Table des matières

Table des figures **6**

Liste des tableaux **13**

1 Introduction et contexte **15**
1.1 Problématique . . . 15
1.2 Les captations du geste d'hier... . . . 17
1.3 ... et d'aujourd'hui . . . 20
1.4 Contexte . . . 22
1.5 Contributions . . . 23
1.6 Plan du document . . . 24

2 Etat de l'Art **27**
2.1 Les grands domaines de l'analyse de gestes . . . 27
2.1.1 Reconnaissance et analyse du geste . . . 27
2.1.2 Synthèse de gestes . . . 28
2.1.3 Segmentation de gestes . . . 28
2.1.4 Évaluation de gestes . . . 28
2.2 Représentation du geste . . . 29
2.2.1 Descripteurs pour résumer un geste . . . 29
2.2.1.1 Les descripteurs reposant sur un modèle du corps humain 29
2.2.1.2 Les descripteurs holistiques . . . 32
2.2.1.3 Les descripteurs locaux . . . 34
2.2.2 Discrimination de l'informativité de différents descripteurs . . . 34
2.3 Techniques d'apprentissage statistique . . . 36
2.3.1 Méthodes reposant sur un codage temporel . . . 37
2.3.1.1 Recherche des plus proches voisins . . . 37
2.3.1.2 Mesures de similarité . . . 37
2.3.1.3 Modèles Markoviens . . . 39
2.3.2 Méthodes reposant sur un codage global . . . 40
2.3.2.1 Machines à vecteurs de support (SVM) . . . 40
2.3.2.2 Forêt d'arbres décisionnels . . . 40
2.3.2.3 Réseaux de neurones . . . 41

2.4 Bilan 43

3 Modélisation de séries temporelles 45
Introduction et contexte 45
3.1 État de l'art 45
3.1.1 Alignement de séries temporelles par DTW 45
3.1.2 Modélisation de séries temporelles 47
3.1.2.1 Moyennage de deux séries temporelles 47
3.1.2.2 Extension au moyennage d'un jeu de séries temporelles . 48
3.1.3 Mise en évidence des chemins pathologiques 51
3.1.4 Le DTW contraint (CDTW) 53
3.1.4.1 Contraintes globales 53
3.1.4.2 Contraintes locales 53
3.2 Moyennage de séries temporelles : le DBA contraint 55
3.3 Modélisation de la variabilité intraclasse : la tolérance 57
3.4 Bases de données utilisées pour la validation 58
3.4.1 Séries temporelles 1D : UCRTSArchive 58
3.4.2 Gestes : ArmGesturesM2S 60
3.5 Moyennage de séries temporelles 60
3.5.1 Procédure de classification 60
3.5.2 Résultats 61
3.6 Modélisation de la variabilité intraclasse 66
3.6.1 Procédure de classification 66
3.6.2 Résultats 66
3.7 Extension à la classification de gestes 68
3.7.1 Procédure de classification 68
3.7.2 Résultats 69
Conclusion 70

4 Mesure de qualité d'un geste sportif 71
Introduction 71
4.1 État de l'art 72
4.1.1 Évaluation de gestes chirurgicaux 72
4.1.2 Évaluation de gestes sportifs 73
4.2 Bases de données et codage du mouvement 74
4.2.1 Notations et codage du mouvement 74
4.2.2 Conditions expérimentales : bases de données 77
4.2.2.1 Le service de tennis 77
4.2.2.2 Le *Zuki* au karaté 79
4.3 Modélisation du mouvement expert 79
4.3.1 Mouvement nominal 79
4.3.2 Tolérance articulaire 81
4.4 Evaluation du mouvement d'un novice 82
4.4.1 Erreurs spatiales 83

4.4.2 Erreurs temporelles . . . 85
4.5 Méthodolologie . . . 87
4.5.1 Annotations . . . 87
4.5.2 Procédure d'évaluation . . . 89
4.6 Résultats . . . 90
4.6.1 Reconnaissance de phases . . . 90
4.6.2 Evaluation spatiale de la qualité d'un geste sportif . . . 92
4.6.3 Evaluation temporelle de la qualité d'un geste sportif . . . 94
Conclusion . . . 96

5 Entraîneur virtuel **99**
Introduction . . . 99
5.1 État de l'art . . . 99
5.1.1 Les différents systèmes d'entraînement . . . 99
5.1.2 Retours d'information . . . 101
5.2 Transposition à un système en ligne . . . 103
5.2.1 Le DTW segmentant (SDTW) . . . 105
5.2.1.1 Approche classique . . . 107
5.2.1.2 Approche temps-réel . . . 109
5.2.2 Recalage de squelettes par transformations locales . . . 111
5.2.3 Mise en place d'une interface adaptée . . . 114
5.2.3.1 Étude de la répétabilité des erreurs d'un novice . . . 114
5.2.3.2 Déroulement d'une séance d'évaluation . . . 115
5.3 Utilisation de l'outil d'évaluation temps-réel . . . 115
Conclusion . . . 123

Conclusion et perspectives **125**

Références bibliographiques **127**

Table des figures

1.1 Théodore Géricault, *Le Derby D'Epsom*, 1821, peinture à l'huile sur toile, 92 × 123 cm, Musée du Louvre, Paris. 17
1.2 Deux systèmes de captation du mouvement, par Eadward Murybridge (a) et Etienne-Jules Marey (b). 18
1.3 Études du mouvement humain par images successives ou chronophotographie, par Eadweard J. Muybridge et Jules-Etienne Marey. 19
1.4 Etienne-Jules Marey, *Course*, épure chronophotographique, 1886. 20
1.5 De l'expérimentation à la reconstruction de mouvements en $3D$. 21
1.6 Franchissement de haie d'un athlète de jambe d'attaque gauche : image de synthèse issue de [1]. 22

