minimale à atteindre afin que le geste soit considéré comme bon. (Source : (Andreas ARISTIDOU *et al.* 2015)).

analyse sur les caractéristiques du mouvement à changer, malgré la précision des descripteurs utilisés. En conséquence, cette évaluation sous forme de score n'aide pas à l'amélioration de l'apprentissage du geste par l'apprenant.

Toujours dans le domaine de la danse, Senecal *et al.* utilisent un ensemble de descripteurs pour l'apprentissage de la salsa, selon trois axes : le rythme, l'interaction entre les partenaires et le style (SENECAL, NIJDAM et THALMANN 2018). Ces descripteurs sont relatifs aux déplacements à effectuer lors de la danse. Ainsi, le rythme est évalué grâce à la position des pieds. Les valeurs successives de la norme de la vitesse des pieds attendues sont obtenues grâce à l'analyse du geste de l'expert. Le décalage entre les points d'inflexion de ce profil, correspondant à une nouvelle étape dans la suite de pas à réaliser, et ceux de l'apprenant permettent de mettre en évidence la synchronisation des pas avec la musique : un décalage temporel est ainsi obtenu. Le spectre du rythme de danse est ensuite extrait, à l'aide d'une transformée de Fourier, pour obtenir le décalage fréquentiel. L'interaction entre les partenaires est analysée sous trois angles : la corrélation du mouvement des jambes des deux partenaires, le décalage de rythme entre les deux partenaires, et enfin le décalage fréquentiel entre les deux partenaires. Enfin, le style est analysé à partir de l'aire couverte par la performance entière et la quantité de mouvement moyenne des hanches. Cette étude nécessite néanmoins d'extraire le tempo de la musique utilisée, dans le but d'obtenir un point de référence pour la synchronisation. Les descripteurs du mouvement sont donc complétés par une information temporelle supplémentaire non-relative au mouvement, celle des temps auxquels les percussions de la musique se font entendre. Ces données ne sont pas analysables sans connaissances en traitement du signal. Ainsi, il est impossible de présenter ces descripteurs sans une phase d'interprétation nécessaire pour tirer des conclusions sur le mouvement des apprenants. De plus, cette étude montre que malgré la sélection des descripteurs *a priori* avec un expert,

seuls quatre d'entre eux peuvent réellement aider à qualifier de manière satisfaisante le niveau des apprenants.

### 3.2.2 Descripteurs pour l'indexation et la recherche de mouvements

Les méthodes d'indexation et de recherche de mouvements cherchent à indexer les mouvements à l'aide de caractéristiques discriminantes, dans le but de les stocker au sein d'une base de données. Le second objectif est de créer un langage de requête servant à retrouver et extraire efficacement et rapidement ces mouvements au sein d'une telle base de données.

Dans le cadre de la recherche de mouvements, Sakurai *et al.* extraient un pentagone servant à représenter les postures du mouvement, à l'aide des données de position des articulations du squelette, dans l'optique de les comparer *a posteriori* (SAKURAI *et al.* 2014). Les données, capturées à l'aide d'une caméra Kinect™, sont prétraitées pour éliminer le bruit inhérent à ce type de capture. Ce traitement consiste à calculer, pour chaque posture, une moyenne sur les cinq postures adjacentes. Les postures sont ensuite normalisées, en harmonisant l'origine des repères, puis en divisant la longueur de chaque membre du corps par la taille du squelette. Un pentagone est ensuite extrait, reliant la tête, la main droite, le pied droit, le pied gauche et la main gauche. De ce polygone, trois aires différentes sont extraites, servant à représenter le mouvement (17). La phase de recherche utilise ensuite l'algorithme de la Déformation Temporelle Dynamique (*Dynamic Time Warping*, *DTW*) pour calculer la distance entre le mouvement effectué et ceux présents au sein de la base de données. Cet algorithme permet de mesurer la similarité entre deux séries temporelles, ainsi que de les aligner, tout en s'affranchissant des contraintes temporelles. Plus formellement, l'algorithme du *DTW* réalise un *matching* entre les valeurs discrètes de deux séries temporelles. Si l'on prend l'exemple de deux personnes effectuant le même parcours, mais marchant à des vitesses différentes, la valeur du *DTW* calculé entre ces deux séries temporelles sera élevée, indiquant une forte similitude entre ces deux séries, là où la distance euclidienne indiquera un écart élevé entre ces séries. Le mouvement stocké ayant la distance au sens du *DTW* la plus petite est considéré comme étant le mouvement recherché. Afin d'évaluer les résultats, la métrique du *F-score* est utilisée. Cette métrique est une combinaison de la Précision (P), qui correspond au ratio de réponses correctes retournées par rapport au nombre total de réponses, et du Rappel (R), qui correspond au ratio de bonnes réponses retournées par rapport au nombre total de bonnes réponses. Les résultats des expérimentations montrent que, dans tous les cas, le descripteur basé sur les l'aire des quatre triangles extrait des mouvements enregistrés par la Kinect™ fournit le meilleur *F-score*, par rapport aux deux autres descripteurs. Cette décomposition du mouvement en aires géométriques permet d'obtenir une performance de retour d'environ 71% (au sens du *F-Score*), environ 30% plus faible que lors de l'utilisation de données capturées avec des systèmes de capture du mouvement plus précis. Les auteurs notent que cette différence peut être due à deux paramètres : la qualité des données obtenues (et donc, par extension, la méthode de filtrage appliquée) ainsi que le nombre d'aires extraites du polygone.

Xiao *et al.* ont développé un système de descripteurs basés sur des postures 2D permettant de retrouver un mouvement à partir de dessins de postures-clés (XIAO *et al.* 2015). Le mouvement est d'abord projeté sur 8 plans différents, de façon à passer d'une représentation 3D à plusieurs représentations 2D. Le système extrait ensuite des lignes reliant certaines articulations entre elles (les articulations extrêmes entre elles et les articulations extrêmes par rapport à la hiérarchie du squelette) servant à caractériser les articulations du corps à partir de leur distance relative, la

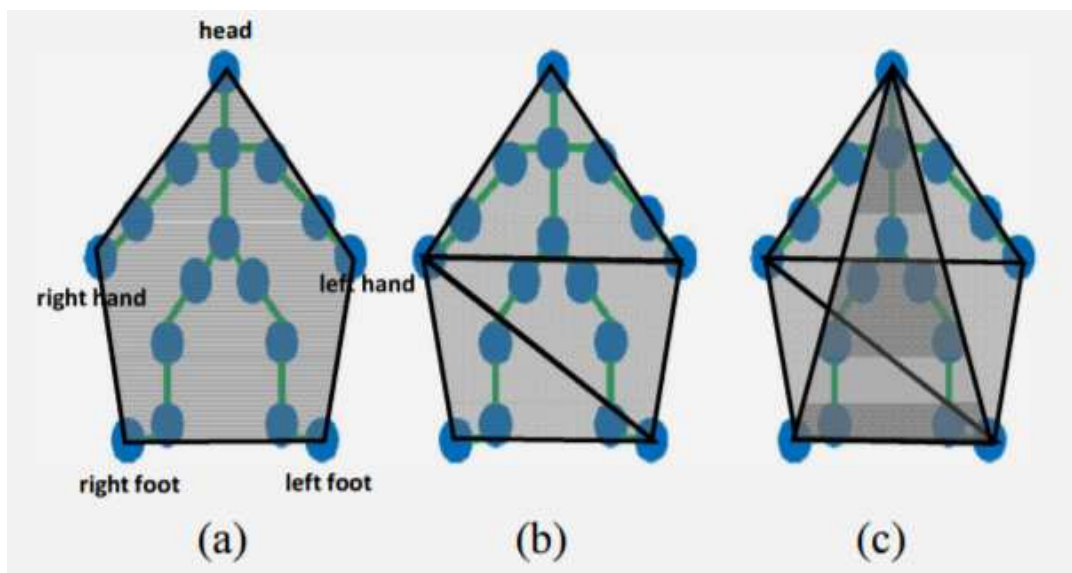


FIGURE 17 – Les trois descripteurs, constitués d’aires extraites du polygone englobant le squelette : (a) l’aire du polygone entier, (b) l’aire de trois polygones, (c) l’aire de quatre polygones (Source : (SAKURAI *et al.* 2014)).

direction de cette distance, la distance des articulations par rapport aux lignes et l’angle formé entre les différentes lignes (Fig. 18).

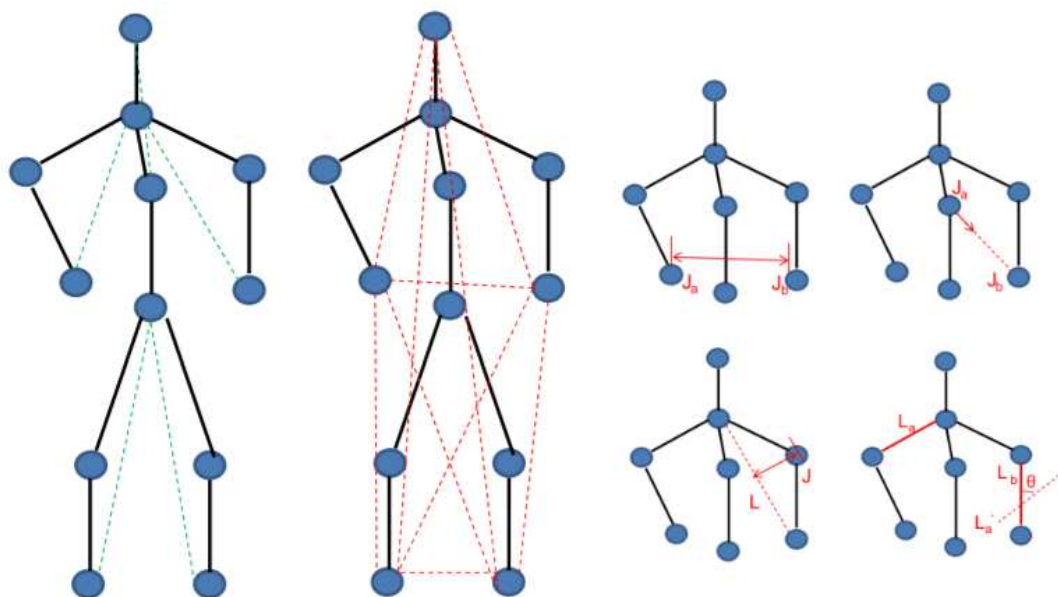


FIGURE 18 – Les descripteurs extraits caractérisent les postures par rapport aux distances relatives des articulations entre elles, ainsi qu’à leur distance et l’angle formé par rapport aux lignes reliant les autres articulations (Source : (XIAO *et al.* 2015)).

Ces descripteurs sont ensuite réduits à un sous-ensemble, afin de limiter la redondance des informations contenues dans les données et d’améliorer les temps d’analyse *a posteriori*. Pour cela, une méthode de sélection de caractéristiques semi-supervisée, basée sur le score Laplacien est utilisée. Ce score se base sur l’hypothèse que la projection des données dans un espace de caractéristiques réduits préserve la structure initiale de ces données (HE, CAI ET NIYOGI 2005). Le processus de recherche du mouvement se base sur un dessin fait à main levée par l’utilisateur. Le nombre de traits est limité à cinq, et le premier trait doit correspondre au torse. Grâce à ces

contraintes, il est possible d'étiqueter automatiquement les articulations une fois le dessin terminé. La suite de postures dessinées par l'utilisateur est ensuite comparée à toutes les projections des mouvements présents dans la base de données à l'aide d'un algorithme des K plus proches voisins, et le mouvement le plus proche du dessin est récupéré (Fig. 19).

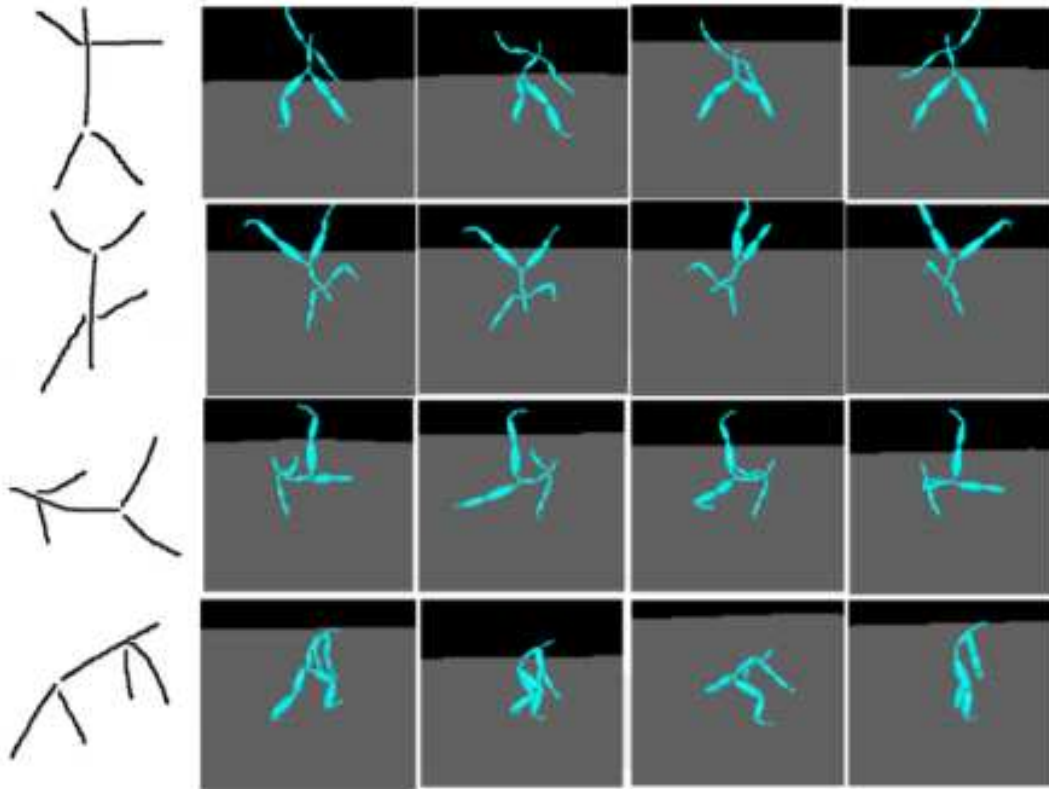


FIGURE 19 – Quatre exemples de mouvement retournés à partir de postures dessinées à la main (Source : (SAKURAI *et al.* 2014)).

Les travaux cités ci-dessus proposent des ensembles de caractéristiques ayant pour but de stocker et de rechercher des mouvements de manière rapide et précise. Ces ensembles utilisent généralement une étape de sélection de caractéristiques, afin de réduire la dimensionnalité des données et ainsi d'accélérer le processus de recherche. Cette réduction peut cependant éliminer des descripteurs utiles à l'analyse du mouvement par un humain, et ainsi priver l'expert d'informations cruciales. En outre, la non prise en compte de l'utilisabilité, ni de la contextualisation de ces descripteurs par un humain fait que leur analyse peut être limitée en termes de retours possibles sur l'apprentissage d'un geste, voire impossible.

### 3.2.3 Descripteurs pour la reconnaissance d'actions

Certains travaux se focalisent sur l'extraction de descripteurs servant à identifier et séparer plusieurs actions. Dans ce cas, l'action peut soit concerner le corps entier (s'asseoir, prendre un objet, *etc.*) ou des parties précises du corps (*e.g.* reconnaissance d'écriture manuscrite).

Ainsi, dans le cadre de la reconnaissance d'écriture, Delaye et Anquetil ont implémenté un ensemble de 49 descripteurs, nommé HBF49, liés à la manière d'écrire (DELAYE et ANQUETIL 2013). Ces descripteurs servent à caractériser les mouvements d'écriture, à l'aide : des points de début et de fin de trait, des angles formés par les différentes parties des lettres, de la proportion de trajectoires dirigées

vers le bas, *etc.* Cet ensemble de descripteurs a pour but de pallier les limites des ensembles de descripteurs existants pour l'écriture manuscrite. En effet, les travaux antérieurs ne permettent pas de caractériser toutes les informations pour ces gestes, car ils répondent souvent à un besoin d'observation et d'analyse précis, et aucune évaluation de référence n'a été conduite afin de les valider. Les tests effectués sur plusieurs jeux de données reconnus montrent que cet ensemble obtient un meilleur taux de reconnaissance de symboles que les jeux de descripteurs existants.

Dans le cadre de la reconnaissance d'actions, les travaux de Boulahia *et al.* se basent en partie sur l'ensemble de descripteurs HBF49 et les généralisent à une trajectoire 3D dans l'environnement (BOULAHIA *et al.* 2016). Ces descripteurs ne posent aucune hypothèse sur les gestes considérés, et peuvent être utilisés pour n'importe-quel type de gestes. La validité de ce jeu de descripteurs pour décrire le mouvement, appelé HIF3D, a été testé sur deux jeux de données (HDMo5 (MÜLLER *et al.* 2007) et UTKinect (XIA, C.-C. CHEN et J. K. AGGARWAL 2012)) constitué de gestes non-techniques, et mis en concurrence avec d'autres ensembles de descripteurs déjà testés sur ces jeux de données (*e.g.* SMIJ, Cov3DJ, HOJ3D). Il en ressort que les résultats sont meilleurs avec le jeu de descripteurs HIF3D.

Rao *et al.* utilisent les moments d'inflexion dans la trajectoire, appelés instants dynamiques, afin de proposer une représentation des changements de trajectoire du geste, permettant ainsi de les segmenter en unités élémentaires (RAO, YILMAZ et SHAH 2002). Les gestes utilisés dans l'étude sont des gestes du quotidien, *e.g.* ouvrir et fermer un placard, prendre un objet, le poser, *etc.* Ces unités élémentaires sont ensuite utilisées pour associer d'autres parties du mouvement, ou d'autres mouvements, à cette action. Ainsi, le système peut ensuite soit reconnaître une unité élémentaire comme étant similaire à une des unités élémentaires déjà visualisée, ou alors créer une nouvelle unité élémentaire si jamais celle-ci n'est pas suffisamment semblable à celles présentes dans le corpus. Cependant, la méthode ne se basant sur aucun apprentissage ou modèle pour segmenter et reconnaître les mouvements, aucune sémantique n'est associée aux groupes d'unités élémentaires ainsi obtenus.

Des ensembles de descripteurs tel que HIF3D proposé par Boulahia *et al.* couvrent beaucoup d'aspects différents du geste, offrant ainsi un bon taux de reconnaissance des actions. L'intégration de ces descripteurs à un système limite les besoins de réingénierie dans le cas de multiple domaines applicatifs, car il sont pensés pour être capable de reconnaître une grande variété de mouvements. Cependant, beaucoup de ces travaux ne se focalisent pas sur l'apprentissage humain de mouvements. Ainsi, l'utilisation de ces descripteurs en tant qu'information descriptive humainement interprétable n'est pas prise en considération. Il n'est pas assuré que les descripteurs soient utilisables dans un processus d'analyse par un expert du geste ne disposant pas de connaissances scientifiques spécifiques à ces descripteurs.

### 3.2.4 Discussion

Les travaux portant sur la classification des descripteurs ont pour but de créer des familles, afin de corréliser chacune d'entre elles à un type d'analyse voulue. Ces classifications présentent des différences dans leur méthode de discrimination. Ainsi, Larboulette et Gibet (LARBOULETTE et GIBET 2015) s'intéressent d'avantage à la séparation des descripteurs en termes d'origine des descripteurs : les descripteurs concernant la cinématique, la dynamique et la géométrie, calculés à partir du mouvement brut, sont dits de bas-niveau, et les descripteurs calculés à partir de descripteurs bas-niveaux sont dits de haut-niveau. Chez Morel MOREL 2017, la considération première est de séparer les descripteurs en fonction de l'objet en mouvement considéré.



Ainsi, la première famille concerne les descripteurs représentant un corps humain en mouvement, la deuxième contient les descripteurs déterminant la dynamique d'un objet en mouvement, et la troisième contient les descripteurs concernant des points précis de l'objet considéré. Aristidou et Chrysanthou proposent une classification en fonction de la caractéristique du mouvement à considérer, entre le mouvement, l'effort, l'espace et la forme (A. ARISTIDOU et CHRYSANTHOU 2014). Dans tous les cas, l'explicabilité des descripteurs calculés et l'utilisation de ces valeurs par des personnes expertes ou non n'est pas considérée. Ainsi, un expert du mouvement sans connaissances scientifiques poussées et un apprenant supposé novice dans le domaine ciblé ne pourront faire un usage direct de ces indicateurs.

Les travaux se focalisant sur un ensemble précis de descripteurs permettent de décrire les aspects du geste considérés dans l'étude (SENECAL, NIJDAM et THALMANN 2018) (Andreas ARISTIDOU *et al.* 2015) (DELAYE et ANQUETIL 2013). Ces ensembles offrent une analyse précise : par exemple, dans le cadre de l'écriture manuscrite, seront considérés non seulement la position de début et de fin de la main, mais également les différents traits effectués, les inflexions du poignet, le nombre de trait allant de haut en bas, *etc.* Les descripteurs présents dans ces études ont souvent une application limitée dans des contextes autres que le cadre applicatif choisi, bien que certains ensembles aient été pensés afin d'être réutilisable dans d'autres contextes (BOULAHIA *et al.* 2016). Quelque-soit la portée des descripteurs calculés, ils sont proposés soit à des systèmes d'analyse automatique (Andreas ARISTIDOU *et al.* 2015), soit à des experts à-même d'en extraire des informations pertinentes dans le cadre de la réalisation ou l'apprentissage d'un geste (SENECAL, NIJDAM et THALMANN 2018). Ainsi, la facilité d'utilisation de ces descripteurs par un non-expert n'est pas garantie, voire peu probable, comme dans les travaux de Senecal *et al.* par exemple (SENECAL, NIJDAM et THALMANN 2018).

L'utilisation des classifications et ensembles de descripteurs proposés dans des contextes annexes n'est pas assurée. En effet, bien que certains travaux proposent des ensembles pouvant couvrir une large gamme de caractéristiques du mouvement, dans le but de faire de la reconnaissance automatique, l'utilisation de ces descripteurs dans une situation d'apprentissage n'est pas considérée. Il en va de même pour les travaux portant sur l'indexation et la recherche de mouvements proposant des ensembles de descripteurs adaptés à ces deux fonctionnalités. De plus, les travaux portant sur l'indexation et la recherche de mouvements au sein de bases de données proposent des ensembles de descripteurs dont l'utilisabilité par un humain n'est pas considéré (SAKURAI *et al.* 2014) (XIAO *et al.* 2015). En outre, ces ensembles peuvent présenter des redondances sémantiques, soit entre eux, soit au sein même du jeu de descripteurs considéré. Le calcul des descripteurs est une tâche qui peut prendre un temps non-négligeable, en fonction du nombre de données à traiter. La réduction des redondances au sein de ces jeux de descripteurs et la prise en compte des capacités de perception et d'interprétation des acteurs de l'apprentissage sont trois éléments importants à prendre en compte pour la conception d'un EIAH dédié à l'apprentissage de geste. La prochaine section propose une classification d'un ensemble de descripteurs réduit, issus de travaux présentés dans cette section, décrivant les différentes caractéristiques d'un mouvement. Cette nouvelle classification propose des cas d'usages et des domaines d'application, ainsi que les connaissances scientifiques nécessaires (ou non à leur utilisation).

### 3.3 CLASSIFICATION DE DESCRIPTEURS ÉLÉMENTAIRES

Une nouvelle classification répondant à des besoins d'observation et d'analyse variés est proposée dans cette partie. Nous ne présentons ici que des descripteurs dit élémentaires, c'est-à-dire des descripteurs qui ne sont pas issus d'un calcul, d'une réinterprétation ou d'un ensemble de descripteurs « bas-niveau » du mouvement. Par exemple, Aristidou et Chrysanthou (A. ARISTIDOU et CHRYSANTHOU 2014) définissent la « hauteur de la hanche » comme descripteur. Cependant, ce descripteur provient de la composante verticale du descripteur « position 3D » de la hanche : Dans ce cas, le descripteur dit élémentaire est celui de la position selon l'axe vertical.

Chaque descripteur est caractérisé par son appartenance à une ou plusieurs des familles présentes dans le tableau 2. Cette classification s'inspire des classifications déjà établies dans les travaux cités précédemment (LARBOULETTE et GIBET 2015) (MOREL 2017). Les différentes familles proposées ne sont pas exclusives. De plus, la nécessité de posséder des connaissances scientifiques pour pouvoir contextualiser et interpréter le descripteur est indiquée, afin de donner des pistes quant à l'utilisation de ces descripteurs dans des contextes de retour à l'expert ou à l'apprenant.

Les colonnes « non-contigu, contigu et global » permettent de donner la possibilité d'obtenir la ou les valeurs du descripteur respectivement(i) sur des points précis (postures-clés), (ii) sur des postures qui se suivent temporellement ou (iii) sur la totalité du mouvement. Dans le premier et le deuxième cas, un vecteur de valeurs correspondant à un descripteur est retourné, dont la taille correspond soit au nombre de posture-clés, soit au nombre de postures successives choisies, et dans le dernier cas, un seul descripteur caractérisant l'ensemble du mouvement sera obtenu.

Les colonnes « cinématique » et « géométrique » donnent une indication sur la provenance du descripteur : un descripteur cinématique est directement dérivé à partir des propriétés du mouvement, alors qu'un descripteur géométrique proposera une information sur des aspects spatiaux du corps, voire son environnement.

La colonne « utilisation courante » met en évidence des domaines où ce descripteur est couramment utilisé, donnant ainsi des pistes quant à l'utilisation d'un descripteur en fonction du contexte applicatif fourni.

Enfin, la colonne « connaissance scientifique requise » indique si l'analyse de ce descripteur par un humain nécessite de posséder des connaissances scientifiques adéquates, ou si, à l'inverse, il est possible pour une personne sans connaissances particulières de comprendre et de tirer des informations utiles pour l'apprentissage du geste à partir de la valeur de ce descripteur.

Un même descripteur élémentaire présenté dans le tableau ci-dessus peut exprimer différentes caractéristiques d'un mouvement, en fonction de la fenêtre temporelle considérée. Ainsi, la vitesse prise sur un intervalle défini permet de suivre l'évolution de cette dernière lors d'un effort particulier, alors que la vitesse moyenne du mouvement prise globalement (moyenne des vitesses instantanées) donne une indication sur la rapidité du mouvement général. L'analyse de la position à un moment donné indique dans quelle position l'apprenant se trouve, alors que l'analyse de l'évolution de la position sur un temps donné résulte en l'obtention de la trajectoire de l'apprenant.

Le sens d'une partie des descripteurs proposés dans le tableau 3 est explicité ici :

- mouvement saccadé : donne une indication sur la fluidité du mouvement
- courbure : mesure la vitesse de changement de la trajectoire
- quantité de mouvement : moyenne pondérée des vitesses d'un ensemble d'articulation du corps

- forme englobante : plus petite forme (boîte, sphère, ellipsoïde...) englobant la totalité des articulations et des membres du corps
- enveloppe convexe : enveloppe minimale englobant le corps (Voir Fig. 20)
- centre de masse : moyenne pondérée (ou non) des positions des articulations (ou d'un sous ensemble des articulations) du corps
- équilibre : valeur booléenne indiquant si la projection du centre de masse au sol est comprise au sein de la projection au sol de la forme englobante du corps
- profilage : décrit l'évolution de la forme du corps au cours du mouvement : extension/rétractation, élargissement/réduction, enflement/creusement
- quantité d'effort : caractérise l'intensité d'un effort fourni sur un intervalle de temps défini, variant de « faible » à « intense »
- durée d'effort : représente la durée d'un effort fourni, caractérisé en « urgent » ou « continu »

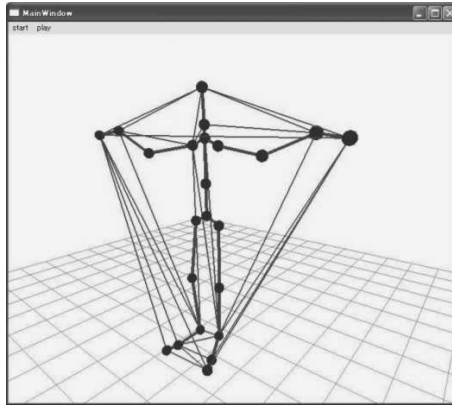


FIGURE 20 – L'enveloppe convexe correspond à la forme qui englobe le corps dans sa totalité, tout en gardant la distance minimale pour chaque segment la composant (Source : (HACHIMURA, TAKASHINA et YOSHIMURA 2005))



Type de descripteur	Description	Exemple
Non contigu	calculé à des moments précis du mouvement (une ou plusieurs valeurs)	Orientations prises à des postures clés
Contigu	calculé sur une suite de postures contigües (possiblement le mouvement entier)	Vitesse au court du temps
Global	une unique valeur qui définit une caractéristique du mouvement sur toute sa durée (une seule valeur)	Périodicité du mouvement
Cinématique	caractérise les aspects du mouvement relatifs au déplacement, la rotation, la vitesse, etc.	Vitesse moyenne du mouvement
Géométrique	caractérise des aspects relatifs aux parties du corps par rapport au corps tout entier ou à l'environnement	Boîte englobante
Utilisation courante	si un descripteur est souvent utilisé dans un domaine applicatif particulier	
Connaissance scientifique requise	indique si le descripteur requiert des connaissances scientifiques (physique, mécanique, géométrie, etc.) ou non pour être compris et utilisé dans un contexte particulier	

TABLE 2 – Différentes familles de descripteurs. Elles proviennent des travaux de Larboulette et Gibet (LARBOULETTE et GIBET 2015) et Morel (MOREL 2017).

Descripteur	Contigu	Non-contigu	Global	Cinématique	Géométrique	Utilisation courante	Connaissance scientifique requise
Position	X	X		X		Polyvalent	Géométrie
Orientation	X	X		X		Polyvalent	Géométrie
Durée			X	X		Polyvalent	Non
Distance effectuée	X		X		X	Polyvalent	Non
Vitesse	X	X	X	X		Polyvalent	Cinématique du mouvement
Accélération	X	X	X	X		Polyvalent	Cinématique du mouvement
Mouvement saccadé	X	X		X		Analyse de la performance	Cinématique du mouvement
Courbure	X	X		X		Reconnaissance d'actions	Géométrie
Quantité de mouvement	X	X	X	X		Analyse de la performance	Dynamique du mouvement
Forme englobante	X	X	X		X	Polyvalent	Géométrie
Enveloppe convexe	X	X	X		X	Indexation et recherche	Géométrie
Centre de masse	X	X	X		X	Analyse de la performance	Dynamique du mouvement
Équilibre	X	X	X		X	Analyse de la performance	Non
Profilage	X	X	X		X	Analyse de la performance	Dynamique du mouvement
Quantité d'effort	X	X	X	X		Analyse de la performance	Dynamique du mouvement
Durée d'effort	X		X			Analyse de la performance	Non

TABLE 3 – Une classification des descripteurs élémentaires de la littérature. Ils proviennent des travaux de Larboulette et Gibet (LARBOULETTE et GIBET 2015) et Morel (MOREL 2017).

En réduisant ainsi les descripteurs à un ensemble élémentaire, il est possible d'exprimer les descripteurs de plus haut-niveau par une combinaison de ces descripteurs. L'utilisation de cet ensemble pourrait ainsi, à terme, mener à la création d'un système de conception et d'intégration de descripteurs de haut niveau, calculables à partir des descripteurs élémentaires proposés et interprétables par l'utilisateur en fonction du contexte et de ses connaissances.

### 3.4 DISCUSSION

Dans ce chapitre, nous avons étudié les descripteurs du mouvement calculables à partir des données capturées en fonction de leurs usages et des objectifs d'analyse. À partir de données de positions et d'orientations décrivant le mouvement de manière précise, il est possible d'extraire des informations caractérisant la forme, la position, la dynamique ou la cinématique du corps humain et du mouvement qui lui est associé. Il existe beaucoup de descripteurs pour caractériser le geste (LARBOULETTE et GIBET 2015). Cependant, ils sont majoritairement issus des descripteurs élémentaires présentés dans le tableau 3, soit d'une combinaison de plusieurs de ces descripteurs. Ainsi, il est en conséquence possible de réduire la large gamme de descripteurs utilisés dans l'analyse du mouvement à l'ensemble élémentaire présenté ici. L'utilisabilité de ces descripteurs par une personne non-experte, ainsi que les domaines applicatifs dans lesquels ils sont le plus souvent utilisés, sont considérés, afin de faciliter le choix des descripteurs en fonction des contextes applicatifs. L'implémentation de cet ensemble élémentaire au sein d'un système dédié à l'analyse de gestes permettrait de couvrir les besoins d'observations et d'analyse des utilisateurs, soit en calculant et en utilisant directement ces descripteurs, soit en combinant certains d'entre eux dans le but d'obtenir de nouveaux descripteurs.

Bien qu'il soit possible d'extraire ces descripteurs à partir de données de mouvement, ils ne sont néanmoins pas tous utiles en fonction du contexte. Ainsi, la pertinence des indicateurs est définie par les besoins d'observation et d'analyse des experts ou apprenants. La formalisation des besoins d'observation et d'analyse en descripteurs calculables est une étape cruciale afin de proposer des indications utiles sur le geste. Cette formalisation peut mener à la proposition d'un ensemble de descripteurs pertinent dans le cadre de l'analyse d'un mouvement. La problématique de formalisation de nouveaux descripteurs dans un système existant est rarement prise en compte dans les travaux existants. Cela peut s'expliquer par le côté *ad-hoc* des systèmes extrayant des descripteurs du mouvement d'une part, et par la non-contextualisation des classifications de descripteurs d'autre part.

Il faut noter que le niveau de connaissance requis pour l'analyse des descripteurs est également lié au contexte. Ainsi, si l'on prend l'exemple d'une course à pied, le calcul et l'analyse conjointe de la durée et de la distance effectuée peuvent être suffisantes pour le pratiquant non spécialiste en biomécanique pour estimer sa performance physique. À l'inverse, l'analyse de la quantité d'effort ainsi que la durée de cet effort pour une partie précise du corps pourra être utilisée par un médecin afin d'évaluer les risques de blessures, mais sera difficile à interpréter pour une personne non-experte. L'analyse de ces données doit se faire dans un contexte précis, par un expert humain ou un système où la connaissance experte a été intégrée, pour être en mesure de donner du sens aux valeurs ainsi obtenues. Le prochain chapitre montre les différentes méthodes et techniques d'analyses possibles à partir de données de mouvement et des descripteurs.

---

## ANALYSE DU MOUVEMENT CAPTURÉ

---

Dans ce chapitre, nous explorons les différentes manières d'analyser le mouvement. En particulier, l'analyse du mouvement humain, dans le contexte d'un apprentissage ou de l'amélioration d'un geste, a principalement deux objectifs non exclusifs : (i) l'apprentissage ou l'amélioration du geste en lui-même, par imitation d'une succession de postures dans le temps ou par reproduction de ses principales caractéristiques (extraites de manière automatique ou non) et (ii), l'apprentissage ou l'amélioration de propriétés du ou des objets manipulés par l'intermédiaire du geste appris (manipulation d'un scalpel lors d'une procédure chirurgicale, par exemple).

L'apprentissage du geste par observation et imitation de démonstrations peut être qualifié de naturelle, simple à mettre en œuvre mais peu efficace sans un grand nombre d'itérations du processus « démonstration-imitation » (Y. YANG et LISBERGER 2014). Le savoir de l'expert permet le plus souvent de juger de la qualité du geste effectué. Ce regard expert sur le geste peut conduire à des propositions d'améliorations, sous la forme d'entraînements relatifs aux défauts du geste effectué. Les observations et l'analyse de l'expert sont cependant limitées par la complexité du geste : un geste qui fait intervenir, rapidement, de nombreuses parties du corps peut être difficile à analyser dans son intégralité, comme pour la danse par exemple (KYAN *et al.* 2015). Il est possible de pallier cette difficulté en analysant le mouvement capturé. En effet, l'utilisation de données de mouvements capturés permet d'avoir une représentation visualisable à volonté, selon différentes vitesses de lecture, sous différents angles de vue (YOSHINAGA et SOGA 2015) et permet également d'utiliser des processus automatiques afin d'aider cette analyse à l'aide de : calculs de descripteurs, détections d'une séquence ordonnée d'actions, l'utilisation de techniques d'apprentissage automatique ou la réduction des données. Les parties suivantes présentent différentes méthodes existantes pour l'analyse des gestes tout en positionnant l'utilisateur et ses contributions (apprenant ou enseignant) au sein de chacune d'elle. Ces catégories ne sont pas exclusives, c'est-à-dire qu'un même système dédié à l'apprentissage de mouvement peut se trouver à l'intersection de deux catégories ou plus.

### 4.1 ANALYSE DU GESTE PAR OBSERVATION HUMAINE

Cette section présente plusieurs études où l'analyse du geste est faite de manière empirique par un humain. Cette analyse peut directement s'effectuer par visualisation du mouvement, ou à l'aide de métriques calculées en fonction des besoins d'observations. Dans les deux cas, les données doivent être analysées par un humain dans un premier temps et les retours sont systématiquement délivrés par un expert à un apprenant.

Le domaine médical est un domaine qui nécessite non seulement de pouvoir étudier la motricité du patient, mais aussi son état physiologique. À l'aide d'un ensemble de capteurs hétérogène (réduit (ALANKUS *et al.* 2010) ou non (VRIES *et al.*

Quantifiable Gait Measures	Gait Disorders
Gait speed	Slow walking
Step length	Parkinson gait, small steps, marche à petits pas (gait with little steps)
Step frequency (cadence)	Slow walking, gait efficiency
Stride-stride variability	Abnormal rhythm of gait
Step width	Cerebellar gait (ataxic gait), wide base, extremely narrow base
Step height	Peripheral neuropathic gait, foot drop, high stepping gait
Transverse plane signal amplitude	Hemiplegic gait, diplegic gait, circumduction, scissor gait
Knee joint angle	Crouch gait, drop foot, equine gait, stiff knee
Ankle joint angle	Equine gait, crouch gait
No. of steps during turning	Difficulty with turning
Hip flexion	Myopathic gait, waddling gait, excessive hip sway, drop of pelvis
Heel-strike amplitude, ground reaction forces	Sensory gait, stomping, stamping
Motion signal distribution	Tremor
Stance time	Antalgic gait, hesitation
Swing time	Difficulty in clearing off at toe off, difficulty in swinging
Double support time	Steadiness
Bilateral sensor comparison	Gait asymmetry
Gait stability measure	Wobbly gait, unstable gait
Gait complexity measure	Choreiform gait, hyperkinetic gait, jerky gait
Gait regularity measure	Reduced gait variability
Moment	Weakness during toe off
Muscle force from EMG	Muscle weakness, abnormal muscle activity

FIGURE 21 – Des exemples de descripteurs extractibles relatifs à la démarche possible et leurs applications pour la détection de pathologies (S. CHEN *et al.* 2016).

2006)), il est possible d’obtenir suffisamment d’informations pour que le médecin soit capable de délivrer un verdict. Bien que de nombreux types de capteurs existent pour l’analyse médicale, nous nous concentrons ici sur les capteurs permettant d’obtenir des données relatives aux mouvements, et non des indications sur l’état physiologique du patient (capteur cardiaque, par exemple).

L’analyse de la démarche fait un usage intensif des données de mouvements capturés (S. CHEN *et al.* 2016). Cette analyse permet d’évaluer le résultat de nombreuses procédures médicales chez un patient : pose de prothèse, chirurgie, rééducation, mais également le suivi de l’évolution de la posture, ainsi que les risques de chute, notamment chez les personnes âgées. Ainsi, il est crucial d’obtenir des données correspondant aux besoins d’observation et d’analyse des médecins (Fig. 21). De tels descripteurs peuvent être extraits à partir d’un squelette 3D.

La chirurgie gagne beaucoup à utiliser les casques de réalité virtuelle afin de simuler des procédures dans un environnement clinique d’apprentissage. En effet, il n’est pas envisageable de s’entraîner sur de vrais patients (KNEEBONE *et al.* 2004). Ainsi, l’utilisation d’environnements virtuels, combinés par exemple à des processus de gamification (PULIJALA, MA et AYOUB 2017), permet d’entraîner plus de chirurgiens en parallèle, ainsi que de permettre une vérification des acquis sans risque. Dans les travaux de Choi *et al.* (CHOI *et al.* 2015), l’utilisation conjointe de retours visuels en temps réel, ainsi que d’indicateurs calculés par rapport à la tâche à effectuer (temps de la procédure, force appliquée, etc.) permet de proposer une visualisation de la performance.

Le sport est un domaine où l’apprentissage du geste est au cœur de la formation. Il est nécessaire de pouvoir évaluer à la fois le résultat du geste, mais également la manière dont il est fait, afin d’éviter des risques de blessure (RAWASHDEH, RAFELDT et UHL 2016). L’utilisation d’experts virtuels est une des possibilités offertes par la



FIGURE 22 – L’environnement virtuel pour le Karaté proposé par Burns *et al.* est utilisé pour l’entraînement de l’apprenant. L’évaluation est faite de manière qualitative par un expert *a posteriori* (BURNS, ANNE-MARIE *et al.* 2011).

capture de mouvements. Bien que ces systèmes permettent parfois de s’affranchir de la présence d’un expert, ils ne présentent pas systématiquement une amélioration de l’apprentissage par rapport aux méthodes d’enseignement traditionnelles (PHILO TAN CHUA *et al.* 2003).

Pour l’apprentissage du karaté, le système développé par Burns *et al.* propose un environnement en réalité virtuelle (BURNS, ANNE-MARIE *et al.* 2011). L’apprenant doit, au fur et à mesure des séances, reproduire à l’identique les gestes de l’expert. Les sessions sont divisées en deux parties : une partie sur l’explication du geste à effectuer, fournie à l’aide d’une vidéo de l’expert effectuant le mouvement et une autre sur des exercices relatifs au geste. L’expert y est représenté sous forme 3D dans un environnement virtuel correspondant à un dojo (Fig. 22). L’évaluation est réalisée par l’expert, une fois au début de l’apprentissage et une fois à la fin, en analysant les vidéos des gestes des apprenants. Ces enregistrements sont réalisés hors du contexte virtuel. L’expert évalue les gestes des apprenants en attribuant un score, en fonction du degré de similitude du geste de l’apprenant par rapport à celui de l’expert. Les résultats montrent que l’avatar virtuel est aussi efficace qu’un expert réel pour l’apprentissage de ces gestes.

Le système développé par Le Naour *et al.*, présenté dans le chapitre 2, permet d’aider à l’apprentissage de mouvement en superposant le modèle de l’expert à celui de l’apprenant au sein d’un environnement virtuel (LE NAOUR, HAMON et BRESCIANI 2019). Le cas de test utilisé était celui du lancer au football américain. À partir d’une démonstration effectuée par l’expert, l’apprenant doit ensuite effectuer une suite de plusieurs lancers, à partir desquels sont calculés la valeur de l’écart-type du DTW entre les mouvements de l’apprenant et de l’expert, ainsi que la distance du lancer par rapport au centre de la cible visée. Le groupe ayant reçu la visualisation du mouvement de l’apprenant superposé à celui de l’expert comme retour a montré une amélioration plus importante du geste que les groupes ne l’ayant pas obtenu.

À l’inverse, Kora *et al.* ont proposé un environnement en réalité augmentée permettant l’apprentissage du golf en autonomie (KORA, SOGA et TAKI 2015). La projection du modèle de l’expert dans un casque porté par l’apprenant, superposé au



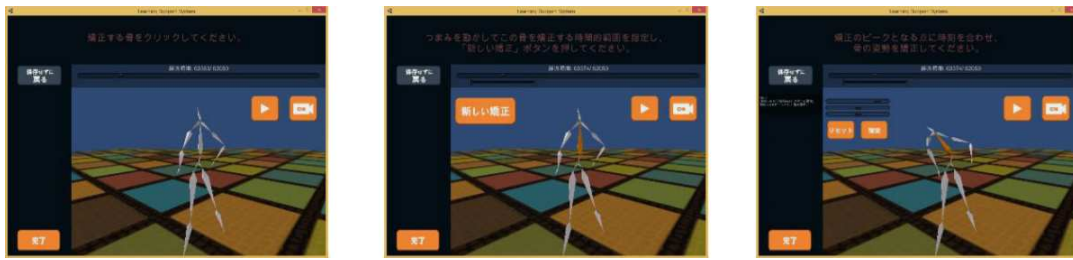


FIGURE 23 – L’environnement proposé par Yoshinaga *et al.* permet à l’expert de modifier directement les positions et les orientations des membres du corps de l’apprenant sur l’enregistrement du geste original, afin de lui proposer des corrections à distance (YOSHINAGA et SOGA 2015).

squelette (capté à l’aide d’une Kinect) de la personne faisant le geste, permet d’aider l’apprenant à corriger son geste au fur et à mesure de l’entraînement. L’apprenant est en auto-évaluation : aucune indication ne lui est fournie, c’est donc à lui d’être capable d’évaluer seul les changements à apporter à son geste afin de parfaire le mouvement du swing.

Des systèmes se basent sur la superposition des données de l’apprenant avec celles des experts afin de proposer une correction par mimétisme (YOSHINAGA et SOGA 2015). Dans ces cas, le mouvement de l’apprenant est rejoué par dessus celui de l’expert, et les corrections peuvent s’effectuer par rapport au décalage temporel ou spatial de l’apprenant par rapport à l’expert. Le système proposé par Yoshinaga et Soga, testé sur l’archerie japonaise, permet d’agréger plusieurs modèles experts afin de proposer soit un expert au choix, soit l’expert le plus proche de l’étudiant morphologiquement parlant, soit une moyenne des experts ou soit l’expert médian (c’est-à-dire l’expert le plus proche de la moyenne des experts) (YOSHINAGA et SOGA 2015). Cette approche permet de limiter les différences morphologiques, dans le cas où un des modèles experts est semblable à celui de l’apprenant. La comparaison du geste de l’apprenant à celui de l’expert est réalisée à l’aide d’une correspondance en programmation dynamique : ainsi, le mouvement est représenté de façon amorphologique, et il est possible de faire correspondre les mouvements temporellement, même en présence de variations de timing, vitesse et d’amplitude des mouvements. Il est possible d’utiliser ce système soit en autonomie, soit à l’aide d’un expert qui analysera le geste de l’apprenant. Dans ce cas, l’expert peut être en mesure d’expliquer les différences entre le geste de l’apprenant et celui des modèles d’experts, ajoutant ainsi son expertise au retour visuel (Fig. 23).

#### 4.2 ANALYSE PAR OBSERVATION D’INDICATEURS CALCULABLES DU GESTE

L’explication des caractéristiques d’un geste, ainsi que son évaluation, peut varier en terme de précision, en fonction du contexte applicatif mais également de la possibilité d’explicitier et de formaliser les caractéristiques importantes en descripteurs calculables. Dans le cas où les caractéristiques du geste à réaliser, ainsi que les propriétés biomécaniques associées, sont connues à l’avance, il est possible de proposer des valeurs précises pour les différentes caractéristiques des différents mouvements à effectuer pour chaque partie du corps concernée, et ainsi permettre de fournir, par exemple, une amplitude acceptable, un intervalle de déplacement correct, une rotation cible, etc. Dans ces cas, l’analyse automatique du geste est possible à condition que ces quantités puissent être mesurées ou calculées à partir des données capturées. Deux grandes stratégies non-exclusives d’analyse automa-

tique peuvent s'appliquer : l'analyse automatique des caractéristiques du geste en lui-même, et l'analyse automatique de la position d'un objet au cours du temps, permettant ainsi de donner des pistes quant aux conseils à donner à l'apprenant.

Dans ce dernier cas, les outils utilisés pour simuler les objets réels peuvent être multiples. L'utilisation de dispositifs haptiques est courante dans le domaine de la chirurgie. La simulation d'outils (trocarts, scalpels, sondes, etc.) intégrés dans des scènes virtuelles réalistes permet de s'entraîner à pratiquer des opérations chirurgicales dans des conditions proches d'une situation *in vivo*.

Pham *et al.* s'intéressent à la position d'un forceps au cours d'une simulation d'un accouchement, et plus particulièrement la trajectoire de la position et celle de l'orientation. Ils comparent celles de l'objet manipulé par l'apprenant par rapport à celles de l'objet manipulé par l'expert (PHAM, MOREAU et BOULANGER 2010). La corrélation entre les signaux correspondants aux trajectoires de l'expert et de l'apprenant est calculée, et sa valeur est utilisée en tant que score, qui correspond à la similitude entre ces trajectoires. Despinoy *et al.* proposent une approche qui analyse également la position et l'orientation d'un porte-aiguilles au cours d'une opération de réalisation de points de suture (DESPINOY *et al.* 2016). Le mouvement est segmenté en plusieurs parties grâce aux points d'inflexion de la trajectoire de l'outil considéré. Ces points correspondent aux moments où le geste subit un changement brusque, soit dans sa linéarité, soit dans sa direction.

Dans le cadre du suivi des patients atteints de la maladie de Parkinson, le système développé par Wang *et al.* permet, à l'aide d'un seul capteur et d'une suite de mouvements à réaliser par le patient, de proposer aux médecins des données permettant d'évaluer la progression de la maladie (R. WANG *et al.* 2013). L'expérimentation proposée cherche à observer et quantifier la dégradation de la posture et des mouvements due à la maladie de Parkinson, à l'aide de la distance faite à chaque pas, le balancement de la posture, le degré de balancement des bras et la constance de l'intervalle de temps entre chaque pas. Ces données étaient relevées au cours de deux mouvements à réaliser par les patients : (i) marcher et (ii) s'asseoir et se lever. La succession de mouvements à réaliser permet aux médecins de vérifier l'état d'avancement de la maladie, et le système permet de faire un suivi à distance, et ainsi d'éviter les déplacements des patients. L'objectif du système n'est cependant pas de proposer une évaluation du geste, mais des métriques qui peuvent être utilisées par un médecin pour évaluer la progression de la maladie.

Le geste a une place prédominante dans l'apprentissage du sport. La technicité, ainsi que la précision des gestes à réaliser en font un domaine d'étude idéal pour le développement de systèmes d'analyse automatique. Au sein de ce domaine, deux cas d'étude principaux se dégagent : l'analyse du geste de l'apprenant en fonction de critères définis au préalable, ou la découverte de la différence entre les gestes experts et ceux des débutants.

Une première approche consiste à effectuer une analyse basée sur les positions relatives des parties du corps pour proposer une évaluation du geste. Yamaoka *et al.* ont proposé un système qui évalue plusieurs modalités du lancer de disque, à différents moments du lancer (pré-lancer, lancer, post-lancer) (Fig. 24) (YAMAOKA *et al.* 2013b). Le mouvement est segmenté selon ces catégories en observant et comparant les positions relatives de plusieurs articulations. Ainsi, pour un droitier, le geste est considéré comme faisant partie du pré-lancer tant que la position de la main est située à la droite de celle du centre des épaules par rapport à l'axe horizontal. La période de lancer est celle où la position de la main, toujours selon le même axe, est située entre le centre des épaules et le coude droit. Le post-lancer commence à partir



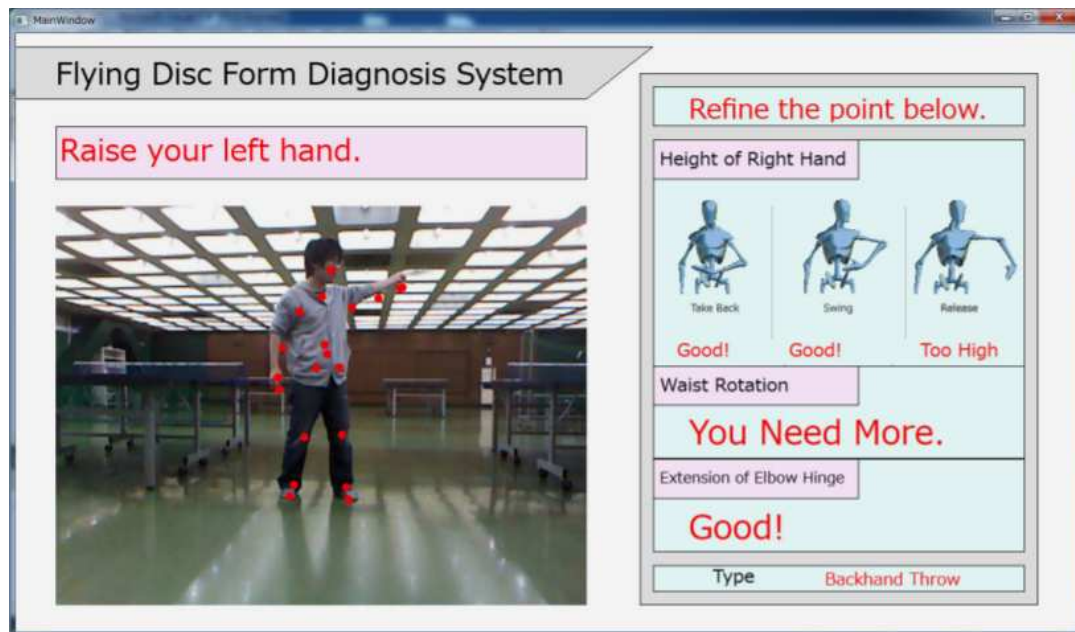


FIGURE 24 – Le système développé par Yamaoka *et al.* pour l'apprentissage du lancer de disque. Les conseils sont donnés sur plusieurs caractéristiques du geste en même temps (YAMAOKA *et al.* 2013b).

du moment où la main droite est à gauche du coude droit. L'évaluation du geste se fait selon 5 critères :

- la présence d'une phase d'élan avant le geste
- l'élévation de la main par rapport à l'épaule
- le changement de hauteur de la main au cours du lancer (afin que le disque reste parallèle au sol durant le vol)
- l'angle formé par le triplet d'articulations épaule/coude/poignet
- la rotation des hanches lors du lancer

En vérifiant que chacune de ces modalités respecte un intervalle ou ne dépasse pas un seuil spatial, le système est en mesure de donner un retour à l'apprenant. Ce retour est donné en temps réel, et il n'est pas possible de rejouer le mouvement ni de visualiser les conseils antérieurement : l'impact sur l'apprentissage s'en trouve diminué.