2.1 Les différents types de codages du geste 30
2.2 Figure extraite de [2]. Les trois articulations en rouge forment un plan symbolisé par un disque vert, le positionnement de l'articulation bleue est évalué relativement à ce plan donnant un résultat binaire. L'association de plusieurs évaluations de positionnement permet d'extraire un vecteur de variables binaires simple et invariant à la morphologie du sujet. 31
2.3 Figure extraite de [3]. Les effecteurs terminaux (tête, pieds, mains) sont utilisés pour former des polygones représentatifs de la pose du squelette. . 32
2.4 Figure extraite de [4]. Énergies du mouvement (MEI) et Historiques du mouvement (MHI) résultant d'un mouvement de bras. Alors que les MEI sont relativement similaires pour les deux derniers mouvements, les MHI, conservant l'information temporelle, permettent de les distinguer. 33
2.5 Figure extraite de [5]. Comparaison de deux métriques de similarité entre deux séries temporelles. À gauche, la distance euclidienne mesure la distance instant après instant ; de fait, elle reflète mal la distance entre deux signaux de temporalités différentes. À droite, l'alignement par DTW permet un recalage non linéaire des deux signaux. Plus intuitive, elle permet d'apparier les événements similaires. 37
2.6 Exemple de représentation graphique d'un HMM à N états. Pour chaque état S_i, les densités de probabilité $b_{i,k}$ de chaque descripteur o_k sont tracées. $a_{i,j}$ représente la probabilité de transiter d'un état S_i à un état S_j. . 40

2.7 Représentation $2D$ des descripteurs, colorés en vert ou bleu selon la classe à laquelle ils appartiennent. L'hyperplan de séparation est indiqué par un trait plein noir, linéaire dans le premier cas et non linéaire dans le second. Plus particulièrement, la figure 2.7b illustre un cas non linéairement séparable dans lequel le *Kernel Trick* permet de séparer les descripteurs par un hyperplan dans l'espace transformé. 41
2.8 Illustration d'un arbre décisionnel aléatoire. A gauche, l'extrémité des branches définit les issues possibles, qui sont atteintes en fonction des décisions prises à chaque étape. Sur la figure de droite est tracé le partitionnement de l'espace (X_1, X_2) qui en résulte. 42

3.1 (a) Les deux signaux (bleu et vert) sont alignés par DTW. La correspondance d'indices est indiquée par des segments gris. (b) La figure de droite superpose la carte de distance cumulée $\mathbf{D}$ et le chemin de déformation (en vert). Le blanc correspond à une grande valeur de distance cumulée $D_{i,j}$ et le noir à une faible valeur. 47
3.2 Illustrations de la méthode de moyennage des deux signaux. Les deux figures correspondent aux deux signaux présentés en figures 3.1 et 3.5. Dans les deux cas, $\tilde{x}(k)$ (en vert) et $\tilde{y}(k)$ (en bleu) de même taille K on été obtenus par rééchantillonnage de $x(i)$ et $y(j)$ relativement à ϕ_{xy}. Le signal moyen résultant est $\mu(k)$ (en noir). 48
3.3 Une itération du DBA sur deux jeux de signaux différents. Le signal noir est le signal moyen $\mu(k)$ résultant de l'itération précédente. Les signaux bleu et vert sont alignés sur le signal noir. Chaque point de $\mu(k)$ est mis à jour comme la moyenne des points qui lui sont appariés (par exemple les points rouges pour l'indice correspondant). 50
3.4 Moyennage de deux jeux de séries par DBA après 4 itérations. Les séries temporelles moyennées sont similaires à celles présentées précédemment. Dans le premier cas (a), le jeu est constitué de 319 exemples, et le second (b) de 101 exemples. 51
3.5 Alignement par chemin de déformation pathologique. (a) Les deux signaux (bleu et vert) sont alignés par DTW. Notez les correspondances pathologiques de plusieurs indices sur un seul. (b) La figure de droite superpose la distance cumulée $\mathbf{D}$ et le chemin de déformation (en vert). La pathologie du chemin de déformation se traduit par de longues zones verticales ou horizontales de stagnation. 52
3.6 Contraintes globales du chemin de déformation couramment rencontrées dans la littérature. Lors de l'alignement de deux signaux, l'espace de recherche du chemin de déformation est restreinte à la zone grise. 53
3.7 Mise en place du CDTW : introduction de nouveaux déplacements élémentaires. Seulement certains déplacements sont autorisés dans le calcul de $D_{i,j}$. (a) Cas particulier de trois déplacements élémentaires admissibles. (b) Cas général de déplacements locaux contraints. 55

3.8 Une itération du CDBA sur deux jeux de signaux différents. Le signal noir est le signal moyen $\mu(k)$ résultant de l'itération précédente. Les signaux bleu et vert sont alignés sur le signal noir. Chaque point de $\mu(k)$ est mis à jour comme la moyenne des points qui lui sont appariés (par exemple les points rouges pour l'indice correspondant). 57

3.9 Signaux moyens obtenus par DBA (en vert) et CDBA (en bleu) après 4 itérations pour les 2 jeux de signaux des figures 3.1a et 3.5a. A l'inverse de ceux obtenus par DBA, les signaux moyens obtenus par CDBA n'engendrent pas de discontinutés. 57

3.10 Séries temporelles moyennes obtenues par CDBA pour deux jeux de signaux distincts. La tolérance, calculée comme $\pm(1 \times \sigma)$, est indiquée par la zone grisée autour de ce signal moyen. 59

3.11 Les 15 gestes du haut du corps de la base de données *ArmGesturesM2S*. Certains mouvements ont des propriétés géométriques très similaires. Illustration issue de [6]. 61

3.12 Taux de classification en fonction de K_p pour deux bases de données : (a) *Lighting7* qui contient des signaux peu décalés et (b) *TwoPatterns* qui contient des signaux très décalés. Il convient de noter que la valeur optimale de K_p dépend des décalages des signaux intraclasses. 62

3.13 Séries temporelles utilisées pour comparer les comportements du CDTW et du DTW. (a) Quatre signaux appartenant à la même classe de la base *TwoPatterns*. (b) Un signal appartenant à la classe $\mathcal{C}_1$ (en bleu) et trois autres appartenant à la classe $\mathcal{C}_2$ (en vert, marron et rouge) de la base *Lighting7*. 63

3.14 Cas particulier de deux signaux très décalés de la base *TwoPatterns*. (a) Les signaux sont alignés par DTW et (b) les signaux sont alignés par CDTW. Dans les deux cas, l'appariement des indices est matérialisé par des segments gris. La seconde ligne présente les séries temporelles moyennes obtenues par (c) DBA et (d) CDBA. 64

3.15 Illustration des alignements par DTW et CDTW pour deux signaux appartenant à des classes différentes. Première ligne : alignement par (a) DTW et (b) CDTW, avec $K_p = 2$. Deuxième ligne : séries temporelles déformées résultant de l'appariement par le chemin de déformation obtenu par (c) DTW et (d) par CDTW. 65

3.16 Taux de classification par DBA et CDBA avec et sans tolérance en fonction de K_p. 67

4.1 (a) Positionnement des marqueurs sur une personne et (b) positionnement des articulations du squelette. On note particulièrement la position de la racine (*root*). Les noms abrègent les articulations correspondantes (par exemple "RElb" abrège *Right Elbow*, *i.e.* coude droit en anglais). 75

4.2 Définition des membres considérés : en gris, les articulations du tronc ; en orange, celles du bras gauche ; en bleu, celles du bras droit ; en rouge, celles de la jambe gauche et en vert, celles de la jambe droite. 76
4.3 Kinogramme d'un service de tennis. 78
4.4 Kinogramme d'un mouvement de *Zuki*. 80
4.5 Deux services de tennis alignés par le DTW. Les deux premières lignes représentent sous forme de kinogrammes deux services de tennis de durées différentes, la troisième ligne superpose ces deux services une fois alignés par DTW. 81
4.6 Tolérance spatiale du poignet gauche (en jaune-vert) et de la hanche droite (en bleu) à un instant k donné. Le mouvement nominal à cet instant est illustré en noir opaque, et correspond à la moyenne des gestes alignés, en gris sur la figure. Pour plus de lisibilité, les tolérances affichées correspondent à $3 \times \boldsymbol{\sigma}_n^a(k)$ à titre illustratif. 82
4.7 Illustration de l'erreur spatiale de la série temporelle $x_l(k)$ dans le cas unidimensionnel, avec $x_n(k)$ et $\sigma_n(k)$ la série temporelle nominale et sa tolérance. À gauche sont représentés les signaux avant recalage. À droite, les signaux ont été déformés selon le chemin de déformation $\phi_{x_l x_n}(k)$. . . . 84
4.8 Mise en place de l'erreur temporelle entre deux membres m_1 et m_2 (ici deux bras) à partir des chemins de déformation qu'ils engendrent. (a) Recalage de deux squelettes basé sur le bras gauche (m_1) uniquement. (b) Recalage de deux squelettes basé sur le bras droit (m_2) uniquement. (c) Chemin de déformations engendrés. 86
4.9 Outil d'annotation spatiale pour le service de tennis. 88
4.10 Instants-clé subdivisant le service de tennis en quatre phases. 89
4.11 Détection des phases d'un mouvement. 90
4.12 Comparaison des estimations instants clés estimés par notre approche avec les annotations de l'entraîneur. 91
4.13 Comparaison des scores spatiaux avec les annotations de l'entraîneur pour chaque phase du service de tennis. L'alignement choisi est le CDTW avec $K_p = K_{pm} + 1$. Le coefficient de corrélation entre les annotations (axe des x) et les estimations (axe des y) est donné en légende. 93
4.14 Comparaison des évaluations spatiales avec les annotations de l'entraîneur de karaté pour le *Zuki*. L'alignement choisi est le CDTW avec $K_p = K_{pm} + 1$. Le coefficient de corrélation entre les estimations et les annotations est $\rho = -0.77774$ ($p = 0.0029$). 94
4.15 Première ligne : un mouvement de *Zuki* exécuté par un expert. Seconde ligne : un mouvement de *Zuki* effectué par un novice. Cet exemple montre une temporalité parfaite des bras de l'expert comparativement au novice qui n'amorce le mouvement de son bras droit que lorsque le bras gauche est pratiquement arrivé à sa position finale (entre t_5 et t_7). 96

4.16 Comparaison des évaluations temporelles avec les annotations de l'entraîneur de karaté concernant le geste du *Zuki*. L'alignement choisi est le CDTW avec $K_p = K_{pm} + 4$. Le coefficient de corrélation entre les estimations et les annotations vaut $\rho = -0.90792$ une fois les essais du novice N_6 enlevés (en gris). 97

5.1 Principe général de communication d'un système d'entrainement 101

5.2 Schéma global d'entraînement du novice. Les traits noirs indiquent les processus hors ligne, les rouges les processus en ligne et répétés lors de l'entraînement. En bleu est indiqué le processus de calibration, qui n'a lieu qu'une fois au début de la séance et qui est utilisé lors de l'alignement et l'évaluation du geste novice. 104

5.3 Contexte du SDTW : à partir d'une séquence non segmentée, le but est de reconnaître et aligner chacun des motifs similaires à un certain motif (représenté en rouge sur la première ligne) 106

5.4 (a) Carte de distance cumulée initiale issue de l'application du SDTW entre les signaux $x(i)$ et $y(j)$. (b) Dernière ligne de la carte de distance cumulée $D_{M,j}$. 107

5.5 (a) Carte de distance cumulée initiale issue de l'application du SDTW entre les signaux $x(i)$ et $y(j)$. En blanc sont tracés les chemins de déformation obtenus. (b) Dernière ligne de la carte de distance cumulée $D_{M,j}$ (en bleu) et seuil choisi de manière à ce que les sous-séquences ne correspondant pas au motif ne soient pas détectées (en rouge). 108

5.6 Matrice de départ $\boldsymbol{S}$ générée par l'application du SDTW "temps-réel" entre le signal motif et la séquence temporelle de la figure 5.3. 109