Comme présenté précédemment, les travaux de Morel proposent l'extraction de descripteurs du mouvement dans le cadre de l'analyse de la performance sportive (MOREL 2017). Après un recalage spatial (à l'aide d'une transformation linéaire) et temporel (à l'aide de l'algorithme de la DTW), les différences spatiales et temporelles entre le geste de l'apprenant et celui de l'expert sont données en calculant d'une part l'alignement spatial d'un membre par rapport au membre de référence, et d'autre part la différence entre les chemins de déformations induits par les alignements des membres de l'apprenant et de l'expert. Ces erreurs sont ensuite traduites en indications pour la position des membres au cours du mouvement.

Le système de Makio *et al.* propose une autre stratégie en se basant sur la trajectoire et la vitesse du poignet lors d'un service au tennis, afin de mettre en lumière les différences entre les gestes des experts et ceux des débutants (MAKIO, TANAKA et UEHARA 2007). L'analyse se base sur une extraction des règles d'associations entre les différentes parties du corps, à partir desquelles sont extraits des motifs fréquents. La distance entre ces motifs est calculée à l'aide d'une distance euclidienne. Bien

que cette analyse permette de mettre en évidence les différences entre les gestes de l'expert et ceux de l'apprenant, et propose également une visualisation des trajectoires de l'expert et de l'apprenant, aucun conseil n'est ensuite donné par le système.

Dans le domaine de la danse, le système proposé par Maes *et al.* permet d'apprendre le geste en trois étapes : la démonstration, l'apprentissage et l'évaluation ludifiée (MAES, AMELYNCK et LEMAN 2012). L'étape de démonstration permet à l'expert de s'enregistrer, tout en configurant plusieurs options telles que la musique, le tempo, le nombre de pas par figure de danse, le nombre d'entraînements requis, *etc.* L'apprentissage se base sur la reproduction des pas effectués par le modèle de l'expert, autant sur l'aspect spatial que temporel, accompagné par les schémas des différents déplacements et rotations à effectuer. Dès cette étape, la corrélation entre le modèle de l'expert et la performance de l'apprenant est calculée, afin de fournir un premier retour sur la performance de ce dernier, à l'aide du coefficient de corrélation croisée de Pearson, calculé sur les données de mouvement de l'expert. Enfin, la partie d'évaluation propose à l'apprenant de danser et effectue une comparaison, toujours selon le même critère, des différents mouvements constituant la danse, permettant ainsi l'évaluation sous la forme d'un score à maximiser. Cette méthode passe cependant outre les spécificités de chaque partie du corps et ne propose qu'une évaluation globale du mouvement, ce qui résulte en une perte d'informations, notamment au niveau des défauts ponctuels du geste.

Toujours dans le domaine de la danse, Kyan *et al.* ont développé un système utilisant la réalité virtuelle afin de faciliter l'évaluation de la performance de l'apprenant (KYAN *et al.* 2015). En utilisant un unique descripteur, à l'aide d'une projection dans un espace sphérique des mouvements de danse, le mouvement est décomposé en plusieurs parties. Chacune de ces parties est ensuite comparée à une base de données de composants gestuels d'expert réalisés au préalable. Cette décomposition permet d'estimer quel mouvement était en train d'être réalisé par l'apprenant. Une fois que le mouvement est reconnu, une comparaison entre des descripteurs angulaires extraits permet de donner un score pour chaque partie du mouvement de danse. Ce score est calculé en fonction de la distance euclidienne entre les signaux de l'apprenant et de l'expert, après alignement temporel à l'aide de l'algorithme du DTW. Une autre partie de ce système permet également de superposer le modèle de l'apprenant à celui de l'expert, afin de corriger précisément la position de chaque partie du corps concernée. L'évaluation conduite montre que le mouvement est correctement segmenté par le système ; cependant, la visualisation des mouvements ne permet qu'une faible amélioration de la performance de l'apprenant. De plus, aucun alignement morphologique, spatial ni temporel n'est effectué.

Toujours dans le domaine des arts, l'analyse des efforts réalisés par les joueurs de violon est le sujet d'étude de Rasamimanana et Bevilacqua (RASAMIMANANA et BEVILACQUA 2008). À partir d'informations cinématiques provenant du geste de la personne, telles que : la vitesse, le tremblement (*jerk*) et l'impulsion (du mouvement), il est possible de déterminer quel a été le mode de jeu utilisé par la personne réalisant le geste : *détaché* ou *martelé*.

Les systèmes calculant des indicateurs à partir du geste de l'apprenant permettent, dans une certaine mesure, de s'affranchir de la présence de l'expert pour l'évaluation et la restitution des conseils à l'apprenant. Leur objectif est souvent de proposer une évaluation de manière automatique, sans nécessiter l'intervention d'un expert pour interpréter les données. La connaissance experte doit être intégrée *a priori*, dès la phase de conception. Dans ce contexte, la généricité des descripteurs extraits n'est pas assurée si elle n'est pas considérée dès la conception du système, et le



FIGURE 25 – Le système développé par Kyan *et al.* fait usage de la réalité virtuelle dans un système CAVE afin de proposer soit une observation côte à côte des mouvements de l'expert et de l'apprenant, soit superposée (cas de l'image) (KYAN *et al.* 2015).

système n'est souvent pas réutilisable dans des contextes différents. De plus, les retours données prennent le plus souvent la forme d'un score à maximiser (indiquant généralement la proximité des données de l'apprenant à celles de l'expert), qui, bien que facilement interprétable en terme de qualité du geste effectué, ne permet pas systématiquement de fournir des indications quant aux améliorations à apporter au geste.

#### 4.3 ANALYSE PAR DÉTECTION D'UNE SÉQUENCE ORDONNÉE D'ACTIONS

Dans les systèmes d'apprentissage du geste, l'analyse peut se focaliser sur la suite d'actions à réaliser afin d'atteindre l'objectif. Dans ce cas, le geste est décomposé en unités considérées comme élémentaires par le système afin de représenter chaque action. La granularité des unités élémentaires est dépendante du système considéré. Ainsi, il est possible de comparer cette suite de mouvements de manière ordonnée à l'enchaînement attendu par l'expert.

Mahdi *et al.* ont développé un environnement virtuel permettant à n'importe-quel enseignant de définir lui-même ses scénarios d'apprentissage à partir de scènes virtuelles pré-existantes (MAHDI *et al.* 2019). L'objectif est de permettre de lier la description du scénario pédagogique aux activités de l'apprenant au sein de l'environnement virtuel. Chaque activité peut être divisée en une séquence d'actions, qui elles-mêmes peuvent être divisées en séquences de comportements élémentaires, appelées " Primitives Virtuelles de Comportement " : (i) observation du monde virtuel, (ii) déplacement dans le monde virtuel, (iii) interaction avec le monde virtuel et (iv) communication avec d'autres personnes ou avec l'application. Les activités sont réalisées par l'apprenant à l'aide d'objets d'intérêt auxquels ont été associés

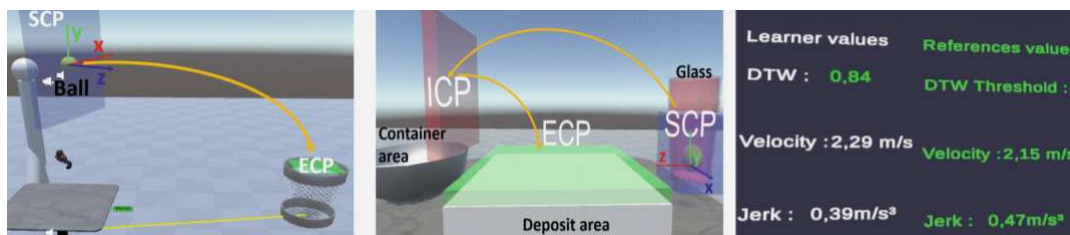


FIGURE 26 – Le système proposé par Delest *et al.* permet à l'expert de générer un parcours à suivre, soit pour la personne, soit pour un objet, dans un ordre prédéfini, et propose l'affichage de plusieurs métriques permettant de comparer le geste de l'apprenant à celui de l'expert (DJADJA, HAMON et GEORGE 2019).

des propriétés techniques (position, forme, couleur, *etc.*) ou pédagogiques (utiliser l'objet, verser, le bouger, *etc.*). L'éditeur de scénario permet ainsi à l'expert de définir une suite d'activités lors de la conception du scénario. Les premiers tests à l'aide d'enseignants ont montrés que l'éditeur leur était utile, et permettait de générer des scénarios de manière aisée.

Les travaux de Delest *et al.* ont mené à la création d'un système où l'expert peut lui-même proposer un parcours, personnalisable à l'aide de points de contrôle (*check-points*), représentés sous la forme d'objets 3D simples (e.g. sphère, parallélépipède rectangle) à passer dans un ordre précis (DJADJA, HAMON et GEORGE 2019) (Fig. 26). Dans ce système, l'expert doit déterminer à l'avance si le parcours concerne le corps de l'apprenant (analyse de la démarche, par exemple) ou un objet présent au sein de la scène. Une fois l'objet d'intérêt choisi, l'expert place ensuite un point de contrôle de départ, possiblement des points de contrôles intermédiaires puis un point de contrôle d'arrivée. Il est possible de modifier la taille et la forme de ces points de contrôles, permettant ainsi de les adapter aux différents objets manipulables (Fig. 26 au milieu). L'expert doit ensuite réaliser le mouvement lui-même, afin de disposer d'un mouvement de référence (mouvement cible). Le mouvement est considéré comme validé lorsque l'objet ou la personne a franchi (c'est-à-dire qu'il y a collision entre le point de contrôle et l'objet ou la personne) tous les points de contrôle dans l'ordre. Le mouvement est ensuite segmenté en fonction des points de contrôle. Chaque mouvement est ensuite comparé au mouvement de référence réalisé par l'expert, selon plusieurs métriques : de manière globale à l'aide de l'algorithme du DTW, mais également à l'aide d'indicateurs cinématiques, tels que la vitesse, la saccade du mouvement (*jerk*), *etc.* (Fig. 26 à droite).

Dans ces systèmes, l'évaluation du geste est principalement centrée sur l'ordre dans lequel est réalisé un ensemble d'unités élémentaires, plutôt que sur l'évaluation en elle-même du geste sous-jacent à ces actions ou unités. Ainsi, les caractéristiques du geste ne sont pas systématiquement analysées (excepté pour (DJADJA, HAMON et GEORGE 2019)). Lors de la manipulation d'objets, c'est l'état de l'objet à des moments clés du mouvement qui va être considéré. Dans ce cas, on considère que le geste qui a mené à l'état cible est correct, sans l'analyser en lui-même.

Ces systèmes fonctionnent par décomposition et analyse des unités élémentaires composant le geste initial. La nature de ces unités élémentaires (durée, granularité de la décomposition, *etc.*) peut varier d'un système à l'autre. La finalité de cette décomposition est de comparer l'enchaînement d'unités élémentaires à l'enchaînement attendu par l'expert. Cela implique de disposer d'un modèle de connaissance du geste à effectuer, et son intégration se fait souvent dès la conception du système, ne permettant pas d'assurer sa généralité. Cependant, des systèmes tels que ceux développés par Delest *et al.* (DJADJA, HAMON et GEORGE 2019) et Mahdi *et al.* (MAHDI

*et al.* 2019), conçus afin être génériques au regard du geste à accomplir, permettent une ré-intégration de la connaissance experte aisée à l'aide d'un éditeur de scénarios pédagogiques intégrés au système. L'apprentissage se focalise néanmoins sur la suite d'action à réaliser, leur enchainement et la manipulation d'un objet, mais pas sur les gestes qui permettent d'arriver aux états ou à la succession de postures désirés.

#### 4.4 ANALYSE FONDÉE SUR DES TECHNIQUES D'APPRENTISSAGE AUTOMATIQUE

Il est également possible d'utiliser des algorithmes d'apprentissage automatique afin d'analyser automatiquement le geste. Dans ce cas, les descripteurs extraits sont le plus souvent choisis en amont du processus d'analyse, puis utilisés dans un algorithme d'apprentissage automatique afin d'observer la similitude entre le mouvement et ceux présents au sein d'une base de données.

Par exemple, les travaux de Pirsiavash *et al.* proposent une approche sur l'évaluation de geste sans *a priori* sur le mouvement en lui-même (PIRSIAVASH, VONDRICK et TORRALBA 2014). Les données sont extraites à partir de vidéos, et sont constituées de descripteurs variés, tels que des gradients de pixels, des vitesses et des trajectoires, ainsi que des postures successives réalisées par les sportifs. Cette étude se base sur les performances d'athlètes olympiques dans deux sports différents, la patinage artistique et le plongeon. Les données extraites sont ensuite associées à un score donné par des experts ; dans ce cas, il s'agit de scores donnés par les juges olympiques. L'étude montre que les juges donnent la même note dans 96% des cas, ce qui suggère qu'il existe un socle de règles communes permettant l'évaluation. Les couples mouvements / score sont ensuite utilisés en entrée d'un SVM. L'algorithme a permis d'extraire les moments les plus déterminants dans l'attribution du score par les juges. Les scores donnés par l'algorithme sont comparés à ceux donnés par l'expert, à l'aide de la corrélation de Spearman. Les valeurs de ce coefficient montre que les scores donnés sont encore assez éloignés de ceux des experts. L'analyse permet néanmoins de mettre en évidence une évaluation de la qualité du geste supérieure à celle d'un non-expert du domaine.

Patrona *et al.* utilisent des algorithmes de logique floue (*fuzzy logic*) afin de proposer une segmentation, une reconnaissance et une évaluation de gestes (PATRONA *et al.* 2018). Il n'y a pas d'*a priori* sur les gestes pouvant être traités par le système. Dans cette études, les gestes considérés sont variés, tous pris au sein de bases de données : *swing* au golf, taper dans ses mains, s'asseoir, lancer, service au tennis, *etc.* Des descripteurs cinématiques sont extraits du mouvement, puis utilisés par des classifieurs binaires, chacun entraîné à reconnaître un mouvement spécifique. Une fois le mouvement reconnu par le système, la phase d'évaluation s'effectue entre le mouvement de l'apprenant et les mouvements cibles contenus dans la base de données de mouvements. Plusieurs étapes sont ensuite réalisées (Fig. 27) : alignement temporel des mouvements (apprenant et cible) à l'aide d'un algorithme de DTW multivarié, normalisation de la longueur des membres afin d'éviter les différences morphologiques, alignement spatial des mouvements à l'aide de la différence de rotation du corps calculée à partir de la position des épaules et du torse, puis utilisation des différences de positions et de vitesses entre les deux mouvements afin de proposer un retour à l'apprenant. Ces retours sont donnés sous la forme d'un score de similarité entre les mouvements (calculé à l'aide de l'algorithme du DTW), mais également les parties du corps et les articulations les plus éloignées de celles de l'expert, au regard des descripteurs utilisés (positions et vitesses). Un exemple de retour donné par le système est le suivant :



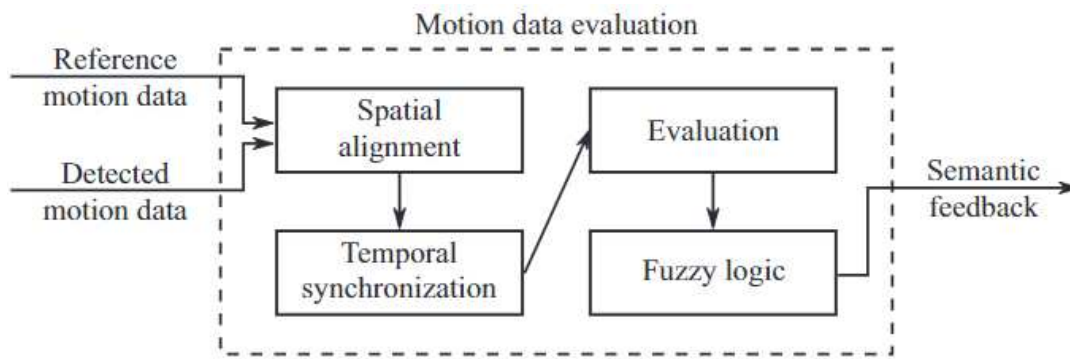


FIGURE 27 – La suite d’étapes utilisée par Patrona *et al.* permet de corriger les différences spatiales et temporelles entre les mouvements de l’expert et de l’apprenant, afin de proposer une évaluation du geste en fonction des descripteurs extraits (PATRONA *et al.* 2018).

« The highest *POSITION* error is detected at **Left Wrist**, at the **Latest** temporal phase (frame 26) of the movement. Please, position your **Left Wrist Left and Down**, at this time instance. »

Ainsi, le retour donné ne considère que des localités temporelles et spatiale, sans tenir compte du geste dans sa globalité. Il peut être difficile pour un apprenant, à partir de ces seules indications, d’être en mesure de modifier les parties du geste suggérées par le système, afin de respecter les conseils donnés par ce dernier, tout en gardant un geste acceptable dans sa globalité.

Dans ces travaux, la contribution de l’expert se situe en amont, lors de la formalisation de ses besoins d’observations en descripteurs calculables à partir du mouvement, voire des conseils à donner. Cependant, les données obtenues en sortie de ces systèmes doivent parfois être analysées et traduites en retour intelligible à l’apprenant par un expert. Dans tous les cas, l’expert n’intervient pas pendant le processus d’apprentissage humain. Lorsque le retour est donné par le système, les informations fournies concernent les propriétés extraites du geste : différence de positions, vitesses, *etc.* Ces informations sont difficilement observables pour un apprenant sans visualisation externe, mais également dans le cas où l’apprenant ne possède pas les connaissances scientifiques requises pour être en mesure de modifier son geste pour prendre en compte les retours du système.

#### 4.5 BILAN DES SYSTÈMES EXISTANTS ET POSITIONNEMENT

Cette section présente les modalités d’usage des différents systèmes présentés auparavant, tout en mettant en lumière leurs avantages et inconvénients par rapport à la place donnée à l’apprenant, ainsi qu’à l’enseignant, en leur sein.

Les travaux où l’analyse des gestes est réalisée par un humain sont ceux où les utilisateurs finaux (experts dans le cas d’un système d’assistance à l’analyse et l’évaluation du geste, ou apprenants) ont une place importante. Les systèmes dédiées à l’analyse médicale (rééducation ou suivi de maladie) nécessitent très souvent de combiner les retours proposés par le système (squelette 3D de démarche, descripteurs spécifiques extraits des parties du corps lors d’un mouvement, *etc.*) au savoir de l’expert, afin de proposer un diagnostic adéquat. Dans ces cas, le système n’a pas pour objectif d’être utilisé par le patient en autonomie, car les informations fournies par le système nécessitent de disposer de la connaissance scientifique permettant d’en tirer des conclusions, et sont ainsi destinées à un expert. L’apprentissage de procédures chirurgicales au travers d’EIAH dédiés permet de

proposer des entraînements sur des simulations d'opérations réelles. Dans ce cas, la connaissance experte est parfois directement intégrée au système *a priori* (ordre des différents mouvement à réaliser, état successifs de l'objet à atteindre, temps total de l'opération, *etc.*), ce qui permet de calculer des indicateurs précis sur le geste, correspondants aux besoins d'observation et d'analyse de l'expert. L'apprentissage du sport nécessite également très souvent un expert afin de proposer des retours compréhensibles sur le geste et exploitables par un apprenant afin de modifier son geste. Le système proposé par Burns *et al.* ne propose aucun retour en lui-même, et ne sert que de support à la visualisation des mouvements de l'expert (soit sous forme de vidéo, soit sous la forme d'un squelette 3D représentant l'expert en train d'effectuer le mouvement, dans un contexte de réalité virtuelle en immersion dans un dojo virtuel). L'évaluation du geste de l'apprenant s'effectue à la fin de l'apprentissage d'un geste par l'expert, en visualisant la vidéo de l'apprenant en train d'effectuer le mouvement. L'environnement virtuel de Kora *et al.*, à l'inverse des systèmes vus précédemment, ne fait intervenir l'expert que dans la phase d'enregistrement du geste cible. Le système est utilisé en autonomie totale par l'apprenant, et l'objectif est d'ajuster le geste de façon à se rapprocher le plus possible visuellement du geste cible de l'expert. Le système ne fournissant aucun conseil ni aucune métrique (indicateurs, score), l'ajustement du geste peut être difficile pour un apprenant non-expert. Enfin, le système de Yoshinaga et Soga utilise plusieurs données expertes différentes afin de proposer une comparaison du geste de l'apprenant à celui des experts les plus proches de l'apprenant au sens morphologique (YOSHINAGA et SOGA 2015). Dans celui-ci, les experts enregistrent non seulement les mouvements cibles, mais peuvent également utiliser le système afin d'analyser le geste de l'apprenant et pour proposer un retour visuel des changements à effectuer sur le geste, en fonction des parties du corps concernées. Le système peut également être utilisé en autonomie par l'apprenant, qui ne dispose alors que de la visualisation de la superposition du mouvement cible choisi avec le sien.

Dans les systèmes où l'analyse est réalisée par un humain, l'implication de l'expert dans la conception dépend souvent de la nécessité ou non de disposer de matériels ou environnements spécifiques à la tâche considérée. Les systèmes proposant de calculer des métriques, à partir du mouvement de l'apprenant, interprétables par les experts nécessitent d'impliquer ce dernier dans le processus de conception afin d'être en mesure de satisfaire ses besoins d'observation. Les valeurs ainsi calculées à partir du geste de l'apprenant ne sont, en général, pas réutilisables dans d'autres contextes applicatifs. De plus, dans les systèmes dédiés à l'apprentissage de procédures chirurgicales, la réutilisabilité est compromise à cause de l'utilisation de matériels spécifiques (bras haptiques, par exemple). À l'inverse, les systèmes ne proposant qu'une visualisation du mouvement (soit de l'expert pendant l'apprentissage de l'apprenant, soit de l'apprenant lors de la phase d'évaluation de l'expert) sont en général très génériques et ne nécessitent ainsi pas d'intégrer l'expert dans le processus de conception, mais ne proposent en général aucune analyse sur le mouvement de l'apprenant, et n'assistent donc l'expert dans sa tâche d'évaluation que de manière marginale.

L'extraction de descripteurs à partir du geste de l'apprenant permet de fournir un ensemble de valeurs afin de l'assister dans son évaluation. Ainsi, le système de suivi de patients atteints de la maladie de Parkinson proposé par Wang *et al.* nécessite de faire réaliser au patient quelques mouvements types (marcher, s'asseoir, se lever), permettant ensuite de calculer des descripteurs relatifs au déplacement du patient. Les experts médicaux peuvent ensuite utiliser ces indications afin de juger de la progression du système. Il a également été pensé de manière à permettre

au patient de réaliser les gestes requis à distance, afin de limiter le déplacement de ce dernier. La place du patient au sein du système se limite à effectuer les gestes demandés par l'expert ou le système. Le système de simulation d'insertion de sonde naso-gastrique développé par Choi *et al.* permet de calculer la force appliquée par l'apprenant, mais également de simuler les butées contre les parois d'un modèle 3D mis à jour en temps réel (CHOI *et al.* 2015). Les indicateurs extraits sont relatifs à la procédure (durée de l'opération, nombre de butées, force maximale appliquée, *etc.*), et bien que leur restitution soit proposée par le système, leur analyse nécessite un expert afin d'être capable d'en tirer des conclusions sur le déroulement de l'opération, et sur les conseils à donner afin d'améliorer le geste de l'apprenant. De par la spécificité du matériel utilisé et des indicateurs calculés, le système n'est pas réutilisable dans d'autres contextes. Les outils d'analyse automatique permettent d'évaluer le geste suivant des descripteurs géométriques, cinématiques, dynamiques et de les comparer. Cependant, des connaissances scientifiques solides sont nécessaires, tant pour l'enseignant que pour l'apprenant afin de pouvoir les interpréter. Les indicateurs utilisés peuvent parfois concerner l'objet manipulé, plutôt que le mouvement en lui-même. Dans le cas de l'EIAH développé par Pham *et al.*, la trajectoire de l'outil de l'apprenant est comparée à celle de l'outil manipulé par l'expert, et cette comparaison fournit un score de similitude entre les gestes (PHAM, MOREAU et BOULANGER 2010). Bien qu'étant facile à interpréter pour un non-expert (la maximisation de ce score étant l'objectif), aucune indication n'est fournie sur les parties du geste à modifier afin de se rapprocher des positions de l'objet ciblées. Le même principe est utilisé par Maes *et al.* dans leur système dédié à l'apprentissage de la danse, en segmentant cette fois le mouvement en plusieurs parties et en assignant un score (correspondant à la similitude des mouvements de l'apprenant par rapport à ceux de l'expert) à maximiser pour chaque partie du mouvement (MAES, AMELYNCK et LEMAN 2012). Il n'y a toujours aucun retour proposé à l'apprenant afin d'améliorer son geste. Le système d'apprentissage de la danse proposé par Kyan *et al.* utilise les mêmes concepts, mais propose en plus de cela une superposition du mouvement de l'apprenant avec celui de l'expert, afin de corriger la position de chaque partie du corps (KYAN *et al.* 2015). Le système proposé par Yamaoka *et al.* évalue le geste du lancer de disque d'un apprenant en comparant les positions relatives des articulations entre elles à différents moments du geste. Ici, la connaissance experte est intégrée dès la conception de l'environnement, et la réutilisabilité du système dans un autre contexte nécessite une re-conception du système. De plus, les retours fournis par le système le sont en temps réel, et il n'est pas possible de rejouer le mouvement, ce qui rend plus difficile la correction et l'apprentissage du geste pour l'apprenant. L'objectif du système développé par Makio *et al.* est de mettre en lumière les différences entre les gestes des apprenants et des experts, sans pour autant fournir des indications ou des retours sur les corrections à effectuer. Ainsi, l'expert et l'apprenant n'ont qu'un rôle minime au sein de la conception, et la phase de restitution en propose qu'une visualisation graphique de ces différences.

Les systèmes se basant sur le calcul d'indicateurs à partir du geste de l'apprenant permettent de s'affranchir dans une certaine mesure de la présence d'un expert pendant l'apprentissage. En effet, ils sont généralement conçus dans l'optique de proposer une évaluation du geste de l'apprenant, souvent en le comparant aux données de l'expert intégrées au préalable lors de la conception du système. Les inconvénients de ces systèmes sont le côté peu explicite des retours ainsi donnés, se limitant souvent à un simple score, et la faible réutilisabilité dans d'autres contextes,



due à la nécessiter d'intégrer la connaissance experte dès la phase de conception du système.

Les outils d'analyse automatique du mouvement sont souvent conçus et développés dans l'objectif d'aider et d'assister l'expert et l'apprenant. En conséquence, la formalisation de la connaissance de l'utilisateur et de ses besoins d'observation est nécessaire, mais difficile à opérationnaliser. Le développement de ce type de systèmes a pour objectif de répondre aux défis d'adaptation au contexte d'apprentissage constitué du profil d'utilisateur (apprenant ou enseignant), du domaine d'application et de la tâche à apprendre. La genericité potentielle du système ainsi développé peut grandement varier suivant le contexte.

Les systèmes s'intéressant à la détection d'une suite d'action ordonnées au sein d'un mouvement vont dans un premier temps segmenter le mouvement en unités élémentaires, pour ensuite proposer une représentation de chacune de ces unités comparable à l'enchaînement attendu par l'expert. Ces travaux nécessitent ainsi l'intégration de la connaissance experte dès la conception du système, et fait qu'ils ne sont habituellement pas génériques. Il est cependant intéressant de noter que les systèmes développés par (MAHDI *et al.* 2019) et (DJADJA, HAMON et GEORGE 2019) ont pris en compte la réutilisabilité du système dès la conception, et proposent ainsi des éditeurs de scénarios pédagogiques utilisables par des non-expert de la réalité virtuelle.

Enfin, les systèmes faisant usage d'apprentissage supervisé afin d'observer la similitude des mouvements de l'apprenant à ceux de l'expert nécessitent également l'intégration de la connaissance experte *a priori*, afin d'utiliser l'ensemble de caractéristiques des gestes les plus pertinents afin d'obtenir la meilleure séparation possible. Le système d'évaluation de gestes développé par Patrona *et al.*, bien que permettant de segmenter des gestes et de les reconnaître sans *a priori* sur ces derniers (PATRONA *et al.* 2018), nécessite une base de données de geste de comparaison suffisamment importante afin de couvrir les différents gestes possibles à reconnaître.

Ce chapitre a présenté plusieurs techniques d'analyse du mouvement, classées en quatre catégories : (i) l'évaluation par observation, (ii) l'évaluation par analyse d'indicateurs calculables et (iii) l'évaluation d'une séquence ordonnée d'actions (iv) l'évaluation utilisant des techniques d'apprentissage automatique. Cette étude a aussi montré l'importance de la prise en compte des connaissances et besoins d'observation de l'expert dans le processus d'observation et d'analyse, ce qui implique de placer l'expert dans le processus de conception du système informatique dédié à l'apprentissage humain.

L'analyse peut se faire à partir du geste en lui-même ou d'un découpage du geste en unités élémentaires (DESPINOY *et al.* 2016) (YAMAOKA *et al.* 2013b) (PIRSIAVASH, VONDRICK et TORRALBA 2014). Les mouvements sont souvent modélisés sous la forme d'un ensemble d'informations cinématiques, telles que les positions, les orientations, les vitesses, les trajectoires, la courbure du mouvement, etc. L'utilisation de tels descripteurs permet d'obtenir des informations précises à traiter. Des descripteurs fournissant des informations globales sur le geste, bien que plus facilement compréhensibles et utilisables pour l'analyse pour un humain non-expert, ne permettent pas de systématiquement mettre en évidence les différences subtiles qui permettent de juger de la qualité d'un geste.

L'apprentissage par observation et imitation est naturelle pour l'humain. Cependant, cette méthode se heurte aux capacités de perception et de connaissances de l'utilisateur qui ne percevra pas nécessairement les propriétés du geste rendant ce dernier comme acceptable. En outre, l'analyse de toutes les caractéristiques d'un geste ainsi que leurs propriétés est difficile, tant ils dépendent des caractéristiques

de la situation d'apprentissage qui demeure très contextuel. L'utilisation d'outils informatiques proposant des indicateurs quant à la performance ou la qualité du geste permet ainsi à l'expert de proposer un retour plus pertinent mais nécessitent des connaissances scientifiques solides pour être interprétés ; ainsi qu'une phase de formalisation des besoins d'observation en amont.

En effet, dans le cas où les descripteurs utilisés sont extraits par un processus automatique, l'analyse en elle-même est souvent réalisée par un humain. De plus, il est difficile de proposer un retour pédagogique pertinent à l'apprenant à partir d'un processus entièrement automatisé. Dans ce cas, il faut disposer à minima d'un modèle d'expertise pédagogique, et d'un modèle de l'apprenant. Au-delà de la difficulté propre à l'extraction de descripteurs pertinents dans un cadre donné, il faut également être en mesure de fournir des indications suffisamment explicites afin que même une personne non-experte du domaine puisse être en mesure de s'auto-corriger. Bien que l'injection de la connaissance experte permet de formaliser et, par exemple, d'intégrer en amont un ensemble de conseils, ce type de systèmes est souvent réalisé de façon « *ad hoc* », contextuel au domaine voire à la tâche à apprendre et peut être limité dans le cas de problèmes non-définis à l'avance.

En outre, l'outil utilisé et la connaissance experte ne se rejoignent pas forcément. Par exemple, le système de (KYAN *et al.* 2015) propose un score pour évaluer chaque sous-partie du geste de danse, mais c'est à l'enseignant de donner les indications et corrections nécessaires afin d'améliorer le geste. Ainsi, l'outil informatique propose une pré-évaluation du geste, utilisable ou non par l'expert. L'analyse d'un geste étant matière à une certaine part de subjectivité de la part de l'expert, une étude purement automatique peut parfois ne pas être en accord avec l'avis de l'expert.

L'association des compétences de l'expert tout au long du processus d'apprentissage et d'évaluation avec l'aide possible apportée par un outil informatique est une approche qui semble être peu explorée. Un système permettant à la fois d'évaluer le geste, mais également de proposer un ensemble d'indicateurs visuels pourrait aider l'expert dans sa tâche d'évaluation lorsque l'observation n'est pas aisée. Il faut cependant qu'un tel système puisse disposer de la connaissance experte en amont, afin de pouvoir formaliser les différentes modalités du geste à analyser, et également qu'il soit en mesure de proposer des visualisations pertinentes et compréhensibles à partir des descripteurs calculés.

Le prochaine chapitre présente un système d'aide à l'observation, l'analyse et l'évaluation de gestes. Ce système est développé de manière à minimiser les travaux de réingénierie en cas de changement du geste cible à apprendre et des besoins d'observation.



Troisième partie

CONTRIBUTIONS



---

## TRAITEMENT DES DONNÉES DE MOUVEMENT

---

Les chapitres précédents montrent que la capture de mouvement peu prendre place dans de nombreux contextes. Les besoins en termes d'observation et d'analyse peuvent varier grandement en fonction du domaine ciblé. Il est nécessaire de choisir un système de capture adapté, au regard des besoins d'observation mais également des contraintes de place, de coûts et de temps d'acquisition. De plus, il est nécessaire de disposer d'un ensemble d'outils logiciels afin de pouvoir extraire, à partir des données, des informations utiles pour l'apprentissage d'un geste. Ces informations peuvent être de plusieurs natures : il est possible de proposer une comparaison précise entre les données de l'expert et celles de l'apprenant (*e.g.* " la position du coude est 4cm trop bas au moment  $T$  du mouvement "), bien que cette méthode soit soumise aux problèmes de différences de morphologie, d'alignement temporel des séquences de mouvements, etc. Il est également possible de fournir des indications de plus haut-niveau (*e.g.* " La vitesse lors du lancer doit être plus élevée "). Dans tous les cas, ces informations peuvent soit être utilisées telles quelles par l'apprenant, soit être utilisées par l'expert afin d'affiner son jugement du geste. Certaines conditions font qu'il n'est pas toujours possible pour un expert d'observer toutes les modalités importantes du geste pendant sa réalisation (trop de parties du corps à regarder, difficile de voir certaines parties du corps lors de l'exécution du geste, *etc.*). La proposition d'outils d'aide à la décision permet ainsi d'aider l'expert dans sa tâche d'évaluation et de remédiation.

Ce chapitre présente l'artefact informatique développé pour répondre aux besoins de la thèse, nommé **Motion Learning Analytics (MLA)**, ainsi que les différentes étapes du processus associé, du traitement des données de mouvements jusqu'à l'aide proposée à l'expert.

### 5.1 MOTION LEARNING ANALYTICS (MLA)

Dans cette partie, nous présentons la plateforme Motion Learning Analytics (MLA), créée pour répondre à nos besoins d'extraction et d'analyse de données de mouvement.

L'objectif de *MLA* est de proposer un ensemble d'outils permettant d'extraire des informations permettant d'analyser un geste à partir de données de mouvement bas-niveaux, afin de pouvoir être utilisées par un humain *a posteriori* afin d'aider au jugement de la qualité d'un geste en fonction du domaine applicatif. Il existe d'ores et déjà de nombreux systèmes d'analyse du mouvement. Ces systèmes ont déjà fait l'objet d'analyses. L'une d'entre elle propose une classification partielle selon l'objectif de ces systèmes (L. WANG, HU et T. TAN 2003) : la détection d'humains, le suivi (*tracking*) d'humains et la reconnaissance d'activité. Bien que ces domaines soient larges et très explorés dans la littérature (H. YANG *et al.* 2011) (J. AGGARWAL et XIA 2014), ils concernent plutôt la segmentation et l'extraction de mouvements au sein de différents types de données, tels que la vidéo ou des enregistrements

partiels à l'aide de capteurs de parties du corps. De plus, le contexte applicatif est souvent secondaire et ces systèmes sont développés de manière à proposer des méthodes automatique innovantes. A l'inverse, notre approche se concentre sur une application particulière, celle de l'apprentissage du geste, afin de proposer des conseils à un apprenant ou à l'expert qui enseigne ce geste. De plus, l'objectif de notre système est d'être indépendant du geste considéré.

Des systèmes tels que ceux présentés dans le chapitre 2 ((FERRARI *et al.* 2018), (MAKIO, TANAKA et UEHARA 2007)) sont eux conçus pour une tâche ou un domaine en particulier. Le fait d'avoir un domaine ciblé dès le début permet d'orienter le développement et l'analyse afin d'offrir un retour à l'expert le plus exact possible. Le problème se pose dans la réutilisabilité de tels systèmes dans des cas proches : le système n'est pas systématiquement suffisamment modifiable pour être utilisé dans d'autres contextes. L'objectif de MLA est, à terme, de proposer un ensemble d'outils et de descripteurs calculables permettant de couvrir une large gamme de domaines applicatifs, à l'aide de l'injection de la connaissance de l'expert. Elle est divisée en trois parties : le traitement des données de mouvement brutes, de la lecture des fichiers jusqu'à l'extraction des descripteurs, l'analyse des données obtenues à l'aide de techniques d'apprentissage semi-supervisé, et le retour fourni à l'expert et/ou l'apprenant. Ces trois parties sont indépendantes l'une de l'autre, et peuvent être utilisées séparément. Chaque partie est découpée en modules, correspondant à un sous-ensemble de fonctions au sein de l'artefact. Le workflow complet est présenté dans la figure 29.

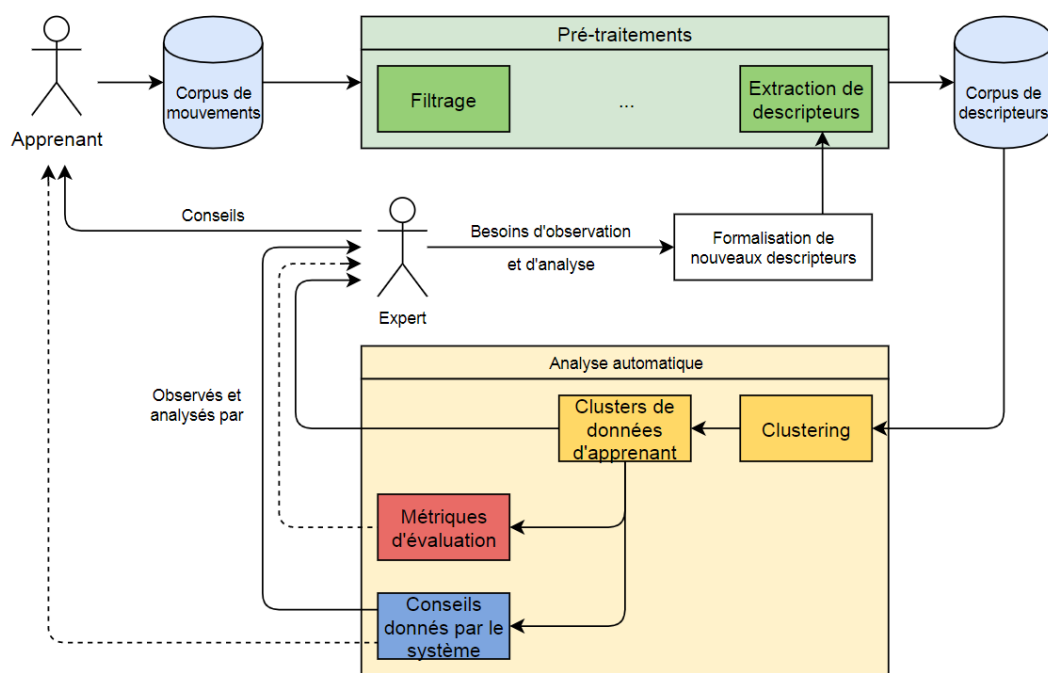


FIGURE 28 – Architecture de la plateforme *Motion Learning Analytics* (MLA).

### 5.1.1 Données de mouvement au format BVH

Le système prend en entrée des fichiers au format **BioVision Hierarchy** (BVH). Ce format de fichier permet de décrire un mouvement, posture par posture. Les fichiers **BVH** sont constitués d'une partie fournissant des informations sur les articulations composant le squelette, sous forme d'une hiérarchie (Fig. 30), et d'une seconde partie donnant les informations pour chaque articulation sur chaque frame du

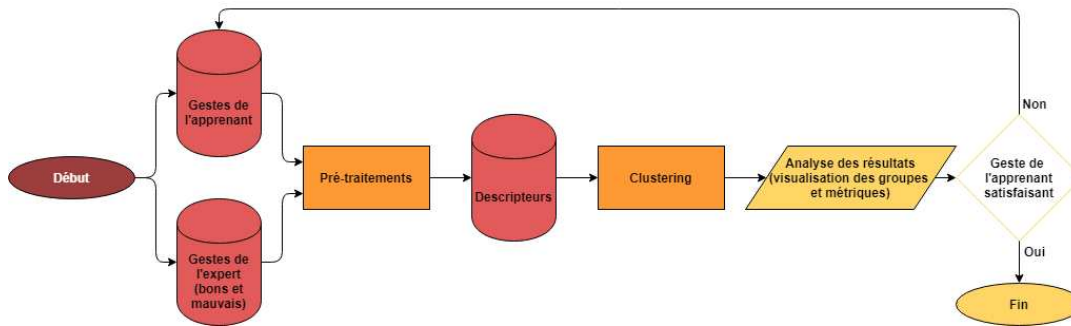


FIGURE 29 – Le workflow général de la plateforme MLA.

mouvement (Fig. 31). Plus précisément, chaque articulation peut avoir jusqu'à six informations : trois valeurs relatives à la position dans l'espace, et trois valeurs relatives à l'orientation dans l'espace. Les informations pour chaque articulation sont données relativement à l'articulation parente.

```

HIERARCHY
ROOT Hips
{
  OFFSET 0.00 0.00 0.00
  CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation Yrotation
  JOINT Chest
  {
    OFFSET 0.00 5.21 0.00
    CHANNELS 3 Zrotation Xrotation Yrotation
    JOINT Neck
    {
      OFFSET 0.00 18.65 0.00
      CHANNELS 3 Zrotation Xrotation Yrotation
    }
  }
}
  
```

FIGURE 30 – Un exemple de la hiérarchie du squelette au sein d'un fichier BVH (source : <http://research.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff/BVH.html>)

```

MOTION
Frames: 2
Frame Time: 0.033333
8.03 35.01 88.36 -3.41 14.78 -164.35 13.09 40.30 -24.60 7.88 43.80 0.00 -3.61 -41.45 5.82
  
```

FIGURE 31 – Un exemple de données relatives au mouvement au sein d'un fichier BVH source : <http://research.cs.wisc.edu/graphics/Courses/cs-838-1999/Jeff/BVH.html>)

### 5.1.2 Pré-traitements des données

La première partie (pré-traitements) a deux objectifs : (i) filtrer les données si nécessaire, afin d'éliminer les artefacts possibles dûs au matériel de captation, et (ii) extraire des descripteurs à partir du mouvement capturé, afin d'obtenir des informations exploitables. Cette partie est intégralement codée en **C++11**. Elle partie du système est subdivisée en six modules (Fig. 32). Le module central de l'architecture contient les classes qui permettent de représenter un mouvement dans son entièreté, à partir d'un fichier **BVH**. La classe **Motion** contient les informations relatives au mouvement en général (nom du mouvement, nombre de frames, temps inter-frame, frame de référence), ainsi qu'une liste d'objets **Frame**. Cette classe représente une frame du mouvement et ne contient que des articulations : un pointeur vers l'articulation racine, et une liste d'objets de type **Joint** (*articulation* en anglais). C'est à partir de l'objet **Motion** que les descripteurs sont extraits (Fig. 33). Une fois cette étape réalisée, les descripteurs sont insérés dans un objet de type **Data**,



qui va ensuite pouvoir être exporté. Cet objet nécessite que les données aient (i) un nom, et aient la forme (ii) d'un vecteur contenant une liste de paires clés (le nom de l'articulation) / valeurs (la valeur du descripteur).

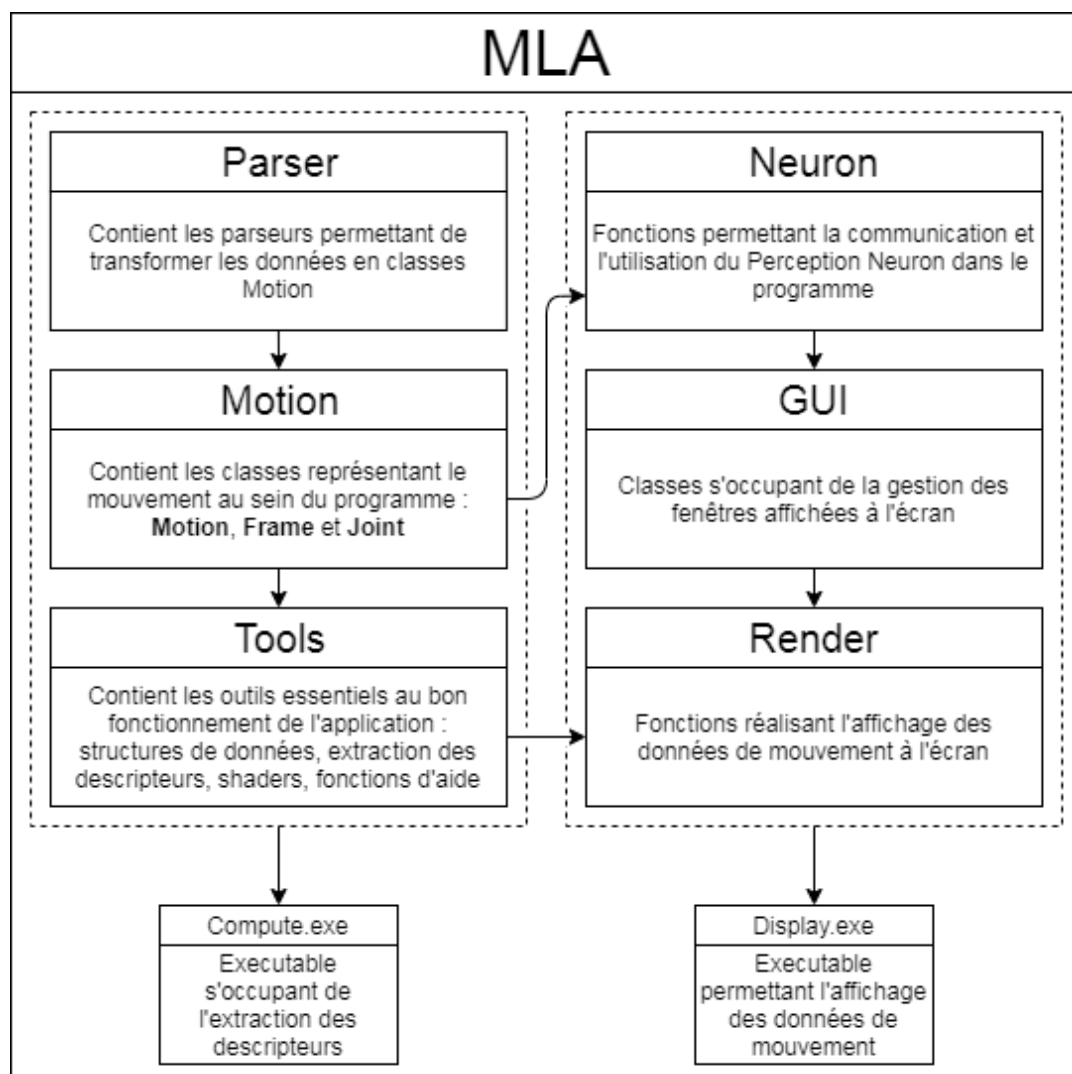


FIGURE 32 – L'architecture des pré-traitements et de l'affichage du mouvement capturé *MLA*, divisée en six modules principaux.

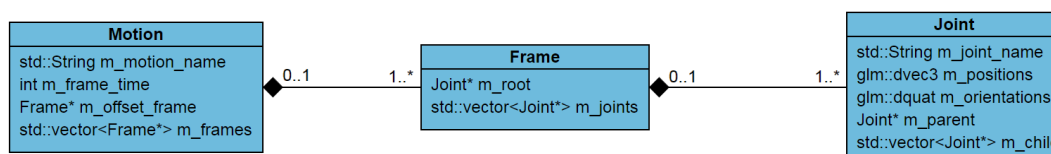


FIGURE 33 – Diagramme de classes des trois classes composant le mouvement.

Les descripteurs extraits du mouvement sont représentés au sein d'une structure **Data**, contenant le nom du descripteur, ainsi qu'un tableau contenant des couples de valeurs [nom de l'articulation] / [valeur du descripteur] (Fig. 34). L'export de ces données se fait au format **JavaScript Object Notation (JSON)** (Fig. 35), selon la hiérarchie suivante : descripteur → articulation → valeur. Le caractère peu-verbeux de ce format est un atout lors de l'export de fichiers de données volumineux. Sa très large utilisation, ainsi que le nombre conséquent d'analyseurs syntaxiques existants pour ce format en font un bon choix lorsque des données ont besoin de transiter d'une application à une autre (GAO et DUAN 2011). Il est ainsi possible de rechercher

de manière naturelle (au sens du langage considéré) les descripteurs, en respectant le cheminement « descripteur → articulation ».

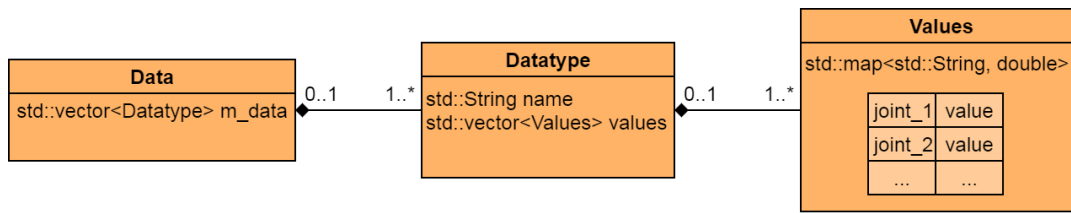


FIGURE 34 – Diagramme de classes des classes composant la structure de stockage des données.

```

"DistanceZ": {
  "distanceRightHandHead": [
    0.241683663821442
  ]
},
"MeanSpeed": {
  "EndHead": [
    0.0693643144687512
  ],
  "EndLeftFoot": [
    0.0596301134473924
  ],
  "EndLeftHandIndex3": [
    0.126122540664891
  ]
}
  
```

FIGURE 35 – Un exemple de descripteurs extraits, au format JSON. Le stockage s’effectue selon la hiérarchie « descripteur → articulation → valeur ».

### 5.1.3 Analyse automatique

La partie Machine learning est codée en Python 3.6+. Le Python a été retenu comme langage pour le large choix de bibliothèques d’analyse de données et de machine learning.

L’objectif de cette partie est de proposer un ensemble d’outils permettant l’analyse des descripteurs extraits du mouvement capturé, à l’aide de techniques d’apprentissage non-supervisé. Il est possible de choisir un sous-ensemble d’articulations et de descripteurs à analyser à partir des données extraites, afin de concentrer l’analyse sur les parties du corps voulues. Un algorithme de *clustering* est ensuite choisi, ainsi que les paramètres de cet algorithme. En fonction de l’algorithme, il est possible de modifier plusieurs paramètres : nombre de clusters (groupes) souhaités, taille du voisinage de recherche, *etc.* Ce choix s’effectue pour l’instant par la personne gérant le programme, car le choix de l’algorithme et de ses paramètres doit se faire en fonction des données utilisées, et nécessite une connaissance scientifique. Des métriques sont ensuite calculées afin de fournir une indication quant à la qualité des regroupements obtenus. Il est possible de comparer les regroupements à un étiquetage des données, afin de vérifier si la séparation obtenue correspond aux *a priori* de l’expert ou non. Cette phase de *clustering* permet également d’obtenir deux

groupes de mouvements : les mouvements cibles, et les mouvements avec défauts. Ces groupes sont ensuite utilisés pour donner un retour à l'apprenant.

#### 5.1.4 *Feedback*

La partie feedback est également codée en Python 3.6+. L'objectif de cette partie du programme est de fournir un module proposant une aide à la décision pour un expert, mais également de proposer un retour compréhensible pour l'apprenant, dans le cas d'une utilisation sans expert. À partir de descripteurs pré-sélectionnés par l'expert lui-même, et d'exemples de mouvements réussis / ratés, le système fournit une aide visuelle afin de visualiser la différence entre les mouvements de l'expert et ceux de l'apprenant. L'objectif pour l'apprenant est de rapprocher le plus possible ses mouvements du groupe des mouvements cibles, en améliorant son geste au fur et à mesure de l'apprentissage, à l'aide des conseils dispensés par l'expert et/ou le système.

La plateforme MLA permet le traitement automatique de données de mouvement capturé, du filtrage au traitement automatique par des méthodes de *clustering*, ainsi qu'à l'affichage synthétique d'information à destination de l'apprenant ou de l'expert. Cette partie a défini le contexte dans lequel s'incluent les travaux présentés dans cette thèse. Avoir une plateforme intégrant le traitement des données brutes, le calcul des descripteurs de mouvement, la comparaison et la visualisation des clusters de mouvements de l'apprenant possède l'avantage de centraliser les différentes étapes, et proposer ainsi un *workflow* plus simple au sein d'un même artefact informatique. La suite de ce chapitre explique en détail le *workflow* d'utilisation système.

## 5.2 CAPTURE DE MOUVEMENT (PERCEPTION NEURON)

Dans le cadre de cette thèse, la solution retenue pour la capture de mouvements a été celle de la combinaison de capteurs " Perception Neuron " (NOITOM 2015b), créée par l'entreprise NOITOM. Ce type de dispositif présente l'avantage d'offrir un compromis entre le coût et la qualité des données obtenues.

Le Perception Neuron est une combinaison de capteurs, basée sur l'utilisation de Neurons, des centrales inertielles (*Inertial Measurement Unit*, ou **IMU** en anglais) composées d'un gyroscope, d'un accéléromètre et d'un magnétomètre. Ce type de capteur possède l'avantage d'être facilement portable par sa taille et son poids, et donc de présenter une gêne réduite pour l'utilisateur. Ces capteurs sont beaucoup utilisés dans le domaine médical (PORCIUNCULA *et al.* 2018), pour l'analyse de la posture ou la rééducation d'une partie spécifique du corps par exemple.

Le Perception Neuron peut utiliser simultanément un ensemble de 32 Neurons au maximum. Les Neurons sont composés d'un accéléromètre, un gyroscope et un magnétomètre. Ces capteurs sont interchangeable, ce qui permet de ne changer que le Neuron défectueux si jamais un problème survient sur l'un d'entre eux. Cet ensemble de capteur est relié à un *hub*, qu'il est possible de connecter à un ordinateur de manière filaire (liaison USB) ou non-filaire (WiFi). Le système est alimenté par un autre port USB, et il est possible d'utiliser une batterie USB rechargeable, afin d'obtenir une mobilité accrue. Il n'est possible d'utiliser le Perception Neuron que dans certaines configurations spécifiques (voir Fig. 36). Les données sont renvoyées à une cadence de 120 postures par secondes lorsque la combinaison est constituée de

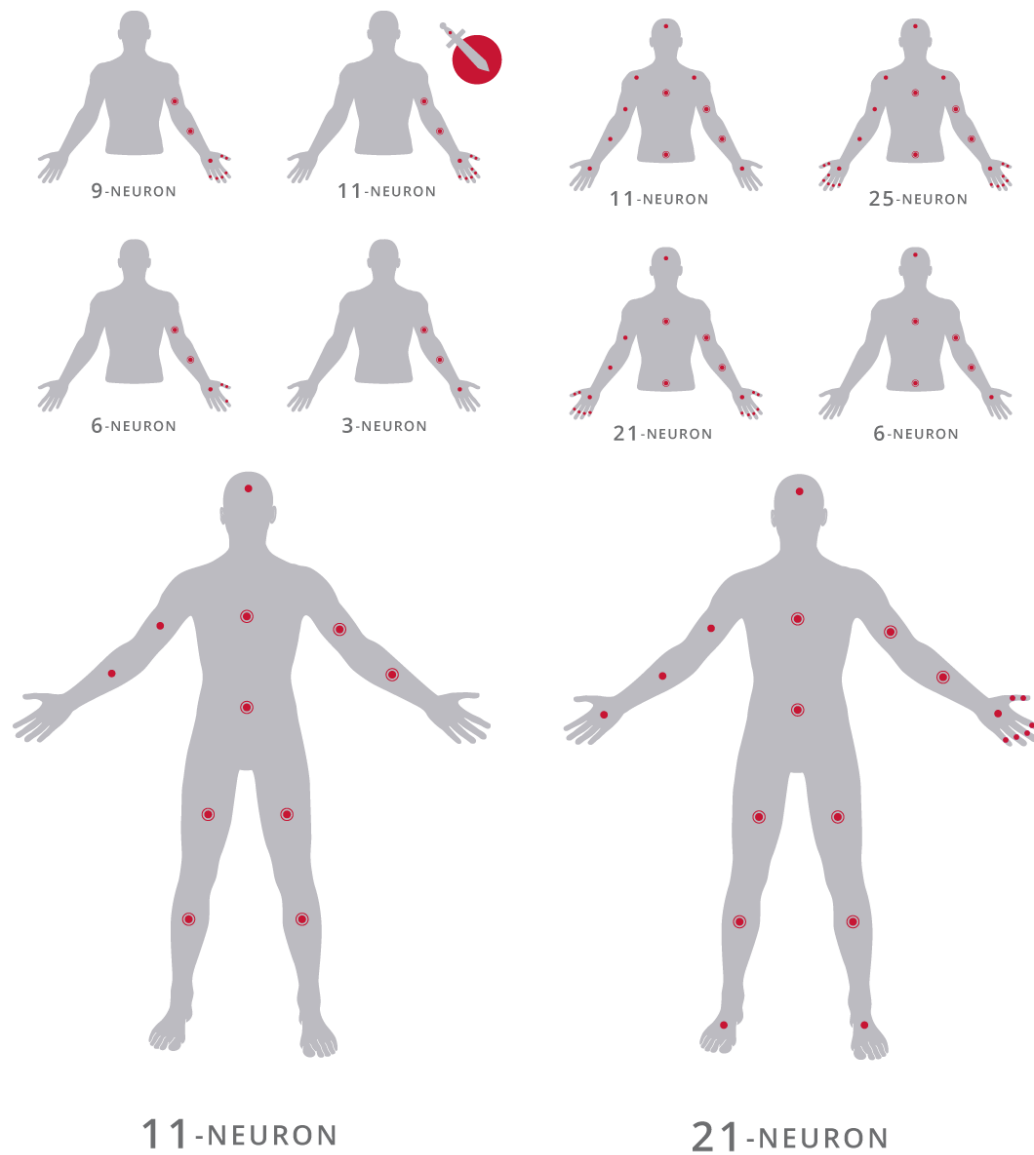


FIGURE 36 – Toutes les configurations possibles pour le Perception Neuron (NOITOM 2015a).

17 Neurones ou moins, et de 60 postures par secondes lorsque le nombre de capteurs est supérieur ou égal à 18.

Cette combinaison permet d'obtenir un squelette humain en 3D et donc d'éviter les problèmes propres aux systèmes de capture vidéo 2D, en particulier celui de l'occlusion.

L'utilisation de cette combinaison nécessite l'utilisation du logiciel Axis Neuron, développé par NOITOM. Ce logiciel permet d'effectuer la communication entre le hub de la combinaison et l'ordinateur. Il est également le seul capable de lire le format de données dans lequel est enregistré le mouvement. Il permet de réaliser les calibrations de la combinaison, de visualiser en temps réel le mouvement réalisé, de rejouer un mouvement (avec possibilité de marche avant / marche arrière, ralentissement, etc.), d'utiliser différents squelettes afin de s'adapter à toutes les morphologies possibles, et d'éditer le mouvement.

### 5.3 PRÉ-TRAITEMENTS DES DONNÉES

Bien que présentant de nombreux avantages, les centrales à inertie ne sont cependant pas dénuées de problèmes (GIANANTI *et al.* 2003). Le phénomène de la dérive (*drift* en anglais) est le plus courant : il s'agit de l'accumulation d'erreurs d'intégration, qui sert à obtenir l'information d'orientation de position à un temps  $T$  à partir de la vitesse, elle-même déduite de l'accélération. Les capteurs magnétiques sont également sensibles aux perturbations extérieures : la présence d'objets métalliques, ou d'objet émettant un champ électromagnétique. Ces phénomènes combinés font que les données obtenues lors des captures de mouvements ne sont pas toujours précises. La différence entre le geste réalisé par une personne et le geste visualisé sur le logiciel Axis Neuron est parfois très grande. De plus, le logiciel Axis Neuron applique des traitements aux données reçues des capteurs. Ils sont au nombre de trois : *Embedded Data Fusion*, *Human Body Dynamics* et *Physical Engine*. Ces algorithmes sont propriétaires, et donc non documentés de manière exhaustive : il est très difficile de savoir dans quelle mesure les données sont modifiées avant d'être présentées dans le logiciel. À cause de cette limitation, des corrections de données ont été effectuées sur les descripteurs extraits du mouvement, et non sur le mouvement original lui-même. Dans les parties suivantes, nous traitons des signaux. Le terme *artefacts* sera ici utilisé pour décrire les erreurs présentes au sein de ces signaux, dues au matériel de capture, et non pour parler d'un artefact informatique.

### 5.4 TRAITEMENTS DES DONNÉES DE MOUVEMENT

Dans le cadre de tests du système, des mouvements de lancers ont été utilisés afin d'en extraire des descripteurs. Ces descripteurs étaient relatifs aux vitesses de la main. Un des premiers tests du système a été la vérification de la cohérence de ces descripteurs par rapport au mouvement réel. Lors de l'extraction d'informations à partir des données de mouvement, il a été remarqué que les vitesses extraites à partir des positions successives présentaient des *artefacts* conséquents (Fig. 37). Ces *artefacts* se retrouvent le plus souvent lorsque des changements brusques de vitesse se produisent, c'est-à-dire quand l'accélération augmente ou décroît rapidement. Il n'est pas possible d'utiliser des données présentant ces défauts sans une étape de correction. L'analyse de l'origine de ces *artefacts* a permis de mettre au jour deux problèmes majeurs des données de mouvement provenant du logiciel Axis Neuron : le changement de la taille des membres du squelette au cours du temps, et les problèmes d'incohérences de changement de position d'une frame à une autre.

#### 5.4.1 Variation de la taille des membres

Un des premiers problèmes rencontrés a été la variation de la taille des membres du corps au cours du mouvement. Il apparaît que la taille des membres ne reste pas fixe tout au long du lancer. Ce problème peut provenir soit de l'imprécision des capteurs, soit des traitements effectués avant la restitution du mouvement.

Une solution simple est de fixer cette taille, qui correspond à l'écartement entre les articulations, pour chaque membre. Pour chaque enregistrement, il existe une posture de référence, qui contient toutes les informations de position et d'orientation pour chaque articulation. Cette *frame* est créée après la phase de calibration, ce qui fait qu'elle est moins sujette aux problèmes inhérents aux capteurs, car ces derniers surviennent au bout d'un certain temps de captation.

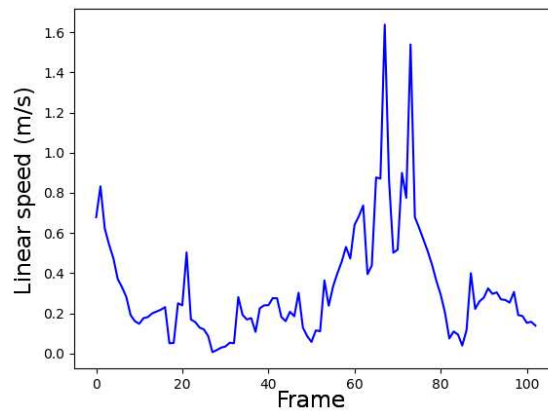


FIGURE 37 – Un exemple de vitesse linéaire extraite d'un mouvement. On remarque bien les pics de vitesse (autant vers le haut que vers le bas) entre les frames 60 et 80 (abscisses), ne correspondant pas à un lancer fluide.

Pour chaque articulation, les informations sont stockées sous la forme d'un vecteur de dimension 3, correspondants aux positions de l'articulation en  $x, y, z$ , et d'un quaternion correspondant à la rotation de cette articulation. Ces informations sont données par rapport à l'articulation parente : les valeurs de position et d'orientation d'une articulation fille  $j_{i+1}$  sont donc données relativement à l'articulation mère  $j_i$ , et la position de l'articulation fille  $p_{j_{i+1}}$  correspond donc à la taille du membre  $m_{j_{i+1}}$ . La première étape du traitement des données par la plateforme MLA consiste donc à copier les positions de l'articulation de la frame de référence sur toutes les autres frames du mouvement. L'orientation reste inchangée.

#### 5.4.2 Incohérences de position des articulations d'une frame à une autre

Un autre défaut présent dans les données a été constaté lorsque les vitesses ont été extraites des mouvements. Dans le cadre de tests, un mouvement de lancer a été enregistré, afin d'en extraire des descripteurs. Il s'est avéré que les valeurs de vitesses pouvaient parfois varier grandement d'une *frame* à une autre, sans pour autant que cette variation ne soit représentative du mouvement réel (Fig. 37). En effet, lors du lancer test, et après observation de l'enregistrement vidéo correspondant à ce lancer, il apparaît que les variations de vitesses très brèves (correspondantes aux pics vers le haut et vers le bas du signal) ne correspondaient pas à la réalité du mouvement. À l'œil nu, lorsque le mouvement est affiché à l'écran, ces erreurs ne sont pas visibles, car ces variations apparaissent sur des temps très courts (de l'ordre du centième ou du millième de seconde). Cependant, lors de l'extraction de la vitesse, il apparaît que la variation est parfois extrême sur un ensemble de frames données.

Nous avons formulé quatre hypothèses pour expliquer ces incohérences de non-linéarité dans le mouvement :

1. Le pas de temps entre les *frames* est constant, et l'interpolation entre ces *frames* est incorrecte, ce qui signifie que l'implémentation de l'interpolation entre les *frames* est mal réalisée au sein du logiciel Axis Neuron.
2. Le pas de temps entre les *frames* est constant, et il n'y a pas d'interpolation, ce qui signifie qu'il manque des *frames*. Ce problème pourrait subvenir lorsque la combinaison n'arrive pas à retourner les données des capteurs, à cause du trop grand nombre de données à traiter en simultané et en temps réel.

3. Le pas de temps entre les *frames* n'est pas constant et il y a interpolation, ce qui signifie que l'interpolation est incorrecte à cause d'un mauvais indice temporel donné à la *frame* reconstituée. Supposons un mouvement linéaire composé de trois frames  $f_0, f_2, f_4$  prises aux temps  $t_0, t_{0+2}, t_{0+4}$  respectivement. Si le système ne renvoie que les informations des frames  $f_0$  et  $f_4$ , l'interpolation entre la première frame  $f_0$  et la deuxième frame  $f_4$  reçue, étant supposée correspondre à l'instant  $t_{0+1}$ , correspondra en fait au moment  $t_{0+2}$ .
4. Le pas de temps entre les frames n'est pas constant, et il n'y a pas d'interpolation, ce qui signifie qu'il manque des frames. Supposons un mouvement linéaire composé de cinq frames  $f_0, f_1, f_2, f_3, f_4$  prises aux temps  $t_0, t_{0+1}, t_{0+2}, t_{0+3}, t_{0+4}$  respectivement. Si le système ne renvoie que les informations des frames  $f_0, f_1, f_3$  et  $f_4$ , il y aura un déplacement entre les frames  $f_1$  et  $f_3$  correspondant à  $2 * t$ .

Le tableau 4 résume ces hypothèses.

	Interpolation	Pas d'interpolation
Intervalle de temps constant	Mauvaise interpolation	<i>Frames</i> manquantes
Intervalle de temps non-constant	Mauvaise posture dû a un indice temporel erroné	<i>Frames</i> affichées au mauvais moment

TABLE 4 – Résumé des hypothèses sur les incohérences de positions

Il est possible que ces erreurs proviennent d'une baisse de la fréquence de retour des données, ou d'un problème d'acquisition des capteurs. En effet, sur le logiciel Axis Neuron®, chaque capteur possède un indicateur visuel permettant de juger de la qualité des données retournées : vert (OK), jaune (moyen), rouge (mauvais). Il n'existe cependant aucune indication sur la nature des dégradations des données. Indépendamment de la cause, il est possible de proposer des solutions permettant de pallier ce problème. Ces constatations font qu'il était nécessaire de trouver une manière de nettoyer les données avant leur utilisation. Plusieurs pistes ont été évoquées. Il serait possible de reconstruire le mouvement par interpolation, avec un pas de temps fixe. En prenant un pas de temps suffisamment grand, on peut espérer obtenir un mouvement recomposé évitant les sauts de distance. Cependant, en faisant ainsi, le nombre de *frames* du mouvement recomposé sera bas, et le mouvement recomposé risque de passer outre des moments clés du mouvement original (flexion d'une partie du corps, variation de l'accélération, etc.). Il existe des techniques d'interpolation se basant sur des *frames* clés (*key-frames*), afin de garder les informations cruciales du mouvement (BROTMAN et NETRAVALI 1988). Cette méthode nécessite de définir ces *frames* clés à l'avance. Or, au vu du nombre de données à traiter, il est difficile de spécifier manuellement ces frames clés pour chaque mouvement enregistré.

Les descripteurs vitesses extraits à partir du mouvement originel présentaient donc des *artefacts*. De par les caractéristiques propres à ces *artefacts* (pics très intenses et de courte durée), un filtre a été appliqué sur les valeurs de vitesses ainsi obtenues. Cette étape est cruciale, car les valeurs de vitesses sont utilisées plus tard dans la chaîne de traitement afin de segmenter le mouvement.

Il existe plusieurs types de filtres pouvant être appliqué à un signal, afin d'en réduire le bruit. Rojas-Lertxundi *et al.* présentent plusieurs de ces filtres dans (ROJAS-LERTXUNDI *et al.* 2015). Parmi eux, le filtre passe-bas a été utilisé par Kristianslund *et*



*al.* (KRISTIANSLUND, KROSSHAUG et BOGERT 2012) afin de filtrer les signaux provenant des mouvements des articulations de joueurs de handball et d'améliorer la détection et la prévention de blessures. Le filtre de Butterworth est un filtre qui possède un gain constant dans sa bande passante. Il a été utilisé par Malfait *et al.* (MALFAIT *et al.* 2014) afin de filtrer des données de test de détente. Les filtres de Chebyshev sont des filtres qui proposent une plus grande ondulation soit en bande passante (Type I) soit en bande atténuée (Type II) que les filtres de Butterworth. Il a été utilisé par Fanis et Stathis (MOSCHAS et STIROS 2011) afin de réduire le bruit provenant de données GPS et d'accéléromètre. Le filtre de Kalman repose sur le principe de l'estimation de données à partir de l'observation de données bruitées. Ces filtres sont très utilisés pour filtrer les données provenant d'IMUs (ZIHACHEZHADI *et al.* 2014) (MARINS *et al.* 2001) (LUNGE, VELTINK et BATEN 1999). Le filtre de Savitsky-Golay est un filtre passe-bas servant à lisser une courbe (SAVITZKY et GOLAY 1964). Il se base sur une fenêtre glissante, et effectue une régression afin de déterminer le polynôme minimisant l'erreur au sens des moindres-carrés pour un point donné. Plus cette fenêtre est large, plus le signal sera lisse, et plus les pics courts seront atténués. Le degré de ce polynôme dépend du cas applicatif : un polynôme de degré 2 permet de prendre en compte une courbure dans le signal, alors qu'un polynôme de degré 3 permet de prendre en compte les points d'inflexion du signal.

Ce filtre a été retenu, de par sa simplicité de compréhension, sa rapidité d'exécution et également son principe de réduction du bruit (à l'inverse du filtre Kalman, qui lui trouve des informations au sein d'un signal bruité). De plus, bien que le filtre de Kalman semble être un candidat judicieux au vu du matériel utilisé, il est habituellement utilisé pour traiter les données de mouvement provenant directement des capteurs, et non les descripteurs extraits de ces derniers.

La détermination des paramètres de ce filtre s'est effectuée de manière empirique. La taille de la fenêtre symétrique correspond à  $\frac{T}{4}$ , avec  $T$  étant la durée totale de l'enregistrement du mouvement. Le polynôme utilisé est de degré 3 : une valeur inférieure lisse trop les données, et résulte en une perte d'information, alors qu'une valeur supérieure crée un sur-ajustement 38. Il a été déterminé empiriquement qu'une taille de fenêtre de  $\frac{nbframes}{4}$  permettait d'obtenir un signal lissé, sans pour autant en gommer les caractéristiques principales.

L'application de ce filtre sur les données de vitesse bruitées (Fig. 39) permet d'obtenir une courbe plus lisse, se rapprochant d'avantage de la réalité de la vitesse du mouvement effectué. La courbe de vitesse lors du lancer correspond aux enregistrements, et il est également possible de distinguer les gestes d'amplitude plus faible précédant le lancer (dans le cas présent, il s'agit de l'étape de balancement de la main avant le lancer, voir Fig. 41).

#### 5.4.3 Extraction du moment d'intérêt du mouvement

Lors de l'analyse d'un mouvement, il n'est pas systématiquement nécessaire de disposer de l'intégralité du mouvement. Par exemple, dans le cadre d'un service au tennis, l'analyse du mouvement se concentre sur les parties hautes du corps, notamment le poignet de la main qui tient la raquette (MAKIO, TANAKA et UEHARA 2007). Il est également possible que le mouvement jugé utile soit contenu au sein d'un enregistrement plus large : il faut alors être capable d'extraire le mouvement d'intérêt. Segmenter manuellement un mouvement est une tâche réalisable par un humain. Cependant, en fonction du nombre de données à traiter, de la durée totale des mouvements, du nombre de passages à segmenter, cette tâche peut devenir chronophage. La segmentation automatique du mouvement peut pallier ce problème.



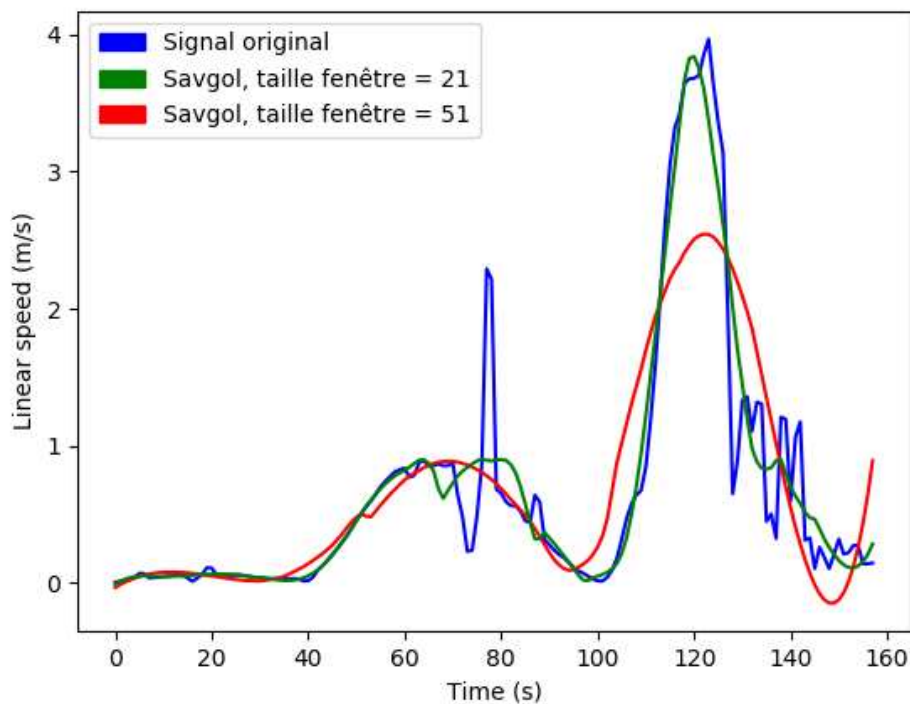


FIGURE 38 – Un exemple de filtrage de données de vitesses par un filtre de Savitsky-Golay. Le signal original est en bleu, suivit par les résultats du filtrage avec une taille de fenêtre de 21 en vert) et une taille de fenêtre de 51 (en rouge). Une taille de fenêtre trop élevée efface tous les artefacts présents, mais gomme également les caractéristiques propres du signal (le changement brusque de vitesse).

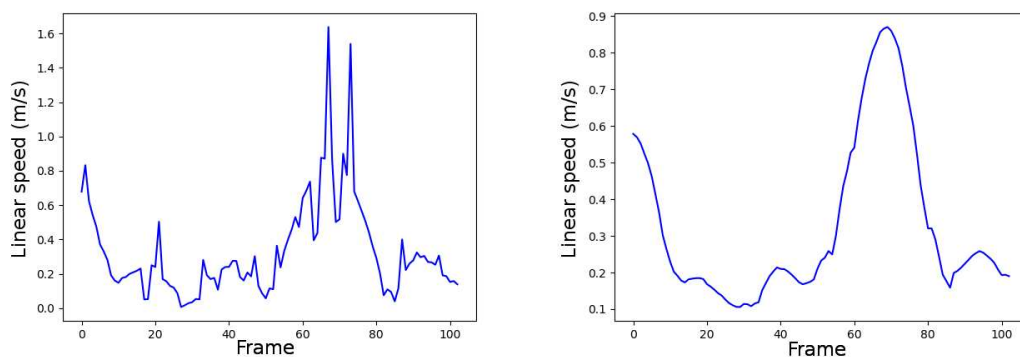


FIGURE 39 – Un exemple de vitesse linéaire filtrée (à droite) à l'aide d'un Savitsky-Golay. Les pics correspondants aux artefacts du mouvement ont disparu, et les caractéristiques initiales du changement de vitesses ont été conservées.

Il existe plusieurs techniques de segmentation automatique du mouvement (ZAPPELLA, LLADÓ et SALVI 2008). L'ensemble des techniques permettant cette segmentation se basent sur différentes méthodes : apprentissage supervisé / non-supervisé, réduction de dimension, etc. Le choix de la méthode se base souvent sur la connaissance *a-priori* des caractéristiques propres au mouvement. Les techniques à base d'apprentissage supervisé supposent que les classes (*i.e.* les types de mouvements présents au sein du corpus) soient connues en amont. À l'inverse, les techniques non-supervisées ne possèdent pas d'*a-priori* sur le type du mouvement : la segmentation se fait sans assignation à une classe.

Les méthodes de segmentation à base d'apprentissage supervisé se basent sur un corpus d'apprentissage correspondant au sous-mouvement à extraire du mouvement principal. Ces méthodes proposent souvent à la fois une segmentation en actions et une classification de ces actions. En entraînant des classifieurs sur des séquences choisies avec soin au préalable, il est possible d'obtenir une segmentation du mouvement, en fonction des différentes actions présentes. Les méthodes peuvent faire appel à des classifieurs SVM (HOAI, LAN et DE LA TORRE 2011), des modèles de markov cachés (*Hidden Markov Models*, HMM) et une combinaison de classifieurs faibles afin de créer un classifieur fort (Multi-class Adaboost) (F. LV et NEVATIA 2006), ou encore des réseaux de neurones (BOUCHARD et BADLER 2007b). Ces méthodes nécessitent toutes un corpus préalablement constitué, étiqueté et sont le plus souvent testées sur des corpus de mouvements existants et connus, tels que CMU (UNIVERSITY 2016) et HDM05 (MÜLLER *et al.* 2007). Ces bases sont constituées de nombreux mouvements, tels que se lever, s'asseoir, prendre un objet, etc. Ces mouvements ne nécessitent pas d'effort cognitif particulier pour un humain, et donc ne sont pas sujet à un quelconque apprentissage. De plus, lorsque le domaine métier ne dispose pas de base de mouvements pré-établie, il faut en constituer une en disposant de suffisamment d'exemples étiquetés, et couvrant une grande variété de cas (différences de morphologie, de durée, de manière de faire, etc.).

La segmentation de mouvement par l'approche à base d'apprentissage non-supervisé permet de s'affranchir de certaines de ces contraintes. L'objectif reste le même, à savoir décomposer un mouvement en plusieurs sous-mouvements, mais cette fois sans *a-priori* sur ces séparations. Trois méthodes ont été proposées par Barbi *et al.* (BARBIČ *et al.* 2004), traitant toutes le mouvement comme étant une séquence ordonnée de *frames*. À partir des données de rotation (quaternions), une analyse en composantes principales (PCA) de ces données est réalisée sur chaque *frame*, afin d'obtenir une représentation de plus basse dimensionnalité. Cette approche se base sur le fait qu'un mouvement simple aura une représentation de basse dimensionnalité. En minimisant l'erreur de projection, la dimensionnalité optimale est trouvée. En calculant ensuite la dérivée des valeurs successives de l'erreur, il est possible de détecter les fractures au sein du mouvement, *i.e.* là où la dérivée augmente fortement. La deuxième méthode se base sur l'analyse en composantes principales probabilistes (PPCA) (TIPPING et BISHOP 1999). Les premières *frames* du mouvement sont modélisées sous forme d'une distribution Gaussienne, en fonction de la dimension intrinsèque estimée. La probabilité d'appartenance des *frames* suivantes à la distribution Gaussienne est ensuite calculée, à l'aide d'une distance de Mahalanobis. En augmentant progressivement le nombre de frames utilisées pour le calcul de la distribution Gaussienne, il est possible d'obtenir une courbe de la distance calculée à chaque fois. Les pics de cette courbe apparaissent lorsqu'un changement de comportement survient : il s'agit du moment où le mouvement change, et donc où la segmentation va prendre place. La dernière technique proposée dans ces travaux se base sur un modèle de mélange de gaussiennes (GMM). L'hypothèse de départ est

que les frames de différents mouvements simples vont former différents clusters. Le PCA est utilisé afin de réduire les temps de traitements de l'algorithme d'estimation, et le nombre de composantes à garder est choisi afin que 90% de la variance de la distribution des données originale soit préservée. Une fois les paramètres du modèle estimés, un cluster d'appartenance est calculé pour chaque *frame*. L'inconvénient de cette méthode est qu'il faut spécifier à l'avance le nombre de clusters.

Une autre technique utilise l'analyse de clusters alignés, afin d'obtenir une segmentation du mouvement (FENG ZHOU, DE LA TORRE et HODGINS 2008). Cette méthode réduit le problème de la segmentation à un problème de *clustering*. Elle est basée sur l'algorithme des *k-means*, et propose deux changements majeurs par rapport à cet algorithme : la taille des caractéristiques d'un cluster peut varier, et la fonction de distance utilisée se base sur la déformation temporelle dynamique (DTW). Cette technique permet de calculer une similarité entre deux séries temporelles, tout en étant robustes aux décalages de phase sur l'axe du temps (BERNDT et CLIFFORD 1994). Afin d'appliquer cette segmentation à de larges corpus, il est nécessaire d'opérer une réduction du nombre de *frames* initiales.

La granularité de la décomposition peut grandement varier. Ainsi, la méthode développée par Krüger *et al.* (KRÜGER *et al.* 2017), en utilisant en premier lieu du regroupement de caractéristiques, puis ensuite le voisinage des graphes constitués par les caractéristiques propres à chaque *frame*, permet non seulement de séparer différentes actions au sein d'un mouvement initial (par exemple, une personne qui marche puis qui se met à courir), mais également de subdiviser encore les sous-mouvements obtenus en primitives (par exemple, les pas individuel lors de la marche et de la course).

Bien que les techniques précédemment présentées atteignent des bons taux de segmentation (par rapport à la segmentation manuelle), la solution retenue est plus simple, et se base sur les pics de vitesses du mouvement. Cet algorithme permet une segmentation rapide des mouvements, dès lors que le mouvement d'intérêt présente une forte variation de vitesse. Les mouvements caractéristiques des expérimentations sont le passage d'une vitesse quasi-nulle (position stationnaire), ou d'une vitesse faible (mouvement de balancier avant le lancer), à une vitesse élevée (mouvement du lancer), puis au retour à une vitesse faible, voire nulle (position de repos). La solution retenue permet ainsi une bonne segmentation du mouvement d'intérêt (*Motion of Interest, MoI*). L'algorithme de segmentation qui a été développé se base sur un repérage de la valeur maximale de la vitesse d'une articulation pré-déterminée. Dans le cas d'étude présenté, la main est l'articulation choisie, car nous nous plaçons dans le cas de mouvements de lancer, et cette articulation est source de la plus grande amplitude en terme de positions. Il serait cependant possible de déterminer automatiquement quelle articulation présente la plus grande amplitude de variations de vitesses, et de trouver ainsi le pic de vitesse maximale lors du geste. En partant de ce maximum, la segmentation s'effectue en choisissant un nombre de minimum locaux à franchir avant et après ce maximum. Ce choix permet d'intégrer au *MoI* les phases de balancement pré-lancer, par exemple. Dans notre cas, nous nous limitons au *MoI* seul. L'algorithme ainsi choisi (Algorithm 1) permet de limiter la consommation de ressources (temps et mémoire).

Cette étape marque la fin de la chaîne de pré-traitement des données de mouvement. En sortie de cette chaîne, un mouvement filtré et segmenté est proposé, afin de n'extraire les descripteurs que des parties d'intérêt du mouvement (Fig. 40 et 41).

**Algorithm 1** Mol extraction algorithm

---

```

1: motion : the full motion
2: lc : the number of desired left cuts
3: rc : the number of desired right cuts
4: procedure MOTION SEGMENTATION(motion, lc, rc)
5:   hs  $\leftarrow$  GetHighestSpeedValue(motion)
6:   p_lm  $\leftarrow$  GetPreviousLocalMin(hs)
7:   n_lm  $\leftarrow$  GetNextLocalMin(hs)
8:   left_found = 0
9:   right_found = 0
10:  seg_motion  $\leftarrow$  segmented motion from p_lm to n_lm
11:  while left_found < lc do
12:    n_lm = p_lm
13:    p_lm  $\leftarrow$  GetPreviousLocalMin(p_lm)
14:    seg_motion  $\leftarrow$  motion from p_lm to n_lm added at the beginning
15:  end while
16:  while right_found < rc do
17:    p_lm = n_lm
18:    n_lm  $\leftarrow$  GetNextLocalMin(n_lm)
19:    seg_motion  $\leftarrow$  motion from p_lm to n_lm added at the end
20:  end while
21: end procedure

```

---

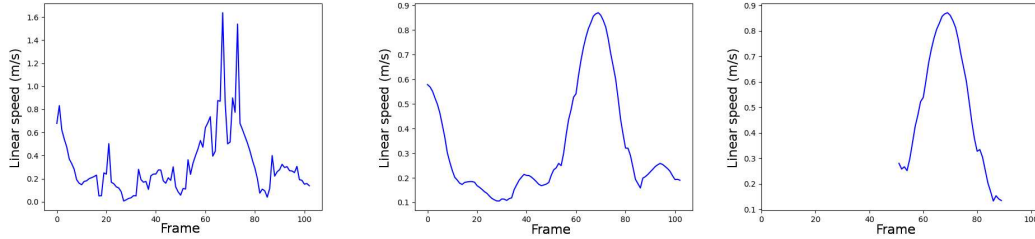


FIGURE 40 – Un exemple de vitesse linéaire filtrée puis segmenté par l’algorithme 1.

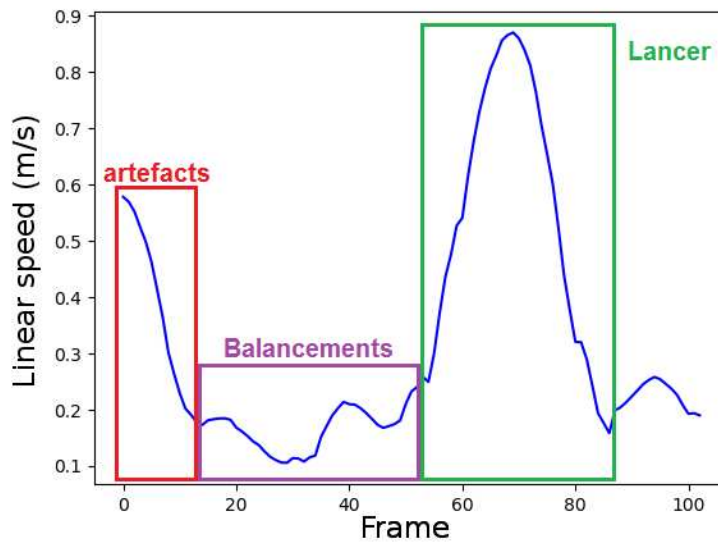


FIGURE 41 – Le signal filtré décomposé. En rouge, des artefacts de vitesses élevées dues aux capteurs. En violet, les balancements pré-lancer, et en vert, le lancer qui est extrait.

#### 5.4.4 Aide à l'expert et retour à l'apprenant

À partir des descripteurs préalablement extraits, cette partie du système se concentre sur la comparaison des données expertes avec celles de l'apprenant, afin de fournir un retour sur les mouvements effectués. L'objectif est d'assister l'expert dans sa tâche d'évaluation du geste, en lui permettant d'obtenir un retour visuel sur les différences entre le geste de l'apprenant et le sien. Cela nécessite d'avoir des données expertes qui correspondent au geste cible, ainsi qu'un jeu de données pour chaque défaut à détecter au sein du mouvement. L'algorithme 2 présente le cheminement. À partir des données de l'expert (réparties en gestes cibles et en mauvais gestes) et de l'apprenant, les données sont normalisées si nécessaire. Ensuite, une liste de descripteurs à observer pour chaque défaut identifié est demandée, afin d'associer les-dits descripteurs aux données de l'expert. Les données de l'expert sont ensuite utilisées dans un processus de *clustering*, afin d'obtenir deux groupes, correspondants aux gestes cibles et aux mauvais gestes. Les données de l'apprenant sont ensuite comparées à celles de l'expert, afin d'être en mesure de déterminer quels sont les défauts à corriger en priorité pour l'apprenant. Enfin, un retour sous forme visuelle, présentant les données de l'expert et celle de l'apprenant, est proposé à l'expert, ainsi que possiblement des conseils des données par le système.

Ce système nécessite des données d'expert qui doivent contenir à minima : (i) des exemples de bons mouvements, et (ii) des exemples de mouvements présentant des défauts. Idéalement, les mouvements avec les défauts de l'expert doivent couvrir le plus d'exemples possibles pour chaque défaut. L'étiquetage marquant les bons et mauvais gestes de ces données doit être explicitement donnée au système, afin qu'il soit possible de faire correspondre un jeu de donnée avec un défaut. Il faut ensuite fournir une liste des descripteurs à observer, pour chaque défaut. Il est possible d'utiliser n'importe quelle combinaison d'articulation en fonction des besoins d'observation.

Il faut ensuite choisir quelles données seront utilisées pour la comparaison, autant pour l'expert que l'apprenant. La spécification de chaque type de données se fait de la manière suivante :

1. le défaut pour lequel ce(s) descripteur(s) est/sont considéré(s)
2. le descripteur à utiliser
3. le(s) articulation(s) à considérer pour ce descripteur, avec pour chaque articulation la prise en compte ou non de la latéralité de la personne (e.g. si il faut remplacer " droite " par " gauche " dans le nom des articulations si la personne est gauchère)

Ainsi, un exemple d'une telle spécification peut ressembler à la hiérarchie suivante :

- Défaut : Se pencher en avant lors du lancer
  - Descripteur : Vitesse moyenne
    - Articulation : Épaule gauche, Latéralité : Non
    - Articulation : Épaule droite Latéralité : Non
- Défaut : Déplacement du coude lors du lancer
  - Descripteur : Vitesse moyenne
    - Articulation : Bras gauche Latéralité : Oui
    - Articulation : Épaule gauche Latéralité : Oui

Les données peuvent être normalisées. En effet, le fait d'avoir des données sur plusieurs échelles différentes fait que certaines données peuvent devenir prédominantes dans le cas d'une analyse. Prenons l'exemple d'un descripteur  $A$  dont les valeurs vont de 0 à 5, et d'un descripteur  $B$  dont les valeurs vont de 500 à 1500. Dans ce cas, les variations du descripteur  $A$  seront toujours insignifiantes comparées aux variations du descripteurs  $B$ , si les valeurs de chaque descripteur couvrent leurs intervalles respectifs de la même manière. L'étape de normalisation permet de mettre les données sur une même échelle, annulant par la même occasion les différences de plages de valeurs. Cette étape est cruciale lorsque les descripteurs utilisés ont une plage de valeur très différente les uns par rapport aux autres. Cette normalisation s'effectue en divisant chaque valeur composant les vecteurs de caractéristiques par la moyenne de ces valeurs au sein des données utilisées. Elle intervient avant le processus de clustering.

Les données de l'expert sont ensuite passées dans une phase de clustering (algorithme des  $k$ -means, avec  $k = 2$ ). Cette étape permet de constituer deux groupes de données : le groupe correspondant aux bons gestes, et le groupe correspondant aux gestes avec des défauts. Une première approche serait de faire ce regroupement à l'aide de l'étiquetage des données. Cependant, un problème de cette approche est qu'elle suggère que l'expert a fourni un ensemble de données où les gestes possédant des défauts sont séparables des bons mouvements. En pratique, il arrive que lorsque l'expert effectue plusieurs gestes avec un même défaut de suite, un phénomène d'auto-correction apparaît : inconsciemment, son geste va se rapprocher du bon geste initial, celui qu'il est habitué à faire. Cela se traduit par la présence de bons mouvements étiquetés comme étant des mouvements avec défauts. Il est possible de pallier en partie ce problème en faisant un regroupement des gestes à l'aide de clustering (Fig. 42).

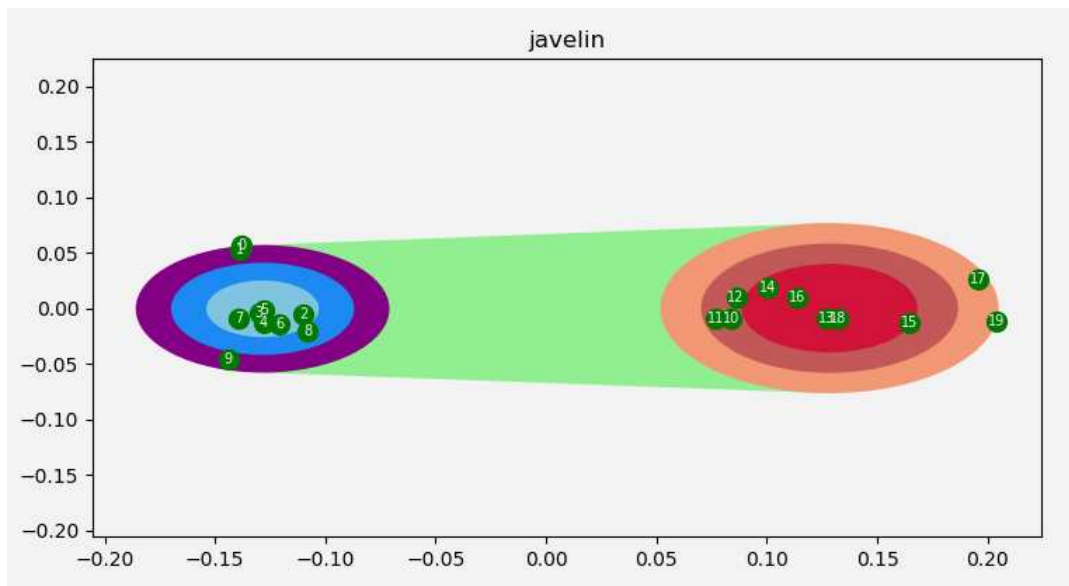


FIGURE 42 – Exemple de clusters formés à partir des données de l'expert (points verts). En bleu, le cluster des bons gestes, en rouge, le cluster des mauvais gestes pour le défaut appelé " Javelin ".

Les limites de cette approche se situent dans l'impossibilité de corriger le problème si jamais deux clusters se superposent. La superposition des clusters implique que les différences entre les bons et mauvais gestes peuvent être plus ou moins floues, en fonction du degré de superposition. Cela implique que les données de l'expert ne sont pas suffisamment significatives par rapport au défaut observé, ou que les



**Algorithm 2** Feedback system

---

```

1: expert : object describing the expert (name, laterality and path to data)
2: student : object describing the student (name, laterality and path to data)
3:  $D_{exp} \leftarrow$  all data from expert
4:  $X_{D_{exp}} \leftarrow$  expert's data repartition
5:  $D_{std} \leftarrow$  all data from student
6:  $P \leftarrow$  all the data to process for each problem
7: normalisation  $\leftarrow$  boolean indicating if the data must be normalised or not
8: procedure FEEDBACK(expert, student)
9:   if normalisation == true then
10:      $P \leftarrow$  normalised( $P$ )
11:   end if
12:   for all  $p_i$  in  $P$  do
13:      $d_{exp} \leftarrow$  get_data( $p_i$ ,  $X_{d_{exp}}$ )
14:     clustering_result  $\leftarrow$  run_clustering( $d_{exp}$ ,  $p_i$ )
15:      $C_{std} \leftarrow$  get_centroid( $D_{std}$ ,  $p_i$ )
16:     if dimension( $C_{std}$ ) > 2 then
17:        $pca = PCA(n_{components} = 2)$ 
18:        $pca(D_{exp}, D_{std}, C_{std})$ 
19:     end if
20:     clustering_problem  $\leftarrow (D_{exp}, X_{D_{exp}}, D_{std}, C_{std}, p_i, \text{clustering\_result})$ 
21:   end for
22:   give_advices(clustering_problem)
23:   display(clustering_problem)
24: end procedure

```

---

descripteurs observés ne sont pas significatifs par rapport au défaut. Une piste serait de vérifier la superposition ou non des deux clusters, et de supprimer automatiquement les points de données dont la distance au centroïde du cluster auquel ils sont assignés est la plus extrêmes, afin de réduire ou d'annuler la superposition des clusters.

Pour la visualisation des données de l'apprenant par rapport à celles de l'expert, une réduction de dimensionnalité des données est nécessaire lorsque celle-ci est supérieure à 2. Pour cette étape, une analyse en composantes principales (PCA) est utilisée, afin d'obtenir la dimensionnalité voulue en sortie. Cette étape est réalisée juste avant l'affichage, afin d'effectuer tous les calculs sur les données non-modifiées. Un exemple de cet affichage est présent sur la Fig. 43. On y voit un exemple avec quatre défauts identifiés, ainsi que les données de l'expert (cercles rouges et bleus). Cet affichage permet d'identifier dans quelle mesure le geste de l'apprenant correspond au geste cible : on remarque que pour les défauts "leaning" et "elbow move", le centroïde de l'apprenant (en rouge) est compris dans le cluster des bons mouvements de l'expert. Par contre, pour les défauts "javelin" et "align arm", les données de l'apprenant sont en moyenne plus proche du cluster de mauvais gestes. Ainsi, l'expert peut se baser sur ce retour visuel pour proposer des exercices permettant d'améliorer ces deux aspects du geste concernés.

Afin de proposer le(s) conseil(s) jugé(s) le(s) plus pertinent(s), le système doit être en mesure de fournir une indication quant à la présence plus ou moins forte des défauts identifiés. Pour répondre à ce besoin, une distance entre le centroïde des données de l'apprenant et les clusters expert est calculé. Soit  $c_g$  le centroïde



du clusters des bons gestes,  $c_b$  le centroïde du clusters des mauvais gestes, et  $c_a$  le centroïde des données de l'apprenant. On définit :

$$A = y_{c_b} - y_{c_g}$$

$$B = x_{c_b} - x_{c_g}$$

$$C = (x_{c_g} * y_{c_b}) - (x_{c_b} * y_{c_g})$$

La distance  $D$  est défini comme il suit :

$$D = \frac{A * x_{c_a} + B * y_{c_a} + C}{\sqrt{A^2 + B^2}}$$

Cette distance correspond à la distance entre le centroïde des données de l'apprenant et le point de la droite reliant les centroïdes des deux clusters experts le plus proche. Elle se base sur l'hypothèse que si les données de l'apprenant se situent dans le trapézoïde (en vert sur les Fig. 43 et 42) reliant les deux clusters de l'expert, le défaut est plus facilement corrigé que s'il est en dehors de cette zone. Elle permet d'évaluer à quel point les descripteurs des données de l'apprenant sont éloignés de ceux de l'expert : plus elle est grande, plus l'apprenant est loin du geste de l'expert pour un défaut donné. Selon les cas d'utilisations, ce conseil peut être utilisé tel quel, ou l'expert peut se baser sur (i) la visualisation proposée par le système et/ou (ii) les conseils fournis par le système. Dans l'exemple montré dans la Figure 43, on peut noter que le centroïde de l'apprenant, pour le défaut « *align arm* », est situé en dehors des clusters et de l'espace situé entre ces clusters. Dans ce cas de figure, l'explication de cette distance peut être problématique. En effet, si le centroïde se situe dans l'alignement des clusters, bien qu'il soit en dehors des clusters et de la zone située entre eux, on peut supposer que le défaut est encore plus accentué chez l'apprenant. Cependant, lorsque le centroïde ne se situe pas dans cet alignement, il nous est pour l'instant impossible de donner une indication précise sur la nature du défaut de l'apprenant seulement à partir des données de mouvement ; ce point devra être travaillé dans le futur. Une piste serait de comparer la distance entre chaque descripteur extrait, afin de donner un retour sur l'aspect du geste étant le plus mauvais.

## 5.5 RÉSUMÉ DES CONTRIBUTIONS

Dans cette partie, une chaîne de traitement adaptée au Perception Neuron, pour l'analyse du mouvement capturé en situation d'apprentissage a été présentée. Les problèmes auxquels cette chaîne de traitement propose des réponses sont (i) la modification de la taille des membres au cours du mouvement, (ii) les incohérences de positions des capteurs entre certaines frames. Plusieurs étapes de filtrage des données ont été présentées afin de palier ces problèmes. L'incohérence de positions des capteurs se traduisait par la présence de pics lors de l'extraction des vitesses. Ces pics ne correspondaient pas à la réalité du geste. Un filtre passe-bas a été appliqué, afin de compenser les artefacts du signal. Le problème de cette méthode est que les paramètres du filtre doivent être estimés (soit empiriquement, soit automatiquement) pour chaque signal, en fonction de ses caractéristiques propres, afin de ne pas trop "lisser" les valeurs réelles du signal.