5.7 Illustration de la nécessité du recalage de squelette par transformations locales. Un même mouvement acquis par une Kinect® (a) ou par un système de capture optoélectronique Vicon® (b) positionne différemment les centres articulaires du sujet. 111

5.8 Mise en place des bases locales $\mathcal{R}_1 = (RElb_1, \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|}, \frac{\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|}, \frac{\boldsymbol{z}_1}{\|\boldsymbol{z}_1\|})$ et $\mathcal{R}_2 = (RElb_2, \frac{\boldsymbol{x}_2}{\|\boldsymbol{x}_2\|}, \frac{\boldsymbol{y}_2}{\|\boldsymbol{y}_2\|}, \frac{\boldsymbol{z}_2}{\|\boldsymbol{z}_2\|})$ de l'articulation du coude obtenues *via* deux systèmes de capture différents. En bleu sont représentées les 3 articulations $RSho_1$, $RElb_1$ et $RWri_1$ du bras droit obtenues grâce au premier outil de capture. En vert celles obtenues grâce au deuxième outil de capture. À partir de ces trois points dans l'espace, on introduit un repère orthonormé direct avec $\boldsymbol{x}$ dirigé selon l'avant-bras et $\boldsymbol{z}$ orthogonal au plan formé par le bras et l'avant-bras. 113

5.9 Projection en z de la trajectoire de la main droite lors d'un service de tennis exécuté par 6 novices distincts (de N_1 à N_6). Chaque novice réalisant 10 essais à la suite, on obtient à chaque fois 10 trajectoires qui sont superposées. 114

5.10 Session d'entraînement d'un novice avec Optitrack®. 37 marqueurs sont positionnés sur le novice, 8 caméras extraient leurs positionnements en $3D$. L'interface d'entraînement est projetée au novice lors de sa session. . 117
5.11 L'utilisateur indique le nombre d'essais qu'il souhaite réaliser. 118
5.12 Des instructions sont données à l'utilisateur quant à la phase de calibration : il doit faire suivre à son squelette le mouvement d'un squelette de référence. 118
5.13 Son squelette est modélisé par des traits rouges, celui à suivre est en noir. Deux vues sont affichées. 119
5.14 Le mouvement attendu est un mouvement très simple d'ouverture et de fermeture des bras. 119
5.15 Une fois la calibration terminée, l'utilisateur peut exécuter son mouvement qui sera enregistré et évalué. 120
5.16 Le sujet termine son mouvement de service de tennis. 120
5.17 L'erreur principale est affichée au joueur avec un code couleur adapté. . . 121
5.18 Le processus se répète autant de fois que l'a décidé l'utilisateur au départ. 121
5.19 À la fin de l'entraînement, un récapitulatif global est donné à l'utilisateur. 122

Liste des tableaux

3.1 Caractéristiques de l'archive *UCR Time-Series Classification Archives* [7] 59
3.2 Caractéristiques de la base de données *ArmGesturesM2S*, introduite dans [6]. 60
3.3 Taux de classification obtenus par DBA et CDBA avec une contrainte de pente K_p variant de 2 à 11. La dernière colonne reporte les meilleurs taux du CDBA pour chaque base. 63
3.4 Taux de classification obtenus par DBA et CDBA avec et sans tolérance. Les résultats par CDBA affichés ont été obtenus après optimisation du coefficient K_p sur chaque base. Le CBDA avec tolérance donne le meilleur taux de classification moyen. 67
3.5 Taux de classification de gestes par DBA et CDBA avec et sans tolérance avec le paramètre K_p variable, de $K_p = K_{pm}$ jusqu'à $K_p = K_{pm} + 6$. . . . 69

4.1 Récapitulatif des annotations faites par les entraîneurs. m représente les membres évalués et p différentes phases constituant le geste. 89
4.2 Validation de l'évaluation spatiale du service de tennis. Coefficient de corrélation pour les différents protocoles de classification : avec ou sans tolérance, par DTW ou CDTW avec K_p variable. 95
4.3 Validation de l'évaluation spatiale du *Zuki*. Coefficient de corrélation pour les différents protocoles de classification : avec ou sans tolérance, par DTW ou CDTW avec K_p variable. 95

Chapitre 1

Introduction et contexte

Cette thèse présente les travaux de trois années de doctorat en France. Elle s'est déroulée entre deux laboratoires de recherche complémentaires :

- le laboratoire *M2S* (Mouvement, Sport Santé) de Ker Lann, proche de Rennes, dans lequel la première année de doctorat d'octobre 2014 à octobre 2015 a été effectuée, sous l'encadrement de Richard Kulpa. Ce laboratoire est rattaché à l'UFR Sciences et Techniques des Activités Physiques et Sportives (STAPS) de Rennes, et est composé de deux axes ("Sport & santé" ; "Sport & performance"). L'expertise en biomécanique de cette dernière équipe ainsi que les dispositifs dont il bénéficie ont permis à la fois de créer une base de données complète et pertinente, mais aussi de positionner la problématique de la thèse dans son contexte sportif, avec tous les enjeux qui en découlent.

- l'*ISIR* (Institut des Systèmes Intelligents et de Robotique) de Paris, Unité Mixte de Recherche 7222 commune à l'Université Pierre et Marie Curie (UPMC) et au Centre National de la Recherche Scientifique (CNRS). Le doctorat s'est poursuivi entre octobre 2015 et octobre 2017 dans ce second laboratoire, sous l'encadrement de Séverine Dubuisson et Catherine Achard dans l'équipe "Intégration Multimodale, Interaction et Signal Social" dirigée par Mohamed Chetouani. Cette fois-ci, l'expertise en traitement du signal et en apprentissage statistique de l'équipe ont permis de formaliser le problème posé et d'y répondre aussi rigoureusement que possible.

La mise en commun des compétences de ces deux laboratoires a permis de donner aux travaux de ce doctorat un champ de vision scientifique très vaste mais a aussi de donné à ces travaux un cadre à la fois théorique et applicatif très appréciable.