Afin de ne garder que la partie utile du mouvement, un algorithme de segmentation a été présenté. L'algorithme de segmentation du mouvement utile permet de ne garder que les parties voulues du mouvement. Cet algorithme se base sur le maximum de la vitesse d'une articulation possédant les amplitudes et/ou variations du

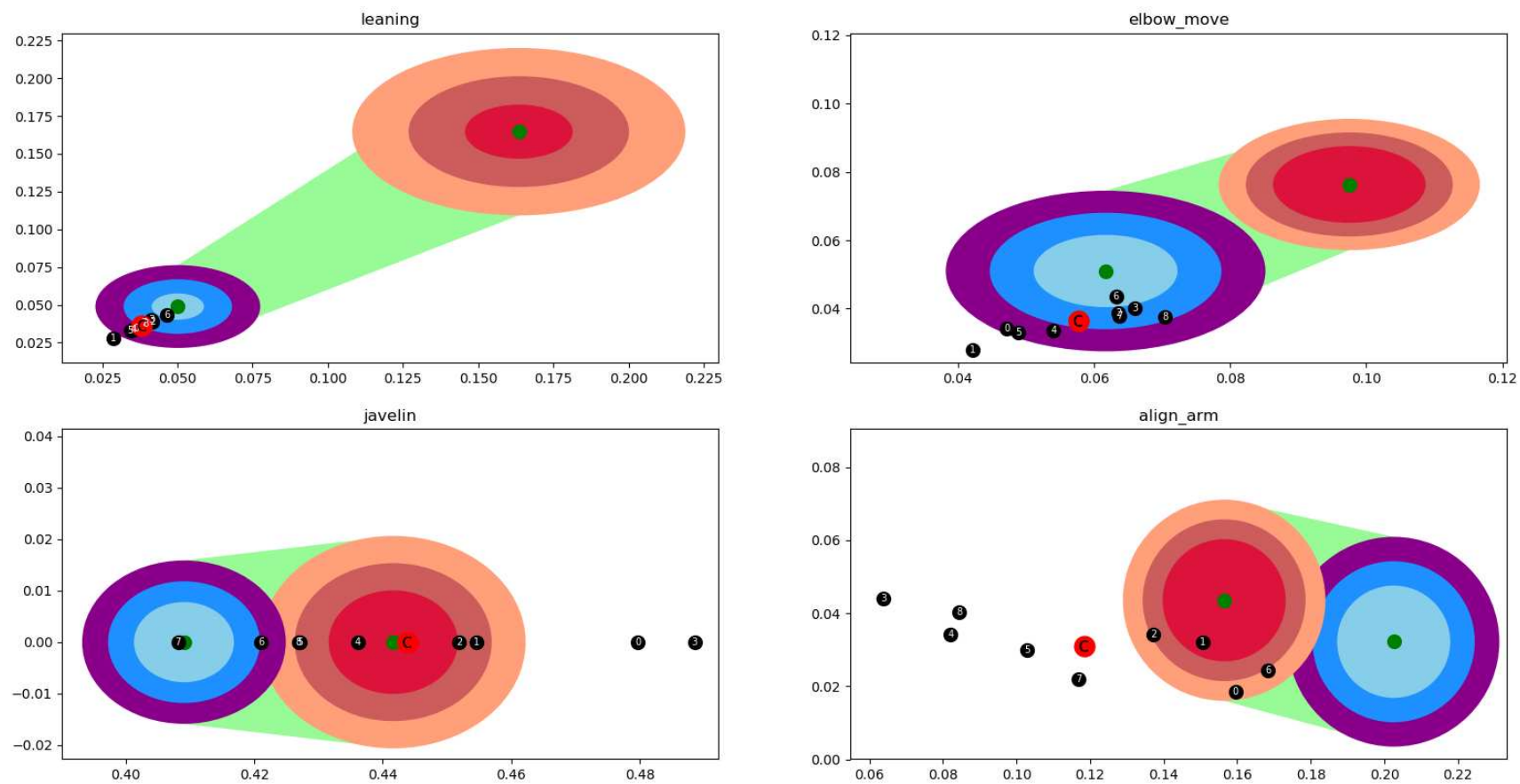


FIGURE 43 – Exemple de feedback visuel proposé à l’expert. Pour chaque défaut identifié (4 dans le cas présent), une visualisation des données de l’apprenant (points noirs) ainsi que du centroïde correspondant (points rouges) est affiché, afin de visualiser la distance de ces derniers par rapport aux données de l’expert pour un défaut identifié.

mouvement les plus importants, puis recherche un nombre prédéfini de minimums locaux les plus proches (à gauche et à droite). Il est possible de choisir le nombre de minimums locaux désirés, afin de garder une plus ou moins grande partie du signal original. Les limitations de cet algorithme se trouvent dans le postulat que le mouvement d'intérêt présente la vitesse la plus élevée.

Cette chaîne de traitement possède l'avantage d'être peu consommatrice de ressources dans le traitement des mouvements par rapport aux méthodes de la littérature, ce qui permet une utilisation en direct sur le terrain. En effet, le temps moyen de traitement d'un mouvement (allant du filtrage à l'extraction de données) est de cinq secondes. Cette étape de filtrage des données est cruciale, afin de pouvoir extraire des informations du mouvement qui seront cohérentes avec la réalité du geste.

Le système de retour à l'apprenant proposé dans cette partie permet d'obtenir une information visuelle de la performance de l'apprenant, par rapport à l'expert. À partir de données d'expert, contenant des bons mouvements et des mouvements présentant des défauts, et des données de l'apprenant, un affichage graphique permet de visualiser les différences entre le geste de l'expert et celui de l'apprenant. L'expert peut ensuite se baser sur ce retour afin d'ajuster ou non son jugement. Ce système permet d'analyser des modalités du gestes qui ne serait pas possible d'observer en temps normal, par exemple de par la nécessité d'observer trop de parties du corps à la fois, ou par l'occlusion de parties du corps résultant du mouvement lui-même.

Ce système a fait l'objet de plusieurs expérimentations, afin de vérifier son bon fonctionnement et sa capacité à fournir un retour à l'expert et à l'apprenant. Le prochain chapitre est consacré à ces expérimentations.



---

## EXPÉRIMENTATIONS

---

La chaîne de traitement précédemment présentée a pu être évaluée au travers de trois expérimentations : une première sur le lancer de balles dans une corbeille, une deuxième sur le *Bottle Flip Challenge*, et une sur le lancer de fléchettes. Chaque expérimentation a pour but de tester une partie du système, cherchant à répondre aux hypothèses précédemment formulées :

- **H1** : Il est possible de regrouper les gestes selon leurs propriétés cinématiques communes.
- **H2** : Il est possible de séparer les gestes des apprenants en deux groupes correspondant à une dichotomie geste réussi / geste raté afin de déterminer, pour une situation d'apprentissage donnée, les propriétés d'un ensemble fini de gestes réussis.
- **H3** : Il est possible de séparer les gestes en fonction de propriétés attendues et identifiées au préalable par l'expert.
- **H4** : Il est possible de corriger chaque défaut du geste de l'apprenant, en lui indiquant les défauts majeurs à corriger en premier. Un défaut majeur est identifié par la plus grande distance séparant le mouvement courant de l'apprenant, du groupe de gestes acceptables ayant éliminé ce défaut.
- **H5** : L'utilisation du système MLA basée sur l'hypothèse 4 permet d'améliorer l'apprentissage du geste par rapport à une situation sans le système MLA.
- **H6** : L'utilisation du système MLA en tant qu'assistant à l'enseignant permet d'améliorer l'apprentissage du geste par rapport à une situation sans MLA, et à une situation avec MLA et sans enseignant.

Ainsi, chaque expérimentation a pour objectif de répondre à une ou plusieurs de ces hypothèses :

- Lancer de balles : **H1, H2**
- Bottle Flip Challenge : **H2**
- Lancer de fléchettes : **H3, H4, H5, H6**

### 6.1 LANCER DE BALLES DANS UNE CORBEILLE

#### 6.1.1 Motivations et objectifs

Afin de vérifier si le système est capable de séparer les gestes selon un étiquetage vérifiable par un humain à l'aide des descripteurs et des algorithmes choisis, une première expérimentation a été réalisée. Les hypothèses correspondantes sont les suivantes :

- **H1** : Il est possible de regrouper les gestes selon leurs propriétés cinématiques communes.

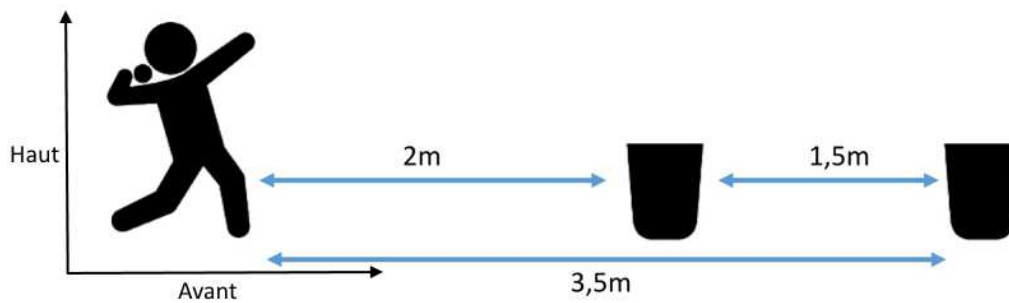


FIGURE 44 – Schéma de l'expérimentation du lancer de balle.

- **H2** : Il est possible de séparer les gestes des apprenants en deux groupes correspondant à une dichotomie geste réussi / geste raté afin de déterminer, pour une situation d'apprentissage donnée, les propriétés d'un ensemble fini de gestes réussis.

Dans cette expérimentation, les personnes doivent lancer une balle dans une corbeille parmi deux disposées devant eux en ligne : l'une est située à 2 mètres devant la personne, et l'autre est située à 3.5 mètres devant la personne (Fig. 44). Trois étiquetages différents ont été établis pour les lancers dans ce contexte :

- (i) le degré de réussite (balle dans la corbeille ou non)
- (ii) la corbeille visée
- (iii) le type de lancer (lancer par le haut ou par le bas)

Concernant les différents types de lancers, seuls deux types différents ont été réalisés par les personnes : le lancer « par le haut » et le lancer « par le bas ». Un geste de lancer « par le haut » est caractérisé par un mouvement de la main allant du haut vers le bas (trajectoire de la balle rectiligne ou légèrement courbée), alors qu'un geste de lancer « par le bas » est caractérisé par un mouvement de la main allant du bas vers le haut (trajectoire en cloche de la balle).

### 6.1.2 Protocole

Des instructions orales sont données quant à la tâche à effectuer : « Votre objectif est de lancer la balle dans une des deux corbeilles. Les cinquante premiers lancers devront viser la corbeille la plus proche, les cinquante autres la corbeille la plus éloignée. Vous êtes libre de lancer comme vous le souhaitez ». La personne retire ensuite tous les objets métalliques présents sur elle (ceinture, bracelets, téléphones, boucles d'oreilles, montres, etc.) afin d'éviter toute perturbation des capteurs (composés d'accéléromètres, de gyroscopes et de magnétomètres).

Le participant est ensuite mesuré selon les préconisations de NOITOM, afin de créer un squelette adapté selon douze mesures de distance suivantes :

- Sommet du crâne jusqu'à la base du lobe de l'oreille / lèvre supérieure
- Base du lobe de l'oreille / lèvre supérieure jusqu'à la vertèbre C7
- Vertèbre C7 jusqu'au haut du fémur
- Largeur des épaules (épaule complète)
- Épaules (centre de rotation) jusqu'au coude (centre de rotation)
- Coude (centre de rotation) jusqu'au poignet (centre de rotation)
- Poignet (centre de rotation) jusqu'au bout du majeur

- Largeur des hanches (du fémur gauche jusqu’au fémur droit)
- Fémur jusqu’au genou (centre de rotation)
- Genou (centre de rotation) jusqu’à la cheville
- Cheville jusqu’au sol
- Talon jusqu’au bout du gros orteil

Une fois le squelette créé, les capteurs sont posés sur la personne. Le sujet se familiarise ensuite avec l’équipement, notamment en se déplaçant, afin de vérifier que la combinaison n’entrave pas ses mouvements. Cinq lancers sont ensuite réalisés avant le début de l’expérimentation, afin de se préparer et de s’habituer à l’équipement.

La combinaison est utilisée dans son intégralité, à l’exception des capteurs spécifiques aux doigts. En effet, nous avons constaté que l’utilisation de ces capteurs génère beaucoup de bruit dans les données de mouvement capturées, et ils ne sont pas nécessaires pour l’analyse réalisée.

Les calibrations sont ensuite effectuées afin d’étalonner les capteurs (position dans l’espace, espacement des capteurs entre eux, mise à zéro sur le logiciel de capture, etc.) pour que le logiciel puisse adapter la combinaison par rapport aux articulations du squelette 3D.

À chaque personne, il a été demandé de réaliser 100 lancers au total. Les 50 premiers lancers devaient viser la première corbeille, et les 50 autres devaient viser la seconde.

Une fois les lancers réalisés, l’export de chaque lancer au format BVH a été effectué. L’intervention humaine se borne à désigner le dossier contenant tous les fichiers de mouvements bruts. En premier lieu, les filtres sont appliqués. L’extraction des descripteurs fait suite à cette étape. La durée du chargement des fichiers BVH, le filtrage des descripteurs, l’extraction des descripteurs utilisés dans l’expérimentation et leur sauvegarde était d’environ 2 secondes par fichier sur une machine équipée d’un processeur 8 coeurs Intel Core i7-4910MQ 2.90GHz, 32Go de mémoire RAM tournant sous Windows 7.

Afin de vérifier la qualité du clustering, plusieurs métriques ont été utilisées en suivant principalement deux objectifs :

- évaluation interne des clusters, en assignant un score de vraisemblance pour chaque cluster, vérifiant la forme des clusters, leur cohésion, leur variance, etc.
- évaluation externe des clusters, comparés les uns par rapport aux autres grâce à la distance les séparant, à l’aide d’une vérité terrain ou même d’une analyse experte (JAIN et DUBES 1988; MAULIK et BANDYOPADHYAY 2002)

Afin de vérifier la séparabilité des données, la métrique utilisée est celle de l’Average Silhouette Score (ASS) (P. J. ROUSSEUW 1987). Cette métrique se base sur le Silhouette Score (SS) de chaque point individuel de donnée.

Soit  $a(i)$  la distance moyenne d’un point  $i$  par rapport à tous les autres points contenus dans le même cluster, définie par

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j)$$

avec  $d(i, j)$  la distance entre les points  $i$  et  $j$  au sein du cluster  $C_i$ . Cette mesure peut être considérée comme une indication de bonne appartenance du point  $i$  au cluster  $C_i$  : plus cette valeur est basse, plus le point  $i$  est proche des autres points du cluster auquel il est assigné.



Soit  $b(i)$  la distance moyenne minimale du point  $i$  par rapport aux points des autres clusters (les distances moyennes étant calculées séparément pour chaque cluster ne contenant pas  $i$ ), définie par

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j)$$

avec  $k$  étant le nombre de clusters.

Le Silhouette Score d'un point est défini par

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \text{ if } |C_i| > 1 |C_i| > 1$$

et

$$s(i) = 0, \text{ si } |C_i| = 1$$

L'Average Silhouette Score est la moyenne des Silhouette Scores de chaque échantillon du problème considéré. Ce score permet de mesurer dans quelle mesure l'échantillon appartient effectivement au cluster qui lui a été assigné, par rapport aux autres clusters. Cette valeur permet d'obtenir une estimation de l'homogénéité des clusters : une valeur élevée indique que les clusters sont bien séparés, alors qu'une valeur faible indique que les clusters sont proches, voire se superposent. Cette valeur varie entre -1 et 1 : une valeur de 1 signifie que tous les échantillons d'un cluster sont proches les uns des autres (les clusters sont bien séparés), alors qu'une valeur de 0 signifie que les clusters sont superposés. Dans ce dernier cas, une explication possible réside dans le nombre trop élevé ou trop faible de clusters. Struyf *et al.* (STRUYF, M. HUBERT et P. ROUSSEEUW 1997) ont proposé quatre paliers pour les valeurs de l'ASS :

- entre 0 et 0.25, aucune structure n'est discernable au sein des données
- entre 0.25 et 0.50, il existe une structure, bien que mal définie, voire artificielle
- entre 0.50 et 0.70, une structure existe au sein des données (séparation correcte des clusters)
- au-delà de 0.70, une structure est clairement définie au sein des données (très bonne séparation des clusters)

Dans le contexte de l'expérimentation, cette métrique permet donc de vérifier si les données sont séparables selon les descripteurs choisis.

S'il y a une séparation acceptable, il est nécessaire de déterminer les propriétés menant à cette séparation. Une approche consiste à corréler cette séparation avec le degré de réussite du geste. Dans le cas présent, ce degré est binaire : soit le geste est réussi (la balle a atterri dans la corbeille désignée), soit le geste est raté (la balle n'a pas atterri dans la corbeille désignée). En outre, il est aussi possible de considérer le type de lancer effectué ou la corbeille visée. Dans tous les cas, il faut disposer *a priori* d'une vérité terrain, afin de pouvoir comparer le résultat du clustering à cet étiquetage. Dans ce contexte, plusieurs métriques ont été retenues : le rappel, la précision, le F1-score et l'Adjusted Rand Index (ARI). Dans un problème de classification, le rappel indique la proportion d'éléments corrects contenus au sein de tous les éléments retournés, alors que la précision indique la proportion d'éléments corrects retournés par rapport au nombre d'éléments corrects total. Le F1-score correspond à la moyenne harmonique de ces deux valeurs. La valeur du F1 score se situe entre 0 et 1, avec une valeur de 1 indiquant une correspondance parfaite entre la séparation obtenue et la vérité terrain. Cependant, ce score n'est qu'une moyenne du F-score de chaque caractéristique de l'échantillon considéré. Ainsi,

la valeur obtenue ne correspond pas au pouvoir discriminant des caractéristiques prises ensemble (information mutuelle) (Y.-W. CHEN et LIN 2006).

L'ARI est une mesure de similarité entre deux partitionnements de données différents (MOREY et AGRESTI 1984). Cette mesure se base sur le Rand Index (RI) (RAND 1971). Soit un ensemble de  $n$  données  $E = e_1, e_2, \dots, e_n$  et deux partitionnements de  $E$  à comparer, avec  $A = A_1, \dots, A_i$  un partitionnement de  $E$  en  $i$  partitions et  $B = B_1, \dots, B_j$  un partitionnement de  $E$  en  $j$  partitions. On définit

- $a$  comme étant le nombre de paires d'éléments dans  $E$  qui se trouvent dans la même partition dans  $A$  et dans  $B$
- $b$  comme étant le nombre de paires d'éléments dans  $E$  qui se trouvent dans des partitions différentes dans  $A$  et dans  $B$
- $c$  comme étant le nombre de paires d'éléments dans  $E$  qui se trouvent dans la même partition dans  $A$  et dans des partitions différentes dans  $B$
- $d$  comme étant le nombre de paires d'éléments dans  $E$  qui se trouvent dans des partitions différentes dans  $A$  et dans la même partition dans  $B$

Le Rand Index est défini par

$$RI = \frac{a + b}{a + b + c + d}$$

Le dénominateur correspond au nombre total de paires d'éléments présents dans  $E$ . L'Adjusted Rand Index est une version normalisée du Rand Index (L. HUBERT et ARABIE 1985). Si on considère les mêmes éléments qu'auparavant (un ensemble de données  $E$ , et deux partitionnements  $A$  et  $B$ ), on peut créer un tableau de contingence comme suit :

$A \setminus B$	$B_1$	$B_2$	$\dots$	$B_j$	Sommes
$A_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1j}$	$a_1$
$A_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2j}$	$a_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$A_i$	$n_{i1}$	$n_{i2}$	$\dots$	$n_{ij}$	$a_j$
Sommes	$b_1$	$b_2$	$\dots$	$b_j$	

Chaque case  $n_{ij}$  indique le nombre d'éléments communs entre  $A_i$  et  $B_j$ . L'Adjusted Rand Index est défini par la formule

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}$$

La valeur maximale est 1, et indique une correspondance parfaite entre les deux partitionnements de données. Une valeur de 0 correspond à une affectation aléatoire des données, et il est possible d'obtenir des valeurs négatives dans certains cas, dû à la normalisation (MEILĂ 2007).

Dans le contexte de l'expérimentation, ces métriques permettent de vérifier que le clustering obtenu correspond ou non à une séparation selon un étiquetage binaire.

### 6.1.3 Données capturées

Pour chaque lancer, le squelette de l'apprenant est capturé. Les mains sont représentées à l'aide d'un seul capteur. La réussite ou non du geste (la balle atterrit dans la corbeille désignée) est notée, ainsi que la corbeille visée et le type de lancer effectué.

#### 6.1.4 Données et métriques utilisées

Pour chaque personne, 100 mouvements ont été enregistrés. Ces données sont enregistrées dans un format brut propriétaire de NOITOM (format RAW), puis exportées au format BVH (format ouvert d'animation de personnages, voir Chapitre 5, section 5.1.1). Plusieurs descripteurs ont été extraits des mouvements, ayant chacun une valeur pour le début, la fin et l'indice de la valeur maximale de la vitesse moyenne de la main dominante :

- la norme du vecteur vitesse (SN),
- la norme du vecteur vitesse en X, Y, Z ( $S_{x/y/z}$ ),
- la direction du vecteur vitesse en X, Y, Z ( $Sd_{x/y/z}$ ),
- la norme du vecteur vitesse ainsi que la direction de ce vecteur en X, Y, Z ( $SND_{xyz}$ ).

L'utilisation des descripteurs liés à la vitesse se base sur l'hypothèse que la vitesse à imprimer à la balle afin de correctement viser la corbeille désignée serait différent selon la corbeille choisie. De plus, viser la corbeille la plus éloignée nécessite une force plus élevée, comparativement à la corbeille la plus proche. Enfin, les principales différences entre les types de lancers observés se situent au niveau de la direction de la balle, sa trajectoire ainsi que la force appliquée.

Plusieurs articulations ont été utilisées (alignée avec la préférence manuelle de la personne, gaucher ou droitier), correspondant aux parties du corps sollicitées lors du mouvement :

- la main,
- l'avant-bras,
- le bras,
- l'épaule.

L'utilisation de plusieurs articulations permet de vérifier s'il existe d'une part une ou des articulations plus importantes pour la lancer, et d'autre part si la combinaison de descripteurs vitesses issus de plusieurs articulations permet d'augmenter la précision du clustering.

#### 6.1.5 Résultats

Les résultats obtenus sont ceux d'une seule personne (100 lancers). L'utilisation de données de mouvement provenant d'une seule personne permet de limiter la variabilité des gestes de lancers : une fois un geste permettant de correctement viser la corbeille trouvé, les lancers suivants essayeront de reproduire les propriétés de ce lancer. Ainsi, on peut s'attendre à des données rapprochées (au sens des caractéristiques extraites) pour les lancers concernant une même corbeille. Le tableau 5 présente les métriques calculées sur les clustering obtenus. Dans le cas des métriques de comparaison à une vérité terrain, les résultats sont donnés pour un clustering avec nombre de clusters  $k = 2$ . Les résultats pour la métrique permettant de vérifier la séparation des clusters sont également donnés pour  $k = 2$ , car c'est la valeur de  $k$  qui donne les meilleurs résultats (quand la valeur  $k = 2$  ne donne pas le meilleur résultat pour l'ASS, les variations de cette valeur sont de l'ordre du centième, quelle que soit la valeur de  $k$ ).

Joints	H					H, FA					H, FA, A				
Metric	ASS	P	R	F1	ARI	ASS	P	R	F1	ARI	ASS	P	R	F1	ARI
	All data (ground truth = success / fail)														
SN	0.59	0.51	0.89	0.64	0.01	0.60	0.50	0.86	0.63	0.00	0.60	0.50	0.86	0.63	0.00
Sxyz	0.56	0.48	0.89	0.62	-0.01	0.54	0.48	0.89	0.62	-0.01	0.54	0.48	0.89	0.62	-0.01
Sdxyz	0.23	0.41	0.25	0.31	-0.01	0.23	0.54	0.77	0.64	0.04	0.24	0.55	0.77	0.64	0.05
SNDxyz	0.26	0.53	0.89	0.67	0.04	0.27	0.55	0.77	0.64	0.05	0.26	0.53	0.77	0.63	0.03
	All data (ground truth = closest / farthest bin)														
SN	0.59	0.65	1.00	0.79	0.21	0.60	0.64	0.98	0.78	0.19	0.60	0.64	0.98	0.78	0.19
Sxyz	0.56	0.54	0.88	0.67	0.01	0.54	0.54	0.88	0.67	0.01	0.54	0.54	0.88	0.67	0.01
Sdxyz	0.23	0.65	0.34	0.45	0.02	0.23	0.68	0.86	0.76	0.20	0.24	0.69	0.86	0.77	0.22
SNDxyz	0.26	0.68	0.98	0.80	0.26	0.27	0.71	0.88	0.79	0.26	0.26	0.69	0.88	0.77	0.22
	All data (ground truth = throwing type)														
SN	0.59	0.87	0.95	0.91	0.83	0.60	0.83	0.95	0.89	0.79	0.60	0.83	0.95	0.89	0.79
Sxyz	0.56	0.06	0.05	0.05	-0.09	0.54	0.06	0.05	0.05	-0.09	0.54	0.06	0.05	0.05	-0.09
Sdxyz	0.23	0.08	0.10	0.09	-0.07	0.23	0.54	0.95	0.69	0.39	0.24	0.53	0.95	0.68	0.37
SNDxyz	0.26	0.74	0.95	0.83	0.68	0.27	0.55	1.00	0.71	0.42	0.26	0.56	0.95	0.70	0.42

TABLE 5 – Métriques calculées sur les résultats du clustering pour le lancer de balle.

L'ajout d'articulations n'améliore pas (ou de manière marginale) la séparation des clusters dans tous les cas (valeurs d'ASS et d'ARI stables ou diminuant lors d'ajout d'autres articulations). Cela montre que la main reste l'articulation la plus discriminante pour la séparation des gestes. Dans le cas de la réussite ou non du geste, bien qu'une séparation acceptable soit possible à l'aide de la norme du vecteur vitesse ( $ASS \approx 0.60$  pour les meilleurs cas), ainsi que les vecteurs vitesses en  $x$ ,  $y$  et  $z$ , cette séparation ne correspondait pas au degré de réussite du geste ( $ARI \approx 0$ ).

Pour la corbeille ciblée, les mêmes tendances se dégagent pour la séparation des clusters. Cette fois, l'ARI obtient des valeurs  $\approx 0.25$ , suggérant une légère correspondance des regroupements avec l'étiquetage. Cependant, les valeurs d'ASS correspondantes sont aux alentours de 0.25, suggérant une mauvaise séparation des clusters.

Il est intéressant de noter qu'il a été possible d'obtenir à la fois une valeur correcte pour l'ASS (0.59) et une valeur élevée pour l'ARI (0.83) pour la séparation par rapport au type de lancer, avec le descripteur de la norme du vecteur vitesse. Cette corrélation entre une bonne valeur pour l'ASS et une forte valeur pour l'ARI suggère qu'il est possible, à l'aide des descripteurs et des méthodes utilisées, de séparer les gestes selon un critère observable par un humain. Ainsi, l'hypothèse **H1** a été validée dans le contexte de cette expérimentation. Cependant, cette séparation n'est pas corrélée avec le degré de réussite du geste, et l'hypothèse **H2** n'a pas pu être validée au cours de cette expérimentation. Une autre expérimentation a été conduite, afin de se focaliser sur la séparation du geste en fonction du résultat obtenu.

## 6.2 BOTTLE FLIP CHALLENGE

Le Bottle Flip Challenge est un défi qui provient de la performance d'un étudiant lors d'un concours de talents américain, au sein du lycée Ardrey Kell (McDERMOTT 2016). L'étudiant de 18 ans, Mike Senatore, a offert au public une performance inattendue : un lancer de bouteille, partiellement remplie, qui réalise un salto arrière, avant de retomber parfaitement droite, sur la table (Fig. 45). Filmé à l'aide d'un smartphone, la vidéo se propage rapidement sur internet, et de nombreuses personnes s'y essayent : le *Bottle Flip Challenge* était né.

Ce mouvement possède l'avantage de ne nécessiter que peu de matériel : une bouteille d'eau partiellement remplie et une table. La durée du mouvement est également très courte, ce qui permet d'enregistrer beaucoup de données en un temps réduit. De plus, le geste ne nécessite pas de connaissance *a priori*, ce qui fait qu'il n'est pas nécessaire d'avoir à disposition des personnes ayant déjà réalisé ce geste.

Le mouvement de la bouteille, ainsi que de l'eau au sein de la bouteille, ont fait l'objet d'une étude récente (DEKKER *et al.* 2018). Il en ressort que les phénomènes physiques les régissant sont complexes et nombreux : dynamique des fluides, mouvement du projectile, moment cinétique, force centripète et gravité. Il ressort de cette étude scientifique que le lancer optimal présente un centre d'inertie bas, une vitesse angulaire faible et que le taux de remplissage optimal pour la bouteille se situe entre 20 et 40 % de sa capacité maximale. Une autre étude (UKESSAYS 2017) montre qu'une bonne manière de lancer pour le Bottle Flip Challenge est de ne donner qu'une faible impulsion vers l'avant, et de lâcher la bouteille lorsqu'elle se trouve perpendiculaire au sol.

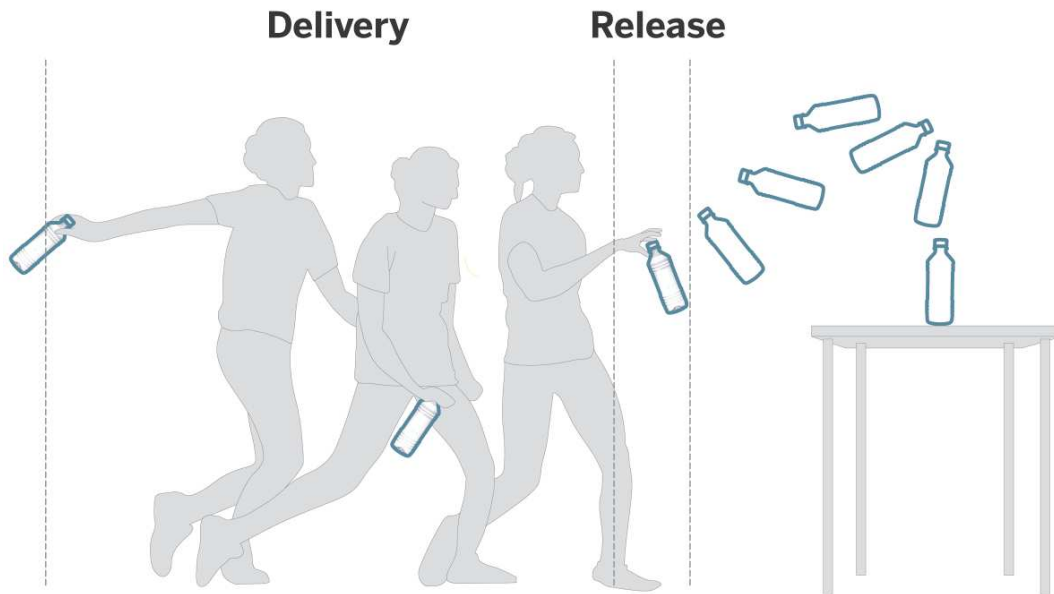


FIGURE 45 – Exemple du mouvement du Bottle Flip Challenge <sup>1</sup>.

#### 6.2.1 Motivations et objectifs

L'objectif de cette expérimentation était de déterminer s'il était possible d'obtenir une séparation des données en deux clusters, correspondant au degré de réussite du geste. Pour que le geste soit considéré comme réussi, il faut que la bouteille soit debout, sur la table, après avoir réalisé une rotation. L'hypothèse correspondante est la suivante :

- **H2** : Il est possible de séparer les gestes des apprenants en deux groupes correspondant à une dichotomie geste réussi / geste raté afin de déterminer, pour une situation d'apprentissage donnée, les propriétés d'un ensemble fini de gestes réussis.

#### 6.2.2 Protocole

La personne participante doit remplir un pré-questionnaire, contenant des informations sur le nom, l'âge, la taille, les connaissances par rapport au Bottle Flip Challenge. Des instructions orales sont données : « Le but de cette expérience est de lancer une bouteille, de manière à ce qu'elle fasse un tour sur son axe horizontal avant de retomber correctement sur la table. Vous êtes libre sur la position de votre corps et le geste de lancer. Cependant, voici le geste communément utilisé pour réussir le lancer [faire une démonstration]. Il vous est demandé de vous mettre droit, les pieds avant la marque au sol, les bras le long du corps, la bouteille tenue dans votre main. Lorsque le premier bip retentira, vous lancerez la bouteille. Une fois le lancer terminé, vous remettrez les bras le long du corps, et attendrez que le deuxième bip retentisse avant de la reprendre, puis de vous remettre en position. Je vous signalerai lorsque vous aurez réalisé suffisamment de lancers. »

La personne retire ensuite tous les objets métalliques présents sur elle (ceinture, bracelets, téléphones, boucles d'oreilles, montres, etc.).

1. <https://wadsworthbruin.com/2016/12/01/bottle-flipping-into-the-heart-of-wadsworth-high/>

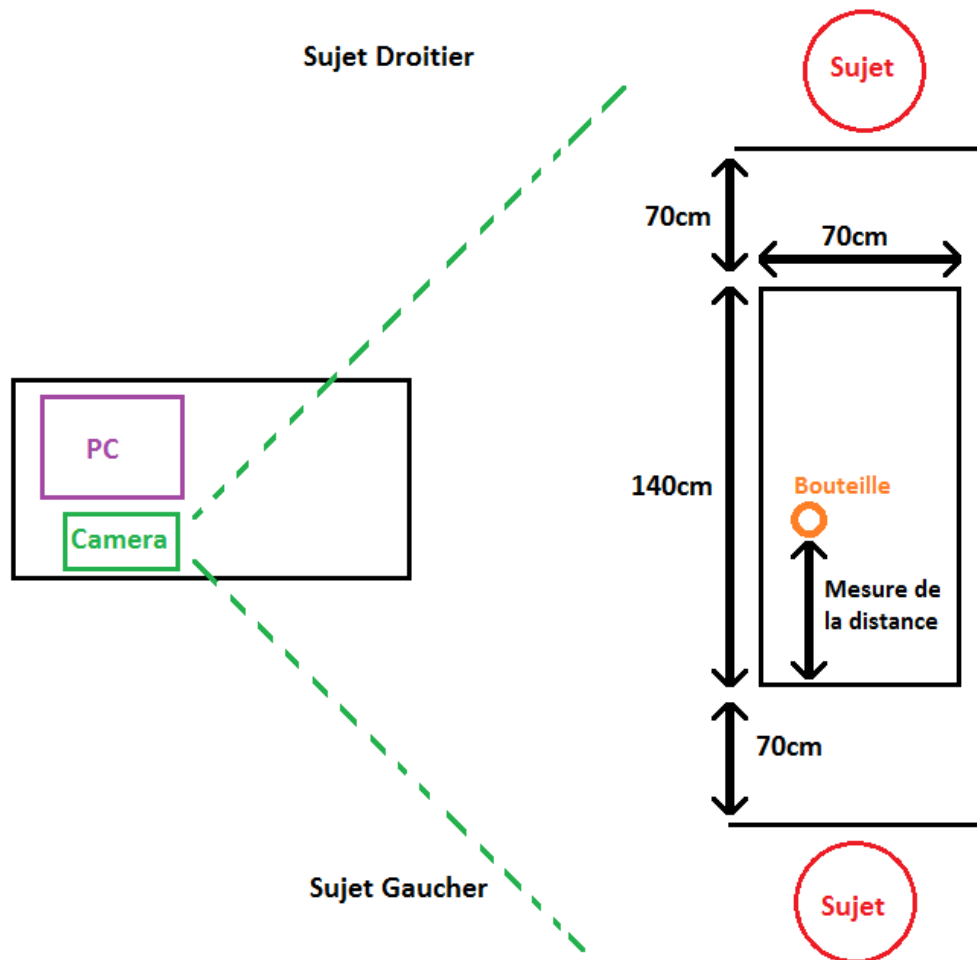


FIGURE 46 – L’installation du matériel lors de l’expérimentation du Bottle Flip Challenge (vue du dessus).

La personne est mesurée selon les préconisations de NOITOM (NOITOM 2015a), pour qu’un squelette adapté soit créé. Une fois le squelette créé, les capteurs sont posés sur la personne. Le sujet se familiarise ensuite avec l’équipement, notamment en se déplaçant, afin de vérifier que la combinaison n’entrave pas le mouvement. Cinq lancers sont ensuite réalisés avant le début de l’expérimentation, pour se préparer, s’habituer à l’équipement et vérifier que le gant de la combinaison ne gêne pas la prise de la bouteille. L’installation complète est visible sur la figure 46.

Une fois les données de mouvement capturées, le même processus que pour la précédente expérimentation est utilisé afin d’importer, de filtrer, d’extraire les descripteurs et de sauvegarder les données ainsi extraites.

Les métriques utilisées pour analyser les résultats sont les mêmes que pour la précédente expérimentation.

### 6.2.3 Données capturées

Pour chaque lancer, le squelette de l’apprenant est capturé, y compris le mouvement des doigts. La réussite ou non du geste (bouteille qui atterrit correctement ou non après le salto arrière) est notée, et pour les gestes réussis, la distance de la bouteille par rapport au bord de la table où se situe la personne est mesurée.



#### 6.2.4 Données et métriques utilisées

13 personnes au total ont participé à l'expérimentation. Pour chacune d'entre elles, 100 mouvements ont été enregistrés. Ces données sont enregistrées au format RAW de NOITOM, puis exportées dans le format BVH. Les mêmes descripteurs que précédemment ont été extraits, et les mêmes combinaisons d'articulations ont été utilisées.

L'hypothèse est que la vitesse de la bouteille est l'indicateur prédominant dans la réussite du geste : une vitesse trop faible ne permettrait pas à la bouteille de faire une rotation complète, et une vitesse trop élevée induirait plusieurs rotations, ainsi qu'un atterrissage non maîtrisé. La direction est utilisée afin de vérifier l'hypothèse que la direction du lancer influait également sur la réussite du geste.

#### 6.2.5 Résultats

Dans le cas des métriques de comparaison à une vérité terrain, les résultats sont donnés pour un clustering avec nombre de clusters  $k = 2$ . Les résultats pour la métrique permettant de vérifier la séparation des clusters sont également donnés pour  $k = 2$ , car c'est la valeur de  $k$  qui donne les meilleurs résultats (quand la valeur  $k = 2$  ne donne pas le meilleur résultat pour l'ASS, les variations de cette valeur sont de l'ordre du centième, quelle que soit la valeur de  $k$ ). Lors de cette expérimentation, la combinaison de capteurs présentait beaucoup plus d'artefacts sur les articulations allant de l'épaule à la main sur le côté gauche. Les données des personnes gauchères présentaient donc plus de bruits que les données des personnes droitères. Ainsi, il a été décidé de faire les tests avec :

- (i) les données mixtes (1300 données),
- (ii) les gauchers seuls (200 données),
- (iii) les droitiers seuls (1100 données).

Les résultats sont présentés dans le tableau 6.

Joints	H					H,FA					H, FA, A				
Metric	ASS	P	R	F1	ARI	ASS	P	R	F1	ARI	ASS	P	R	F1	ARI
	Left and Right-Handed														
SN	0.48	0.25	0.33	0.29	0.04	0.44	0.25	0.33	0.29	0.04	0.43	0.25	0.33	0.29	0.04
Sxyz	0.54	0.18	0.67	0.30	0.05	0.52	0.27	0.32	0.29	0.05	0.51	0.18	0.68	0.29	0.05
Sdxyz	0.24	0.21	0.53	0.30	0.00	0.27	0.25	0.25	0.25	0.04	0.22	0.18	0.72	0.27	0.04
SNDxyz	0.21	0.18	0.47	0.30	0.00	0.27	0.25	0.26	0.26	0.04	0.22	0.26	0.28	0.27	0.04
	Left Handed														
SN	0.41	0.39	0.39	0.39	0.02	0.42	0.38	0.39	0.39	0.01	0.41	0.31	0.61	0.39	0.01
Sxyz	0.35	0.32	0.57	0.39	0.00	0.34	0.32	0.57	0.39	0.00	0.33	0.35	0.43	0.39	0.00
Sdxyz	0.31	0.34	0.48	0.40	0.00	0.27	0.34	0.54	0.39	0.00	0.23	0.34	0.48	0.40	0.00
SNDxyz	0.27	0.34	0.49	0.40	0.00	0.25	0.33	0.48	0.39	0.00	0.22	0.34	0.52	0.41	0.00
	Right Handed														
SN	0.42	0.18	0.29	0.22	0.00	0.36	0.17	0.28	0.21	0.00	0.34	0.17	0.28	0.21	0.00
Sxyz	0.73	0.19	0.12	0.15	0.01	0.71	0.19	0.12	0.15	0.01	0.71	0.19	0.12	0.15	0.01
Sdxyz	0.28	0.15	0.45	0.28	0.00	0.20	0.16	0.49	0.27	0.00	0.26	0.19	0.13	0.15	0.01
SNDxyz	0.26	0.16	0.45	0.28	0.00	0.19	0.19	0.52	0.27	0.00	0.26	0.17	0.87	0.15	0.01

TABLE 6 – Métriques calculées sur les résultats du clustering pour le Bottle Flip Challenge.

	Beginning	Maximum	End
X (Side)	0.0398	0.5071	0.0110
Y (Upward)	0.0415	1.7497	0.0998
Z (Forward)	0.0847	2.0477	0.0536

TABLE 7 – Distance relative entre les centroïdes des clusters, pour la main droite, avec les vecteurs vitesse en X, Y, Z pris au début, au maximum et à la fin du lancer, pour  $k = 2$ .

Dans tous les cas, l'ajout d'autres articulations que celle de la main ne fait que rajouter du bruit (les valeurs de l'ASS et l'ARI décroissent). Les plus fortes valeurs d'ASS (séparation des clusters) ont été obtenues pour les descripteurs vecteur vitesse en X, Y, Z (ASS des descripteurs Sxyz = 0.73). On peut également remarquer que les résultats de l'ASS sont meilleurs sur les données des droitiers seuls pour ces descripteurs, par opposition aux données mixtes ou des gauchers seulement (dernière partie du tableau 6). Cette tendance tend à confirmer les problèmes de capteurs du côté gauche de la combinaison. Le tableau 7 montre que les valeurs les plus discriminantes pour la séparation sont le vecteur vitesse en Z (en avant) et en Y (en haut) au moment de la vitesse maximale. Cette séparation est cohérente avec les études réalisées par (DEKKER *et al.* 2018) et (UKESSAYS 2017) qui suggèrent que les différences entre un lancer réussi et raté se situent au niveau de la force et de la direction du lancer.

Dans tous les cas, les valeurs de l'ARI (correspondance avec la vérité terrain) sont restées proches de 0, ce qui indique que les séparations obtenues ne correspondaient pas au degré de réussite du geste. Ainsi, l'hypothèse **H2** n'a pas été validée au cours de cette expérimentation.

Les descripteurs utilisés conjointement avec les algorithmes de clustering ne peuvent donc pas proposer un regroupement corrélé avec le degré de réussite du geste, dans le cas du *Bottle Flip Challenge*. Le peu de variabilité du geste d'un lancer à un autre (et, par extension, d'un lancer réussi à un lancer raté) est une piste pour expliquer que l'algorithme des *k-means* appliqué non récursivement n'est pas suffisant pour obtenir une telle séparation. De plus, les descripteurs utilisés dans cette expérimentation sont tous basés sur la vitesse. Ce type d'indicateurs semble ne pas être suffisant pour caractériser la réussite ou non du geste. Plusieurs autres descripteurs auraient pu être utilisés, tels que la courbure (indication sur la vitesse de changement d'une courbe au cours du temps), la vitesse angulaire des articulations sollicitées (poignet et main), ainsi que le déplacement du centre de masse de la personne, *etc.* La complexité des phénomènes physiques régissant le lancer (autant pour le mouvement du bras et de la main que celui de l'eau dans la bouteille) peuvent également indiquer que les descripteurs basés vitesses ne sont pas suffisant pour caractériser tous les aspects du geste important pour un lancer réussi.

Plusieurs problèmes ont également émané de l'expérimentation elle-même. Premièrement, la distance entre la table et la personne lançant n'était pas systématiquement constante. En effet, certaines personnes prenaient un léger élan, ou se reculaient légèrement avant de lancer par rapport à la position limite imposée (voir figure 46). De plus, il n'était pas possible de mesurer le point d'impact exact de la bouteille sur la table, car cette dernière glissait lors d'un lancer réussi. Cette distance n'a ainsi pas été considérée pour notre analyse. Bien que corrigée à l'aide de pré-traitements, l'imprécision des données de capture récoltées (cf. chapitre 5 section 5.4) est également un facteur à prendre en compte dans les résultats obtenus.

### 6.2.6 Discussion

Ces deux premières expérimentations ont permis de vérifier la capacité de la plateforme MLA à traiter de manière automatique les données, une fois paramétrée selon les besoins d'analyse. Ces expérimentations ont également permis de vérifier s'il était possible d'obtenir automatiquement une séparation entre les gestes jugés réussis et ceux jugés ratés, pour un cas précis. Les descripteurs choisis, ainsi que la méthode utilisée, ont permis d'aboutir à une séparation des données acceptable dans les deux expérimentations.

Cependant, pour le *Bottle Flip Challenge*, la séparation obtenue ne correspond pas à une dichotomie geste réussi/raté. Le manque d'analyse experte est un problème qui est apparu au cours de l'expérimentation. Bien que le *Bottle Flip Challenge* ait fait l'objet d'articles scientifiques sur la physique régissant le lancer et la bouteille (DEKKER *et al.* 2018 ; UKESSAYS 2017), il ne semble pas y avoir d'expertise formalisée du geste au moment où ces lignes sont écrites. En ce qui concerne l'étude du geste, la rotation du poignet aurait pu être une piste intéressante à explorer. Un sous-ensemble réduit de descripteurs a été utilisé pour cette expérimentation, ce qui a limité les résultats obtenus. Le calcul d'autres descripteurs (tels que ceux présentés dans (LARBOULETTE et GIBET 2015) par exemple) aurait pu mener à de meilleurs résultats.

L'analyse du geste, dans le cas de ces deux expérimentations, se fait de manière binaire : soit le geste est acceptable, soit il ne l'est pas. Dans une situation d'apprentissage, il n'est pas toujours possible de regrouper ces mouvements dans ces deux catégories. Le degré de succès peut avoir une granularité plus fine, ou dépendre de l'acceptabilité binaire (ou non) de plusieurs autres propriétés du geste.

De plus, dans un tel processus, l'humain est écarté du processus d'analyse. Son expertise n'est utilisée qu'*a priori*, et son intervention n'est pas requise dans le processus d'évaluation et *in fine* de conseil pour l'apprentissage. Bien qu'un tel système puisse présenter des avantages, l'objectif est de proposer une aide à la décision pour l'enseignant/l'expert, et non pas un remplacement de ce dernier. Le système de retours de MLA, basé sur une analyse et une évaluation par regroupement des besoins d'observation du geste, permet de fournir un retour graphique à l'expert ou à l'apprenant. Ce système est l'objet de l'expérimentation suivante.

## 6.3 EXPÉRIMENTATION : LANCER DE FLÉCHETTES

Cette section présente l'évaluation de la partie de retours à l'apprenant du système MLA au travers d'une expérimentation portant sur le lancer de fléchettes. Il s'agit d'un sport où la technicité du geste est formalisée<sup>2</sup>. Il existe, pour chaque partie du corps impliquée, des explications précises quant à la position et au mouvement. Après une analyse de ces explications, nous proposons une représentation du geste sous la forme d'un ensemble de descripteurs représentant chaque besoin d'observation. Les valeurs d'acceptabilité de ces descripteurs sont ensuite estimées à l'aide d'un clustering basé sur les démonstrations de l'expert. Enfin, ces valeurs sont comparées à celles de l'apprenant afin de lui fournir un retour.

2. <https://www.dartslive.com/enjoy/en/throw/>

### 6.3.1 Motivations et objectifs

Les objectifs de cette expérimentation sont multiples. Il s'agit d'une part de montrer qu'il est possible, à l'aide du système de retours à l'apprenant basé sur les démonstrations de l'expert et l'estimation de valeurs d'acceptabilité empiriques des propriétés du geste, de donner des conseils pertinents lors de la réalisation de ces derniers. D'autre part, il s'agit également de montrer qu'une amélioration du geste, en termes de précision du tir et de la « forme » du mouvement résultait de ces conseils. Enfin, l'efficacité du système selon les modalités précédemment citées est évaluée en termes d'usage : utilisation du système seul, ou utilisation conjointe avec l'analyse de l'expert.

Les hypothèses correspondantes sont les suivantes :

- **H3** : Il est possible de séparer les gestes en fonction de propriétés attendues et identifiées au préalable par l'expert.
- **H4** : Il est possible de corriger chaque défaut du geste de l'apprenant, en lui indiquant les défauts majeurs à corriger en premier. Un défaut majeur est identifié par la plus grande distance séparant le mouvement courant de l'apprenant, du groupe de gestes acceptables ayant éliminé ce défaut.
- **H5** : L'utilisation du système MLA basée sur l'hypothèse 4 permet d'améliorer l'apprentissage du geste par rapport à une situation sans le système MLA.
- **H6** : L'utilisation du système MLA en tant qu'assistant à l'enseignant permet d'améliorer l'apprentissage du geste par rapport à une situation sans MLA, et à une situation avec MLA et sans enseignant.

En conséquence, l'objectif pour les personnes réalisant l'expérimentation est de viser le centre de la cible, et également de suivre les conseils donnés par l'expert utilisant le système ou non.

Afin d'évaluer les résultats obtenus, trois groupes ont été formés : un groupe " contrôle ", où seul l'expert donne des conseils (système absent), un groupe où seul le système donne les conseils (évaluation experte absente), et un groupe où l'expert peut s'appuyer sur le système afin de décider des conseils à donner.

### 6.3.2 Fléchettes : principes et description

#### 6.3.2.1 Matériel

Le lancer de fléchettes est un sport qui se pratique à l'aide d'une cible (Figures 47 et 48) et de fléchettes (Fig. 49). La cible est constituée de vingt zones distinctes et de tailles égales, numérotées de 1 à 20, ainsi que d'un cercle au centre (Fig. 47). Chacune des zones est également divisée (Fig. 48) :

- les simples sont les deux plus grandes parties d'une zone,
- les doubles se situent sur la couronne extérieure des zones,
- les triples se situent sur la couronne intérieure des zones.

Le double centre compte comme un " double " du centre simple ( $2 \times 25$ ).

La *World Darts Federation* définit des critères à respecter pour l'installation et les parties de fléchettes. Ainsi, la cible doit respecter les dimensions suivantes :

- largeur intérieure des doubles et des triples : 8 mm
- diamètre du centre (bulle) : 12,7 mm
- diamètre du demi-centre (demi-bulle) : 31,8 mm

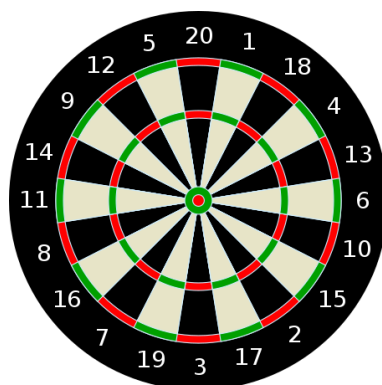


FIGURE 47 – Une cible de fléchettes classique<sup>1</sup>.

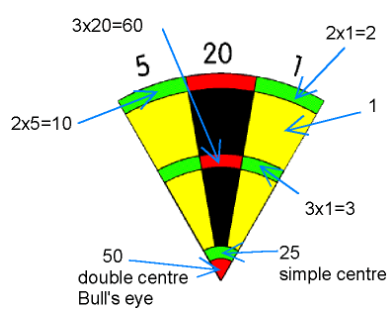


FIGURE 48 – Détail des zones d'une cible de fléchettes classique<sup>1</sup>.



FIGURE 49 – Exemple de fléchette respectant les normes officielles<sup>1</sup>.

- rayon du cercle extérieur de la couronne des doubles : 170 mm
- rayon du cercle extérieur de la couronne des triples : 107,4 mm
- diamètre total de la cible : 451 mm
- épaisseur des fils : minimum 1,6 mm - maximum 1,8 mm

Le centre de la cible doit se trouver à 172,72 centimètres du sol, et la limite de tir est située à 2,37 mètres de la cible.

Les fléchettes doivent avoir une largeur inférieure à 3,38 centimètres et peuvent être de n'importe quelle longueur, tant que la distance allant de pointe au bout de l'empennage ne dépasse pas 20 centimètres. La masse d'une fléchette ne doit pas dépasser 50 grammes.

Les règles des différents jeux existants ne seront pas détaillées ici, car elles n'ont pas été utilisées dans le contexte de l'expérimentation. L'installation réalisée pour cette expérimentation respecte les dimensions et les distances imposées par la *World Darts Federation*.

#### 6.3.2.2 Caractéristiques du mouvement et descripteurs considérés

Des entretiens ont été réalisés avec un expert des fléchettes : Aurélien Orsini, président du *Dart dart club de fléchettes mayennais* et champion régional en 2014. Ces entretiens ont pour but de prendre connaissance des modalités d'évaluation du geste de l'apprenant par l'expert lors d'une séance d'apprentissage, ainsi que des conseils prodigués pour améliorer le geste. Cette étape a permis de vérifier que les parties du corps et les modalités du mouvement regardées par l'expert sont formalisables en descripteurs opérationnalisables et en conséquence calculables à partir du mouvement enregistré. De ces entretiens sont ressortis quatre défauts majeurs souvent présents chez les débutants.

Le premier défaut est lié au fait de se pencher en avant (*leaning*) durant le lancer. Lorsqu'un débutant lance une fléchette, son corps accompagne naturellement le geste en se penchant en avant (Fig. 50). Ce défaut peut entraîner une visée plus basse que prévu, car le mouvement du corps imprime une rotation vers le bas au bras, par extension. Le conseil donné par l'expert est de rester fixe sur ses appuis, une fois en position de tir, tout en gardant le haut du corps (des hanches jusqu'à la tête) le plus immobile possible. Les descripteurs utilisés pour caractériser cet aspect du geste est la vitesse moyenne des épaules au cours du lancer : plus elle est basse, moins le haut du corps a bougé.

Le deuxième défaut identifié est celui du mouvement du coude (*elbow move*). Lors du lancer de fléchettes, le corps, ainsi que le bras, doivent rester le plus statique possible. Le mouvement de lancer doit s'effectuer à l'aide d'une rotation du coude, ce qui n'est possible que lorsque le bras est placé suffisamment en position de pré-lancer. Le problème usuel du geste d'un novice est que le bras et l'avant-bras forment un « V » avant le lancer, et finissent par former une ligne après le lancer, ce qui implique un déplacement du coude au cours du lancer. Le mouvement est ainsi moins maîtrisé que si le coude est la seule articulation qui pivote (Fig. 51). Le conseil donné par l'expert, dans ce cas, est de se concentrer sur l'immobilité du bras, afin que le coude effectue un pivot et que seul l'avant-bras ne bouge. Cet aspect du geste est caractérisé par les descripteurs de vitesse moyenne du coude, ainsi que de l'épaule (du côté de la préférence manuelle de l'utilisateur). Plus ces valeurs sont basses, moins le bras a bougé.

---

2. source : <https://fr.wikipedia.org/wiki/Fl%C3%A9chettes>





FIGURE 50 – Défaut du « leaning » illustré : l'apprenant se penche en avant lors du lancer.



FIGURE 51 – Défaut du « elbow move » illustré : le coude de l'apprenant effectue un déplacement lors du lancer, au lieu de seulement effectuer une rotation.



FIGURE 52 – Défaut du « javelin » illustré : la main de l'apprenant passe à côté (voire derrière) la tête lors de la prise d'élan du lancer.

Un autre défaut possible est celui du " lancer de javelot " (*javelin*). Lors d'un lancer, le bras, l'avant-bras et la main doivent rester devant le corps, dans l'alignement de l'axe tête-cible. Du début à la fin du mouvement, le bras, ainsi que la main, doivent toujours rester devant le corps. Or, chez les novices, le bras est souvent décalé (sur la gauche pour les gauchers, sur la droite pour les droitiers), et la prise d'élan du geste fait aller le bras à côté de la tête, voire même derrière (Fig. 52). Afin de détecter ce problème, la distance en  $x$ ,  $y$ , et  $z$  de la main par rapport à la tête avant le lancer est utilisée.

Le dernier défaut est celui de l'alignement du bras (*align arm*). Un phénomène naturel lors du lancer est d'avoir un décalage entre l'alignement initial du bras et l'alignement final (lors du lancer). En effet, le bras a naturellement tendance à revenir vers le centre du corps, quelle que soit la préférence manuelle de la personne (droitier ou gaucher) (Fig. 53). Il faut consciemment compenser cet effet, en essayant de garder un alignement équivalent du début à la fin du lancer. Afin de vérifier cet aspect du geste, une boîte englobante est calculée sur l'ensemble du geste. La moyenne de la largeur de cette boîte, ainsi que l'écart type, sont ensuite utilisés.

Chaque défaut n'apparaît pas forcément indépendamment des autres : dans la réalité, les apprenants cumulent très souvent plusieurs de ces défauts. Il existe également d'autres défauts, tels que la mauvaise prise d'une fléchette, un mauvais relâchement de la fléchette (qui lui imprime un mouvement erratique), etc. Ces problèmes ne sont cependant pas détectables à l'aide du mouvement de l'apprenant seul. Ils nécessitent d'autres données (telles que la position de la fléchette, la connaissance du moment du lâcher de fléchette, etc.) qu'il n'est pas possible d'obtenir à l'aide du matériel utilisé pour cette expérimentation.



FIGURE 53 – Défaut du « align arm » illustré : le bras de l'apprenant se dirige légèrement vers la droite lors du lancer (dans le cas d'un gaucher). L'effet inverse se produit pour un droitier.

### 6.3.3 *Protocole*

#### 6.3.3.1 *Acquisition des données de l'expert*

La première étape est la captation des données de l'expert, afin d'intégrer les besoins d'observation de l'expert au sein du système. L'objectif est d'obtenir des données correspondant à des mouvements jugés corrects par l'expert, et des données avec les défauts pris séparément. Ainsi, l'expert a réalisé dix lancers de fléchettes bons, puis dix lancers pour chacun des quatre défauts identifiés. Puis, pour chaque défaut, les descripteurs correspondants sur les bons mouvements, et sur les mouvements présentant le défaut considéré sont extraits.

#### 6.3.3.2 *Participants*

Trois groupes de 15 personnes chacun ont participé à l'expérimentation. Dans ces trois groupes, aucune personne n'était experte des fléchettes : le plus souvent, les seules connaissances de ce sport provenaient d'une pratique peu fréquente en tant que loisir. Le premier groupe n'a reçu que des conseils de l'expert, le deuxième n'a reçu que des conseils provenant du système MLA, et le troisième recevait des conseils de l'expert qui s'appuyait sur une visualisation des défauts donnée par le système pour donner son jugement. L'installation finale de l'expérimentation est présentée sur la Fig. 54.

#### 6.3.3.3 *Déroulement*

Avant l'expérimentation, un pré-questionnaire était rempli par chaque participant, afin de recueillir sa taille (pour fournir un squelette adapté dans le logiciel de capture de mouvements), le niveau d'expertise autoévalué aux fléchettes (échelle de Likert à 7 réponses), ainsi que toute pratique actuelle d'un sport (voir Annexe 1).

Une fois ce pré-questionnaire rempli, la combinaison de capteurs était installée sur la personne. Les calibrations étaient ensuite effectuées, afin d'étalonner les capteurs (position dans l'espace, espacement des capteurs entre eux, mise à zéro sur le logiciel de capture, etc.) pour que le logiciel puisse adapter la combinaison par rapport aux articulations du squelette 3D.

Chaque participant a reçu comme instruction de viser le centre de la cible à l'aide de fléchettes. La position de lancer étant importante, il a été demandé de se placer perpendiculairement à la marque au sol (Fig. 55).

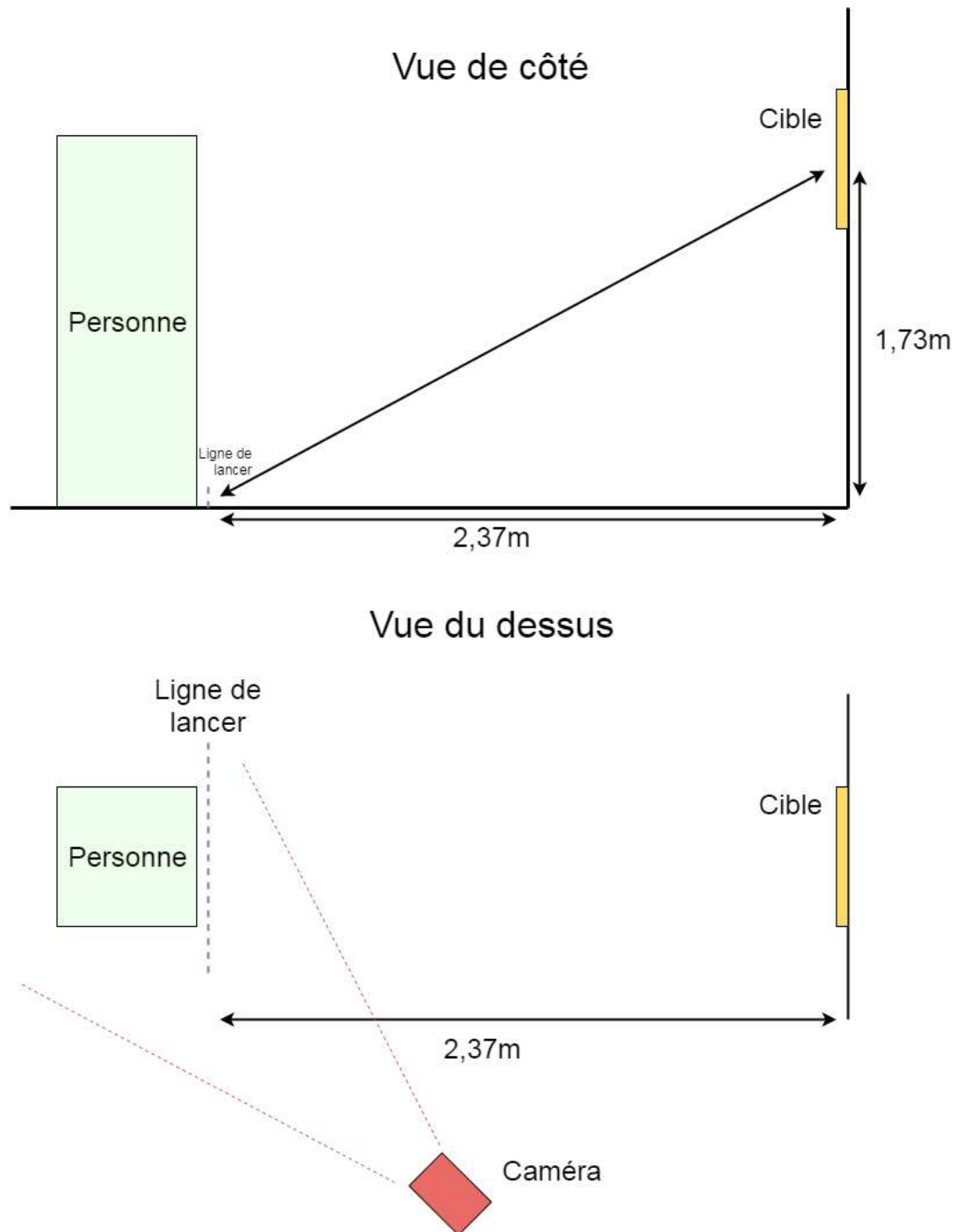


FIGURE 54 – Installation de l'expérimentation.



FIGURE 55 – Position latérale imposée aux participants.

Les participants devaient ensuite effectuer neuf lancers de test, non enregistrés. Ces lancers avaient pour buts de permettre à la personne de s'habituer à la combinaison et au lancer, ainsi que de vérifier que le squelette ne présentait pas de problèmes lors du mouvement (dérive des capteurs, mauvais positionnement des capteurs, parties de la combinaison qui ne fonctionnent pas, etc.).

Chaque personne a été affectée à l'un des trois groupes expérimentaux, sans avoir connaissance de cette assignation. Afin d'éviter toute différence de perception, il a été dit à chaque personne que les conseils étaient donnés par le système (même si ce n'était pas le cas pour tous les groupes).

Avant chaque lancer, la personne doit se placer en position de lancer, dire « prêt », puis attendre le signal sonore de départ du logiciel (qui correspond au début de l'enregistrement du mouvement). La personne lance ensuite sa fléchette, puis garde la position finale de son bras jusqu'au deuxième signal sonore (correspondant à l'arrêt de l'enregistrement). L'intervalle de temps entre le premier signal sonore et le lancer varie grandement d'une personne à l'autre ; cependant, l'algorithme de segmentation permet de ne garder que le geste utile pour chaque participant (cf. chapitre 5 section 5.4.3). À chaque lancer, la distance de la fléchette par rapport au centre est relevée (Fig. 56).

Chaque personne a effectué quatre séries de neuf lancers, soit trente-six lancers au total. Au bout d'une série de neuf lancers, les mouvements étaient analysés par le système (indépendamment du fait que les conseils donnés proviennent du système ou non). Pour le premier groupe, les données étaient analysées mais non utilisées par l'expert pour le jugement du geste. Chaque personne recevait alors deux conseils par rapport à son geste. En fonction de son groupe (conseils de l'expert pour le groupe 1, conseils du système pour le groupe 2, conseils de l'expert assisté du système pour le groupe 3). Les conseils donnés (par le système ou l'expert) sont les suivants :

- corps penché : « Votre corps et vos épaules ne doivent pas bouger pendant votre lancer. »



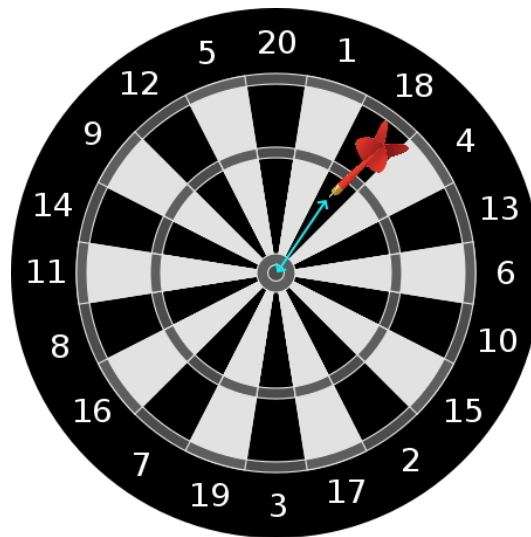


FIGURE 56 – En bleu, la distance relevée après chaque lancer de fléchette.

- lancer type javelot : « Votre main doit toujours rester devant votre corps lorsque vous lancez. »
- alignement du bras : « Votre bras doit rester aligné (de la main à l'épaule) lorsque vous lancez. »
- mouvement du coude : « Votre coude ne doit pas bouger lors du lancer. »

Dans le cas du groupe deux, les conseils sont choisis comme il suit :

- La distance entre le centroïde des données de l'apprenant et du cluster de gestes cibles de l'expert est calculée, pour chaque défaut (voir chapitre 5 section 5.4.4)
- Si deux défauts sont situés soit (i) dans le trapèze reliant les deux clusters, soit (ii) dans le cluster de mauvais gestes de l'expert, les deux conseils correspondant à ces défauts sont donnés
- S'il n'y en a qu'un seul, le conseil correspondant au défaut situé dans la zone susmentionnée est donné, puis le conseil correspondant au défaut dont la distance est la plus éloignée du bon cluster est donné
- S'il n'y en a aucun, les deux conseils correspondant aux défauts dont les distances sont les plus éloignées du bon cluster sont donnés

Dans le cas du groupe 3, une visualisation des données de l'apprenant par rapport à celles de l'expert est proposée (Fig. 57). Ainsi, l'expert peut s'appuyer sur cette visualisation pour prendre sa décision. Cette visualisation permet également de compenser un cas particulier : si les données de l'apprenant sont très dispersées, mais centrées autour du cluster de bons mouvements de l'expert, le centroïde des données de l'apprenant sera situé dans le cluster de bons gestes de l'expert. Ainsi, il est possible pour l'expert de repérer ce cas particulier dans le cas où il s'appuie sur le système pour donner ses conseils, et de ne pas considérer que cet aspect des gestes de l'apprenant est réussi. Cependant, dans le cadre de l'expérimentation, l'expert s'est basé sur les valeurs de distance affichées directement, plutôt que sur la visualisation. Il procède comme il suit :

- s'il détecte deux défauts, les deux conseils correspondant à ces défauts sont donnés à l'apprenant,
- s'il détecte un seul défaut, il donne le conseil correspondant à ce défaut, puis se base sur les valeurs de distance affichées par le système afin de donner le conseil sur le défaut étant le plus éloigné du cluster de bons gestes,

## Apprenant VS expert

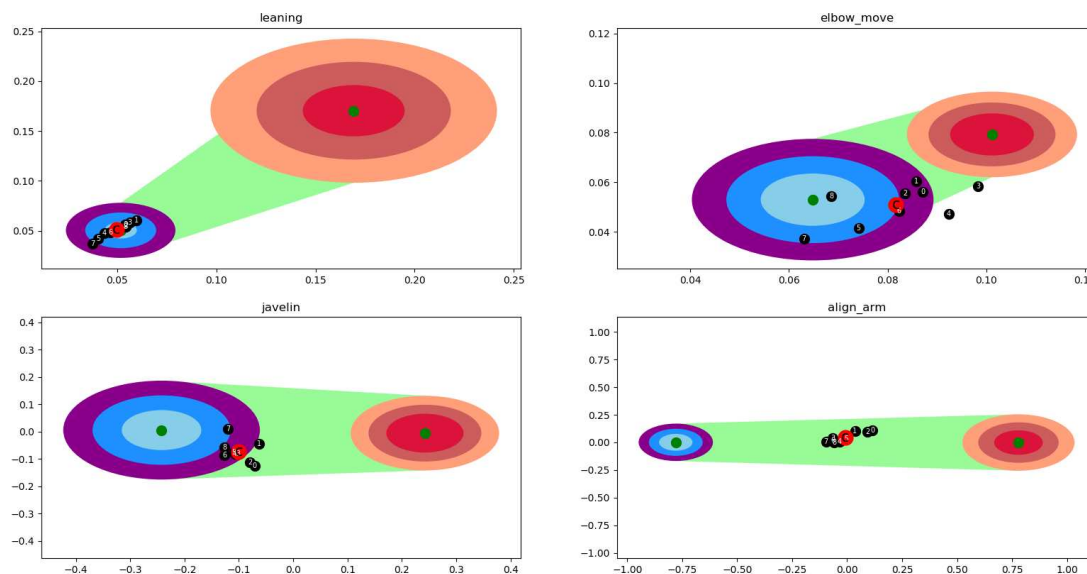


FIGURE 57 – Un exemple de visualisation des données de l'apprenant fournie à l'expert à la fin d'une série de neuf lancers.

- s'il n'arrive pas à déceler de défauts, il se base sur les valeurs de distance affichées par le système, et choisit les deux défauts les plus éloignés du bon cluster.

La personne refaisait ensuite 9 lancers, en essayant de tenir compte de ces deux conseils. Il n'y a pas de conseils donnés à la fin de la dernière série.

Une fois l'expérimentation finie, la personne remplit un post-questionnaire (voir Annexe 1). Ce post-questionnaire mesure le ressenti de la personne par rapport à la combinaison, à la progression de son geste et à l'auto-évaluation de sa performance. Les questions sont toutes basées sur l'échelle de Lickert, allant de 1 à 7. Les questions portent sur :

- le niveau d'aisance avec le matériel,
- le niveau de stress de l'apprenant lors de l'expérimentation,
- le niveau de compréhension des instructions reçues,
- l'intérêt ressenti des conseils donnés,
- l'auto-évaluation de la performance sur toute l'expérimentation,
- l'auto-évaluation de la progression de la performance sur toute l'expérimentation,
- l'auto-évaluation de la qualité du geste sur toute l'expérimentation,
- l'auto-évaluation de la progression de la qualité du geste sur toute l'expérimentation,
- l'expression libre (remarques diverses).

#### 6.3.3.4 Métriques calculées sur le clustering des données de l'expert

L'algorithme utilisé afin de regrouper les données de l'expert est un *k-means*, avec  $k = 2$ . Les métriques d'ASS et d'ARI ont été calculées sur les regroupements obtenus (voir tableau 8).



	Silhouette Score	Adjusted Rand Index
Leaning	0.85	1
elbow move	0.58	0.47
javelin	0.79	1
align arm	0.53	1

TABLE 8 – Valeurs de l’Average Silhouette Score et de l’Adjusted Rand Index pour les données de l’expert sur les quatre défauts.

Les valeurs pour l’ASS sont supérieures ou égales à celles obtenues pour les deux expérimentations précédentes, suggérant une bonne (voire très bonne) séparation des données. L’hypothèse  $H_3$  est validée dans le contexte de cette expérimentation. Les valeurs pour l’ARI suggèrent que le regroupement des données d’un défaut (mouvement du coude) ne correspond pas exactement à la vérité terrain relevée. Cela implique que certaines des données de l’expert ne sont pas séparables de manière fidèle à la vérité terrain à l’aide des descripteurs utilisés. Cela peut s’expliquer soit par (i) un mauvais choix de descripteurs, soit par (ii) des mouvements de l’expert où la différence entre le geste correct et le défaut n’est pas assez prononcée. Dans ces cas, le clustering permet d’obtenir un regroupement des données selon leur similitude, plutôt que l’étiquetage relevé.

#### 6.3.4 Résultats

Les résultats de l’expérimentation ont été analysés sous deux angles : l’amélioration de l’objectif du mouvement (c’est-à-dire la distance des fléchettes par rapport au centre de la cible) et l’amélioration des propriétés du mouvement de l’apprenant (c’est-à-dire le rapprochement du centroïde de l’apprenant du cluster de l’expert pour chaque défaut). L’analyse a été réalisée sous deux aspects : l’analyse intergroupes pour chaque jeu de lancers, et l’analyse intragroupe, sur les quatre jeux pour chaque groupe. L’objectif est double : (i) vérifier qu’il y a une amélioration significative du geste (au sens des tests statistiques effectués) au cours de l’expérimentation, que le système soit utilisé ou non et (ii) vérifier l’existence d’une différence significative (au sens des tests statistiques effectués) de l’amélioration du geste en fonction de l’utilisation faite du système : expert assisté, système seul ou sans système.

##### 6.3.4.1 Tests utilisés

Deux tests de normalité sont utilisés pour évaluer si la distribution des données suit une loi normale : le test de Shapiro-Wilk, et le test d’Agostino-Pearson. Plus précisément, le test de Shapiro-Wilk teste l’hypothèse nulle selon laquelle la population est normalement distribuée. Bien que ce test soit parfois recommandé pour tester la normalité de données (THODE 2002) et soit très largement utilisé, il n’est que peu efficace lorsque plusieurs données ont la même valeur. Ainsi, le test d’Agostino-Pearson est également utilisé, permettant de pallier ce problème (D’AGOSTINO, BELANGER et D’AGOSTINO 1990). Le niveau alpha de comparaison utilisé est 0.05 dans les deux cas.

La distribution de la population selon une loi normale permet d’utiliser des tests paramétriques, plus précis que leurs contreparties non-paramétriques (qui ne font aucune hypothèse sur la distribution des données). Dans le cas des données analysées, la distribution ne suit pas systématiquement une loi normale. Les résultats

des tests paramétriques et non-paramétriques sont présentés dans les tableaux, mais seuls les tests pertinents au regard de la distribution des données sont analysés.

Pour l'analyse intergroupe, 15 personnes différentes pour chaque groupe ont réalisé l'expérimentation. Nous supposons donc que les échantillons de données résultants de l'activité humaine, dans le contexte de cette expérimentation, sont indépendants entre les groupes.

**TESTS POUR LES RÉSULTATS INTERGROUPE** Dans le cas d'une distribution normale de la population, où l'on compare deux échantillons issus de populations différentes, les tests utilisés sont des tests paramétriques. Le test choisi ici est l'ANOVA (*ANalysis Of VAriance*). Son objectif est de déterminer si les distributions suivent une même loi normale (hypothèse nulle) ou si au moins une des moyennes des distributions est différente de celles des autres échantillons (hypothèse alternative). Si l'hypothèse nulle est vérifiée, l'homogénéité des variances des groupes est alors vérifiée à l'aide du test de Levene. Dans ce test, l'hypothèse nulle stipule que les variances de population sont égales. Enfin, un test de Student paire-à-paire est effectué, afin d'identifier les différences significatives entre les groupes pris deux-à-deux.

Dans le cas où la distribution de la population ne suit pas une loi normale, les tests non-paramétriques sont utilisés. Ces tests nécessitent plus de données que leur contrepartie paramétrique. Le premier test utilisé est celui de Kruskal-Wallis ; il est la contrepartie non-paramétrique du test d'ANOVA. L'hypothèse nulle est que les médianes de chaque groupe sont égales, l'hypothèse alternative étant qu'au moins une médiane d'un groupe est différente d'au moins l'une de celles des autres groupes. Si l'hypothèse nulle est vérifiée, un test de Wilcoxon-Mann-Whitney est utilisé, afin de tester si les distributions sont égales (hypothèse nulle).

Pour l'analyse intragroupe, dans chacun des groupes pris indépendamment, les 15 mêmes personnes réalisent chacun des 4 jeux de lancers à la suite. Nous supposons donc que les échantillons de données résultants de l'activité humaine, dans le contexte de cette expérimentation, sont dépendants entre les jeux.

**TESTS POUR LES RÉSULTATS INTRAGROUPES** Dans le cas où la distribution de la population suit une loi normale, le test de Student sur deux échantillons est utilisé. L'hypothèse nulle de ce test stipule que la moyenne des deux populations est égale.

Dans le cas où la distribution de la population ne suit pas une loi normale, le test non-paramétrique utilisé est le test des rangs signés de Wilcoxon (la contrepartie non-paramétrique du test de Student sur deux échantillons). L'hypothèse nulle de ce test est que la médiane de deux échantillons est égale.

Dans tous les cas considérés, il faut noter que le rejet ou la validation de l'hypothèse nulle n'est pas une preuve formelle que les résultats sont reproductibles et valides (HALSEY *et al.* 2015). Lorsque l'hypothèse nulle est rejetée, le résultat est considéré comme significatif et n'est pas dû à un phénomène aléatoire. Cependant, ce n'est pas une preuve formelle que la différence significative obtenue entre les échantillons soit liée à leur condition expérimentale respective identifiée par le chercheur. L'interprétation des résultats significatifs obtenus en comparant les différences dans les conditions expérimentales doit être menée avec une grande prudence et se limiter au cadre expérimental défini. La reproductibilité des résultats ne peut être alors qu'envisagée dans les limites du cadre, tandis que la généralisation peut

	Jeu 1		Jeu 2		Jeu 3		Jeu 4	
	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type
Groupe 1	10.1481	3.8336	14.4170	5.4708	10.9926	3.2721	10.2481	3.5990
Groupe 2	10.2481	3.6957	13.5741	6.4499	11.9674	4.6060	10.9296	3.6617
Groupe 3	9.7393	2.3020	11.9667	4.0647	12.6889	5.7258	11.0074	5.4516

TABLE 9 – Moyennes et écarts-types des données des apprenants (distance au centre de la cible).

être uniquement supposée. À l'inverse, lorsque l'hypothèse nulle est validée, il n'est généralement pas utile de pousser plus loin l'interprétation des résultats obtenus.

Cette partie ne présentera que les résultats pertinents au regard des tests considérés. Les tableaux complets d'analyse sont disponibles dans l'annexe 2.

Pour chaque analyse, les données sont affichées sous forme de diagrammes en boîte. Ces diagrammes, également appelés boîtes à moustaches, permettent de visualiser :

- la moyenne (losange noir),
- la médiane (barre noire),
- les valeurs du premier et troisième quartile (l'extrémité basse et haute des boîtes, respectivement),
- la différence entre  $q_3$  et  $q_1$  est appelée écart interquartile, et est défini par la différence  $q_3 - q_1$ ,
- les limites (moustaches) représentées respectivement par les points les plus proches de  $q_1 - \frac{q_3 - q_1}{1.5}$  et  $q_3 + \frac{q_3 - q_1}{1.5}$ ,
- la dispersion, représentée par l'écart entre les limites (moustaches),
- les données aberrantes, situées en deçà ou au-delà des limites (points rouges).

#### 6.3.4.2 Distance des fléchettes par rapport au centre de la cible

Le tableau 9 présente les moyennes et les écarts-types des distances des fléchettes par rapport au centre de la cible lors des lancers des participants. On remarque une dégradation systématique de la moyenne, ainsi que de l'écart-type, entre le jeu 1 et le jeu 2. Ces valeurs tendent à diminuer pour les jeux suivants, sauf dans le cas du groupe 3, où la valeur augmente entre le jeu 2 et le jeu 3, puis diminue ensuite. La différence entre le jeu 1 et le jeu 4 montre que ces valeurs se rapprochent des valeurs initiales. Ces résultats suggèrent que les premiers conseils donnés entraînent une perte de précision du geste, puis que les participants réussissent à améliorer la précision de leurs lancers.

Les résultats intergroupes sont présentés dans la figure 59. Pour chaque jeu (un jeu par graphique), les données des trois groupes sont présentées en parallèle. La figure 58 montre la progression de la précision en intra-groupe pour chaque groupe (un groupe par graphique).

Sur la figure 58, on peut observer une augmentation systématique de la valeur de la médiane ainsi que de la dispersion des données entre le jeu 1 et le jeu 2 pour tous les groupes, correspondant à une augmentation de la distance par rapport au centre. S'en suit une réduction de la dispersion des valeurs pour les jeux suivants (sauf pour le jeu 4 du groupe 3), suggérant ainsi une précision des lancers en hausse. Cependant, la précision ne revient pas à une valeur inférieure ou égale à celle du premier jeu (en terme de médiane, moyenne et précision). Cette tendance est la plus visible pour le groupe 1. Des résultats significatifs ont été obtenus entre les

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
ANOVA					
Pairwise T-test					
Chi-square		0.1875	1.8023	0.4432	0.3937
Kruskal-Wallis ( $p < 0.05$ )	p	0.9907	0.4061	0.8012	0.8213
	df	2	2	2	2
Pairwise Mann-Whitney					

TABLE 10 – Résultats des tests statistiques en intergroupe (distance par rapport au centre).

jeux 1 et 2 et les jeux 2 et 4 pour chaque groupe (tableau 11), montrant d’une part une dégradation systématique de la distance par rapport au centre de la cible à la suite des premiers conseils donnés, quel que soit le groupe considéré, et une amélioration progressive de la précision du lancer au fil de l’expérimentation. Des résultats significatifs ont également été obtenus entre le jeu 1 et le jeu 3 pour les groupes 2 et 3, le jeu 2 et le jeu 3 pour le groupe 1 et le jeu 3 et le jeu 4 pour le groupe 3. Cette dégradation semble correspondre au « désapprentissage » du geste initial au profit de l’apprentissage du geste cible. On constate la présence de données aberrantes sur les trois groupes, dont au moins une persistante sur les quatre jeux pour les groupes 2 et 3. Il y a plus de données aberrantes pour le groupe 3 (entre une et trois, selon le jeu considéré), malgré le fait que ce groupe soit celui présentant les meilleures valeurs de médianes, écart interquartile et dispersion.

Pour l’analyse intergroupe (figure 59), on observe que le groupe 3 est celui dont l’écart interquartile est le plus faible sur tous les jeux sauf le quatrième (groupe 1), tout en ayant le plus de données aberrantes. Le groupe 1 est celui ayant la médiane la plus basse sur tous les jeux, sauf le deuxième (groupe 3). Dans les deux cas, les résultats ne présentent pas de différences significatives (tableau 10).

#### 6.3.4.3 Amélioration des défauts du mouvement

Le tableau 12 montre la fréquence à laquelle sont donnés les conseils entre chaque jeu. On remarque qu’après le premier jeu, le défaut du corps penché est plus détecté que les autres, à l’inverse du défaut de l’alignement du bras. Cela peut s’expliquer par la facilité d’observation du défaut du corps penché par l’expert, à l’inverse de l’alignement du bras à cause de la position d’observation (visible sur la figure 54). À l’inverse, pour les conseils donnés à la fin du jeu 3, le défaut du corps penché est moins donné, car souvent corrigé dès les premiers jeux.

Le tableau 13 présente la fréquence des conseils donnés pour chaque groupe, entre chaque jeu et pour chaque groupe. Ainsi, on peut constater que le défaut du corps penché est très souvent détecté comme étant le défaut à corriger en premier, quel que soit le groupe considéré. Ce défaut est en fait une caractéristique des lancers de manière générale : se pencher lors d’un lancer permet d’imprimer plus de force au projectile. Pour le groupe 1, le conseil le plus donné est celui du mouvement du coude. Ce défaut est facilement observable lors du geste d’un apprenant, et donc l’expert tend à se concentrer sur ce défaut tant qu’il n’estime pas le mouvement satisfaisant. À l’inverse, le défaut du lancer de type javelot est rarement considéré

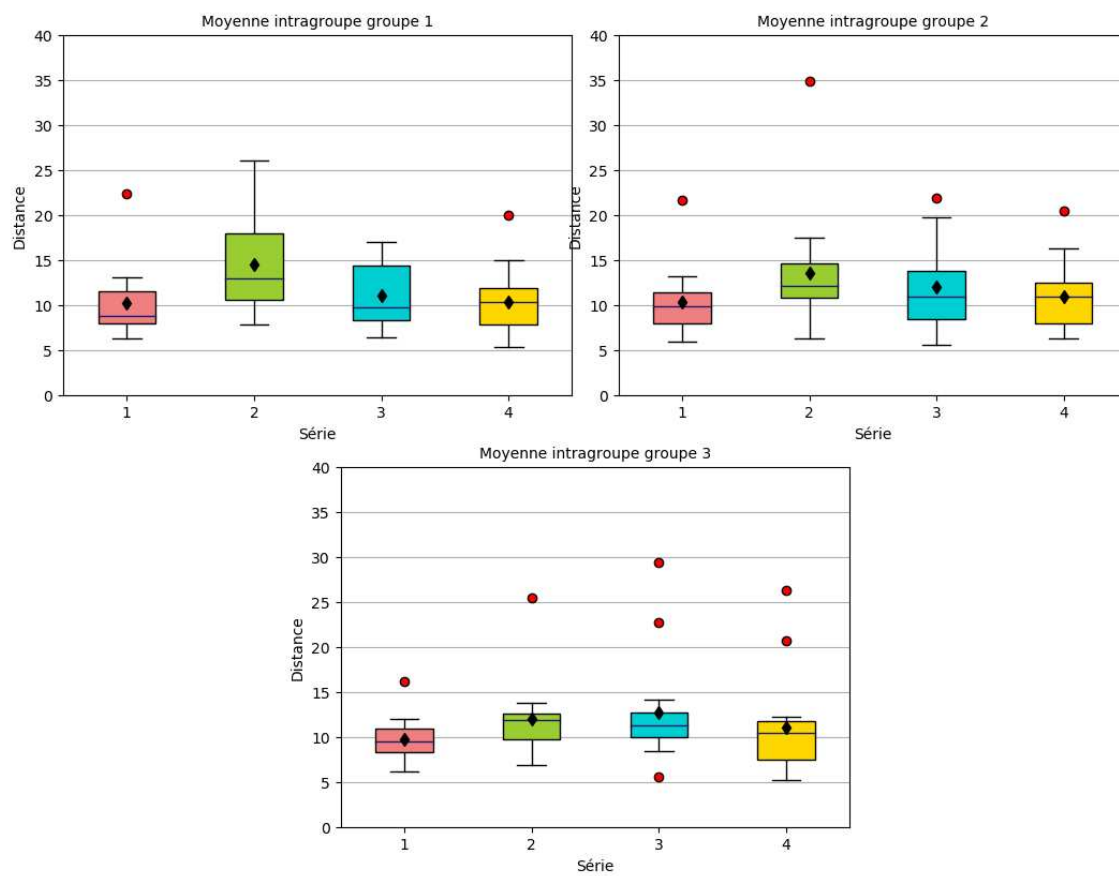


FIGURE 58 – Moyennes des distances des fléchettes par rapport au centre de la cible en intra-groupe.

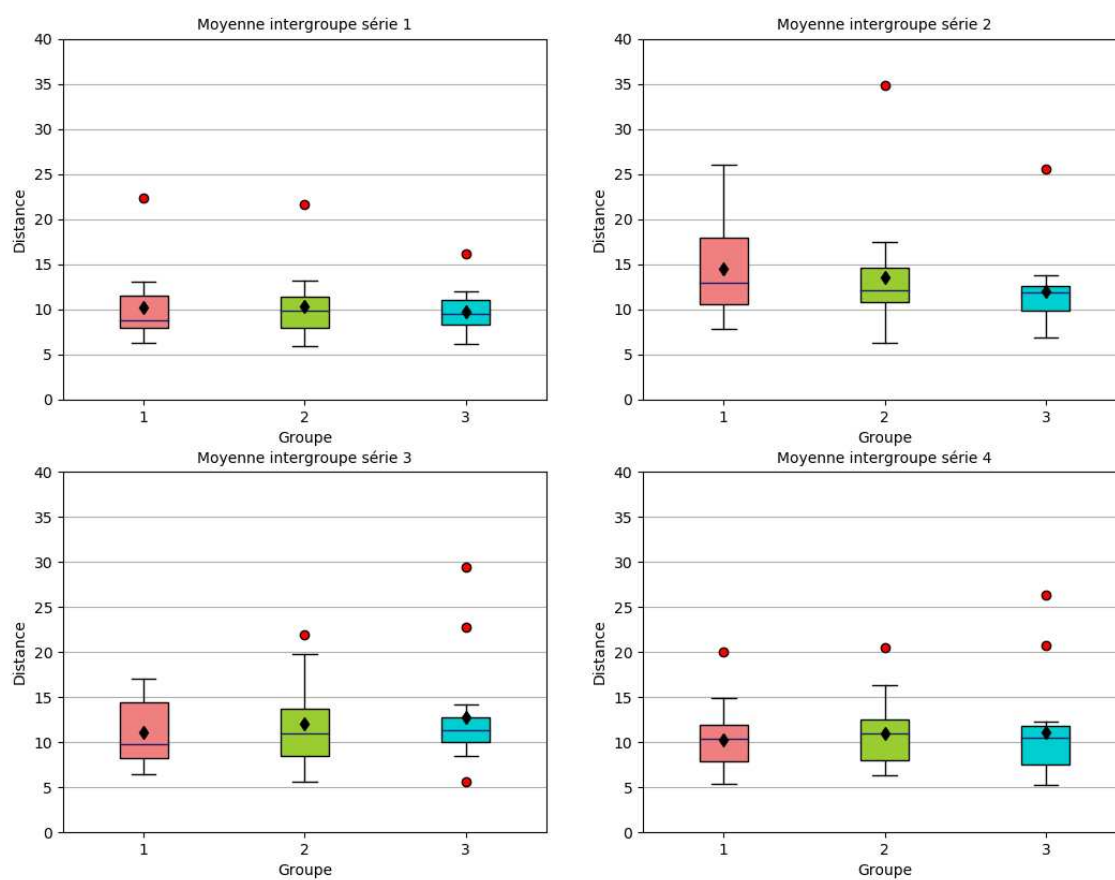


FIGURE 59 – Moyennes des distances des fléchettes par rapport au centre de la cible en inter-groupe.

		Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
		One tail		Two tails		T	One tail	Two tails	t	df
		p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	0.0057	0.0042	0.0115	0.0084	15				
	G2	0.0021	0.0010	0.0041	0.0020	9				
	G3	0.0038	0.0021	0.0076	0.0043	12.5				
Jeu 2-3	G1	0.0049	0.0034	0.0098	0.0067	14				
	G2	0.1006	0.1039	0.2013	0.2078	37				
	G3	0.5	0.4890	1	0.9780	59.5				
Jeu 3-4	G1						0.1514	0.3028	1.0698	14
	G2	0.1221	0.1262	0.2443	0.2524	39	0.1439	0.2877	1.1052	14
	G3	0.0144	0.0128	0.0288	0.0256	21				
Jeu 1-3	G1	0.1601	0.1651	0.3202	0.3303	42				
	G2	0.1110	0.1146	0.2220	0.2293	38				
	G3	0.01659	0.0151	0.0332	0.0301	22				
Jeu 2-4	G1	0.0005	$9.1553e^{-5}$	0.0011	0.0002	2				
	G2	0.0073	0.0051	0.0146	0.0102	16.5				
	G3	0.0250	0.0240	0.0500	0.0479	25				
Jeu 1-4	G1	0.4660	0.4670	0.9321	0.9341	58				
	G2	0.0864	0.0844	0.1728	0.1688	35.5				
	G3	0.2051	0.2106	0.4102	0.4212	45				

TABLE 11 – Résultats des tests statistiques en intragroupe (distance par rapport au centre).



	jeu 1-2	jeu 2-3	jeu 3-4
align arm	0.09	0.23	0.27
elbow move	0.27	0.29	0.30
javelin	0.26	0.25	0.24
leaning	0.38	0.23	0.19

TABLE 12 – Fréquences des conseils donnés en fonction des jeux.

		align arm	elbow move	javelin	leaning
G1	Jeu 1-2	0.10	0.43	0.10	0.37
	Jeu 2-3	0.33	0.33	0.20	0.14
	Jeu 3-4	0.37	0.40	0.06	0.17
G2	Jeu 1-2	0.06	0.10	0.47	0.37
	Jeu 2-3	0.13	0.16	0.33	0.37
	Jeu 3-4	0.20	0.20	0.37	0.23
G3	Jeu 1-2	0.10	0.27	0.23	0.40
	Jeu 2-3	0.23	0.37	0.20	0.20
	Jeu 3-4	0.23	0.30	0.30	0.17

TABLE 13 – Fréquences des conseils donnés pour chaque groupe, en fonction des jeux.

par l'expert. En effet, il est plus difficilement observable que les autres dans le contexte de cette expérimentation. Le système seul tend à donner les conseils sur les mêmes deux défauts pour les deux premiers jeux, à savoir le lancer de type javelot et le corps penché. Pour le troisième jeu, la fréquence des conseils s'harmonise, suggérant ainsi une stabilisation au niveau de tous les défauts (sans pour autant systématiquement atteindre une correction acceptable). Enfin, on remarque que le groupe 3 est celui présentant la répartition de conseils la plus homogène.

**ALIGN ARM** Le tableau 14 présente les moyennes et les écarts-types de la distance des données des participants par rapport au centroïde du cluster de bons gestes de l'expert, pour le défaut de l'alignement du bras. On remarque une amélioration des moyennes pour le groupe 1 entre les jeux 1 et 2, suivi d'une dégradation de ces valeurs. L'écart-type suit les mêmes tendances pour les trois groupes, suggérant ainsi que la correction des autres défauts se fait au détriment de l'alignement du bras, ou que le défaut d'alignement du bras n'est détecté qu'entre le jeu 1 et le 2, puis non-détecté par la suite.

	Jeu 1		Jeu 2		Jeu 3		Jeu 4	
	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type
Groupe 1	0.0836	0.0548	0.0547	0.0486	0.0656	0.0439	0.0727	0.0496
Groupe 2	0.0836	0.0556	0.0567	0.0375	0.0497	0.0452	0.0490	0.0391
Groupe 3	0.0677	0.0342	0.0606	0.0283	0.0578	0.0316	0.0592	0.0356

TABLE 14 – Moyennes et écarts-types des données des apprenants (alignement du bras).

Les résultats intergroupes pour le défaut d'alignement du bras sont présentés sur la figure 61, et les résultats en intragroupe sont présentés sur la figure 60.

Pour le groupe 1, la figure 60 montre que le défaut se réduit au jeu 2, mais reprend ensuite de l'ampleur dans les jeux suivants (et dépasse au final les valeurs initiales), tant au niveau de l'écart interquartile que de la valeur de la médiane. Le défaut se réduit également pour le groupe 2, sauf pour le 4ème jeu où la valeur de la médiane et l'écart interquartile augmentent. Enfin, les valeurs du groupe 3 sont stables pour

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
ANOVA					
Pairwise T-test					
Chi-square		0.4970	1.9223	3.2518	2.5979
Kruskal-Wallis ( $p < 0.05$ )	p	0.78000	0.3824	0.1967	0.2728
	df	2	2	2	2
Pairwise Mann-Whitney					

TABLE 15 – Résultats des tests statistiques en intergroupe (alignement du bras).

la valeur de la médiane, avec une dégradation de la valeur de la médiane sur le jeu 2, suivie par une amélioration de cette valeur, un écart interquartile qui reste sensiblement le même sur les quatre jeux et une dispersion des données variant grandement entre chaque jeu. Aucun résultat significatif n'a été trouvé pour le groupe 3, ce qui rend difficile l'explication de la rapide stabilité de ce défaut pour ce groupe. Les résultats sont significatifs pour le groupe 2 entre les jeux 1-2, 1-3 et 1-4 (table 17), et pour le groupe 1 entre les jeux 1-2, mais le conseil n'est que peu donné, suggérant ainsi que le défaut est corrigé par les apprenants au travers d'un autre conseil. Il y a, pour chaque groupe, au moins une donnée aberrante.

En analyse intergroupe (figure 61), les écarts interquartiles les plus faibles sont obtenus par les groupes 1 et 3 sur les deux premiers jeux, puis par les groupes 2 et 3 pour les deux derniers jeux. Les moyennes sont également sensiblement meilleures pour ces mêmes groupes. La dispersion est cependant meilleure pour le groupe 1 lors des deux premiers jeux. Aucun de ces résultats ne présente de différences significatives (table 15).

**ELBOW MOVE** Le tableau 16 présente les moyennes et les écarts-types de la distance des données des participants par rapport au centroïde du cluster des bons gestes de l'expert, pour le défaut du mouvement du coude. Il y a une amélioration systématique de ces valeurs entre le jeu 1 et le jeu 2, puis une stabilisation, sauf pour le groupe 3, où l'écart-type diminue encore entre le jeu 2 et le jeu 3. Cela indique une bonne prise en compte de ce conseil lorsqu'il est donné, ou qu'un conseil sur un autre défaut permet également de corriger le mouvement du coude.

	Jeu 1		Jeu 2		Jeu 3		Jeu 4	
	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type
Groupe 1	0.1174	0.0653	0.0527	0.0275	0.0544	0.0247	0.0547	0.0266
Groupe 2	0.0932	0.0519	0.0566	0.0274	0.0453	0.0236	0.0453	0.0267
Groupe 3	0.1092	0.0655	0.0575	0.0432	0.0492	0.0296	0.0552	0.0325

TABLE 16 – Moyennes et écarts-types des données des apprenants (mouvement du coude).

Les résultats intergroupes pour le défaut du mouvement du coude sont présentés dans la figure 63, et les résultats en intragroupe sont présentés dans la figure 62.

Lors de l'analyse intragroupe (figure 62), on peut constater une forte réduction de la valeur de la médiane, de l'écart interquartile et de la dispersion dès le jeu 2 pour tous les groupes. Les valeurs restent ensuite sensiblement les mêmes, autant pour la

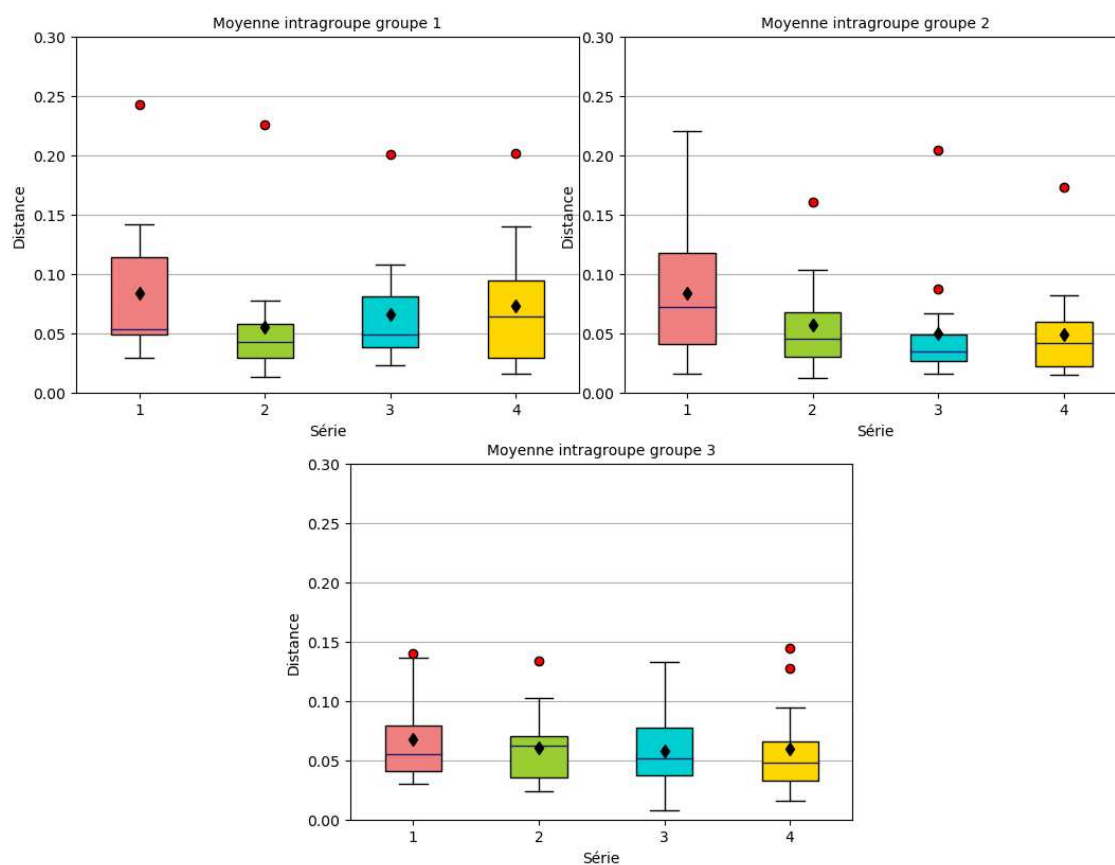


FIGURE 60 – Moyenne des distances du défaut align arm par rapport au centroïde du cluster des bons mouvements de l'expert en intra-groupe.

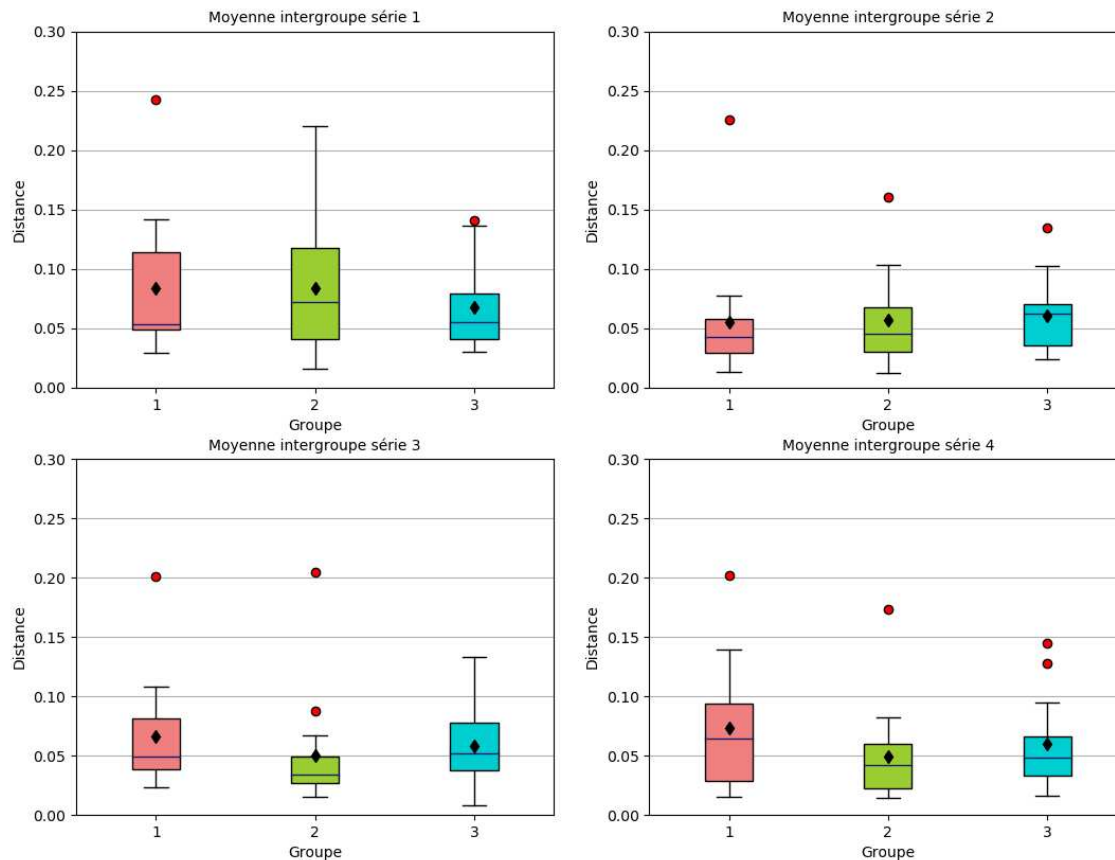


FIGURE 61 – Moyenne des distances du défaut align arm par rapport au centroïde du cluster des bons mouvements de l'expert en inter-groupe.

		Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
		One tail		Two tails		T	One tail	Two tails	t	df
		p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	0.0035	0.0021	0.0070	0.0043	12				
	G2	0.0285	0.0277	0.0571	0.0553	26				
	G3	0.2389	0.2443	0.4777	0.4887	47				
Jeu 2-3	G1	0.1110	0.1146	0.2220	0.2293	38				
	G2	0.2568	0.2622	0.5136	0.5245	48				
	G3						0.3905	0.7809	0.2835	14
Jeu 3-4	G1	0.2216	0.2271	0.4432	0.4543	46				
	G2	0.4660	0.4670	0.9321	0.9341	58				
	G3	0.3774	0.3808	0.7547	0.7615	54				
Jeu 1-3	G1	0.0661	0.0677	0.1323	0.1354	33				
	G2	0.0144	0.0128	0.0288	0.0256	21				
	G3	0.2216	0.2271	0.4432	0.4543	46	0.2134	0.4268	0.8185	14
Jeu 2-4	G1	0.0820	0.0844	0.1641	0.1688	35				
	G2	0.2389	0.2243	0.4777	0.4887	47				
	G3	0.2216	0.2271	0.4432	0.4543	46				
Jeu 1-4	G1	0.1467	0.1514	0.2934	0.3028	41				
	G2	0.0219	0.0206	0.0438	0.0413	24				
	G3	0.2216	0.2271	0.4432	0.4543	46				

TABLE 17 – Résultats des tests statistiques en intragroupe (alignement du bras).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
ANOVA					
Pairwise T-test					
	Chi-square	0.9793	0.2319	1.0321	0.5969
Kruskal-Wallis ( $p < 0.05$ )	p	0.6128	0.8905	0.5969	0.4460
	df	2	2	2	2
Pairwise Mann-Whitney					

TABLE 18 – Résultats des tests statistiques en intergroupe (mouvement du coude).

valeur de la médiane que de l'écart interquartile et la dispersion. Cette réduction peut être interprétée de deux manières : soit le défaut est systématiquement décelé par le système dès les premiers lancers, et sa correction est facilement effectuée par l'apprenant, soit un autre défaut souvent détecté par le système lors du premier jeu influence le défaut du mouvement du coude. Des valeurs significatives sont obtenues pour tous les groupes entre les jeux 1-2, 1-3 et 1-4 (table 20), confirmant ainsi l'amélioration nette du défaut entre le premier et le dernier jeu. Le groupe 2 obtient des résultats significatifs entre les jeux 2-3 et 2-4, suggérant ainsi que l'amélioration entre le jeu 3 et le jeu 4 est plus marginale. Enfin, le groupe 3 obtient des résultats significatifs entre le jeu 3 et le jeu 4, montrant une amélioration jusqu'au dernier jeu. Le groupe 1 présente trois données aberrantes pour le premier jeu, qui disparaissent ensuite pour les jeux suivants. Le groupe 3 semble contenir quelques apprenants ayant des difficultés avec le mouvement du coude (présence d'une ou deux données aberrantes sur chaque jeu).

L'analyse intergroupe (figure 63) permet d'observer que la valeur de la médiane et l'écart interquartile sont les plus faibles pour les groupes faisant intervenir le système (groupes 2 et 3). La dispersion reste sensiblement la même pour tous les groupes pour les jeux 2 et 3. L'observation effectuée sur l'analyse intragroupe se confirme ici, montrant que les différences intergroupes sont faibles dès le jeu 2. Les données intergroupes ne présentent pas de différences significatives (table 18).

**JAVELIN** Le tableau 19 présente les moyennes et les écarts-types de la distance des données des participants par rapport au centroïde des bons gestes de l'expert, pour le défaut du lancer type javelot. Il y a une amélioration de ces valeurs entre le jeu 1 et le jeu 2 pour tous les groupes, puis une stabilisation ensuite, sauf pour le groupe 2, où la moyenne diminue encore entre le jeu 3 et le jeu 4. Ces résultats suggèrent que la correction du défaut se fait dès le jeu 2, sans montrer d'amélioration par la suite, que le conseil soit donné ou non.

	Jeu 1		Jeu 2		Jeu 3		Jeu 4	
	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type
Groupe 1	0.1791	0.0720	0.1445	0.0798	0.1492	0.0658	0.1467	0.0689
Groupe 2	0.1789	0.0594	0.1452	0.0718	0.1511	0.0888	0.1141	0.0513
Groupe 3	0.1330	0.0513	0.1175	0.0321	0.1261	0.0314	0.1207	0.0375

TABLE 19 – Moyennes et écarts-types des données des apprenants (mouvement du coude).

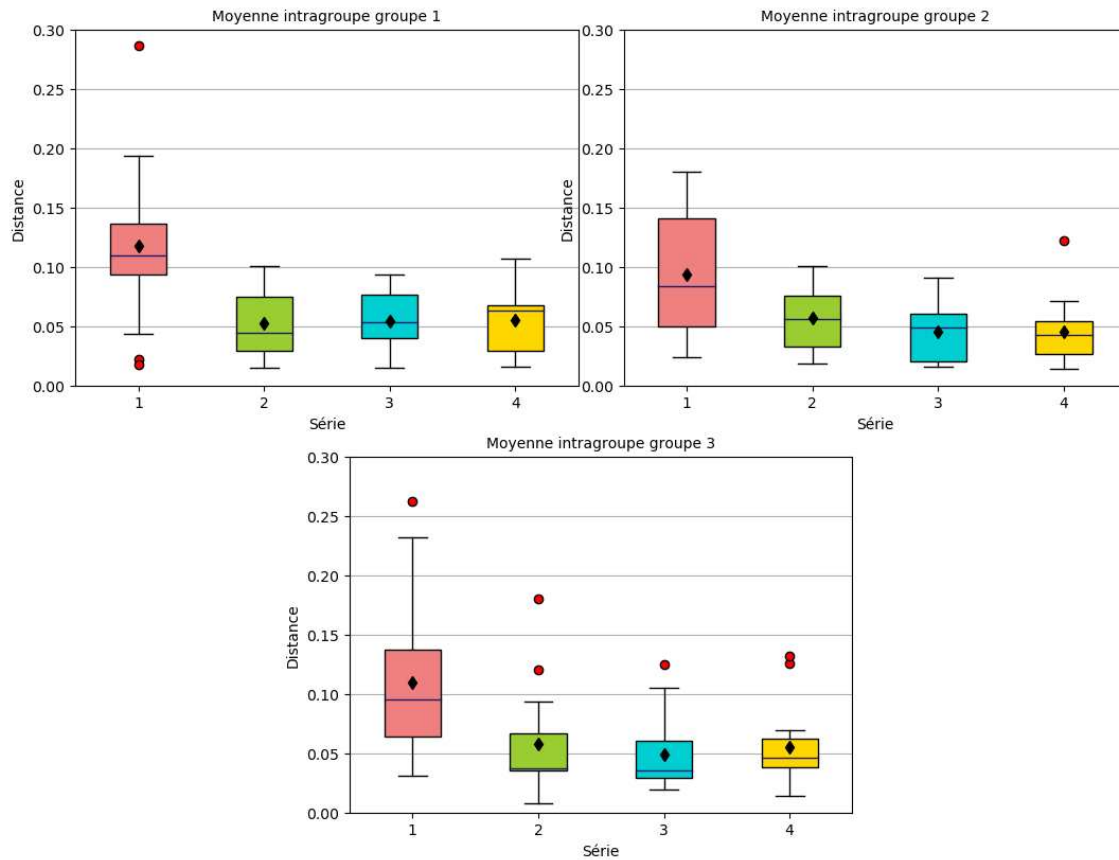


FIGURE 62 – Moyenne des distances du défaut *elbow move* par rapport au centroïde du cluster des données de l'expert en intra-groupe

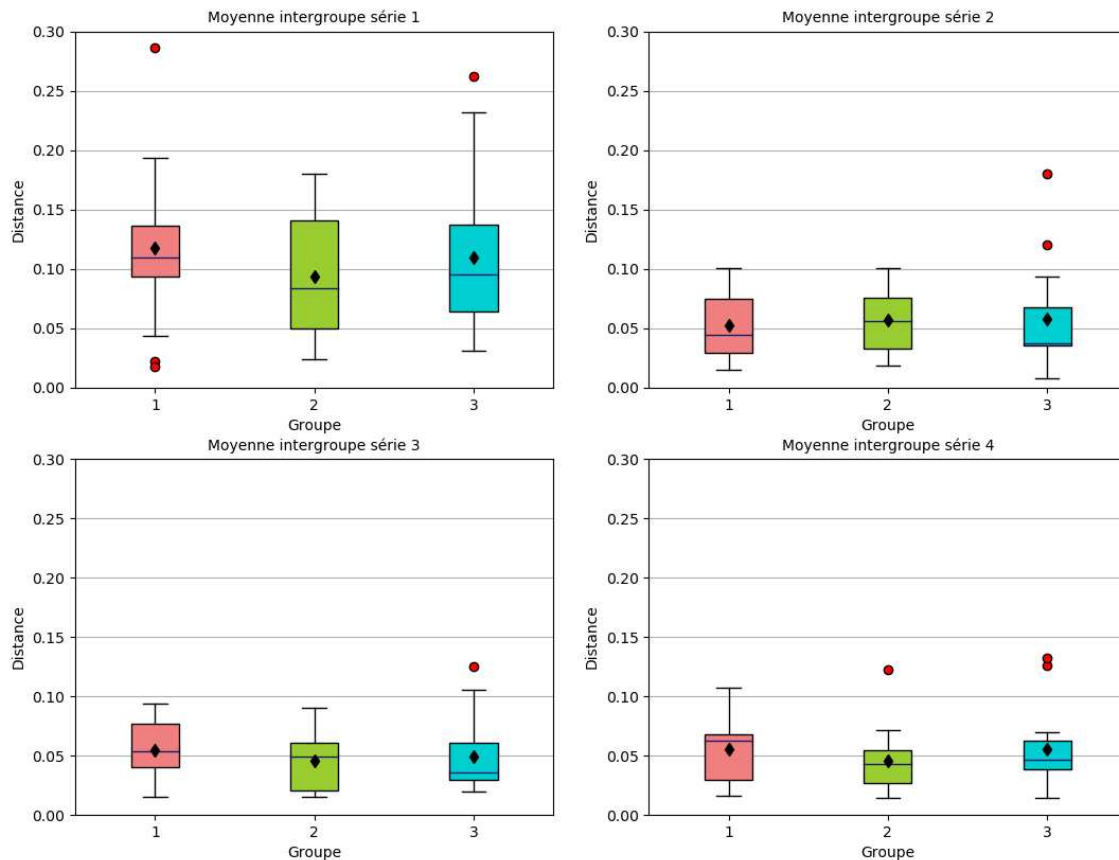


FIGURE 63 – Moyenne des distances du défaut *elbow move* par rapport au centroïde du cluster des données de l'expert en inter-groupe.

		Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
		One tail		Two tails		T	One tail	Two tails	t	df
		p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	$6.6602e^{-4}$	$1.5259e^{-4}$	0.0013	$3.0518e^{-4}$	3				
	G2	$9.8276e^{-4}$	$3.0518e^{-4}$	0.0020	$6.1035e^{-4}$	5	0.0026	0.0052	3.3014	14
	G3	$4.4595e^{-4}$	$6.1035e^{-5}$	$8.9190e^{-4}$	$1.2207e^{-4}$	1				
Jeu 2-3	G1						0.3617	0.7235	-0.3910	14
	G2	0.0738	0.0757	0.1475	0.1514	34	0.0482	0.0964	1.7820	14
	G3	0.0661	0.0677	0.1323	0.1354	33				
Jeu 3-4	G1						0.4712	0.9425	-0.0735	14
	G2	0.3991	0.4020	0.7983	0.8039	55				
	G3	0.0416	0.0416	0.0832	0.0832	29				
Jeu 1-3	G1	$3.6326e^{-4}$	$3.0518e^{-5}$	$7.2651e^{-4}$	$6.1035e^{-5}$	0				
	G2	0.0038	0.0021	0.0076	0.0043	12.5	0.0018	0.0036	3.4882	14
	G3	0.0014	$5.7983e^{-4}$	0.0029	0.0012	7				
Jeu 2-4	G1						0.3837	0.7675	-0.3015	14
	G2						0.0263	0.0525	2.1182	14
	G3	0.4435	0.4452	0.8871	0.8904	57	0.3351	0.6702	0.4350	14
Jeu 1-4	G1	$6.6602e^{-4}$	$1.5259e^{-4}$	0.0013	$3.0518e^{-4}$	3				
	G2	$6.6602e^{-4}$	$1.5259e^{-4}$	0.0013	$3.0518e^{-4}$	3	0.0015	0.0030	3.5854	14
	G3	$5.4581e^{-4}$	$9.1553e^{-5}$	0.0011	$1.8311e^{-4}$	2				

TABLE 20 – Résultats des tests statistique en intragroupe (mouvement du coude).



		Jeu 1	Jeu 2	Jeu 3	Jeu 4
ANOVA ( $p < 0.05$ )	F (2, 42)	2.6135			1.4252
	p	0.0852			0.2518
Pairwise T-test ( $p < 0.05$ )					
Kruskal-Wallis ( $p < 0.05$ )	Chi-square		0.8371	0.4924	
	p		0.6580	0.7818	
	df		2	2	
Pairwise Mann-Whitney ( $p < 0.05$ )					

TABLE 21 – Résultats des tests statistiques en intergroupe (lancer de type javelot).

Les résultats intergroupes pour le défaut du lancer type javelot sont présentés dans la figure 65, et les résultats en intragroupe sont présentés dans la figure 64.

Pour l'analyse intragroupe, la figure 64 montre que la valeur de la médiane augmente sensiblement par au jeu 2 pour le groupe 1, suggérant une variabilité assez forte pour ce défaut. À l'inverse, pour le groupe 3, la valeur de la médiane, ainsi que l'écart interquartile, sont bas dès le jeu 1, et restent stables sur les autres. Pour le groupe 2, la valeur de la médiane baisse de manière continue, et l'écart interquartile, après une augmentation au jeu 2, diminue également au fur et à mesure des jeux. L'analyse statistique montre que les résultats sont significatifs pour le groupe 2 entre les jeux 1-3, 2-4, 1-4 et 3-4 (table 23) et pour le groupe 1 entre le jeu 1 et le jeu 2, confirmant ainsi le fait que le conseil est souvent donné par le système (voir tableau 13) et que les apprenants réussissent à le corriger efficacement. Les données aberrantes présentes sur le jeu 1 du groupe 1 disparaissent dès le jeu 3. Le groupe 2 possède une donnée aberrante sur chaque jeu.

L'analyse intergroupe (figure 64) montre que le groupe 3 obtient systématiquement la meilleure dispersion pour l'ensemble des jeux, et les meilleures valeurs de médianes pour les deux premiers jeux. Le groupe 2 obtient les meilleures valeurs de médianes pour les deux derniers jeux. La stabilité de la valeur de la médiane, ainsi que l'écart interquartile bas dès le premier jeu suggère que le défaut n'était pas présent pour le groupe 3. La dispersion reste élevée pour le groupe 1 tout au long de l'expérimentation. Ces résultats ne présentent cependant pas de différences significatives (table 21).

**LEANING** Le tableau 22 présente les moyennes et les écarts-types de la distance des données des participants par rapport au centroïde des bons gestes de l'expert, pour le défaut du corps qui se penche. Il y a une forte amélioration de ces valeurs entre le jeu 1 et le jeu 2 pour tous les groupes, puis une stabilisation. Ces résultats suggèrent que la correction du défaut est effective dès le jeu 2.

	Jeu 1		Jeu 2		Jeu 3		Jeu 4	
	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type
Groupe 1	0.0813	0.0489	0.0359	0.0194	0.0359	0.0170	0.0354	0.0175
Groupe 2	0.0557	0.0411	0.0351	0.0182	0.0265	0.0159	0.0270	0.0146
Groupe 3	0.0725	0.0473	0.0370	0.0344	0.0322	0.0238	0.0369	0.0253

TABLE 22 – Moyennes et écarts-types des données des apprenants (corps penché).

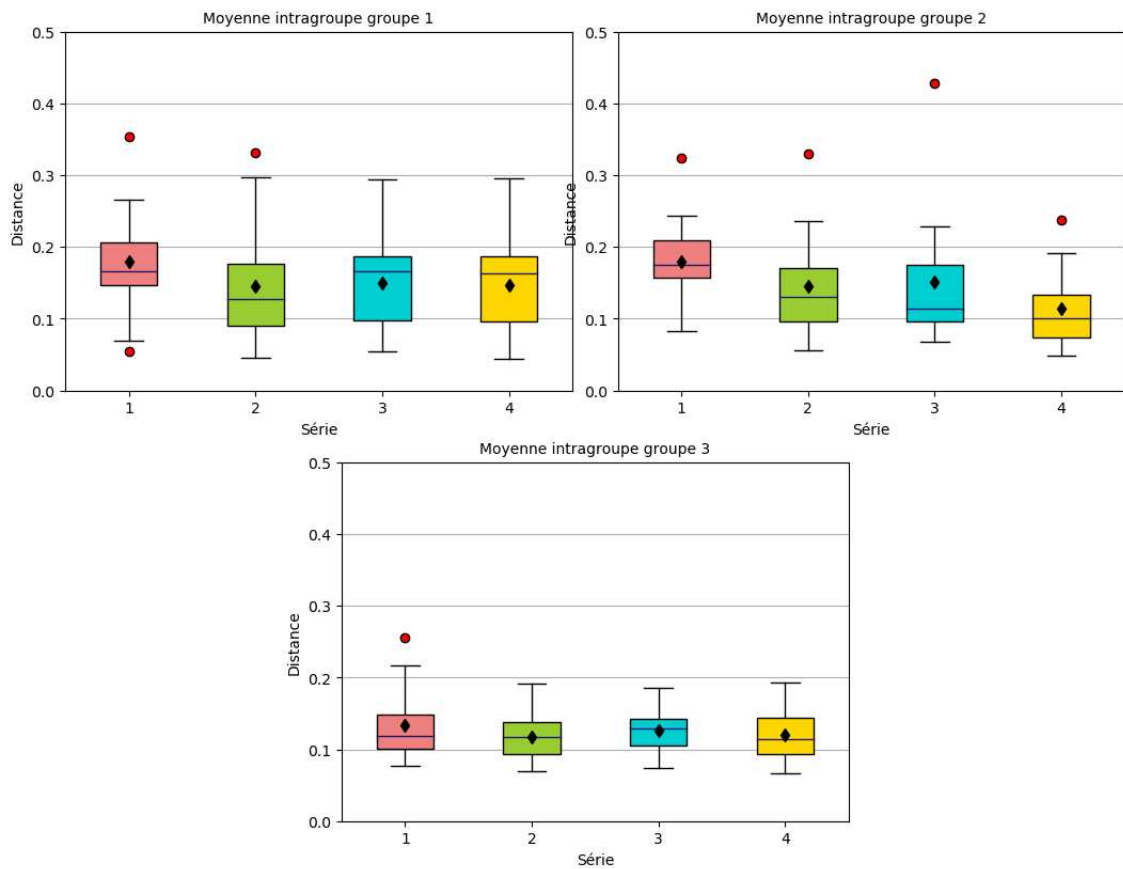


FIGURE 64 – Moyenne des distances du défaut *javelin* par rapport au centroïde du cluster des données de l'expert en intra-groupe

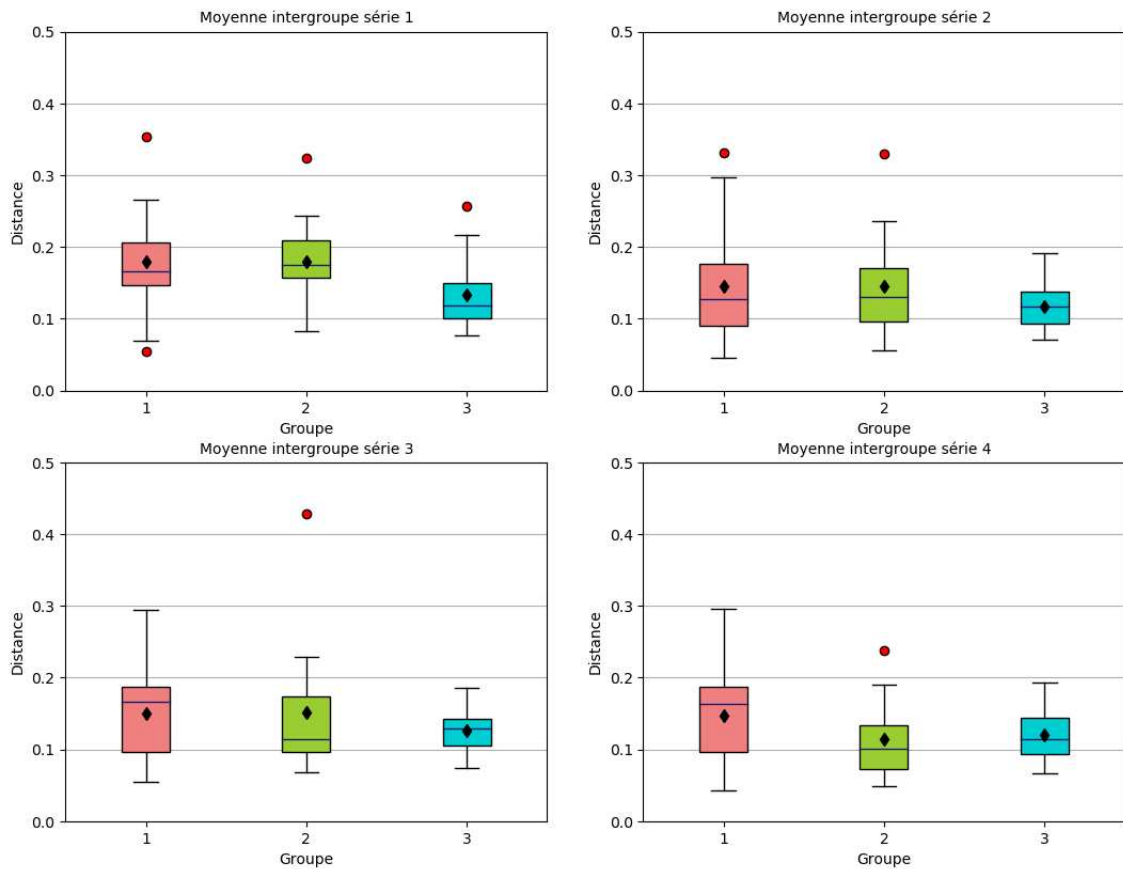


FIGURE 65 – Moyenne des distances du défaut *javelin* par rapport au centroïde du cluster des données de l'expert en inter-groupe.

		Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
		One tail		Two tails		T	One tail	Two tails	t	df
		p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1						0.0226	0.0452	2.1984	14
	G2						0.0596	0.1192	1.6598	14
	G3	0.2755	0.2807	0.5509	0.5614	49				
Jeu 2-3	G1						0.3370	0.6741	-0.4295	14
	G2	0.2568	0.2622	0.5136	0.5245	48				
	G3						0.1105	0.2209	-1.2812	14
Jeu 3-4	G1						0.2455	0.4911	0.7071	14
	G2	0.0029	0.0017	0.0059	0.0034	11				
	G3						0.2817	0.5635	0.5917	14
Jeu 1-3	G1						0.0739	0.1478	1.5322	14
	G2	0.0191	0.0177	0.0381	0.0353	23				
	G3	0.4435	0.4452	0.8871	0.8904	57				
Jeu 2-4	G1						0.4212	0.8424	-0.2025	14
	G2	0.0057	0.0042	0.0115	0.0084	15				
	G3						0.3671	0.7341	-0.3465	14
Jeu 1-4	G1						0.0530	0.1061	1.7275	14
	G2						0.0030	0.0061	3.2280	14
	G3	0.2389	0.2443	0.4777	0.4887	47				

TABLE 23 – Résultats des tests statistiques en intragroupe (lancer de type javelot).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
ANOVA ( $p < 0.05$ )	F (2, 42)	1.1188			
	p	0.3362			
Pairwise T-test ( $p < 0.05$ )					
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	2.5484	0.7258	2.8197	1.3503
	p	0.2796	0.6957	0.2442	0.5091
	df	2	2	2	2
Pairwise Mann-Whitney ( $p < 0.05$ )					

TABLE 24 – Résultats des tests statistiques en intergroupe (corps penché).

Les résultats intergroupes pour le défaut du corps qui se penche lors du lancer sont présentés dans la figure 67, et les résultats en intragroupe sont présentés dans la figure 66.

Dans le cas de l'analyse intragroupe, la figure 66 montre une forte baisse de l'écart interquartile, de la dispersion et de la valeur de la médiane pour les 3 groupes, et ce dès le jeu 2. Les valeurs restent ensuite stables, sauf pour le groupe 3, où l'écart interquartile se resserre progressivement jusqu'au jeu 3, à partir duquel il s'élargit de nouveau. Les résultats sont significatifs pour tous les groupes entre les jeux 1-3 et 1-4 (table 26), pour les groupes 1 et 3 entre le jeu 1 et le jeu 2, montrant une amélioration dès les premiers conseils donnés, et pour le groupe 2 entre les jeux 2-3 et 2-4, indiquant une progression plus étalée sur le temps. Ces résultats confirment ainsi la haute fréquence de ce conseil (tableau 13) et l'amélioration systématique de ce défaut dans les mouvements des apprenants. Seule une donnée aberrante est présente (groupe 3), visible après le jeu 1.

Pour l'analyse intergroupe, la figure 67 permet d'observer que l'écart interquartile est le meilleur pour le groupe 2 sur tous les jeux, ainsi que les valeurs des médianes sur les jeux 1 et 4 (très proches des valeurs des médianes du groupe 3). Le groupe 3 obtient une médiane sensiblement meilleure sur les jeux 2 et 4. La dispersion reste sensiblement la même quel que soit le groupe ou le jeu considéré. Aucun résultat ne présente une différence significative (tableau 24).

**TOUS LES DÉFAUTS PRIS ENSEMBLE** Pour ce dernier cas, les gestes de l'apprenant sont comparés à ceux de l'expert dans leur entièreté (c'est-à-dire en utilisant tous les descripteurs calculés dans les autres cas). Les données sont concaténées, puis réduites à l'aide d'une analyse en composantes principales (avec  $n = 2$ ), afin d'être cohérentes avec la visualisation en 2D proposée à l'expert.

Le tableau 25 présente les moyennes et les écarts-types de la distance des données des participants par rapport au centroïde des bons gestes de l'expert, pour tous les défauts pris ensemble. Il y a une amélioration de ces valeurs entre le jeu 1 et le jeu 2 pour tous les groupes, puis une stabilisation ensuite, sauf pour le groupe 2, où la moyenne continue de baisser pour les jeux suivants. Ces résultats suggèrent que la correction des défauts est la plus grande dès les premiers conseils, puis que les ajustements sont discutables par la suite.

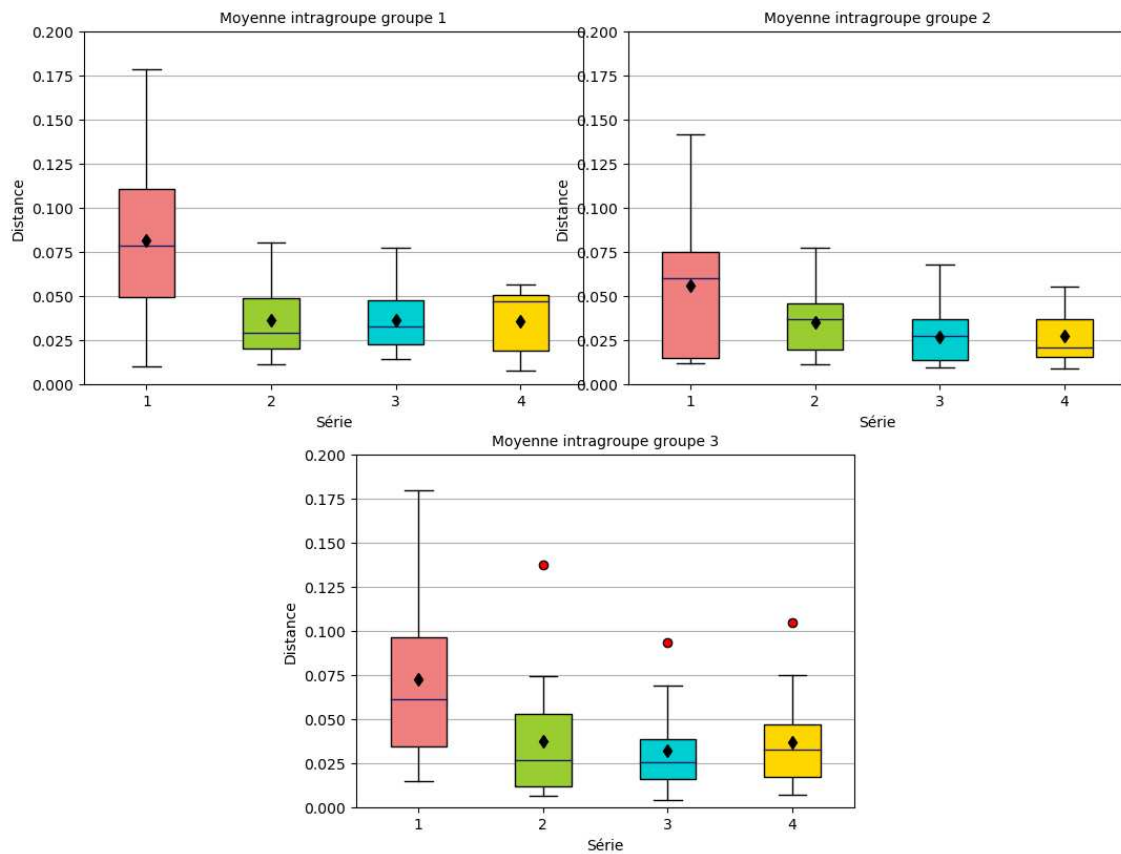


FIGURE 66 – Moyenne des distances du défaut *leaning* par rapport au centroïde du cluster des données de l'expert en intra-groupe

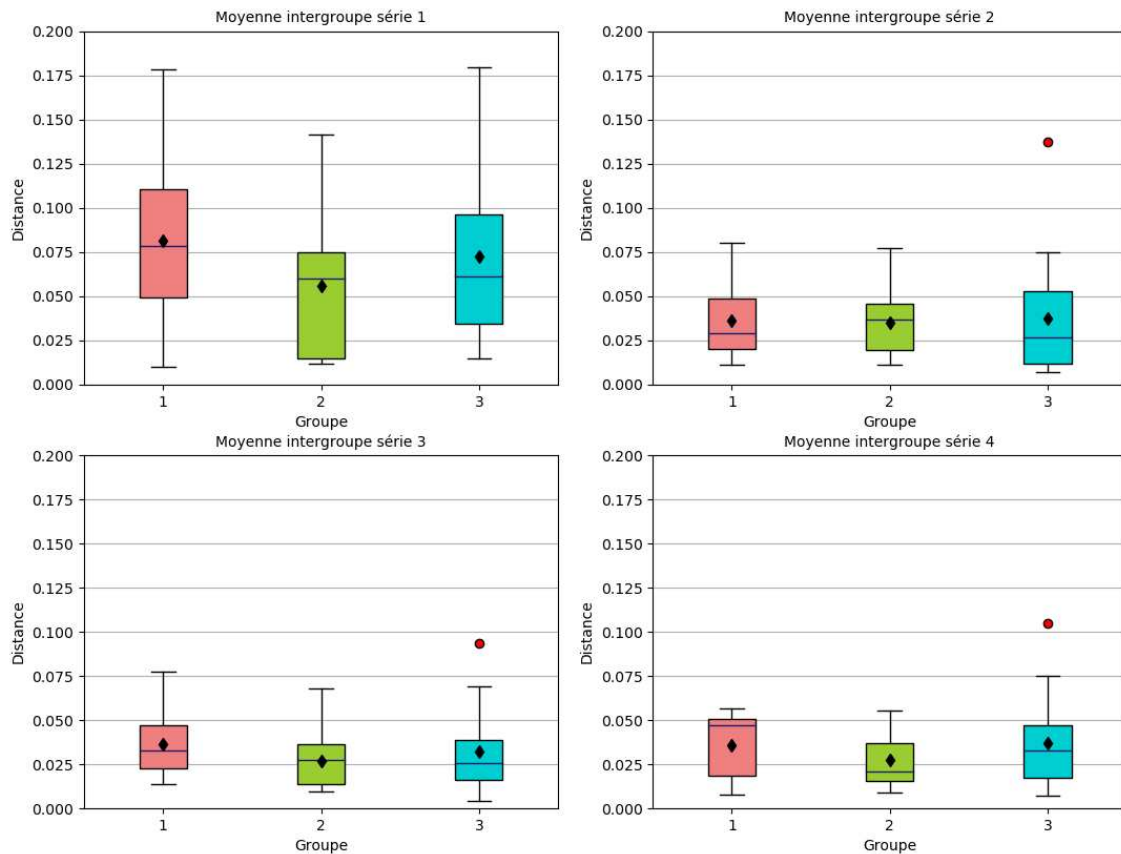


FIGURE 67 – Moyenne des distances du défaut *leaning* par rapport au centroïde du cluster des données de l'expert en inter-groupe.

		Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
		One tail		Two tails		T	One tail	Two tails	t	df
		p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	$8.1026e^{-4}$	$2.1362e^{-4}$	0.0016	$4.2725e^{-4}$	4				
	G2	0.0592	0.0603	0.1183	0.1205	32				
	G3	$9.8276e^{-4}$	$3.0518e^{-4}$	0.0020	$6.1035e^{-4}$	5				
Jeu 2-3	G1						0.4993	0.9986	0.0018	14
	G2	0.0469	0.047	0.0938	0.0946	30	0.0336	0.0671	1.9845	14
	G3	0.2755	0.2807	0.5509	0.5614	49				
Jeu 3-4	G1						0.4445	0.8891	0.1420	14
	G2	0.4435	0.4452	0.8871	0.8904	57	0.4438	0.8875	-0.1440	14
	G3	0.0592	0.0603	0.1183	0.1205	32				
Jeu 1-3	G1	$9.8276e^{-4}$	$3.0518e^{-4}$	0.0020	$6.1035e^{-4}$	5				
	G2	0.0219	0.0206	0.0438	0.0413	24				
	G3	0.0014	$5.7983e^{-4}$	0.0029	0.0012	7				
Jeu 2-4	G1	0.4660	0.4670	0.9321	0.9341	58	0.4633	0.9266	0.0937	14
	G2	0.0219	0.0206	0.0438	0.0413	24	0.0166	0.0332	2.3622	14
	G3	0.4435	0.4452	0.8871	0.8904	57				
Jeu 1-4	G1	0.0014	$5.7983e^{-4}$	0.0029	0.0012	7				
	G2	0.0191	0.0177	0.0382	0.0353	23				
	G3	0.0017	$7.6294e^{-4}$	0.0034	0.0015	8				

TABLE 26 – Résultats des tests statistiques en intragroupe (corps penché).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
ANOVA ( $p < 0.05$ )	F (2, 42)	1.0611		0.5728	1.8710
	p	0.3551		0.5683	0.1666
Pairwise T-test ( $p < 0.05$ )					
Kruskal-Wallis ( $p < 0.05$ )	Chi-square		0.0216		
	p		0.9892		
	df		2		
Pairwise Mann-Whitney ( $p < 0.05$ )					

TABLE 27 – Résultats des tests statistiques en intergroupe (tous les défauts).

	Jeu 1		Jeu 2		Jeu 3		Jeu 4	
	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type	Moyenne	Écart-Type
Groupe 1	2.0027	0.7263	1.1203	0.4826	1.1103	0.3991	1.1337	0.4221
Groupe 2	1.6542	0.6728	1.0988	0.4676	0.9560	0.4468	0.8767	0.4177
Groupe 3	1.7852	0.4966	1.0702	0.4587	1.0344	0.2784	1.1228	0.3482

TABLE 25 – Moyennes et écarts-types des données des apprenants (tous les défauts).

Les résultats intergroupes pour l'ensemble des défauts sont présentés dans la figure 69, et les résultats en intragroupe sont présentés dans la figure 68.

Dans le cas de l'analyse intragroupe, la figure 68 montre une baisse de l'écart interquartile, de la dispersion et de la valeur de la médiane pour les 3 groupes, et ce dès le jeu 2. Les valeurs restent ensuite stables, sauf pour le groupe 3, où l'écart interquartile se resserre progressivement jusqu'au jeu 3, à partir duquel il s'élargit de nouveau, ainsi que pour les groupes 1 et 2 entre le jeu 3 et le jeu 4. Le groupe 2 montre une forte réduction de la dispersion au jeu 2, et varie peu par la suite. La réduction de la dispersion est plus lente pour les groupes 1 et 3, et se stabilise au niveau du jeu 3. Enfin, l'écart interquartile augmente entre le jeu 3 et le jeu 4, pour tous les groupes. Les résultats sont significatifs pour tous les groupes entre les jeux 1-2, 1-3 et 1-4, ainsi que pour le groupe 2 entre le jeu 2 et le jeu 4, et pour le groupe 3 entre le jeu 3 et le jeu 4, montrant une progression plus lente (table 28).

Pour l'analyse intergroupe, la figure 69 permet d'observer que bien que l'écart interquartile varie largement entre les différents groupes pour le jeu 1, les valeurs des médianes sont sensiblement les mêmes. Elles restent proches les unes des autres pour tous les jeux excepté le jeu 4, où la médiane du groupe 2 baisse alors que les deux autres augmentent. La médiane du groupe 2 est la meilleure dans tous les cas, sauf pour le jeu 2. Aucun résultat ne présente une différence significative (tableau 27).

#### 6.3.4.4 Corrélation entre les défauts et la précision

Afin de vérifier s'il y a une corrélation entre la précision des lancers et l'amélioration des gestes, le  $\tau$  de Kendall (tau de Kendall) a été calculé entre la précision et chaque défaut constaté du geste. Ce coefficient calcule la corrélation de rang entre deux variables  $X$  et  $Y$ , et produit des valeurs pouvant varier de -1 à 1. Une valeur de 1 indique une corrélation positive totale, 0 indique qu'il n'y a aucune corrélation



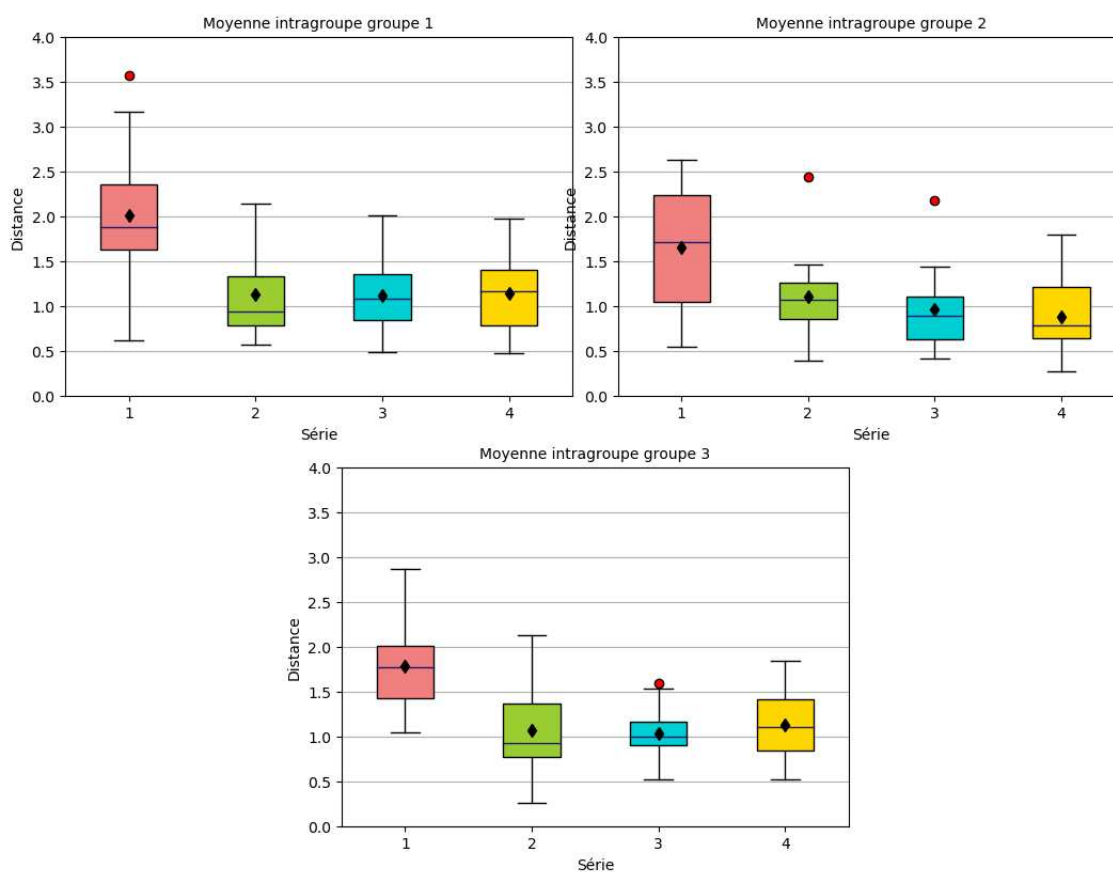


FIGURE 68 – Moyenne des distances de tous les défauts par rapport au centroïde du cluster des données de l'expert en intra-groupe

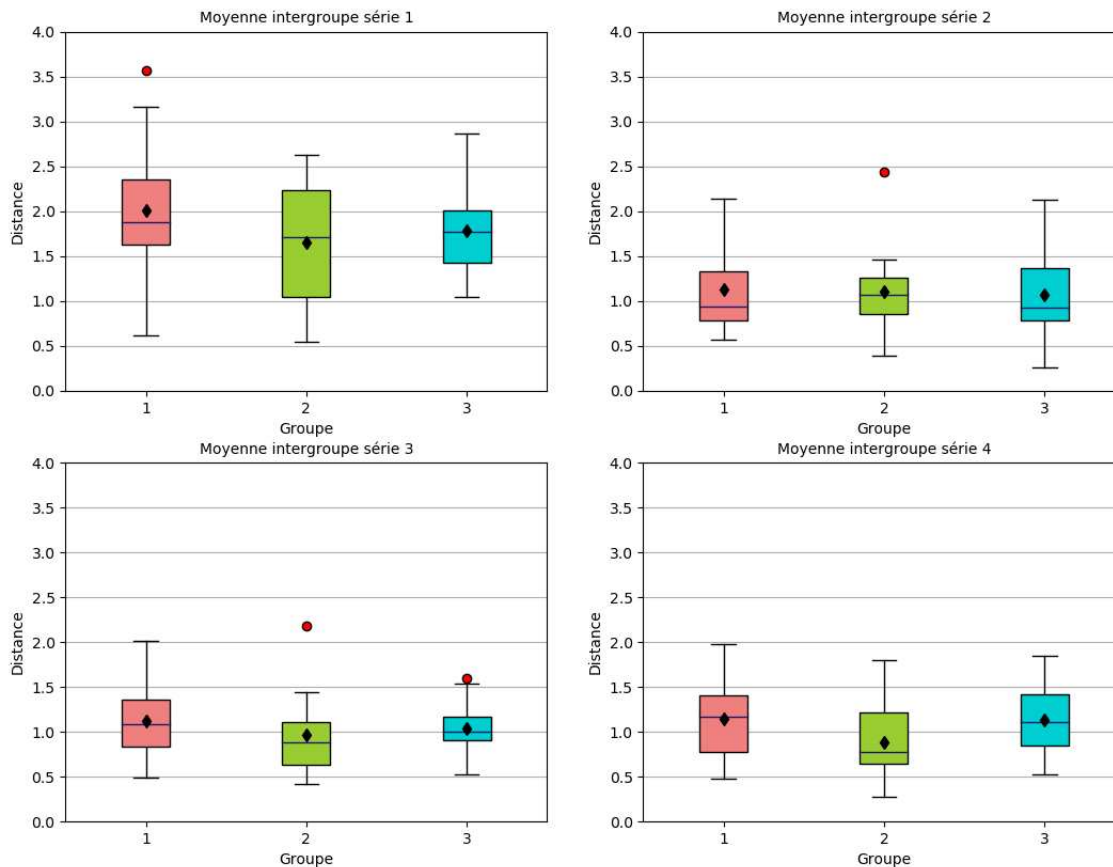


FIGURE 69 – Moyenne des distances de tous les défauts par rapport au centroïde du cluster des données de l'expert en inter-groupe.

		Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
		One tail		Two tails		T	One tail	Two tails	t	df
		p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1						$5.2824e^{-5}$	0.0001	5.3330	14
	G2	0.0067	0.0051	0.0135	0.0102	16	0.0031	0.0064	3.2028	14
	G3						0.0003	0.0007	4.3281	14
Jeu 2-3	G1	0.3991	0.4020	0.7983	0.8039	55	0.4617	0.9234	0.0978	14
	G2	0.0528	0.0535	0.1055	0.1070	31				
	G3						0.3554	0.7109	0.3783	14
Jeu 3-4	G1						0.3567	0.7135	-0.3747	14
	G2	0.1221	0.1262	0.2443	0.2524	39	0.1262	0.2523	1.1940	14
	G3						0.0435	0.0870	-1.8404	14
Jeu 1-3	G1						$8.2423e^{-5}$	0.0002	5.0895	14
	G2	0.0092	0.0075	0.0184	0.0151	18	0.0037	0.0073	3.1343	14
	G3						0.0001	0.0002	4.9613	14
Jeu 2-4	G1	0.3351	0.3394	0.6701	0.6788	52	0.4607	0.9214	-0.1005	14
	G2	0.0191	0.0177	0.0382	0.0353	23	0.0117	0.0235	2.5419	14
	G3						0.2657	0.5314	-0.6418	14
Jeu 1-4	G1						0.0001	0.0003	4.7662	14
	G2	0.0029	0.0017	0.0059	0.0034	11	0.0012	0.0023	3.7136	14
	G3						0.0003	0.0005	4.4723	14

TABLE 28 – Résultats des tests statistiques en intragroupe (tous les défauts).

et -1 indique qu'il y a une corrélation négative totale. Les notations utilisées sont les suivantes :

- P : précision (distance par rapport au centre de la cible),
- A : alignement du bras,
- E : mouvement du coude,
- J : lancer type javelot,
- L : corps qui se penche,
- T : tous les descripteurs.

Les valeurs (ainsi que les p-values) des  $\tau$  de Kendall sont données dans les tableaux 29, 30 et 31. L'hypothèse nulle ( $p > 0.05$ ) indique que les variables sont indépendantes.

	P	A	E	J	L	T
P	X	-0.1675 (0.0590)	-0.2049 (0.0209)	-0.0175 (0.8432)	-0.1211 (0.1722)	-0.1754 (0.0480)
A	-0.1675 (0.0590)	X	-0.0441 (0.6189)	0.1079 (0.2231)	-0.0983 (0.2671)	-0.0407 (0.6461)
E	-0.2049 (0.0209)	-0.0441 (0.6189)	X	0.3012 (0.0007)	0.7333 (0.0000)	0.6554 (0.0000)
J	-0.0175 (0.8432)	0.1079 (0.2231)	0.3012 (0.0007)	X	0.2266 (0.0105)	0.4244 (0.0000)
L	-0.1211 (0.1722)	-0.0983 (0.2671)	0.7333 (0.0000)	0.2266 (0.0105)	X	0.6056 (0.0000)
T	-0.1754 (0.0480)	-0.0407 (0.6461)	0.6554 (0.0000)	0.4244 (0.0000)	0.6056 (0.0000)	X

TABLE 29 – Valeurs et p-values (entre parenthèses) des tau de Kendall entre les défauts du Groupe 1.

	P	A	E	J	L	T
P	X	0.0322 (0.7162)	-0.0639 (0.4711)	0.0187 (0.8333)	-0.0718 (0.4179)	-0.1012 (0.2536)
A	0.0322 (0.7162)	X	0.1254 (0.1568)	0.1345 (0.1290)	0.0520 (0.5574)	0.2056 (0.0203)
E	-0.0639 (0.4711)	0.1254 (0.1568)	X	0.3243 (0.0003)	0.7322 (0.0000)	0.6938 (0.0000)
J	0.0187 (0.8333)	0.1345 (0.1290)	0.3243 (0.0003)	X	0.2418 (0.0063)	0.4814 (0.0000)
L	-0.0718 (0.4179)	0.0520 (0.5574)	0.7322 (0.0000)	0.2418 (0.0063)	X	0.6565 (0.0000)
T	-0.1012 (0.2536)	0.2056 (0.0203)	0.6938 (0.0000)	0.4814 (0.0000)	0.6565 (0.0000)	X

TABLE 30 – Valeurs et p-values (entre parenthèses) des tau de Kendall entre les défauts du Groupe 2.

	P	A	E	J	L	T
P	X	0.0368 (0.6784)	-0.0708 (0.4252)	0.1274 (0.1512)	-0.0804 (0.3650)	-0.1240 (0.1624)
A	0.0368 (0.6784)	X	-0.1887 (0.0332)	0.0034 (0.9695)	-0.1057 (0.2330)	0.1751 (0.0480)
E	-0.0708 (0.4252)	-0.1887 (0.0332)	X	-0.0554 (0.5319)	0.6889 (0.0000)	0.4734 (0.0000)
J	0.1274 (0.1512)	0.0034 (0.9695)	-0.0554 (0.5319)	X	-0.1532 (0.0839)	-0.1209 (0.1723)
L	-0.0804 (0.3650)	-0.1057 (0.2330)	0.6889 (0.0000)	-0.1532 (0.0839)	X	0.5883 (0.0000)
T	-0.1240 (0.1624)	0.1751 (0.0480)	0.4734 (0.0000)	-0.1209 (0.1723)	0.5883 (0.0000)	X

TABLE 31 – Valeurs et p-values (entre parenthèses) des tau de Kendall entre les défauts du Groupe 3.

On remarque qu'aucune corrélation positive ou négative significative n'a été détectée entre la variation d'un défaut et la distance des fléchettes par rapport au centre de la cible, sauf pour le groupe 1 pour le mouvement du coude ( $\tau = -0.2049$  et  $p = 0.0209$ ) et la totalité des descripteurs ( $\tau = -0.1754$  et  $p = 0.0480$ ). Les valeurs des tau sont cependant faibles dans les deux cas. Cela indique que, dans le cadre de cette expérimentation, l'amélioration ou la détérioration d'un aspect du geste de l'apprenant n'est pas corrélée à l'amélioration ou la détérioration des autres aspects. Il y a cependant une corrélation forte entre le défaut du mouvement du coude et du corps qui se penche en avant pour tous les groupes ( $\tau \approx 0.7$  et  $p = 0$ ). Cela peut s'expliquer par la présence d'un descripteur commun aux deux défauts : la

vitesse moyenne de l'épaule (du côté de la main dominante de l'apprenant). Ainsi, la correction du défaut du corps qui se penche en avant implique la diminution de cette valeur pour les deux défauts. Il existe également une forte corrélation entre le défaut du lancer type javelot et du mouvement du coude en avant pour les groupes 1 et 2, ainsi qu'entre le corps qui se penche et le lancer type javelot pour ces deux groupes.