1.1 Problématique

Le mouvement peut être vu comme la variation temporelle du positionnement d'un corps dans l'espace. De fait, un mouvement naît de l'association subtile d'une variabilité

spatiale et temporelle. Dans le cadre de cette thèse, c'est le mouvement sportif qui sera au centre de nos préoccupations.

Ces dernières années ont vu apparaître le développement intensif d'objets connectés et d'applications mobiles dédiés au mouvement sportif du grand public. Effectivement, les nouvelles technologies de pointe ont permis de mettre sur le marché des dispositifs informatiques de moins en moins coûteux, de moins en moins encombrants, de plus en plus précis. Majoritairement développés pour le jeu vidéo, ces dispositifs peu coûteux de captation du geste peuvent être embarqués (télécommande de la console Wii, multiples applications mobiles basées sur l'utilisation d'accéléromètres) ou fixes (Kinect®, Balance Board de la Wii®, *etc.*). L'expert préférera souvent les dispositifs de captation, plus onéreux mais aussi plus précis, tels que la *Motion Capture* ou les capteurs inertiels. D'autres outils plus spécifiques au sport sont également développés, tels que le Cyclus2 pour le cyclisme, le Plane Swing Training System pour le golf ou encore le Smart Basket-Ball pour le basket. Ainsi, il ne manque aujourd'hui pas de capteurs permettant d'acquérir des informations sur un geste. L'utilisation de cette information gestuelle peut-être multiple.

Elle est ludique dans le cadre du jeu vidéo : vue comme un outil interactif elle permet souvent le déplacement d'un objet, d'un personnage dans un espace virtuel. L'information gestuelle peut également être utilisée à des fins de détection de comportements dangereux en télésurveillance. De manière semblable, par la reconnaissance d'activités, elle peut être utilisée dans la domotique pour automatiser des pilotages d'appareils de la vie quotidienne par exemple. Enfin, elle peut faire l'objet d'une évaluation, par exemple dans un cadre médical pour permettre à un chirurgien d'améliorer son geste, ou encore dans un cadre sportif pour optimiser la performance d'un athlète.

Néanmoins, il est encore un pas à franchir quant à l'utilisation de capteurs à des fins de progression sportive, même si la tendance est à son expansion. Très peu d'entraîneurs ou de sportifs utilisent aujourd'hui des capteurs pour acquérir des informations sur les gestes, bien qu'ils seraient probablement très appréciés. Peut-être est-ce à cause d'une fidélité aveugle aux méthodes anciennes établies, un manque de recul ou tout simplement une ignorance ou une appréhension du procédé. Toujours est-il qu'il reste beaucoup à faire, à la fois pour faire connaître les ressources actuelles, leurs apports mais aussi pour les adapter au public visé.

L'objectif de cette thèse est d'amener une réponse à ce besoin *via* la mise en place d'un outil d'évaluation automatique et générique d'un quelconque geste sportif individuel.

Avant d'en venir au cœur même du sujet, revenons quelques décennies en arrière, aux origines de la captation du mouvement. À l'époque, la notion de mouvement agitait les plus curieux, puisqu'elle représentait quelque chose qui ne pouvait être enregistré, de part sa variabilité spatio-temporelle créant un déplacement souvent incompris à l'oeil nu.

1.2 Les captations du geste d'hier...

Comprendre le geste peut, de nos jours, paraître assez banal voire ordinaire. Qu'il soit utile au physicien, au joueur de jeux vidéos ou au commentateur sportif, il ne manque pas d'outils pour capter, rejouer, figer ce qui ne peut être enregistré à la simple observation visuelle, comme on vient de l'évoquer.

Il faut cependant se rappeler qu'il y a 200 ans, la compréhension d'un geste, qu'il soit humain ou animal, était au cœur des discussions scientifiques et artistiques puisqu'il était alors impossible de comprendre l'invisible. De fait, l'instantané d'un mouvement ne pouvait qu'être imaginé, donnant lieu à l'époque à nombre de peintures irréalistes telles que celle très répandue de Théodore Gericault, *Le Derby D'Epsom*, représentée en figure 1.1. Deux hommes, aux destins qu'on peut penser liés, puisque tous deux nés en

FIGURE 1.1 – Théodore Géricault, *Le Derby D'Epsom*, 1821, peinture à l'huile sur toile, 92 × 123 cm, Musée du Louvre, Paris.

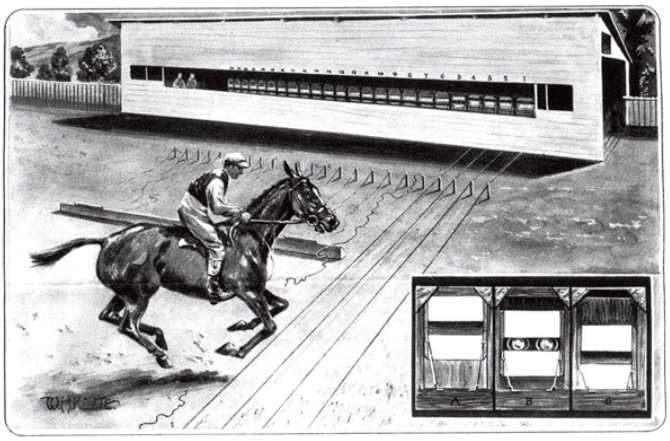
1830, morts en 1904 et portant les mêmes initiales, sont à l'origine des premiers travaux révolutionnaires sur la captation du mouvement. Tous deux auraient pu inventer le cinéma, seulement ils n'y voyaient pas d'intérêt : "Pourquoi montrer ce qu'on voit déjà ?". L'important était pour eux de reculer les limites du visible, de l'impalpable.

Le premier, Etienne-Jules Marey, français, était issu du milieu médical. Toujours intéressé par le mouvement, en particulier celui du sang, une de ses principales préoccupations était de mesurer la circulation. Il met alors au point le sismographe. Technicien, il décide très rapidement de se focaliser sur la locomotion terrestre et aérienne. Une de ses grandes obsessions est alors de parvenir à saisir le mouvement rapide du cheval et de traduire ce mouvement sous forme synthétique. Il s'interroge également sur le vol d'un oiseau, cherchant avant tout à comprendre comment un objet plus lourd que l'air peut

se maintenir dans l'air.