Concernant la corrélation de tous les descripteurs ramenés en dimension 2 grâce à l'analyse en composantes principales, on remarque pour le groupe 1 une corrélation ( $p = 0$ ) avec le mouvement du coude ( $\tau = 0.6554$ ), le lancer javelot ( $\tau = 0.4244$ ) et le corps qui se penche ( $\tau = 0.6056$ ). Pour le groupe 2, la corrélation ( $p = 0$ ) est avec le mouvement du coude ( $\tau = 0.6938$ ), le lancer type javelot ( $\tau = 0.4814$ ) et le corps penché lors du lancer ( $\tau = 0.6565$ ). Enfin, pour le groupe 3, la plus grande corrélation ( $p = 0$ ) est avec le mouvement du coude ( $\tau = 0.4734$ ) et le corps penché lors du lancer ( $\tau = 0.5883$ ). Ces résultats suggèrent que les caractéristiques les plus significatives pour l'analyse en composantes principales des mouvements varient pour chaque groupe, avec cependant en commun le mouvement du coude, ainsi que le corps penché lors du lancer.

#### 6.3.4.5 Réponses aux questionnaires

La figure 70 montre la répartition des réponses des participants, classées par ordre de satisfaction de très satisfait (7) à pas du tout satisfait (1). Le tableau 32 montre la moyenne des réponses des apprenants au pré-questionnaire et au post-questionnaire, par groupe et de manière globale. Parmi les 45 participants, aucun n'était expert du lancer de fléchettes, et la majorité d'entre eux avaient déjà joué aux fléchettes en tant que loisir (niveau aux fléchettes :  $2.33 \pm 0.85$ , 41 personnes ayant donné une réponse de 3 ou moins). Le niveau de stress des apprenants était systématiquement bas ( $6,111 \pm 1,335$ ), avec un écart-type élevé dû à quelques apprenants stressés par le contexte de l'expérimentation. Les instructions données étaient claires et compréhensibles (niveau de compréhension des instructions reçues :  $6.467 \pm 0.815$ ). L'auto-évaluation de la progression de la performance, ainsi que de la progression de la qualité du geste, présentent un écart-type élevé, montrant ainsi une difficulté à quantifier la progression par l'apprenant ( $4.11 \pm 1.45$  pour la progression de la performance et  $4.378 \pm 1.27$  pour la progression de la qualité du geste). Il semble que cet écart corresponde à une sous-évaluation quasi-systématique des apprenants. La corrélation linéaire calculée entre les réponses des différents groupes (tableau 33) montre que, quel que soit le groupe, les réponses suivent toutes la même tendance. Ainsi, le ressenti de la progression du geste et de la précision, ainsi que de l'intérêt des conseils reçus, était le même, quel que soit le groupe considéré.

		Niveau Fléchettes	Aisance Matériel	Stress (7 = no stress)	Compréhension	Intérêt des conseils	Performance globale	Progression Performance	Qualité Globale	Progression Qualité
G1	Moyenne	2,533	5,733	6,067	6,400	5,533	4,400	3,867	3,600	3,933
	Écart-type	1,125	0,884	1,163	0,737	0,743	1,121	1,506	0,828	1,100
G2	Moyenne	2,200	5,400	6,133	6,600	5,200	4,600	4,133	3,933	4,533
	Écart-type	0,775	1,056	1,457	0,828	1,014	1,242	1,727	0,799	1,506
G3	Moyenne	2,267	5,867	6,133	6,400	5,667	4,133	4,333	4,333	4,667
	Écart-type	0,594	0,640	1,457	0,910	1,397	0,834	1,113	0,488	1,113
Tous	Moyenne	2,333	5,667	6,111	6,467	5,467	4,378	4,111	3,956	4,378
	Écart-type	0,853	0,879	1,335	0,815	1,079	1,072	1,449	0,767	1,267

TABLE 32 – Moyenne et écart-type des réponses aux questions en fonction de chaque groupe.

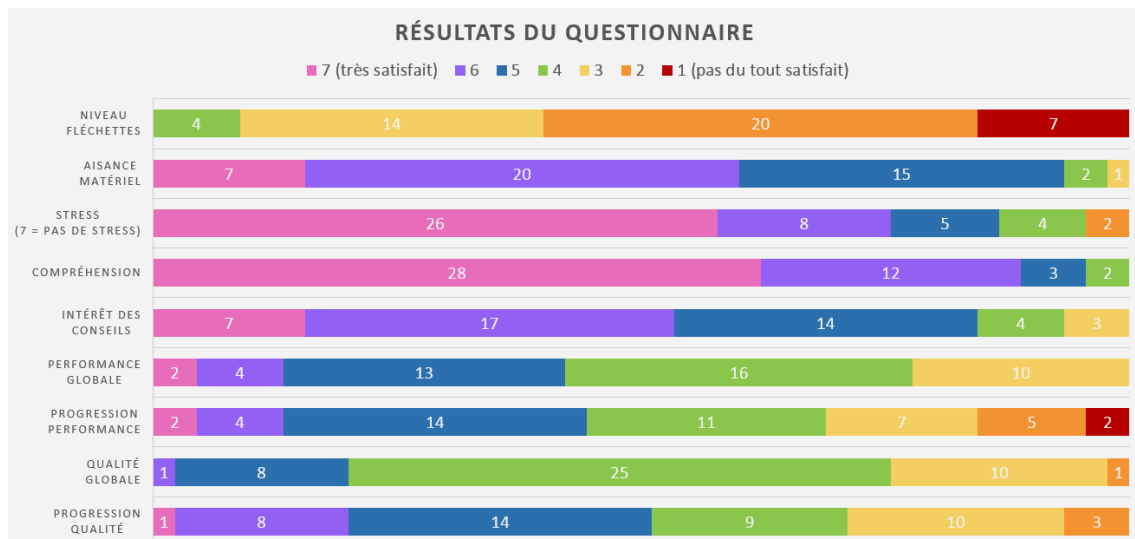


FIGURE 70 – Le résultat des pré-/post-questionnaires par nombre de réponses pour chaque question, classées par ordre de satisfaction de très satisfait (7) à pas du tout satisfait (1).

Corrélations	G1	G2	G3
G1	1	0,9668	0,9582
G2	0,9668	1	0,9709
G3	0,9582	0,9709	1

TABLE 33 – Indices de corrélations entre les moyennes des réponses aux questionnaires pour chaque groupe.

#### 6.3.4.6 Discussion

Des résultats significatifs ont été obtenus pour l'amélioration de certains défauts entre le jeu 1 et le jeu 4, à savoir les défauts *leaning* et *elbow move* pour tous les groupes, et *javelin* et *align arm* pour le groupe 2 seulement. Une explication possible est que ces deux derniers conseils n'ont pas été donnés par l'expert (c'est-à-dire la personne faisant passer les expérimentations), soit car d'autres défauts étaient plus proéminents, soit car ces défauts sont difficilement observables à l'œil nu. Ce dernier cas est plus susceptible d'arriver pour le groupe 1, où le système n'intervient pas. Il n'est pas possible d'utiliser une aide pour déterminer quels sont les défauts à corriger, lorsque l'observation à l'œil nu devient plus difficile. Ces résultats indiquent que les conseils donnés sont pertinents au regard des défauts considérés, et que les indicateurs calculés, ainsi que l'ensemble de règles choisies pour la sélection des conseils à donner, permettent de mettre en évidence les défauts du geste des apprenants, et ainsi les corriger. L'hypothèse **H4** est validée dans le contexte de cette expérimentation.

Les résultats de l'analyse de la corrélation des défauts entre eux, ainsi que de la distance par rapport au centre de la cible, montrent que l'amélioration de la distance des fléchettes par rapport au centre de la cible n'est pas corrélée avec l'amélioration du geste en lui-même. Cette observation est cohérente avec les résultats obtenus par d'autres études analysant le geste (LE NAOUR, HAMON et BRESCIANI 2019; ASHFORD, BENNETT et DAVIDS 2006). L'amélioration du geste se fait, dans un premier temps, au détriment de la précision du lancer. Cependant, il y a une dépendance, pour tous les groupes, entre le défaut du mouvement du coude et le défaut du corps qui se penche

en avant, à cause de l'utilisation d'un descripteur commun. Ainsi, la correction d'un de ces deux défauts entraîne probablement une amélioration de l'autre.

La présence de données aberrantes persistantes sur l'ensemble des jeux pour certains défauts peut provenir d'un manque de conseil pour les défauts dans lesquels ils se trouvent. Ainsi, si d'autres défauts sont plus proéminents, l'expert ou le système ne donnera pas le conseil du défaut en question, et il n'y a pas d'amélioration du geste au regard de ce critère. Cela peut également indiquer que ce défaut n'est pas dépendant des autres, et que la correction d'autres défauts n'influence pas le défaut considéré.

Le défaut du lancer de type javelot présente une grande variabilité dans ses valeurs pour le groupe 1 (Fig. 64 et 65). Cette variabilité pourrait s'expliquer par le fait que ce conseil n'est que très peu donné par l'expert, car plus difficilement visualisable que les autres.

Les analyses intergroupes ne permettent cependant pas de mettre en valeur une différence significative dans l'évolution du geste des apprenants au cours de l'expérimentation. Ainsi, l'expérimentation ne permet pas de conclure quant à l'amélioration de l'apprentissage du geste à l'aide du système développé en utilisation seule ou conjointement avec l'expert. Les hypothèses **H5** et **H6** n'ont pas été validées au cours de cette expérimentation.

#### 6.3.4.7 *Limites de l'expérimentation*

Cette expérimentation, bien qu'ayant permis de montrer que le système permettait d'améliorer la correction des défauts des gestes des apprenants, présente plusieurs limites.

Tout d'abord, lors de la captation des gestes de l'expert, un phénomène d'auto-correction peut apparaître. Lorsque l'expert réalise des gestes présentant des défauts, l'enchaînement des captations peut mener à un certain manque d'attention et à un geste qui se rapproche du geste fait par l'expert en temps normal (c'est-à-dire le bon geste). Ainsi, la différence entre certains gestes bons et d'autres supposés mauvais peut apparaître comme étant très faible, voire invisible à l'œil nu. Ce problème est en partie pallié par l'utilisation du clustering, permettant de regrouper les gestes similaires ensemble.

Le nombre de lancers effectués (36) peut ne pas être suffisant pour pouvoir juger d'une évolution du mouvement sur la durée. Les processus d'apprentissage du geste nécessitent généralement plus de temps et de séances. Aussi, il peut être difficile d'obtenir des améliorations significatives lorsque l'on se base sur une seule séance de 36 lancers. De plus, entre chaque série de neuf lancers, il y avait une interruption afin que les données soient exportées puis traitées par le système. Cette limitation est propre au logiciel de traitement du mouvement Axis Neuron qui doit être utilisé lors de la captation avec la combinaison de capteurs, et non au système MLA. L'interruption des répétitions pendant quelques minutes peut mener à une « perte » de perception des caractéristiques du bon mouvement de l'apprenant, résultant ainsi en une perte de précision ou une dégradation de certains aspects du geste.

Une des limites du protocole expérimental exposé dans ce chapitre est qu'il en résulte systématiquement deux conseils à donner, même dans le cas où un seul (ou aucun) défaut est présent dans le geste. Ainsi, cela peut mener à une dégradation du geste sur la série suivante, car l'apprenant tente de corriger un défaut qui n'est pas ou plus présent. En pratique, ce cas ne s'est présenté que pour une seule série d'un apprenant.

Dans le cadre de l'expérimentation, l'analyse de l'expert est également plus difficile que dans un cadre d'apprentissage classique. En effet, le lancement et l'arrêt de la capture se font pour chaque lancer par l'expert, et cela implique une position proche du système informatique. Cette position ne permet pas de se déplacer autour de l'apprenant lors de la réalisation des gestes.

Dans un cadre d'apprentissage du lancer de fléchettes, l'expert analyse également des données extérieures au mouvement. Par exemple, le mauvais alignement du bras est détecté par l'expert en regardant la position des fléchettes sur la cible. L'expert donne comme tâche à l'apprenant de viser la partie « 20 » de la cible (localisé au milieu, allant du centre au bord supérieur de la cible) : si toutes les flèches tombent d'un côté, l'expert déduit que le bras penche vers ce côté lors du tir. Notre étude portant sur le geste en lui-même, nous avons dû définir des descripteurs provenant du geste de l'apprenant capable de mettre en relief ce défaut. D'autres analyses peuvent porter également sur le moment du lâcher de la fléchette, mais aussi sur la manière de la saisir avec les doigts.

Enfin, l'évaluation du système suppose que l'apprenant a systématiquement appliqué les conseils à la lettre ; en pratique, ce n'est pas toujours le cas. Les conseils peuvent soit être difficiles à comprendre pour l'apprenant, soit difficile à mettre en œuvre, et l'apprenant peut juger le conseil comme étant inutile pour l'amélioration de l'objectif du geste.

#### 6.4 SYNTHÈSE DES CONTRIBUTIONS

Ce chapitre a présenté les trois expérimentations conduites dans le cadre de la validation du système MLA. Chaque expérimentation visait à vérifier une des facettes du système :

- la capacité du système à séparer les gestes selon un étiquetage vérifiable par un humain à l'aide des descripteurs et des algorithmes choisis,
- la possibilité d'obtenir une séparation nette des données en deux clusters, correspondant au degré de réussite du geste,
- la capacité du système à assister l'enseignant dans sa tâche d'apprentissage d'un geste par l'analyse de ses besoins d'observation et démonstrations afin de déterminer et opérationnaliser les :
  - descripteurs représentatifs du geste,
  - propriétés de séparation des gestes acceptables de ceux qui ne le sont pas.

Ces expérimentations ont ainsi permis de valider les hypothèses suivantes :

- **H1** : Il est possible de regrouper les gestes selon leurs propriétés cinématiques communes.
- **H3** : Il est possible de séparer les gestes en fonction de propriétés attendues et identifiées au préalable par l'expert.
- **H4** : Il est possible de corriger chaque défaut du geste de l'apprenant, en lui indiquant les défauts majeurs à corriger en premier. Un défaut majeur est identifié par la plus grande distance séparant le mouvement courant de l'apprenant, du groupe de gestes acceptables ayant éliminé ce défaut.

Les hypothèses non-validées sont les suivantes :

- **H2** : Il est possible de séparer les gestes des apprenants en deux groupes correspondant à une dichotomie geste réussi / geste raté afin de déterminer, pour une situation d'apprentissage donnée, les propriétés d'un ensemble fini de gestes réussis.



- **H5** : L'utilisation du système MLA basée sur l'hypothèse 4 permet d'améliorer l'apprentissage du geste par rapport à une situation sans le système MLA.
- **H6** : L'utilisation du système MLA en tant qu'assistant à l'enseignant permet d'améliorer l'apprentissage du geste par rapport à une situation sans MLA, et à une situation avec MLA et sans enseignant.

La première expérimentation a permis de montrer qu'il était possible d'obtenir une séparation acceptable (au sens des métriques calculées) des gestes de l'apprenant, en fonction des caractéristiques propres aux gestes observables par un humain. Bien qu'il ait été possible d'obtenir une séparation correcte dans la deuxième expérimentation, cette séparation ne correspondait pas à la dichotomie geste réussi / raté. Plusieurs pistes ont été évoquées pour expliquer ce résultat, par exemple l'absence d'une expertise du geste considéré, l'utilisation de descripteurs uniquement basés sur des propriétés cinématiques du geste. La troisième expérimentation a permis de vérifier si le système était capable d'assister un enseignant dans sa tâche d'enseignement d'un geste en intégrant ses besoins d'observation en tant que propriétés caractérisant les gestes d'une part, et permettant de regrouper les gestes acceptables et les gestes présentant les défauts préalablement identifiés d'autre part. Il y a eu une amélioration du geste dans tous les cas d'utilisation du système MLA (expert seul, système seul, expert assisté du système), avec une amélioration sur les quatre défauts considérés pour l'utilisation du système en autonomie. Cependant, il n'y a pas eu de différences significatives dans l'amélioration du geste entre les groupes.

L'amélioration de la distance par rapport au centre de la cible n'était cependant pas corrélée avec l'amélioration des défauts du geste. Ces résultats sont en accord avec ceux de la littérature (LE NAOUR, HAMON et BRESCIANI 2019; ASHFORD, BENNETT et DAVIDS 2006), qui montrent que l'amélioration du geste implique une perte de précision quant à l'objectif visé, dans le cadre du geste de lancers.

Les résultats obtenus au cours de la troisième expérimentation montrent que le système permet l'apprentissage du geste de manière autonome, ou d'assister l'expert dans sa tâche d'évaluation du geste. Cependant, dans le contexte de l'expérimentation, l'amélioration du geste n'était pas significativement meilleure (au sens des tests statistiques effectués) selon les différents groupes de participants. L'agrandissement de la population, afin d'obtenir des résultats plus fiables, ainsi que le test du système sur un autre geste, permettraient de confirmer ou d'infirmer ces tendances.



Quatrième partie

CONCLUSION



---

## BILAN ET PERSPECTIVES

---

Ce travail de thèse s'inscrit dans le cadre de la conception et l'utilisation d'Environnements Informatiques pour l'Apprentissage Humain afin d'aider à l'apprentissage de gestes. Les EIAH dédiés à l'apprentissage de gestes existants sont souvent limités dans leur capacité à être réutilisés dans des contextes applicatifs variés. La recherche de la généricité d'un tel système doit se faire dès sa conception, et il doit en particulier être en mesure de proposer une intégration aisée de la connaissance experte. Ainsi, notre objectif a été de créer un EIAH générique au regard du domaine considéré, permettant d'assister l'apprenant dans sa tâche d'apprentissage du geste, et l'enseignant dans son analyse du mouvement et sa restitution à l'apprenant.

### 7.1 SYNTHÈSE DES CONTRIBUTIONS

Nous avons proposé un EIAH dédié à l'apprentissage de gestes et conçu de manière à être générique. Cet EIAH est divisé en deux grandes parties. La première permet d'importer, de filtrer et d'extraire des descripteurs à partir de fichiers de mouvements au format BVH (BioVision Hierarchy). La chaîne de traitement proposée n'agit que sur les descripteurs extraits du mouvement, et non sur le mouvement en lui-même, corrigeant ainsi les irrégularités directement sur les valeurs des descripteurs. En effet, la correction de ces erreurs en amont nécessiterait de faire des hypothèses quant aux causes des problèmes présents au sein des données. Particulièrement adaptée aux gestes possédant au moins une accélération soudaine (comme les gestes de lancers), elle permet également de ne conserver automatiquement qu'une partie du mouvement, appelée *Motion of Interest (MoI)*, afin d'éviter une segmentation manuelle dans le cas où plusieurs gestes similaires sont à traiter. L'extraction des descripteurs requis se fait à l'aide de structures de données internes adaptées à l'export au format JSON.

La deuxième partie du système permet la comparaison des mouvements de l'apprenant à ceux de l'expert. L'affichage des différences entre ces mouvements apporte une aide à l'enseignant conseillant l'apprenant en vue de l'amélioration de son geste. Le système permet d'afficher et de quantifier l'importance des défauts du geste de l'apprenant, et permet d'établir une stratégie de conseil en ciblant les défauts à corriger en premier. Les données de l'expert, servant de base pour la comparaison, peuvent être regroupées à l'aide de techniques de clustering, afin d'identifier les propriétés des groupes constitués : (i) majoritairement de gestes acceptables ou (ii) majoritairement des gestes non acceptables. La visualisation des données de l'apprenant et de l'expert permet de rapidement déterminer quels sont les défauts les plus proéminents dans les gestes de l'apprenant, et ainsi de proposer une correction adaptée. L'intégration des conseils donnés par l'expert pour chacun des défauts identifiés est également possible et permet, dans une certaine mesure, d'utiliser le système en autonomie.

Les expérimentations menées ont permis de tester plusieurs aspects du système. La première expérimentation, concernant le lancer d'une balle dans deux corbeilles différentes, a montré qu'il était possible de séparer les données à l'aide de descripteurs basés sur la vitesse dans le cas de ce lancer. La deuxième expérimentation, utilisant le geste du Bottle Flip Challenge, cherchait à vérifier s'il était possible d'obtenir une séparation correspondant au degré de réussite du geste, tout en utilisant des descripteurs basés sur la vitesse. Les résultats n'ont pas été concluants sur le degré de réussite, bien qu'un regroupement acceptable a été obtenu. La troisième expérimentation, portant sur le lancer de fléchettes, avait pour objectif de vérifier la capacité du système à être utilisé afin de prodiguer des conseils pertinents au regard des défauts du geste des apprenants. Les participants à l'expérimentation ont été séparés en trois groupes : (a) un groupe où seul l'expert analysait le geste et donnait des conseils pour son amélioration, (b) un groupe où l'expert pouvait s'appuyer, si besoin, sur le système afin de donner des conseils et (c) un groupe où l'expert restituait les conseils donnés par le système. L'analyse s'est effectuée sous deux angles principaux : l'amélioration de l'objectif du geste (distance des fléchettes par rapport au centre), et l'amélioration du geste en lui-même (distance des données de l'apprenant par rapport au centroïde des bon gestes de l'expert). Ainsi, bien que la précision n'ait pas augmenté, une amélioration du geste a été constatée sur chaque aspect du geste considéré (en termes de correction de défauts), ainsi que sur le geste dans sa globalité au fur et à mesure de l'expérimentation. L'utilisation du système par un expert pour l'assister dans sa tâche d'enseignement a également conduit à une amélioration du geste. Cependant, aucune différence significative au regard de l'amélioration du geste n'a été obtenue entre les différents groupes, ne permettant pas ainsi de conclure quant à la valeur ajoutée apportée par le système dans le cadre d'une session courte d'apprentissage de cette tâche.

Ce travail présente des limites, et ouvre la porte à de nombreuses perspectives d'ingénierie et de recherche.

## 7.2 LIMITES DES TRAVAUX ET DU SYSTÈME MLA

Une première limite du travail réside dans le manque d'interface graphique pour l'utilisation de la plateforme *MLA*. En l'état, la sélection du dossier contenant les données de l'apprenant pour la phase de pré-traitement se fait directement au sein du code. Ainsi, il est nécessaire de re-compiler le code pour chaque nouvelle personne. De plus, la sélection des données des personnes à traiter, ainsi que les descripteurs associés aux défauts à analyser se font également à la main dans le code Python. Bien qu'une compilation ne soit pas nécessaire pour cet aspect, ce processus est tout autant fastidieux. Une interface graphique, soit séparée selon les deux parties du système, soit unifiée, permettrait de grandement faciliter la manipulation de l'outil, en limitant les compétences techniques demandées à l'utilisateur. Une perspective pour combler cette limite actuelle est proposée dans la section suivante.

La pertinence des indicateurs à visualiser (descripteurs du mouvement, métrique de clustering, *etc.*), ainsi que la visualisation proposée à l'expert ou l'apprenant reste à déterminer, par le biais d'expérimentation avec des personnes non-expertes dans l'utilisation du système. La visualisation actuellement proposée répond aux besoins d'observations de l'expérimentation réalisée, mais sa généralisation à d'autres domaines applicatifs n'est pas garantie.

L'utilisation de la combinaison Perception Neuron pour la capture des données nécessite des étapes de filtrage des données, afin qu'elles soient exploitables. Bien que les techniques de filtrage des données mises en places permettent d'obtenir des

données acceptables dans le contexte des expérimentations réalisées, ces méthodes n'ont pas été conçues pour traiter n'importe quels types de gestes. L'utilisation de méthodes de captation de meilleure qualité permettrait d'obtenir des données plus précises, et de s'affranchir des filtrages mis en place.

Le faible nombre de lancers, ainsi que la réalisation d'une seule séance d'apprentissage du geste, pour le lancer de fléchettes ne permet pas de juger correctement de la capacité du système à assister un apprenant ou un expert sur le long terme. En effet, dans une situation d'apprentissage classique, l'apprentissage s'effectue sur plusieurs séances, et le nombre de lancers par séance est beaucoup plus conséquent.

### 7.3 PERSPECTIVES SUR LE PROJET MLA

#### 7.3.1 Interface graphique

La partie en charge du pré-traitement des données doit bénéficier d'une interface graphique, afin de choisir quels descripteurs sont à extraire, ainsi que les paramètres des différents algorithmes : filtrage, mouvement à extraire (possibilité de directement choisir sur un graphique), *etc.* Cette partie nécessite pour l'instant des connaissances (i) sur les données de mouvement, (ii) scientifiques pour le choix des paramètres des algorithmes et des filtres et (iii) techniques pour la modification de ces paramètres. En proposant une visualisation du résultat du filtrage et du mouvement à extraire, il est possible de diminuer la complexité d'utilisation du système et ainsi permettre à des non-experts de modifier ces paramètres.

Dans l'optique de faciliter l'utilisation du système *MLA*, une première maquette d'une interface graphique pour la partie donnant des retours sur le mouvement a été réalisée (Fig. 71). La configuration de l'analyse se fait selon l'ordre établi auparavant, à savoir descripteur puis articulation.

L'intégration des différents retours proposés à l'enseignant dans les dernières versions de *MLA* (visualisation de la distance des données de l'apprenant par rapport aux données de l'expert, métriques calculées sur les groupements obtenus, *etc.*) sera l'objet d'une autre itération de cette interface.

#### 7.3.2 Explication des données non-alignées de l'apprenant

Une des difficultés d'interprétation du retour visuel fourni à l'expert se situe dans le cas où les données de l'apprenant ne sont pas situées dans un des clusters de mouvements de l'expert et qu'elles n'appartiennent pas au trapèze formé par les deux clusters (Fig. 72, partie verte claire).

Cette différence, lorsque les données utilisées sont de dimension 2, peut s'expliquer visuellement, sous réserve de posséder les connaissances scientifiques nécessaires à l'analyse de ces données. En effet, dans ce cas, il est possible d'obtenir directement quelles sont les différences entre les valeurs des descripteurs de l'expert et ceux de l'apprenant. Cependant, lorsque la dimensionnalité des données est plus élevée, le non-alignement des données de l'apprenant avec celles de l'expert ne peut s'expliquer aussi facilement. En effet, l'affichage des données est actuellement réalisé sur un graphique possédant uniquement deux dimensions. Il serait possible d'étendre cette visualisation en 3D, mais le problème se pose toujours pour les données de dimension supérieure. Il est donc nécessaire de réduire la dimensionnalité des données lorsque celle-ci est supérieure à deux. Cette étape implique une réduction des données, et les valeurs représentées sur les deux axes du graphique ne peuvent



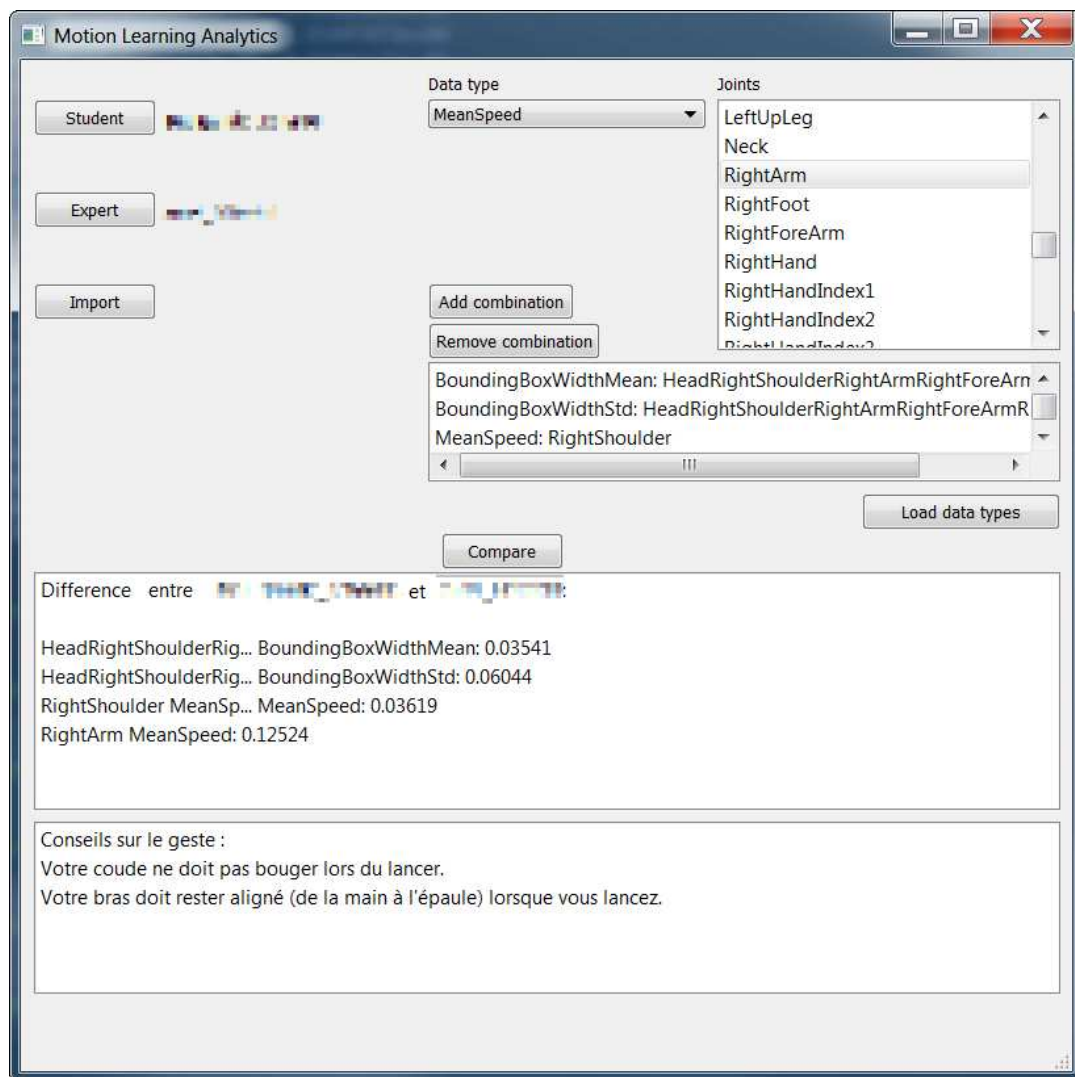


FIGURE 71 – Une première itération de l’interface graphique des modules d’analyse de MLA. Les types de données possibles, ainsi que les articulations disponibles pour les mouvements capturés sont directement extraits des données retournées par les modules de pré-traitements au format *JSON*.

être corrélées à une caractéristique du geste si elle est représentée par un vecteur à plus de deux dimensions dans le système actuel. En effet, elles peuvent être issues de la fusion de plusieurs caractéristiques possédant un pouvoir de discrimination entre les mouvements plus faible, ou encore de la suppression de certaines de ces caractéristiques, en fonction de la méthode choisie.

Ainsi, une des pistes de travail est d’étudier la sémantique et l’origine des données non-alignées. L’objectif sera de proposer une comparaison des différents descripteurs utilisés. Une première étape réside dans le calcul de la moyenne des écarts de chaque dimension des données de l’apprenant avec chaque dimension des données de l’expert étiquetée comme geste cible, afin d’obtenir une valeur correspondant à la différence brute entre ces données. Ainsi, il serait possible de proposer une explication sur les différences entre les gestes de l’expert et ceux de l’apprenant. La deuxième étape serait ensuite de travailler sur le sens de cette différence, autant pour l’expert que pour l’apprenant. Cette étape implique d’être capable de donner une sémantique pédagogiquement utile.

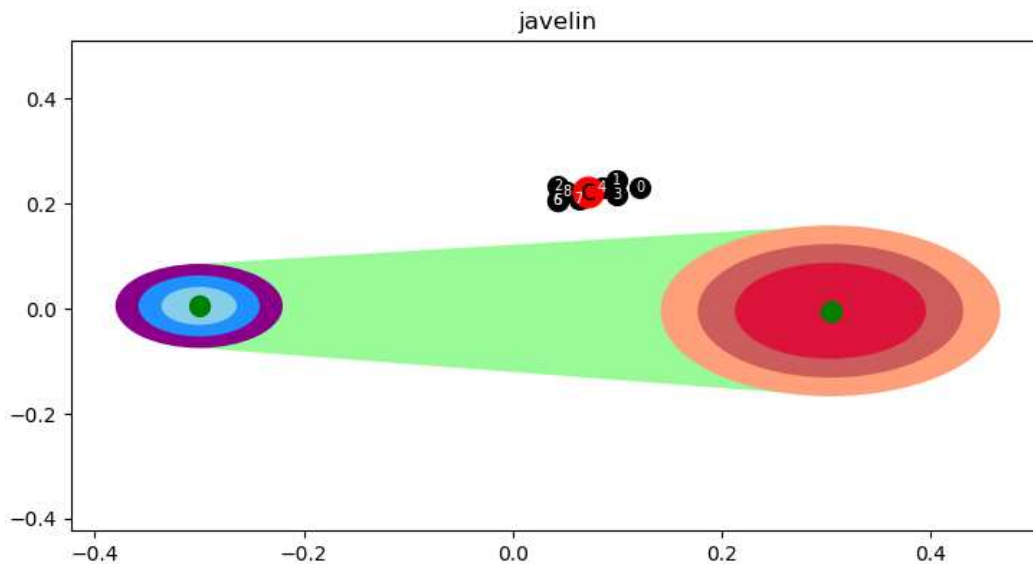


FIGURE 72 – Dans cet exemple, on peut constater que les données de l'apprenant ne se situent pas dans le trapèze formé par les deux clusters de l'expert. L'explication en détail du défaut est difficile dans ce cas.

### 7.3.3 Création de nouveaux descripteurs à l'aide d'un système fondé sur des patrons de conception

L'implémentation des descripteurs élémentaires proposés au Chapitre 3 permet de limiter le nombre de descripteurs à implémenter, tout en permettant de couvrir des besoins d'observation et d'analyse variés. En pratique, il est possible que les besoins de l'expert ne concernent qu'une dimension d'un descripteur élémentaire, ou au contraire une combinaison de plusieurs de ces descripteurs. Par exemple, la hauteur de la hanche est la dimension verticale du descripteur position de la hanche. Une des solutions possibles est de proposer l'extraction de composantes de descripteurs existants (la valeur en  $x$  de du vecteur vitesse par exemple), ou la fusion de descripteurs déjà implémentés. Cette méthode pourrait s'appuyer sur une approche basée sur un langage créé pour ce cas d'usage, comme présentée dans le chapitre 3, section 3.3 : en spécifiant un ou une combinaison de descripteurs, l'intervalle ou les postures sur lesquelles il est choisi, il serait possible de créer de nouveaux descripteurs. Plusieurs exemples :

```
DESCRIPTOR Hauteur hanche: POSITION_ALL[ 2 ]
```

pour la hauteur de la hanche (valeur en  $y$ ) calculée sur tout le mouvement.

```
DESCRIPTOR Vitecceleration: SPEED_ALL[ 2 ] +  
ACCELERATION_ALL[ 1 ]
```

pour calculer l'addition de la composante en  $y$  de la vitesse et de la composante en  $x$  de l'accélération.

DESCRIPTOR Vitecceleration constante:  $\text{SPEED\_ALL}[3] * \text{CONSTANT\_ACCELERATION\_SPECIFIC}[15][1]$

permettrait de calculer la vitesse sur l'axe  $z$  sur tout le mouvement, multipliée par la valeur d'accélération sur l'axe  $x$  de la frame 15.

#### 7.3.4 *Expérimentation à plus grande échelle*

Les expérimentations présentées dans ce manuscrit ont pris place dans un espace de temps très court pour chaque participant. Ainsi, pour le *Bottle Flip Challenge*, les cent lancers ont été réalisés à la suite. Pour le lancer de fléchettes, seuls trente-six lancers ont été réalisés par personne, ce qui est très peu au regard d'une séance d'apprentissage classique. Dans ces conditions, l'amélioration du geste peut être marginale, voire inexistante, à cause de la fenêtre de temps très réduite pour l'évaluation. En conséquence, le système devra être testé sur une multitude de séances d'entraînement, de durée plus conséquente, afin de prouver ou non sa capacité à assister l'expert dans son évaluation du geste de l'apprenant dans un contexte d'apprentissage réel. De plus, une telle évaluation permettrait de tester l'utilisabilité du système au travers des interfaces à développer évoquées précédemment.

#### 7.3.5 *Expérimentation d'apprentissage en autonomie*

Il est également intéressant de se pencher sur l'utilisation de ce système en autonomie, c'est-à-dire un apprenant seul conseillé par le système. L'intégration des retours de l'expert, en fonction des différences du geste de l'apprenant par rapport à ceux de l'expert, pourra s'effectuer en même temps que la sélection des descripteurs à observer. Le système pourra donc être utilisé en autonomie par l'apprenant. L'analyse de la progression du geste à travers une utilisation en autonomie et la comparaison avec la progression en présence d'un expert peut être à même de donner des pistes d'améliorations pour les conseils donnés par le système. La prolongation de l'expérimentation du lancer de fléchettes sur plusieurs séances pour chaque personne permettrait de vérifier l'évolution du geste, ainsi que la rapidité de l'évolution du geste, au fur et à mesure des séances d'apprentissage, et permettrait ainsi de comparer ces évolutions entre les différents groupes (expert / expert assisté du système / système).

#### 7.3.6 *Extension des expérimentations à d'autres gestes*

Bien que pensé pour être générique dès sa conception, la généricité du système devra être améliorée afin de répondre à plusieurs cas d'études impliquant des gestes différents dans plusieurs domaines. Les domaines applicatifs retenus devront cependant respecter plusieurs critères :

- Existence d'expertise dans le domaine : la présence d'un expert reconnu (soit par des diplômes, soit par ses pairs) permet de s'assurer de l'existence d'une connaissance experte.
- Possibilité de formaliser la connaissance experte en descripteurs : cette étape est cruciale, car c'est elle qui déterminera quelles seront les caractéristiques du

geste à analyser. La sélection des descripteurs à calculer et à utiliser se fait à cette étape, avec l'aide de l'expert.

- Séances d'apprentissage contenant les mêmes gestes : la similarité des gestes effectués entre les différentes séances d'apprentissage permet d'assurer la continuité des conseils donnés.
- Possibilité d'acquérir suffisamment de données au cours d'une seule séance d'apprentissage : le système nécessite un certain nombre de gestes (autant de la part de l'expert que de celle de l'apprenant) afin de proposer un retour visuel représentatif de la performance de l'apprenant. Il peut néanmoins être intéressant de tester le système sur une donnée de l'apprenant à la fois, afin de vérifier si la visualisation des défauts, ainsi que les conseils donnés par le système restent pertinents.

### 7.3.7 *Clustering récursif*

Une piste explorée, mais non concrétisée, au cours de la thèse a été celle de l'utilisation d'algorithmes de clustering de manière récursive sur chaque groupe obtenu lors d'un premier clustering. Cette idée provient de la volonté d'obtenir une séparation des données en groupes de données correspondants aux gestes réussis et aux gestes ratés. Au terme de l'expérimentation du *Bottle Flip Challenge*, il n'a pas été possible d'obtenir cette séparation. Cependant, une hypothèse soulevée était que la granularité de la séparation (en deux groupes) n'était pas assez fine et qu'il existait différents stratégies ou groupes de mouvements, possédant des caractéristiques différentes, conduisant au succès de la tâche. L'observation des caractéristiques les plus discriminantes entre ces deux groupes a montré que la séparation se faisait sur la vitesse du lancer. De là est venue l'idée d'appliquer récursivement un algorithme de clustering sur les deux groupes obtenus, afin de vérifier si, au sein de ces groupes, il était possible d'obtenir un groupement des données qui correspondait au degré de réussite du geste. Ainsi, nous faisons l'hypothèse qu'il est possible d'obtenir un ensemble défini de sous-groupes composés majoritairement de gestes réussis, ou majoritairement de gestes ratés. Un travail de recherche devra être conduit pour confirmer ou infirmer cette hypothèse.

La méthode testée se base pour l'instant sur une approche semi-supervisé d'apprentissage automatique. En effet, bien que l'objectif à terme soit d'utiliser des métriques calculées sur les clusters afin de déterminer si les sous-groupes obtenus sont corrects, le critère d'arrêt actuel utilise le taux de gestes du même type au sein des sous-groupes obtenus. Dans notre cas, l'étiquetage des données correspondant à la réussite ou non du geste.

En partant initialement d'un ensemble constitué de la totalité des données, l'algorithme effectue plusieurs clustering sur ces données, avec un nombre  $k = 2, 3, 4, \dots, 8$  de clusters. Seul le meilleur clustering, au sens du Silhouette Score, est conservé. Pour chaque sous-groupe ainsi obtenu, la répartition des données en son sein est alors calculée. Un sous-groupe est considéré comme homogène dès lors qu'il contient au moins 75% de gestes du même type (réussi ou raté), et l'algorithme ne fait pas de passe supplémentaire dessus. Enfin, lorsqu'un groupe ne contient plus que quatre données, aucun clustering n'est appliqué dessus.

Les premiers résultats obtenus ont montrés qu'il était possible d'obtenir un taux de même geste au sein des clusters supérieur ou égal au seuil défini. Le travail de recherche devra se concentrer dans un premier temps sur l'analyse de la pertinence des paramètres choisis : taille minimale du sous-groupe, seuil de répartition des

gestes, algorithmes utilisés. Un travail de recherche doit ensuite être effectué afin d'analyser les spécificités communes des gestes entre eux, au sein de chaque sous-groupe obtenu, afin de vérifier si (i) les propriétés communes des gestes au sein de chaque sous-groupe sont pertinentes au regard du geste considéré, et si (ii) des sous-groupes ont en communs ces propriétés, afin de les réunir dans une approche ascendante, analogue au clustering hiérarchique.

---

## BIBLIOGRAPHIE

---

- AGGARWAL, J.K. et Lu XIA (2014). "Human activity recognition from 3D data : A review". In : *Pattern Recognition Letters* 48. Celebrating the life and work of Maria Petrou, p. 70-80. ISSN : 0167-8655.
- AHMED, Faisal, Padma POLASH PAUL et Marina L GAVRILOVA (2015). "Kinect-based gait recognition using sequences of the most relevant joint relative angles". In : ALANKUS, Gazihan *et al.* (2010). "Towards Customizable Games for Stroke Rehabilitation". In : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA : ACM, p. 2113-2122.
- AMINIAN, Kamiar et Bijan NAJAFI (2004). "Capturing human motion using body-fixed sensors : outdoor measurement and clinical applications". In : *Computer Animation and Virtual Worlds* 15.2, p. 79-94.
- ARISTIDOU, A. et Y. CHRYSANTHOU (jan. 2014). "Feature extraction for human motion indexing of acted dance performances". In : *2014 International Conference on Computer Graphics Theory and Applications (GRAPP)*, p. 1-11.
- ARISTIDOU, Andreas *et al.* (août 2015). "Folk Dance Evaluation Using Laban Movement Analysis". In : *J. Comput. Cult. Herit.* 8.4, 20 :1-20 :19. ISSN : 1556-4673.
- ASHFORD, Derek, Simon J. BENNETT et Keith DAVIDS (mai 2006). "Observational Modeling Effects for Movement Dynamics and Movement Outcome Measures Across Differing Task Constraints : A Meta-Analysis". In : *Journal of Motor Behavior* 38.3, p. 185-205.
- BALDOMINOS, Alejandro, Yago SAEZ et Cristina García del Pozo (2015). "An Approach to Physical Rehabilitation Using State-of-the-art Virtual Reality and Motion Tracking Technologies". In : *Procedia Computer Science* 64. Conference on ENTERprise Information Systems/International Conference on Project MANagement/-Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN / HCist 2015 October 7-9, 2015, p. 10-16. ISSN : 1877-0509.
- BARBIČ, Jernej *et al.* (2004). "Segmenting Motion Capture Data into Distinct Behaviors". In : *Proceedings of Graphics Interface 2004*. GI '04. London, Ontario, Canada : Canadian Human-Computer Communications Society, p. 185-194. ISBN : 1-56881-227-2.
- BASSO, Filippo *et al.* (2013). "Fast and Robust Multi-people Tracking from RGB-D Data for a Mobile Robot". In : *Intelligent Autonomous Systems 12 : Volume 1 Proceedings of the 12th International Conference IAS-12, held June 26-29, 2012, Jeju Island, Korea*. Sous la dir. de Sukhan LEE *et al.* Berlin, Heidelberg : Springer Berlin Heidelberg, p. 265-276. ISBN : 978-3-642-33926-4.
- BAUDOUIN, Cyrille *et al.* (2008). "Recueil de traces pour le suivi de l'activité d'apprenants en travaux pratiques dans un environnement de réalité virtuelle". In : *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation* 14, en ligne.
- BERNDT, Donald J. et James CLIFFORD (1994). "Using Dynamic Time Warping to Find Patterns in Time Series". In : *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAIWS'94. Seattle, WA : AAAI Press, p. 359-370.

- BEVLY, David M (2004). "Global positioning system (GPS) : A low-cost velocity sensor for correcting inertial sensor errors on ground vehicles". In : *Journal of dynamic systems, measurement, and control* 126.2, p. 255-264.
- BOUCHARD, Durell et Norman BADLER (2007a). "Semantic Segmentation of Motion Capture Using Laban Movement Analysis". In : *Intelligent Virtual Agents*. Sous la dir. de Catherine PELACHAUD *et al.* Berlin, Heidelberg : Springer Berlin Heidelberg, p. 37-44. ISBN : 978-3-540-74997-4.
- (2007b). "Semantic Segmentation of Motion Capture Using Laban Movement Analysis". In : *Intelligent Virtual Agents*. Sous la dir. de Catherine PELACHAUD *et al.* Berlin, Heidelberg : Springer Berlin Heidelberg, p. 37-44. ISBN : 978-3-540-74997-4.
- BOULAHIA, S. Y. *et al.* (déc. 2016). "HIF3D : Handwriting-Inspired Features for 3D skeleton-based action recognition". In : *2016 23rd International Conference on Pattern Recognition (ICPR)*, p. 985-990.
- BROTMAN, Lynne Shapiro et Arun N. NETRAVALI (juin 1988). "Motion Interpolation by Optimal Control". In : *SIGGRAPH Comput. Graph.* 22.4, p. 309-315. ISSN : 0097-8930.
- BURNS, ANNE-MARIE *et al.* (2011). "Using virtual humans and computer animations to learn complex motor skills : a case study in karate". In : *BIO Web of Conferences* 1, p. 00012.
- CHAMARET, Damien *et al.* (mai 2010). "Integration and evaluation of haptic feedbacks : from CAD models to virtual prototyping". In : *International Journal on Interactive Design and Manufacturing (IJIDeM)* 4.2, p. 87-94. ISSN : 1955-2505.
- CHAN, J. C. P. *et al.* (avr. 2011). "A Virtual Reality Dance Training System Using Motion Capture Technology". In : *IEEE Transactions on Learning Technologies* 4.2, p. 187-195. ISSN : 1939-1382.
- CHANG, Chuan-Yu *et al.* (2013a). "Physiological emotion analysis using support vector regression". In : *Neurocomputing* 122, p. 79-87. ISSN : 0925-2312.
- (2013b). "Physiological emotion analysis using support vector regression". In : *Neurocomputing* 122. Advances in cognitive and ubiquitous computing, p. 79-87. ISSN : 0925-2312.
- CHELLALI, Amine *et al.* (2016). "Achieving interface and environment fidelity in the Virtual Basic Laparoscopic Surgical Trainer". In : *International journal of human-computer studies* 96, p. 22-37.
- CHEN, S. *et al.* (nov. 2016). "Toward Pervasive Gait Analysis With Wearable Sensors : A Systematic Review". In : *IEEE Journal of Biomedical and Health Informatics* 20.6, p. 1521-1537. ISSN : 2168-2194.
- CHEN, Yi-Wei et Chih-Jen LIN (2006). "Combining SVMs with Various Feature Selection Strategies". In : *Feature Extraction : Foundations and Applications*. Sous la dir. d'Isabelle GUYON *et al.* Berlin, Heidelberg : Springer Berlin Heidelberg, p. 315-324. ISBN : 978-3-540-35488-8.
- CHOI, Kup-Sze *et al.* (2015). "A virtual reality based simulator for learning nasogastric tube placement". In : *Computers in Biology and Medicine* 57, p. 103-115. ISSN : 0010-4825.
- CHOQUET, Christophe et Sébastien IKSAL (avr. 2007). "Modeling Tracks for the Model Driven Re-engineering of a TEL System". In : *Journal of Interactive Learning Research* 18.2, p. 161-184.
- COOPER, Helen, Brian HOLT et Richard BOWDEN (2011). "Sign language recognition". In : *Visual Analysis of Humans*. Springer, p. 539-562.
- CORRÊA, Cléber G. *et al.* (2019). "Haptic interaction for needle insertion training in medical applications : The state-of-the-art". In : *Medical Engineering & Physics* 63, p. 6-25. ISSN : 1350-4533.



- D'AGOSTINO, Ralph B., Albert BELANGER et Ralph B. D'AGOSTINO (nov. 1990). "A Suggestion for Using Powerful and Informative Tests of Normality". In : *The American Statistician* 44.4, p. 316.
- DEKKER, P. J. et al. (2018). "Water bottle flipping physics". In : *American Journal of Physics* 86.10, p. 733-739.
- DELAYE, Adrien et Eric ANQUETIL (2013). "HBF49 feature set : A first unified baseline for online symbol recognition". In : *Pattern Recognition* 46.1, p. 117-130. ISSN : 0031-3203.
- DESPINOY, Fabien et al. (juin 2016). "Unsupervised Trajectory Segmentation for Surgical Gesture Recognition in Robotic Training". In : *IEEE Transactions on Biomedical Engineering* 63.6, p. 1280-1291.
- DJADJA, Djadja Jean Delest, Ludovic HAMON et Sébastien GEORGE (2019). "Modeling and Evaluating of Human 3d+t Activities in Virtual Environment". In : *Transforming Learning with Meaningful Technologies*. Sous la dir. de Maren SCHEFFEL et al. Cham : Springer International Publishing, p. 696-700.
- ERAUT, Michael (2000). "Non-formal learning and tacit knowledge in professional work". In : *British journal of educational psychology* 70.1, p. 113-136.
- FENG ZHOU, F. DE LA TORRE et J. K. HODGINS (sept. 2008). "Aligned Cluster Analysis for temporal segmentation of human motion". In : *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, p. 1-7.
- FERRARI, Emilio et al. (2018). "Motion Analysis System for the digitalization and assessment of manual manufacturing and assembly processes". In : *IFAC-PapersOnLine* 51.11. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018, p. 411-416. ISSN : 2405-8963.
- GAMEIRO, João, Tiago CARDOSO et Yves RYBARCZYK (2014). "Kinect-Sign, Teaching Sign Language to "Listeners" through a Game". In : *Procedia Technology* 17. Conference on Electronics, Telecommunications and Computers – CETC 2013., p. 384-391. ISSN : 2212-0173.
- GAO, Jing et Hui-chuan DUAN (2011). "Research on data transmission efficiency of JSON". In : *Computer Engineering and Design* 32.7, p. 2267-2270.
- GIANSANTI, D. et al. (avr. 2003). "Is it feasible to reconstruct body segment 3-D position and orientation using accelerometric data?" In : *IEEE Transactions on Biomedical Engineering* 50.4, p. 476-483. ISSN : 0018-9294.
- GIBET, Sylvie et al. (2016). "Interactive Editing in French Sign Language Dedicated to Virtual Signers : Requirements and Challenges". In : *Universal Access in the Information Society* 15.4, p. 525-539.
- GILLESPIE, R Brent (2018). "Haptic interface to virtual environments". In : *Robotics and Automation Handbook*. CRC Press, p. 463-485.
- HACHIMURA, K., K. TAKASHINA et M. YOSHIMURA (2005). "Analysis and evaluation of dancing movement based on LMA". In : *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. IEEE.
- HALABI, Osama et Yasmin HALWANI (2018). "Design and implementation of haptic virtual fixtures for preoperative surgical planning". In : *Displays* 54, p. 9-19. ISSN : 0141-9382.
- HALSEY, Lewis G et al. (2015). "The fickle P value generates irreproducible results". In : *Nature methods* 12.3, p. 179.
- HE, Xiaofei, Deng CAI et Partha NIYOGI (2005). "Laplacian Score for Feature Selection". In : *Proceedings of the 18th International Conference on Neural Information Processing Systems*. NIPS'05. Vancouver, British Columbia, Canada : MIT Press, p. 507-514.

- HOAI, M., Z. LAN et F. DE LA TORRE (juin 2011). "Joint segmentation and classification of human actions in video". In : *CVPR 2011*, p. 3265-3272.
- HTC (2016). *HTC Vive Specifications*. URL : <https://www.vive.com/fr/product/#vive-spec> (visité le 07/10/2019).
- HUANG, Jie *et al.* (juin 2015). "Sign Language Recognition using 3D convolutional neural networks". In : *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, p. 1-6.
- HUBERT, Lawrence et Phipps ARABIE (déc. 1985). "Comparing partitions". In : *Journal of Classification* 2.1, p. 193-218. ISSN : 1432-1343.
- IZADI, Shahram *et al.* (2011). "Kinectfusion : real-time 3D reconstruction and interaction using a moving depth camera". In : *In Proc. UIST*, p. 559-568.
- JAIN, Anil K. et Richard C. DUBES (1988). *Algorithms for Clustering Data*. Upper Saddle River, NJ, USA : Prentice-Hall, Inc. ISBN : 0-13-022278-X.
- JIE HUANG *et al.* (juin 2015). "Sign Language Recognition using 3D convolutional neural networks". In : *2015 IEEE International Conference on Multimedia and Expo (ICME)*, p. 1-6.
- KAPSOURAS, Ioannis et Nikos NIKOLAIDIS (2014a). "Action recognition on motion capture data using a dynemes and forward differences representation". In : *Journal of Visual Communication and Image Representation* 25.6, p. 1432-1445. ISSN : 1047-3203.
- (2014b). "Action recognition on motion capture data using a dynemes and forward differences representation". In : *Journal of Visual Communication and Image Representation* 25.6, p. 1432-1445. ISSN : 1047-3203.
- KASSAB, Randa et Jean-Charles LAMIREL (2008). "Feature-based cluster validation for high-dimensional data". In : *Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications*. ACTA Press, p. 232-239.
- KNEEBONE, R L *et al.* (2004). "Simulation and clinical practice : strengthening the relationship". In : *Medical Education* 38.10, p. 1095-1102.
- KOBAYASHI, Yuichi (oct. 2007). "The EMOSIGN - analyzing the emotion signature in human motion". In : *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, p. 1171-1176.
- KORA, Takahiro, Masato SOGA et Hirokazu TAKI (2015). "Golf Learning Environment Enabling Overlaid Display of Expert's Model Motion and Learner's Motion Using Kinect". In : *Procedia Computer Science* 60. Knowledge-Based and Intelligent Information & Engineering Systems 19th Annual Conference, KES-2015, Singapore, September 2015 Proceedings, p. 1559-1565. ISSN : 1877-0509.
- KRISTIANSLUND, Eirik, Tron KROSSHAUG et Antonie J. van den BOGERT (2012). "Effect of low pass filtering on joint moments from inverse dynamics : Implications for injury prevention". In : *Journal of Biomechanics* 45.4, p. 666-671. ISSN : 0021-9290.
- KRÜGER, B. *et al.* (avr. 2017). "Efficient Unsupervised Temporal Segmentation of Motion Data". In : *IEEE Transactions on Multimedia* 19.4, p. 797-812. ISSN : 1520-9210.
- KUIPERS, Jack B. (2000). "Quaternions and Rotation Sequences". In : *Proceedings of the International Conference on Geometry, Integrability and Quantization*. Sofia, Bulgaria : Coral Press Scientific Publishing, p. 127-143.
- KYAN, Matthew *et al.* (mar. 2015). "An Approach to Ballet Dance Training Through MS Kinect and Visualization in a CAVE Virtual Reality Environment". In : *ACM Trans. Intell. Syst. Technol.* 6.2, 23 :1-23 :37. ISSN : 2157-6904.
- LARBOULETTE, Caroline et Sylvie GIBET (août 2015). "A Review of Computable Expressive Descriptors of Human Motion". In : *MOCO 2015 : 2nd International Workshop on Movement and Computing*. Proceedings of the International Workshop on Movement and Computing. Vancouver, Canada, p. 21-28.

- LE NAOUR, Thibaut, Ludovic HAMON et Jean-Pierre BRESCIANI (2019). "Superimposing 3D Virtual Self + Expert Modeling for Motor Learning : Application to the Throw in American Football". In : *Frontiers in ICT* 6, p. 16.
- LEWIS, J. P., Matt CORDNER et Nickson FONG (2000). "Pose space deformation". In : *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH '00*. ACM Press.
- LIU, Meiyin, SangUk HAN et SangHyun LEE (juil. 2016). "Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis". In : *Construction Innovation* 16.3, p. 348-367.
- LUINGE, Hendrik J., Petrus H. VELTINK et Christian T.M. BATEN (1999). "Estimating orientation with gyroscopes and accelerometers". Undefined. In : *Technology and health care* 7.6, p. 455-459. ISSN : 0928-7329.
- LV, Fengjun et Ramakant NEVATIA (2006). "Recognition and Segmentation of 3-D Human Action Using HMM and Multi-class AdaBoost". In : *Computer Vision – ECCV 2006*. Sous la dir. d'Aleš LEONARDIS, Horst BISCHOF et Axel PINZ. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 359-372. ISBN : 978-3-540-33839-0.
- LV, Zhuowen et al. (jan. 2015). "Class Energy Image Analysis for Video Sensor-Based Gait Recognition : A Review". In : *Sensors* 15.1, p. 932-964.
- MAES, Pieter-Jan, Denis AMELYNCK et Marc LEMAN (2012). "Dance-the-Music : an educational platform for the modeling, recognition and audiovisual monitoring of dance steps using spatiotemporal motion templates". In : *EURASIP Journal on Advances in Signal Processing* 2012.1, p. 35. ISSN : 1687-6180.
- MAHDI, Oussema et al. (2019). "Towards an Editor for VR-Oriented Educational Scenarios". In : *Transforming Learning with Meaningful Technologies*. Sous la dir. de Maren SCHEFFEL et al. Cham : Springer International Publishing, p. 756-760. ISBN : 978-3-030-29736-7.
- MAKIO, Kosuke, Yoshiki TANAKA et Kuniaki UEHARA (2007). "Discovery of Skills from Motion Data". In : *New Frontiers in Artificial Intelligence*. Sous la dir. d'Akito SAKURAI, Kôiti HASIDA et Katsumi NITTA. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 266-282.
- MALFAIT, B. et al. (2014). "How reliable are lower-limb kinematics and kinetics during a drop vertical jump?" In : *Medicine and science in sports and exercise* 46.
- MARINS, J. L. et al. (oct. 2001). "An extended Kalman filter for quaternion-based orientation estimation using MARG sensors". In : *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180)*. T. 4, 2003-2011 vol.4.
- MAULIK, U. et S. BANDYOPADHYAY (déc. 2002). "Performance evaluation of some clustering algorithms and validity indices". In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.12, p. 1650-1654. ISSN : 0162-8828.
- MCDERMOTT, Maeve (2016). *Watch the simple water bottle flip that dominated this teen's talent show*. URL : <https://eu.usatoday.com/story/life/entertainthis/2016/05/26/teen-water-bottle-flip-viral-video-mike-senatore/84964358/> (visité le 10/07/2019).
- MEILĂ, Marina (2007). "Comparing clusterings—an information based distance". In : *Journal of Multivariate Analysis* 98.5, p. 873-895. ISSN : 0047-259X.
- MOREL, Marion (nov. 2017). "Multidimensional time-series averaging. Application to automatic and generic evaluation of sport gestures". Theses. Université Pierre et Marie Curie - Paris VI. URL : <https://tel.archives-ouvertes.fr/tel-01703849>.
- MOREY, Leslie C. et Alan AGRESTI (1984). "The Measurement of Classification Agreement : An Adjustment to the Rand Statistic for Chance Agreement". In : *Educational and Psychological Measurement* 44.1, p. 33-37.

- MOSCHAS, Fanis et Stathis STIROIS (2011). "Measurement of the dynamic displacements and of the modal frequencies of a short-span pedestrian bridge using GPS and an accelerometer". In : *Engineering Structures* 33.1, p. 10-17. ISSN : 0141-0296.
- MÜLLER, M. *et al.* (2007). *Motion Capture Database HDM05*. URL : <http://resources.mpi-inf.mpg.de/HDM05/> (visité le 11/07/2019).
- NG, Kia (2008). "Technology-Enhanced Learning For Music With I-Maestro Framework And Tools". In : *EVA 2008 London Conference*.
- NOITOM (2015a). *Axis Neuron User Guide*.
- (2015b). *Perception Neuron*. URL : [https://neuronmocap.com/products/perception\\_neuron](https://neuronmocap.com/products/perception_neuron) (visité le 03/07/2019).
- NUNES, João F. et Pedro M. MOREIRA (2016). "Handbook of Research on Computational Simulation and Modeling in Engineering." In : chap. Human Motion Analysis and Simulation Tools : A Survey. P. 359-388.
- PATRONA, Fotini *et al.* (2018). "Motion analysis : Action detection, recognition and evaluation based on motion capture data". In : *Pattern Recognition* 76, p. 612-622. ISSN : 0031-3203.
- PEPLEY, David F. *et al.* (2017). "Training Surgical Residents With a Haptic Robotic Central Venous Catheterization Simulator". In : *Journal of Surgical Education* 74.6, p. 1066-1073. ISSN : 1931-7204.
- PFISTER, Alexandra *et al.* (2014). "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis". In : *Journal of Medical Engineering & Technology* 38.5, p. 274-280.
- PHAM, M T, R MOREAU et P BOULANGER (août 2010). "Three-dimensional gesture comparison using curvature analysis of position and orientation". In : *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE.
- PHILO TAN CHUA *et al.* (mar. 2003). "Training for physical tasks in virtual environments : Tai Chi". In : *IEEE Virtual Reality, 2003. Proceedings*. P. 87-94.
- PIRSIAVASH, Hamed, Carl VONDRICK et Antonio TORRALBA (2014). "Assessing the Quality of Actions". In : *Computer Vision – ECCV 2014*. Sous la dir. de David FLEET *et al.* Cham : Springer International Publishing, p. 556-571.
- PORCIUNCULA, Franchino *et al.* (2018). "Wearable Movement Sensors for Rehabilitation : A Focused Review of Technological and Clinical Advances". In : *PM&R* 10.9, Supplement 2. Innovations Influencing Physical Medicine and Rehabilitation, S220-S232. ISSN : 1934-1482.
- PULIJALA, Yeshwanth, Minhua MA et Ashraf AYOUB (2017). "VR Surgery : Interactive Virtual Reality Application for Training Oral and Maxillofacial Surgeons using Oculus Rift and Leap Motion". In : *Serious Games and Edutainment Applications : Volume II*. Sous la dir. de Minhua MA et Andreas OIKONOMOU. Cham : Springer International Publishing, p. 187-202. ISBN : 978-3-319-51645-5.
- RAND, William M. (1971). "Objective Criteria for the Evaluation of Clustering Methods". In : *Journal of the American Statistical Association* 66.336, p. 846-850. ISSN : 01621459.
- RAO, Cen, Alper YILMAZ et Mubarak SHAH (nov. 2002). "View-Invariant Representation and Recognition of Actions". In : *Int. J. Comput. Vision* 50.2, p. 203-226. ISSN : 0920-5691.
- RASAMIMANANA, Nicolas et Frédéric BEVILACQUA (déc. 2008). "Effort-based Analysis of Bowing Movements : Evidence of Anticipation Effects". In : *Journal of New Music Research* 37.4, p. 339-351.
- RAWASHDEH, Samir, Derek RAFELDT et Timothy UHL (nov. 2016). "Wearable IMU for Shoulder Injury Prevention in Overhead Sports". In : *Sensors* 16.11, p. 1847. ISSN : 1424-8220.

- REGAZZONI, Daniele, Giordano de VECCHI et Caterina RIZZI (2014). "RGB cams vs RGB-D sensors : Low cost motion capture technologies performances and limitations". In : *Journal of Manufacturing Systems* 33.4, p. 719-728. ISSN : 0278-6125.
- RISSANEN, J (1998). *Stochastic Complexity in Statistical Inquiry*. WORLD SCIENTIFIC.
- ROETENBERG, Daniel et al. (2005). "Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation". In : *IEEE Transactions on neural systems and rehabilitation engineering* 13.3, p. 395-405.
- ROJAS-LERTXUNDI, Sendoa et al. (2015). "Motion Capture Systems for Jump Analysis". In : *Hybrid Artificial Intelligent Systems*. Sous la dir. d'Enrique ONIEVA et al. Cham : Springer International Publishing, p. 111-124. ISBN : 978-3-319-19644-2.
- ROUSSEEUW, Peter J. (1987). "Silhouettes : A graphical aid to the interpretation and validation of cluster analysis". In : *Journal of Computational and Applied Mathematics* 20, p. 53-65. ISSN : 0377-0427.
- SAKURAI, K. et al. (oct. 2014). "Retrieval of similar behavior data using Kinect data". In : *2014 14th International Conference on Control, Automation and Systems (ICCAS 2014)*, p. 1368-1372.
- SALAMAH, Samer, Liang ZHANG et Guido BRUNETT (2015). "Hierarchical Method for Segmentation by Classification of Motion Capture Data". In : *Virtual Realities : International Dagstuhl Seminar, Dagstuhl Castle, Germany, June 9-14, 2013, Revised Selected Papers*. Sous la dir. de Guido BRUNETT et al. Cham : Springer International Publishing, p. 169-186.
- SARAFIANOS, Nikolaos et al. (2016). "3D Human pose estimation : A review of the literature and analysis of covariates". In : *Computer Vision and Image Understanding* 152, p. 1-20. ISSN : 1077-3142.
- SAVITZKY, Abraham. et M. J. E. GOLAY (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." In : *Analytical Chemistry* 36.8, p. 1627-1639.
- SENECAL, Simon, Niels A. NIJDAM et Nadia Magnenat THALMANN (2018). "Motion Analysis and Classification of Salsa Dance Using Music-related Motion Features". In : *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games*. MIG '18. Limassol, Cyprus : ACM, 11 :1-11 :10. ISBN : 978-1-4503-6015-9.
- SHAPIRA, L., J. AMORES et X. BENAVIDES (sept. 2016). "TactileVR : Integrating Physical Toys into Learn and Play Virtual Reality Experiences". In : *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, p. 100-106.
- SREELAKSHMI, M. et T.D. SUBASH (2017). "Haptic Technology : A comprehensive review on its applications and future prospects". In : *Materials Today : Proceedings* 4.2, Part B. International Conference on Computing, Communication, Nanophotonics, Nanoscience, Nanomaterials and Nanotechnology, p. 4182-4187. ISSN : 2214-7853.
- STRUYF, Anja, Mia HUBERT et Peter ROUSSEEUW (1997). "Clustering in an Object-Oriented Environment". In : *Journal of Statistical Software, Articles* 1.4, p. 1-30. ISSN : 1548-7660.
- TAN, Benedict, Abdul Rashid AZIZ et Teh Kong CHUAN (2000). "Correlations between physiological parameters and performance in elite ten-pin bowlers". In : *Journal of Science and Medicine in Sport* 3.2, p. 176-185. ISSN : 1440-2440.
- TCHOUNIKINE, Pierre (juin 2009). "Précis de recherche en Ingénierie des EIAH". 109 pages.
- THODE, Henry C (2002). *Testing for normality*. T. 164. CRC press.
- TIPPING, M. E. et C. M. BISHOP (1999). "Probabilistic principal component analysis". In : *Journal of the Royal Statistical Society, Series B* 61, p. 611-622.



- TOUSSAINT, Ben-Manson (2015). "Apprentissage automatique a partir de traces multi-sources heterogenes pour la modelisation de connaissances perceptivo-gestuelles". Thèse de doct. Universite de Grenoble.
- UKESSAYS (2017). *Physics of the Bottle Flip Challenge*. URL : <https://www.ukessays.com/essays/physics/physics-bottle-flip-challenge-2329.php> (visité le 11/08/2019).
- UNIVERSITY, Carnegie Mellon (2016). *CMU Graphics Lab Motion Capture Database*. URL : <http://mocap.cs.cmu.edu/> (visité le 11/07/2019).
- VRIES, Sanne I. de *et al.* (juil. 2006). "Clinimetric review of motion sensors in children and adolescents". In : *Journal of Clinical Epidemiology* 59.7, p. 670-680.
- WANG, Liang, Weiming HU et Tieniu TAN (2003). "Recent developments in human motion analysis". In : *Pattern Recognition* 36.3, p. 585-601. ISSN : 0031-3203.
- WANG, R. *et al.* (juin 2013). "Home Monitoring Musculo-skeletal Disorders with a Single 3D Sensor". In : *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, p. 521-528.
- WHITMIRE, Eric *et al.* (2018). "Demonstration of Haptic Revolver : Touch, Shear, Texture, and Shape Rendering on a VR Controller". In : *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada : ACM, D109 :1-D109 :4. ISBN : 978-1-4503-5621-3.
- WIKIPEDIA (2019). *Cinétographie Laban*. URL : [https://fr.wikipedia.org/wiki/Cin%C3%A9tographie\\_Laban](https://fr.wikipedia.org/wiki/Cin%C3%A9tographie_Laban) (visité le 02/09/2019).
- XIA, Lu, Chia-Chih CHEN et J. K. AGGARWAL (2012). *UTKinect-Action3D Dataset*. URL : <http://cvrc.ece.utexas.edu/KinectDatasets/H0J3D.html> (visité le 18/11/2019).
- XIAO, Jun *et al.* (2015). "Sketch-based human motion retrieval via selected 2D geometric posture descriptor". In : *Signal Processing* 113, p. 1-8. ISSN : 0165-1684.
- XU, Mingliang *et al.* (2019). "Personalized training through Kinect-based games for physical education". In : *Journal of Visual Communication and Image Representation* 62, p. 394-401. ISSN : 1047-3203.
- YAMAOKA, Koji *et al.* (2013a). "Feedback of Flying Disc Throw with Kinect and its Evaluation". In : *Procedia Computer Science* 22. 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013, p. 912-920. ISSN : 1877-0509.
- (2013b). "Feedback of Flying Disc Throw with Kinect and its Evaluation". In : *Procedia Computer Science* 22. 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013, p. 912-920. ISSN : 1877-0509.
- YANG, C. *et al.* (2016). "Human Upper Limb Motion Analysis for Post-Stroke Impairment Assessment Using Video Analytics". In : *IEEE Access* 4, p. 650-659.
- YANG, Hanxuan *et al.* (2011). "Recent advances and trends in visual tracking : A review". In : *Neurocomputing* 74.18, p. 3823-3831. ISSN : 0925-2312.
- YANG, Yan et Stephen G. LISBERGER (2014). "Role of Plasticity at Different Sites across the Time Course of Cerebellar Motor Learning". In : *Journal of Neuroscience* 34.21, p. 7077-7090.
- YOSHINAGA, Toshihiro et Masato SOGA (2015). "Development of a Motion Learning Support System Arranging and Showing Several Coaches' Motion Data". In : *Procedia Computer Science* 60. Knowledge-Based and Intelligent Information & Engineering Systems 19th Annual Conference, KES-2015, Singapore, September 2015 Proceedings, p. 1497-1505. ISSN : 1877-0509.

- YU, Zhibin et Minhoo LEE (sept. 2015a). "Human motion based intent recognition using a deep dynamic neural model". In : *Emerging Spatial Competences : From Machine Perception to Sensorimotor Intelligence*. T. 71, p. 134-149.
- (2015b). "Human motion based intent recognition using a deep dynamic neural model". In : *Robotics and Autonomous Systems* 71. *Emerging Spatial Competences : From Machine Perception to Sensorimotor Intelligence*, p. 134-149. ISSN : 0921-8890.
- YUTCHI KOBAYASHI (oct. 2007). "The EMOSIGN - analyzing the emotion signature in human motion". In : *2007 IEEE International Conference on Systems, Man and Cybernetics*, p. 1171-1176.
- ZAPPALA, Joseph *et al.* (2017). "Influence of Posture-Cuing Shirt on Tennis Serve Kinematics in Division III Tennis Players". In : *Journal of Chiropractic Medicine* 16.1, p. 49-53. ISSN : 1556-3707.
- ZAPPELLA, Luca, Xavier LLADÓ et Joaquim SALVI (2008). "Motion Segmentation : A Review". In : *Proceedings of the 2008 Conference on Artificial Intelligence Research and Development : Proceedings of the 11th International Conference of the Catalan Association for Artificial Intelligence*. Amsterdam, The Netherlands, The Netherlands : IOS Press, p. 398-407. ISBN : 978-1-58603-925-7.
- ZIHAJEHZADEH, S. *et al.* (août 2014). "A cascaded two-step Kalman filter for estimation of human body segment orientation using MEMS-IMU". In : *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, p. 6270-6273.







---

## ANNEXES

---

### A.1 ANNEXE 1 : PRÉ-QUESTIONNAIRE ET POST-QUESTIONNAIRE DE L'EXPÉRI- MENTATION DU LANCER DE FLÉCHETTES

# Pré-questionnaire

Nom / Prénom

Age

Sexe

Taille

Veuillez qualifier votre niveau aux fléchettes

*1. Je n'y ai jamais joué ... 7. Je suis un joueur professionnel*

1	2	3	4	5	6	7

Pratiquez-vous un autre sport (loisir, amateur, pro) ? Si oui, lequel ?

Oui / Non

# Post-questionnaire

Veillez qualifier le niveau d'aisance ressenti avec le matériel de captation de mouvement.

*1. Pas du tout à l'aise ... 7. Totalelement à l'aise*

1	2	3	4	5	6	7

Etiez-vous stressé lors de l'expérimentation ?

*1. Très stressé ... 7. Pas du tout stressé*

1	2	3	4	5	6	7

Veillez qualifier le niveau de compréhension des instructions reçues.

*1. Je n'ai pas du tout compris ... 7. J'ai parfaitement compris*

1	2	3	4	5	6	7

Veillez évaluer l'intérêt des conseils qui vous ont été donnés durant l'ensemble de l'expérimentation.

*1. Pas du tout utiles / je n'en ai pas tenu compte ... 7. Très utiles*

1	2	3	4	5	6	7

Veillez évaluer votre performance globale sur l'ensemble de l'expérimentation.

*1. Très mauvais ... 7. Très bon*

1	2	3	4	5	6	7

Veillez évaluer la progression de votre performance globale sur l'ensemble de l'expérimentation.

*1. Aucune progression ... 7. Très forte progression*

1	2	3	4	5	6	7

Veillez évaluer la qualité de votre geste de lancer sur l'ensemble de l'expérimentation.

*1. Geste mauvais ... 7. Geste excellent*

1	2	3	4	5	6	7

Veillez évaluer la progression de la qualité de votre geste de lancer sur l'ensemble de l'expérimentation.

*1. Aucune progression ... 7. Très forte progression*

1	2	3	4	5	6	7

Remarques diverses :

## A.2 ANNEXE 2 : TABLEAUX D'ANALYSES COMPLETS DES RÉSULTATS DE L'EXPÉRIMENTATION DES FLÉCHETTES

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
Shapiro-Wilk ( $p > 0.05$ )		$4.29e^{-5}$	$1.43e^{-6}$	$2.06e^{-4}$	$4.43e^{-5}$
d'Agostino-Pearson ( $p > 0.05$ )		$1.27e^{-7}$	$8.38e^{-7}$	$1.79e^{-5}$	$1.62e^{-5}$
ANOVA ( $p < 0.05$ )	$F(2, 42)$	0.0907	0.7392	0.4704	0.1307
	p	0.9134	0.4836	0.6280	0.8779
Pairwise T-test ( $p < 0.05$ )	g1 - g2	0.9445	0.7121	0.5244	0.6233
	g1 - g3	0.7354	0.1902	0.3462	0.6675
	g2 - g3	0.6659	0.4380	0.7162	0.9650
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	0.1875	1.8023	0.4432	0.3937
	p	0.9907	0.4061	0.8012	0.8213
	df	2	2	2	2
Pairwise Mann-Whitney ( $p < 0.05$ )	g1 - g2	0.9174	0.6186	0.9357	0.6041
	g1 - g3	0.9834	0.2210	0.5068	0.9669
	g2 - g3	0.9339	0.3615	0.7244	0.6041

TABLE 34 – Résultats complets des tests statistiques en intergroupe (distance par rapport au centre).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
Shapiro-Wilk ( $p > 0.05$ )		$1.0716e^{-4}$	$1.0445e^{-6}$	$1.5109e^{-6}$	$5.0037e^{-5}$
d'Agostino-Pearson ( $p > 0.05$ )		$2.0430e^{-4}$	$4.8091e^{-9}$	$1.5579e^{-7}$	$1.6463e^{-4}$
ANOVA ( $p < 0.05$ )	$F(2, 42)$	0.4877	0.0820	0.5330	1.1240
	p	0.6175	0.9214	0.5908	0.3346
Pairwise T-test ( $p < 0.05$ )	g1 - g2	0.9972	0.9012	0.3538	0.1728
	g1 - g3	0.3651	0.6990	0.5972	0.4170
	g2 - g3	0.3724	0.7628	0.5855	0.4774
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	0.4970	1.9223	3.2518	2.5979
	p	0.78000	0.3824	0.1967	0.2728
	df	2	2	2	2
Mann-Whitney ( $p < 0.05$ )	g1 - g2	0.9339	0.6783	0.0890	0.1585
	g1 - g3	0.5338	0.1466	0.8357	0.5614
	g2 - g3	0.5897	0.4807	0.1844	0.2290

TABLE 35 – Résultats complets des tests statistiques en intergroupe (alignement du bras).

		Shapiro-Wilk ( $p > 0.05$ )	d'Agostino-Pearson ( $p > 0.05$ )	Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
				One tail		Two tails		T	One tail	Two tails	t	df
				p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	$1.2329e^{-3}$	$1.6380e^{-2}$	0.0057	0.0042	0.0115	0.0084	15	0.0044	0.0087	-3.0431	14
	G2	$1.8799e^{-5}$	$2.8416e^{-8}$	0.0021	0.0010	0.0041	0.0020	9	0.0023	0.0046	-3.3623	14
	G3	$3.8069e^{-5}$	$2.5155e^{-8}$	0.0038	0.0021	0.0076	0.0043	12.5	0.0038	0.0076	-3.1152	14
Jeu 2-3	G1	$9.8712e^{-3}$	$3.1072e^{-2}$	0.0049	0.0034	0.0098	0.0067	14	0.0016	0.0032	3.5433	14
	G2	$3.5333e^{-4}$	$1.8565e^{-6}$	0.1006	0.1039	0.2013	0.2078	37	0.1151	0.2303	1.2543	14
	G3	$7.6972e^{-6}$	$3.0779e^{-6}$	0.5	0.4890	1	0.9780	59.5	0.2068	0.4135	-0.8427	14
Jeu 3-4	G1	$1.1170e^{-1}$	$1.6271e^{-1}$	0.1743	0.1796	0.3487	0.3591	43	0.1514	0.3028	1.0698	14
	G2	$2.1607e^{-2}$	$9.0891e^{-2}$	0.1221	0.1262	0.2443	0.2524	39	0.1439	0.2877	1.1052	14
	G3	$4.5776e^{-5}$	$1.1167e^{-4}$	0.0144	0.0128	0.0288	0.0256	21	0.0107	0.0214	2.5907	14
Jeu 1-3	G1	$2.6205e^{-3}$	$1.6650e^{-3}$	0.1601	0.1651	0.3202	0.3303	42	0.2149	0.4298	-0.8131	14
	G2	$3.3477e^{-3}$	$1.5668e^{-2}$	0.1110	0.1146	0.2220	0.2293	38	0.0662	0.1323	-1.5981	14
	G3	$5.6502e^{-6}$	$5.7737e^{-8}$	0.01659	0.0151	0.0332	0.0301	22	0.0131	0.0261	-2.4868	14
Jeu 2-4	G1	$1.2009e^{-2}$	$3.0490e^{-2}$	0.0005	$9.1553e^{-5}$	0.0011	0.0002	2	0.0003	0.0005	4.4525	14
	G2	$2.9579e^{-5}$	$3.8042e^{-8}$	0.0073	0.0051	0.0146	0.0102	16.5	0.0149	0.0297	2.4199	14
	G3	$4.2739e^{-5}$	$3.4339e^{-5}$	0.0250	0.0240	0.0500	0.0479	25	0.1345	0.2691	1.1508	14
Jeu 1-4	G1	$8.1752e^{-4}$	$1.2342e^{-4}$	0.4660	0.4670	0.9321	0.9341	58	0.4635	0.9270	-0.0933	14
	G2	$1.7507e^{-3}$	$9.6526e^{-4}$	0.0864	0.0844	0.1728	0.1688	35.5	0.1538	0.3076	-1.0589	14
	G3	$2.1970e^{-5}$	$5.3364e^{-7}$	0.2051	0.2106	0.4102	0.4212	45	0.1415	0.2830	-1.1164	14

TABLE 37 – Résultats complets des tests statistiques en intragroupe (distance par rapport au centre).

		Shapiro-Wilk ( $p > 0.05$ )	d'Agostino-Pearson ( $p > 0.05$ )	Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
				One tail		Two tails			One tail	Two tails	t	df
				p-norm	p-exact	p-norm	p-exact	T				
Jeu 1-2	G1	$7.6220e^{-6}$	$9.9203e^{-6}$	0.0035	0.0021	0.0070	0.0043	12	0.0033	0.0066	3.1837	14
	G2	0.0037	0.0058	0.0285	0.0277	0.0571	0.0553	26	0.0199	0.0398	2.2659	14
	G3	0.0031	0.0348	0.2389	0.2443	0.4777	0.4887	47	0.2783	0.5566	0.6024	14
Jeu 2-3	G1	$1.836e^{-6}$	$2.2924e^{-7}$	0.1110	0.1146	0.2220	0.2293	38	0.0729	0.1459	-1.5398	14
	G2	$1.1879e^{-5}$	$1.3753e^{-6}$	0.2568	0.2622	0.5136	0.5245	48	0.1927	0.3853	0.8961	14
	G3	0.1079	0.0980	0.1467	0.1514	0.2934	0.3028	41	0.3905	0.7809	0.2835	14
Jeu 3-4	G1	$6.6718e^{-4}$	0.0011	0.2216	0.2271	0.4432	0.4543	46	0.1425	0.2850	-1.1117	14
	G2	$1.3259e^{-6}$	$1.1894e^{-7}$	0.4660	0.4670	0.9321	0.9341	58	0.4521	0.9041	0.1226	14
	G3	0.0218	0.0435	0.3774	0.3808	0.7547	0.7615	54	0.4253	0.8506	-0.1919	14
Jeu 1-3	G1	$5.2763e^{-5}$	$5.9935e^{-5}$	0.0661	0.0677	0.1323	0.1354	33	0.0406	0.0813	1.8786	14
	G2	$1.0753e^{-4}$	0.0011	0.0144	0.0128	0.0288	0.0256	21	0.0050	0.0100	2.9752	14
	G3	0.0301	0.1060	0.2216	0.2271	0.4432	0.4543	46	0.2134	0.4268	0.8185	14
Jeu 2-4	G1	$3.6799e^{-5}$	$2.4848e^{-5}$	0.0820	0.0844	0.1641	0.1688	35	0.0488	0.0977	-1.7748	14
	G2	$1.6698e^{-4}$	$1.0856e^{-4}$	0.2389	0.2243	0.4777	0.4887	47	0.1556	0.3112	1.0506	14
	G3	0.0040	0.0152	0.2216	0.2271	0.4432	0.4543	46	0.4276	0.8552	0.1859	14
Jeu 1-4	G1	0.0017	0.0014	0.1467	0.1514	0.2934	0.3028	41	0.2167	0.4333	0.8067	14
	G2	$6.7686e^{-4}$	0.0029	0.0219	0.0206	0.0438	0.0413	24	0.0128	0.0256	2.4962	14
	G3	0.0020	0.0432	0.2216	0.2271	0.4432	0.4543	46	0.2807	0.5615	0.5948	14

TABLE 38 – Résultats complets des tests statistiques en intragroupe (alignement du bras).



		Shapiro-Wilk ( $p > 0.05$ )	d'Agostino-Pearson ( $p > 0.05$ )	Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
				One tail		Two tails			One tail	Two tails	t	df
				p-norm	p-exact	p-norm	p-exact	T				
Jeu 1-2	G1	0.0042	$9.6266e^{-4}$	$6.6602e^{-4}$	$1.5259e^{-4}$	0.0013	$3.0518e^{-4}$	3	$3.9473e^{-4}$	$7.8946e^{-4}$	4.2621	14
	G2	0.0244	0.0162	$9.8276e^{-4}$	$3.0518e^{-4}$	0.0020	$6.1035e^{-4}$	5	0.0026	0.0052	3.3014	14
	G3	$5.4414e^{-4}$	0.0030	$4.4595e^{-4}$	$6.1035e^{-5}$	$8.9190e^{-4}$	$1.2207e^{-4}$	1	$4.2438e^{-5}$	$8.4875e^{-5}$	5.4545	14
Jeu 2-3	G1	0.1226	0.0587	0.3351	0.3394	0.6701	0.6788	52	0.3617	0.7235	-0.3910	14
	G2	0.0467	0.1537	0.0738	0.0757	0.1475	0.1514	34	0.0482	0.0964	1.7820	14
	G3	$4.4354e^{-5}$	$8.3421e^{-5}$	0.0661	0.0677	0.1323	0.1354	33	0.0897	0.1794	1.4131	14
Jeu 3-4	G1	0.1857	0.3624	0.4887	0.4890	0.9773	0.9780	59	0.4712	0.9425	-0.0735	14
	G2	0.0219	0.0302	0.3991	0.4020	0.7983	0.8039	55	0.4998	0.9995	$-5.9803e^{-4}$	14
	G3	$4.3685e^{-4}$	0.0050	0.0416	0.0416	0.0832	0.0832	29	0.0499	0.0997	-1.7628	14
Jeu 1-3	G1	0.0051	$4.9284e^{-4}$	$3.6326e^{-4}$	$3.0518e^{-5}$	$7.2651e^{-4}$	$6.1035e^{-5}$	0	$4.8085e^{-4}$	$9.6171e^{-4}$	4.1605	14
	G2	0.0038	0.0728	0.0038	0.0021	0.0076	0.0043	12.5	0.0018	0.0036	3.4882	14
	G3	$1.8900e^{-4}$	$3.5421e^{-4}$	0.0014	$5.7983e^{-4}$	0.0029	0.0012	7	$1.9457e^{-4}$	$3.8914e^{-4}$	4.6306	14
Jeu 2-4	G1	0.1147	0.1421	0.2947	0.2997	0.5895	0.5995	50	0.3837	0.7675	-0.3015	14
	G2	0.0984	0.2806	0.0285	0.0277	0.0571	0.0554	26	0.0263	0.0525	2.1182	14
	G3	$1.0791e^{-4}$	$3.3700e^{-4}$	0.4435	0.4452	0.8871	0.8904	57	0.3351	0.6702	0.4350	14
Jeu 1-4	G1	0.0055	$7.2099e^{-4}$	$6.6602e^{-4}$	$1.5259e^{-4}$	0.0013	$3.0518e^{-4}$	3	$5.2838e^{-4}$	0.0011	4.1121	14
	G2	0.0027	0.0791	$6.6602e^{-4}$	$1.5259e^{-4}$	0.0013	$3.0518e^{-4}$	3	0.0015	0.0030	3.5854	14
	G3	$3.6654e^{-4}$	$4.8664e^{-4}$	$5.4581e^{-4}$	$9.1553e^{-5}$	0.0011	$1.8311e^{-4}$	2	$1.1619e^{-4}$	$2.3239e^{-4}$	4.9044	14

TABLE 39 – Résultats complets des tests statistique en intragroupe (mouvement du coude).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
Shapiro-Wilk ( $p > 0.05$ )		0.0291	$7.8979e^{-4}$	0.0224	$1.5553e^{-3}$
d'Agostino-Pearson ( $p > 0.05$ )		0.0324	$2.4829e^{-4}$	0.1270	0.0105
ANOVA ( $p < 0.05$ )	F (2, 42)	0.5691	0.0828	0.4350	0.5286
	p	0.5703	0.9207	0.6501	0.5933
Pairwise T-test ( $p < 0.05$ )	g1 - g2	0.2861	0.7085	0.3233	0.3562
	g1 - g3	0.7416	0.7254	0.6167	0.9693
	g2 - g3	0.4787	0.9455	0.6993	0.3870
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	0.9793	0.2319	1.0321	0.5969
	p	0.6128	0.8905	0.5969	0.4460
	df	2	2	2	2
Mann-Whitney ( $p < 0.05$ )	g1 - g2	0.4306	0.7715	0.4068	0.2134
	g1 - g3	0.4807	0.9339	0.4068	0.6187
	g2 - g3	0.5614	0.6187	0.6783	0.4807

TABLE 36 – Résultats complets des tests statistiques en intergroupe (mouvement du coude).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
Shapiro-Wilk ( $p > 0.05$ )		0.1308	$2.1392e^{-4}$	$5.2743e^{-5}$	0.0902
d'Agostino-Pearson ( $p > 0.05$ )		0.0872	$2.2381e^{-4}$	$1.3488e^{-7}$	0.1036
ANOVA ( $p < 0.05$ )	F (2, 42)	2.6135	0.8371	0.6180	1.4252
	p	0.0852	0.4400	0.5438	0.2518
Pairwise T-test ( $p < 0.05$ )	g1 - g2	0.9922	0.9792	0.9484	0.1667
	g1 - g3	0.0619	0.2552	0.2491	0.2280
	g2 - g3	0.0374	0.2017	0.3335	0.6985
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	5.2877	0.8371	0.4924	1.9695
	p	0.0711	0.6580	0.7818	0.3735
	df	2	2	2	2
Mann-Whitney ( $p < 0.05$ )	g1 - g2	0.7400	0.8682	0.7715	0.2455
	g1 - g3	0.0512	0.5338	0.4068	0.2998
	g2 - g3	0.0512	0.3837	1	0.4807

TABLE 40 – Résultats complets des tests statistiques en intergroupe (lancer de type javelot).

		Jeu 1	Jeu 2	Jeu 3	Jeu 4
Shapiro-Wilk ( $p > 0.05$ )		0.0050	$6.3275e^{-5}$	0.0017	$7.1272e^{-4}$
d'Agostino-Pearson ( $p > 0.05$ )		0.1405	$3.7124e^{-6}$	0.0034	0.0020
ANOVA ( $p < 0.05$ )	F (2, 42)	1.1188	0.0211	0.8588	1.0277
	p	0.3362	0.9791	0.4310	0.3667
Pairwise T-test ( $p < 0.05$ )	g1 - g2	0.1463	0.9068	0.1401	0.1803
	g1 - g3	0.6315	0.9178	0.6399	0.8557
	g2 - g3	0.3267	0.8536	0.4597	0.2183
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	2.5484	0.7258	2.8197	1.3503
	p	0.2796	0.6957	0.2442	0.5091
	df	2	2	2	2
Mann-Whitney ( $p < 0.05$ )	g1 - g2	0.1249	0.9010	0.0890	0.3195
	g1 - g3	0.7089	0.4553	0.3615	0.7400
	g2 - g3	0.2808	0.5069	0.5614	0.3615

TABLE 41 – Résultats complets des tests statistiques en intergroupe (corps penché).

	Jeu 1	Jeu 2	Jeu 3	Jeu 4	
Shapiro-Wilk ( $p > 0.05$ )		0.7801	0.0117	0.0794	0.3802
d'Agostino-Pearson ( $p > 0.05$ )		0.5604	0.0178	0.0533	0.4258
ANOVA ( $p < 0.05$ )	F (2, 42)	1.0611	0.0400	0.5728	1.8710
	p	0.3551	0.9608	0.5683	0.1666
Pairwise T-test ( $p < 0.05$ )	g1 - g2	0.1985	0.9057	0.3436	0.1166
	g1 - g3	0.3638	0.7805	0.5648	0.9409
	g2 - g3	0.5629	0.8714	0.5827	0.1020
Kruskal-Wallis ( $p < 0.05$ )	Chi-square	1.8419	0.0216	2.3961	4.5921
	p	0.3981	0.9892	0.3018	0.1006
	df	2	2	2	2
Mann-Whitney ( $p < 0.05$ )	g1 - g2	0.2455	0.9339	0.1711	0.0745
	g1 - g3	0.2998	0.8682	0.5069	0.9669
	g2 - g3	0.6187	0.9669	0.2808	0.0620

TABLE 42 – Résultats complets des tests statistiques en intergroupe (tous les défauts).

		Shapiro-Wilk ( $p > 0.05$ )	d'Agostino-Pearson ( $p > 0.05$ )	Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
				One tail		Two tails			One tail	Two tails	t	df
				p-norm	p-exact	p-norm	p-exact	T				
Jeu 1-2	G1	0.01367	0.1962	0.0219	0.2006	0.0438	0.0413	24	0.0226	0.0452	2.1984	14
	G2	0.01983	0.2513	0.0124	0.0108	0.0249	0.0215	20	0.0596	0.1192	1.6598	14
	G3	0.0039	0.0040	0.2755	0.2807	0.5509	0.5614	49	0.1727	0.3453	0.9766	14
Jeu 2-3	G1	0.0573	0.1503	0.1743	0.1796	0.3487	0.3591	43	0.3370	0.6741	-0.4295	14
	G2	$4.5656e^{-4}$	$5.7237e^{-5}$	0.2568	0.2622	0.5136	0.5245	48	0.2265	0.4531	-0.7718	14
	G3	0.4498	0.6546	0.1340	0.1384	0.2681	0.2768	40	0.1105	0.2209	-1.2812	14
Jeu 3-4	G1	0.1498	0.7804	0.3774	0.3808	0.7547	0.7615	54	0.2455	0.4911	0.7071	14
	G2	$6.2827e^{-5}$	$4.9680e^{-7}$	0.0029	0.0017	0.0059	0.0034	11	0.0065	0.0129	2.8463	14
	G3	0.3668	0.6230	0.3774	0.3808	0.7547	0.7615	54	0.2817	0.5635	0.5917	14
Jeu 1-3	G1	0.2698	0.3782	0.0910	0.0938	0.1820	0.1876	36	0.0739	0.1478	1.5322	14
	G2	0.0028	$3.7064e^{-4}$	0.0191	0.0177	0.0381	0.0353	23	0.1562	0.3125	1.0477	14
	G3	0.0147	0.0110	0.4435	0.4452	0.8871	0.8904	57	0.3199	0.6398	0.4784	14
Jeu 2-4	G1	0.0891	0.1857	0.2051	0.2106	0.4102	0.4212	45	0.4212	0.8424	-0.2025	14
	G2	0.0119	0.0066	0.0057	0.0042	0.0115	0.0084	15	0.0038	0.0076	3.1159	14
	G3	0.2446	0.4480	0.3774	0.3808	0.7547	0.7615	54	0.3671	0.7341	-0.3465	14
Jeu 1-4	G1	0.9655	0.4871	0.0738	0.0757	0.1475	0.1514	34	0.0530	0.1061	1.7275	14
	G2	0.2532	0.3443	0.0035	0.0021	0.0070	0.0043	12	0.0030	0.0061	3.2280	14
	G3	0.0161	0.0243	0.2389	0.2443	0.4777	0.4887	47	0.2189	0.4378	0.7987	14

TABLE 43 – Résultats complets des tests statistiques en intragroupe (lancer de type javelot).

		Shapiro-Wilk ( $p > 0.05$ )	d'Agostino-Pearson ( $p > 0.05$ )	Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
				One tail		Two tails		T	One tail	Two tails	t	df
				p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	0.0027	0.0181	$8.1026e^{-4}$	$2.1362e^{-4}$	0.0016	$4.2725e^{-4}$	4	$3.6924e^{-4}$	$7.3848e^{-4}$	4.2965	14
	G2	$8.8603e^{-4}$	0.0032	0.0592	0.0603	0.1183	0.1205	32	0.0166	0.0332	2.3622	14
	G3	0.0013	0.0135	$9.8276e^{-4}$	$3.0518e^{-4}$	0.0020	$6.1035e^{-4}$	5	$1.1282e^{-4}$	$2.2564e^{-4}$	4.9202	14
Jeu 2-3	G1	0.0772	0.2829	0.4212	0.4235	0.8425	0.8469	56	0.4993	0.9986	0.0018	14
	G2	0.0243	0.1288	0.0469	0.047	0.0938	0.0946	30	0.0336	0.0671	1.9845	14
	G3	$1.022e^{-4}$	$8.5817e^{-5}$	0.2755	0.2807	0.5509	0.5614	49	0.1708	0.3416	0.9845	14
Jeu 3-4	G1	0.0715	0.4646	0.3351	0.3394	0.6701	0.6788	52	0.4445	0.8891	0.1420	14
	G2	0.0150	0.0968	0.4435	0.4452	0.8871	0.8904	57	0.4438	0.8875	-0.1440	14
	G3	0.0024	0.0062	0.0592	0.0603	0.1183	0.1205	32	0.0349	0.0698	-1.9630	14
Jeu 1-3	G1	0.0018	0.0110	$9.8276e^{-4}$	$3.0518e^{-4}$	0.0020	$6.1035e^{-4}$	5	$8.0872e^{-4}$	0.0016	3.8950	14
	G2	$1.2268e^{-4}$	0.0011	0.0219	0.0206	0.0438	0.0413	24	0.0084	0.0168	2.7140	14
	G3	$5.7966e^{-4}$	0.0021	0.0014	$5.7983e^{-4}$	0.0029	0.0012	7	$4.0077e^{-4}$	$8.0155e^{-4}$	4.2542	14
Jeu 2-4	G1	0.0376	0.3332	0.4660	0.4670	0.9321	0.9341	58	0.4633	0.9266	0.0937	14
	G2	0.0352	0.1804	0.0219	0.0206	0.0438	0.0413	24	0.0166	0.0332	2.3622	14
	G3	$3.0491e^{-4}$	$2.4575e^{-4}$	0.4435	0.4452	0.8871	0.8904	57	0.4895	0.9790	0.0268	14
Jeu 1-4	G1	0.0023	0.0137	0.0014	$5.7983e^{-4}$	0.0029	0.0012	7	$6.9935e^{-4}$	0.0014	3.9689	14
	G2	$1.2774e^{-4}$	$7.9260e^{-4}$	0.0191	0.0177	0.0382	0.0353	23	0.0081	0.0163	2.7300	14
	G3	0.0010	0.0026	0.0017	$7.6294e^{-4}$	0.0034	0.0015	8	$2.1431e^{-4}$	$4.2861e^{-4}$	4.5798	14

TABLE 44 – Résultats complets des tests statistiques en intragroupe (corps penché).

		Shapiro-Wilk ( $p > 0.05$ )	d'Agostino-Pearson ( $p > 0.05$ )	Wilcoxon Signed-Rank ( $p < 0.05$ )					T Test ( $p < 0.05$ )			
				One tail		Two tails		T	One tail	Two tails	t	df
				p-norm	p-exact	p-norm	p-exact					
Jeu 1-2	G1	0.0722	0.1662	0.0004	$3.0518e^{-5}$	0.0007	$6.1035e^{-5}$	0	$5.2824e^{-5}$	0.0001	5.3330	14
	G2	0.0286	0.2356	0.0067	0.0051	0.0135	0.0102	16	0.0031	0.0064	3.2028	14
	G3	0.7993	0.6119	0.0014	0.0006	0.0029	0.0012	7	0.0003	0.0007	4.3281	14
Jeu 2-3	G1	0.0402	0.1294	0.3991	0.4020	0.7983	0.8039	55	0.4617	0.9234	0.0978	14
	G2	0.0117	0.0035	0.0528	0.0535	0.1055	0.1070	31	0.0713	0.1426	1.5535	14
	G3	0.6875	0.2135	0.3146	0.3193	0.6293	0.6387	51	0.3554	0.7109	0.3783	14
Jeu 3-4	G1	0.4668	0.6112	0.3351	0.3394	0.6701	0.6788	52	0.3567	0.7135	-0.3747	14
	G2	0.0754	0.0428	0.1221	0.1262	0.2443	0.2524	39	0.1262	0.2523	1.1940	14
	G3	0.6169	0.6673	0.0528	0.0535	0.1055	0.1070	31	0.0435	0.0870	-1.8404	14
Jeu 1-3	G1	0.1183	0.0835	0.0007	0.0001	0.0013	0.0003	3	$8.2423e^{-5}$	0.0002	5.0895	14
	G2	0.0165	0.1820	0.0092	0.0075	0.0184	0.0151	18	0.0037	0.0073	3.1343	14
	G3	0.1135	0.0997	0.0010	0.0003	0.0020	0.0006	5	0.0001	0.0002	4.9613	14
Jeu 2-4	G1	0.0417	0.1981	0.3351	0.3394	0.6701	0.6788	52	0.4607	0.9214	-0.1005	14
	G2	0.0877	0.0150	0.0191	0.0177	0.0382	0.0353	23	0.0117	0.0235	2.5419	14
	G3	0.8220	0.6408	0.3560	0.3599	0.7120	0.7197	53	0.2657	0.5314	-0.6418	14
Jeu 1-4	G1	0.1496	0.1067	0.0004	$6.1035e^{-5}$	0.0009	0.0001	1	0.0001	0.0003	4.7662	14
	G2	0.0393	0.2506	0.0029	0.0017	0.0059	0.0034	11	0.0012	0.0023	3.7136	14
	G3	0.3415	0.1977	0.0017	0.0008	0.0034	0.0015	8	0.0003	0.0005	4.4723	14

TABLE 45 – Résultats complets des tests statistiques en intragroupe (tous les défauts).

**Titre :** Contribution à l'apprentissage humain de gestes à l'aide de techniques de clustering pour l'analyse de mouvements capturés

**Mots clés :** Apprentissage humain, EIAH, Mouvement capturé, Apprentissage de mouvements, Apprentissage automatique

**Résumé :** Cette thèse s'inscrit à la croisée des domaines des Environnements Informatiques pour l'Apprentissage Humain (EIAH) et de l'apprentissage du mouvement. Les EIAH utilisant le geste sont nombreux, et présents dans des domaines variés. Les EIAHs de la littérature sont le plus souvent conçus de façon *ad-hoc* ou se concentrent sur une tâche particulière dans un domaine d'apprentissage donné. Leur réutilisation dans d'autres situations d'apprentissage est souvent impossible ou nécessiterait un travail de réingénierie conséquent. Le développement d'EIAH support à l'enseignement de gestes, extensibles au-delà de la tâche pour laquelle elles ont été conçues et ayant un coût minimal en termes de réingénierie, représente un défi qui pose plusieurs verrous techniques et questions scientifiques. Afin de répondre à ces questions, le système *Motion Learning Analytics* (MLA) a été développé.

Ce système permet la captation, le filtrage, le traitement et l'analyse automatique et semi-supervisée du mouvement d'un apprenant. Ce système a été testé, sur des gestes de lancers, au travers de trois expérimentations, permettant d'évaluer quatre facettes du système proposé et fondé sur des propriétés cinématiques et géométriques du mouvement: (i) la possibilité d'obtenir une séparation des gestes en différents groupes correspondant à plusieurs stratégies de lancers, (ii) la possibilité d'obtenir une séparation en fonction de la réussite ou non de ces gestes, (iii) l'intégration des besoins d'observation de l'enseignant, en tant que critères de séparation des gestes et d'évaluation de la progression de l'apprenant et enfin, (iv) une analyse de la capacité du système à fournir une aide à la décision pertinente pour un expert, dans l'objectif d'améliorer le geste de l'apprenant.

**Title :** A Contribution to Human Motion Learning through Clustering Techniques for Captured Motion Analysis

**Keywords :** Human learning, TEL, Captured motion, Motion learning, Machine learning

**Abstract :** This PhD thesis lies at the crossroads of the Technology Enhanced Learning (TEL) and human motion learning fields. A lot of TEL systems for motion learning already exist, and they are used in numerous application domains. While efficient in the task they were designed for, they are usually *ad-hoc* by design, focusing on a specific task and learning context. Reusing such systems in other learning contexts is impossible or requires a heavy re-engineering process. The design of TEL systems for motion learning, expandable beyond the task they were created for and needing a minimal amount of re-engineering, represents a challenge from which arise several technical issues and scientific questions.

To tackle these challenges, the Motion Learning Analytics (MLA) system was developed. This system was tested on throwing motions through three different experimentations, designed to test four aspects of the system : (i) the possibility to achieve a good separation of the motions into multiple groups corresponding to different throwing strategies, (ii), the possibility to achieve a good separation corresponding to the degree of success of these motions, (iii), the integration of the expert's observations needs as criteria for the separation and evaluation of the learner's progression and (iv), an analysis on the ability of the system to provide an efficient and relevant assistance to the expert in order to improve the learner's gesture.