Son contemporain britannique, Eadweard J. Muybridge, suit un parcours différent puisqu'il est avant tout artiste. À la suite d'un accident gravissime de cheval, celui-ci s'initie à la photographie lors de sa convalescence. La qualité de son matériel et le soin qu'il accorde à ses clichés font rapidement de lui un photographe de prestige. Il est capable de **saisir l'espace** comme peu d'hommes de son époque. Sa renommée le conduit auprès de Leland Stanford, célèbre mécène à qui on doit l'université du même nom. Celui-ci, obsédé par le cheval, prenant part aux débats hippiques et s'intéressant au trot et au galop de l'animal, fait appel à Muybridge pour saisir la course d'un cheval. À cette époque, le cheval est partout, il est l'objet de discussions de passion, de représentations esthétiques, *etc.*

Eadweard J. Muybridge est alors sollicité pour résoudre la question du cheval à haute vitesse. Le photographe doute, pense que l'enjeu est trop difficile, ses premières prises de vue ne sont d'ailleurs pas très concluantes. Il persiste cependant et met en place un système très sophistiqué mais aussi très encombrant contenant 24 appareils photos et 24 fils qui les déclenchent lorsque le cheval les heurtent en courant. Son système est illustré sur la figure 1.2a, sur laquelle on peut voir un cheval déclencher un à un les 24 fils reliés aux 24 appareils photo alignés dans le box blanc. Suite à cette prodigieuse invention,



(a) Système de caption du mouvement équestre mis en place par Eadweard J. Muybridge pour son œuvre "The Horse In Motion". Impression publiée dans *London News* le 18 juillet 1931.

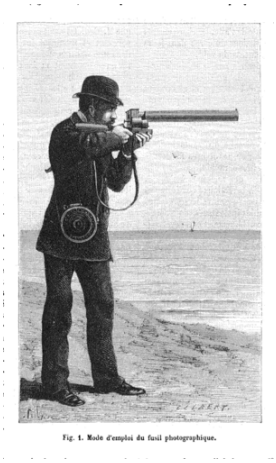
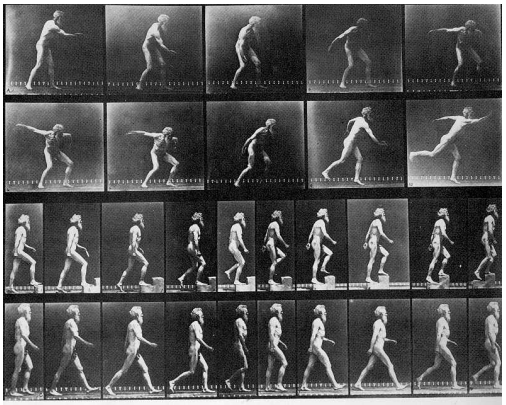


Fig. 1. Mode d'emploi du fusil photographique.

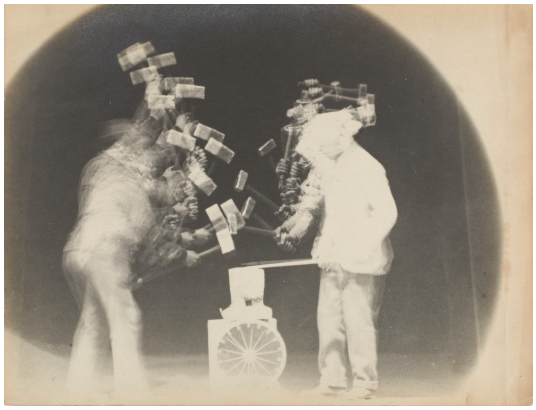
(b) Fusil photographique, illustration d'article d'Étienne-Jules Marey paru dans *la Nature* numéro 464 du 22 avril 1882.

FIGURE 1.2 – Deux systèmes de captation du mouvement, par Eadward Murybridge (a) et Etienne-Jules Marey (b).

le britannique devient très célèbre. Il met également en œuvre un appareil qui projette les images de façon séquentielle ; c'est, pourtant quelques décennies avant l'invention du cinéma, une première idée de film qui germe. Muybridge tient des conférences durant



(a) Eadweard J. Muybridge, *Man throwing discus, walking up steps, walking*, Animal Locomotion, 1887.



(b) Etienne-Jules Marey, *Le coup de marteau*, chronophotographie sur plaque fixe, 16.3 × 20.2 cm, 1895.

FIGURE 1.3 – Études du mouvement humain par images successives ou chronophotographie, par Eadweard J. Muybridge et Jules-Etienne Marey.

lesquelles il présente ses travaux, et rapidement la peinture est concurrencée par ce dispositif bien plus vrai que l'imaginaire artistique. De fait, certains artistes tels que le renommé Ernest Meissonier reprennent leurs peintures et les rectifient conformément au travaux de Muybridge.

La réputation et les travaux du photographe s'étendant outre-manche, ils arrivent rapidement aux oreilles de Etienne-Jules Marey, qui décide alors de contacter le désormais très célèbre britannique. Dans sa lettre, le technicien sollicite de Eadweard J. Muybridge la captation du mouvement des ailes des oiseaux. Pas vraiment convaincu par ce que lui propose l'anglais (son système à base de fils a forcément des limites lorsqu'il s'agit de capter un animal volant), Marey parvient par lui-même à fixer de mieux en mieux et de plus en plus précisément le mouvement de la mouette. Pour ce faire, il met en place un fusil photographique de 25 images qui vise et accompagne l'oiseau lors de son vol, comme le montre la gravure de la figure 1.2b. Cette brillante idée recyclée du système séquentiel du fusil dans un but photographique lui permet, lui aussi, de **saisir l'instantané**.

L'un comme l'autre se tournent alors vers le mouvement humain. Cette fois, c'est le monde de la caricature qui s'approprie l'instantané grâce à la technique développée. Bénéfice de l'entraîneur de sport ou simple œuvre esthétique, la chronophotographie se développe alors avec des sujets de toute sorte, comme en témoignent les œuvres de la figure 1.3.

Alors que le britannique voit dans ses clichés une œuvre esthétique et prône la nudité de ses sujets, pour le français, l'image ne suffit pas en elle-même, il faut l'expliquer. C'est là que la ligne apparaît. On ne se contente plus du corps, on fait apparaître les tracés. Pour ce faire, Marey habille ses sujets tout en noir et peint une ligne blanche le long du squelette de la personne. Enregistrant le mouvement du sujet devant un fond sombre, il vise alors à faire ressortir uniquement le tracé du squelette, comme le montre la figure

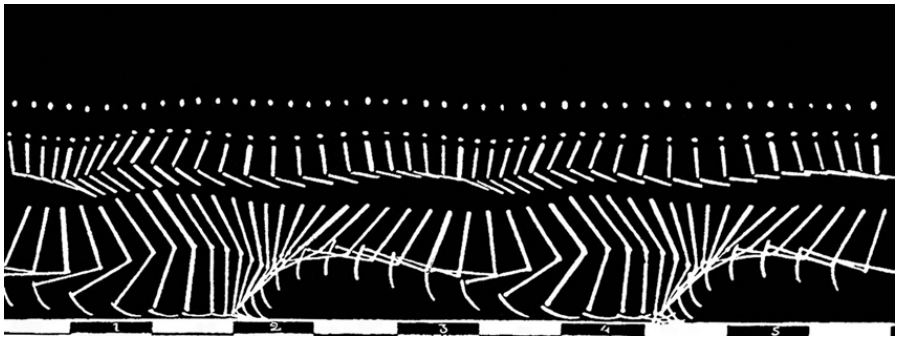


FIGURE 1.4 – Etienne-Jules Marey, *Course*, épure chronophotographique, 1886.

épurée 1.4 : c'est le tout début de la *Motion Capture*.

Les techniques actuelles, certes bien plus poussées, précises, et permettant une reconstruction dans l'espace d'un squelette humain en mouvement, se basent cependant sur un principe qui n'a pas tellement évolué par rapport à celui proposé par Etienne-Jules Marey à la fin du XIXème siècle, comme il est temps de l'exposer maintenant.

1.3 ... et d'aujourd'hui

Effectivement, aujourd'hui, une multitude d'outils permettent d'acquérir un geste, plus ou moins précis, plus ou moins onéreux et plus ou moins encombrants. D'abord, plusieurs auteurs ont analysé des gestes avec de simples caméras vidéo. Cette méthode est certes peu onéreuse et très simple à mettre en place, elle ne donne cependant pas directement accès à la mesure mouvement. Pour évaluer un mouvement acquis par caméra vidéo, on pourra typiquement extraire le squelette par différentes techniques de suivi résumées dans [8]. Ceci dit, la précision obtenue est discutable selon le degré d'exactitude que l'on veut donner à l'évaluation. D'autres outils plus spécifiques sont développés, notamment les caméras de profondeur telles que la Kinect V2® (Microsoft Corporation, Washington, Etats-Unis), qui permet d'extraire un squelette approximatif du sujet placé devant elle. Certains chercheurs travaillent à améliorer la précision de l'estimation de pose à partir de données issues d'une Kinect tels que Plantard *et al.* [9]. Pléthore d'autres outils pourraient être cités ici, comme par exemple les accéléromètres de la télécommande Wiimote® ou plus récemment les Joycon de la Nintendo Switch® (Nintendo, Kyoto, Japon), le Perception Neuron (Noitom Ltd., Pékin, Chine), la Wii Balance Board® (Nintendo, Kyoto, Japon) ou encore des outils plus spécifiques tels que le Cyclus 2 (RBM elektronik-automation GmbH, Leipzig, Allemagne) dédié à l'étude du mouvement du cylciste. Tous permettent, d'une façon ou d'une autre, embarqués ou non, d'acquérir une information sur un mouvement, localisés sur quelque(s) membre(s) ou sur le corps entier.

Un outil très précis et qui est au cœur de nos travaux est le système de capture optoélectronique (*Motion Capture System*), utilisé notamment dans [10, 11, 12, 13, 14]. Cet outil d'acquisition de mouvement s'est répandu ces dernières années dans l'industrie cinématographique et les jeux vidéo, puisqu'il permet de capturer avec une très grande

précision des positionnements $3D$. Sa mise en place est plus complexe puisqu'elle nécessite de fixer sur le sujet des marqueurs réfléchissants dont les positions tridimensionnelles sont enregistrées par des caméras infrarouges situées autour du sujet exécutant un geste. Une fois les positionnements enregistrés, une étape d'étiquetage (ou *labellisation*) est alors nécessaire pour identifier chacun des marqueurs au cours du temps. Selon l'application, un squelette peut être extrait au cours du temps. Les différentes étapes de l'acquisition du geste par *motion capture* sont illustrées en figure 1.5, dans une situation d'écran au basket-ball enregistrée par 10 caméras infrarouges `Vicon MX-40` (Oxford Metrics Inc., Oxford, Grande Bretagne) au laboratoire M2S. Les sujets sont alors tous équipés de marqueurs positionnés sur des repères anatomiques bien précis, de manière à estimer le plus précisément possible les squelettes.

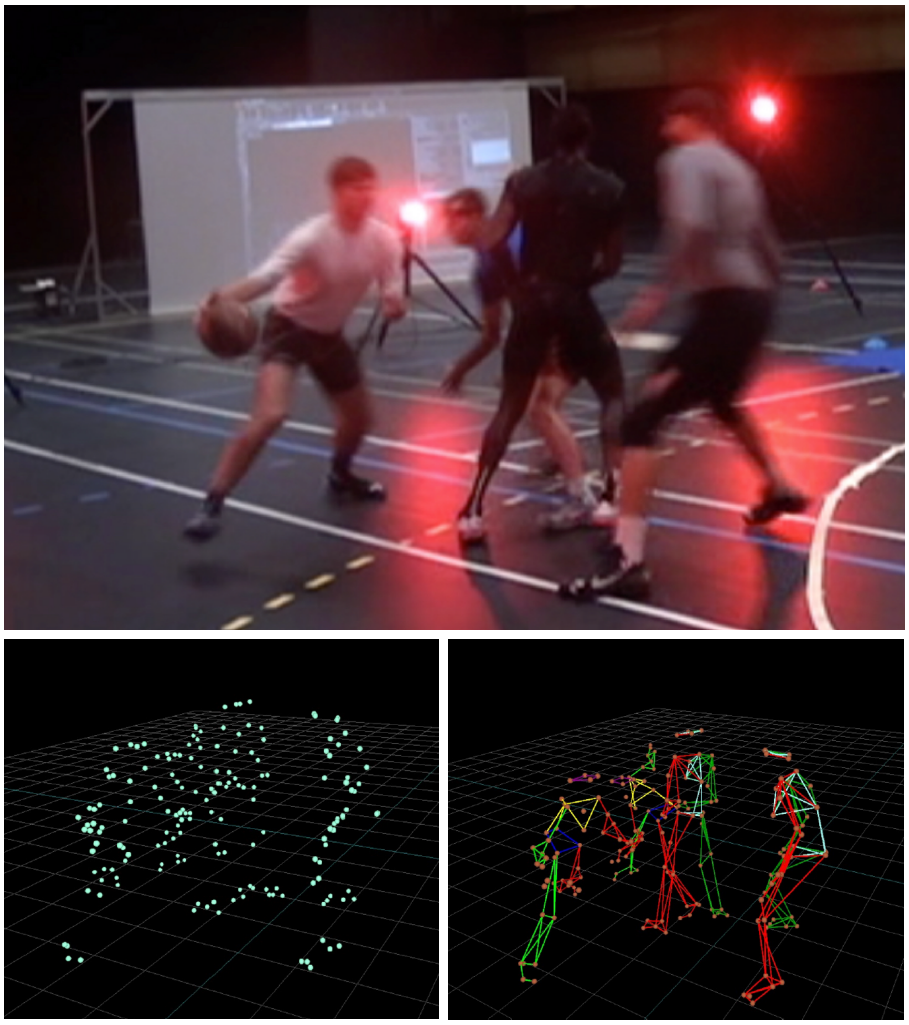


FIGURE 1.5 – De l'expérimentation à la reconstruction de mouvements en $3D$.

1.4 Contexte

Améliorer un mouvement sportif est complexe pour plusieurs raisons. D'abord, le geste réalisé est très souvent rapide. Il est difficile pour un entraîneur à l'œil nu de localiser les erreurs d'exécution en un très court laps de temps. Cette difficulté est encore plus accentuée par la multidimensionnalité du mouvement global. Considérons par exemple un franchissement de haie en athlétisme pour un athlète de jambe d'attaque gauche, illustré en figure 1.6. Non seulement ce geste doit être très rapide, mais il combine des mouvements complexes et subtiles des membres. Le centre de masse du corps doit être maintenu le plus bas possible lors du franchissement pour conserver une vitesse maximale, la jambe gauche se tend à l'horizontale au dessus de la haie pour l'attaquer avec la plante de pied. Simultanément, le bras droit aide au maintient de l'équilibre en poussant le torse en avant pour ne pas perdre de vitesse. Le bras droit fait contrepoids à l'arrière, et la jambe droite dite d'esquive est ramenée rapidement sous le bras de manière à franchir la haie. Elle doit permettre une amorce rapide et énergique de la foulée suivante. Ce mouvement combine donc des positionnements complexes des membres. De plus, le point de vue considéré risque de masquer une partie du mouvement, à l'œil nu comme au ralenti après enregistrement vidéo. La combinaison subtile des membres rend donc l'analyse délicate, et le conseil d'autant plus, notamment parce que le geste dit "parfait" n'existe pas.

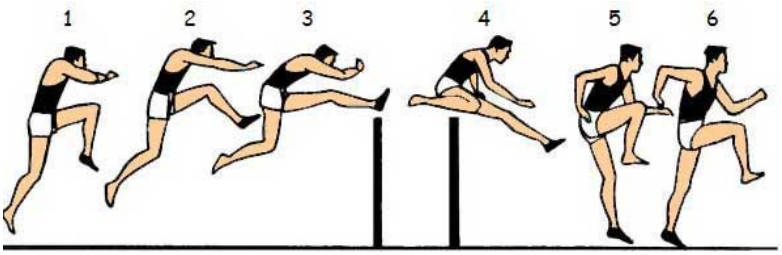


FIGURE 1.6 – Franchissement de haie d'un athlète de jambe d'attaque gauche : image de synthèse issue de [1].

À l'image du personnage inventé par Jacques Tati en 1953 dans le film *Les Vacances de Monsieur Hulot*, dont le personnage principal invente un geste de service de tennis *a priori* très grossier mais pourtant très performant, le geste "meilleur que les autres" n'est pas figé. Néanmoins, au cours de cette thèse, on supposera qu'il existe des mouvements menant plus favorablement à une bonne performance que d'autres, mais qu'au sein de ces mouvements il réside tout de même une variabilité intrinsèque admissible. C'est sur ce postulat que se base le métier d'entraîneur sportif. On gardera en tête la notion de style qui peut varier d'une personne à l'autre, comme Dick Fusbury a pu nous le démontrer en 1968 en remportant la compétition de saut en hauteur lors des Jeux Olympiques de Mexico, en sautant avec une technique de rouleau dorsal alors que l'unique technique répandue jusqu'alors était le rouleau ventral.

Le rendu de l'entraîneur virtuel doit être facile d'accès et approprié à la personne

souhaitant perfectionner son geste. La difficulté de cet enjeu réside notamment dans la généricité du procédé, qui le contraint à une indépendance face au mouvement exécuté. Afin d'être pertinent quant au retour d'informations donné au sportif, une modélisation par apprentissage du geste sera effectuée à partir d'une base de données de gestes experts. Il est important de noter qu'aucune information relative au mouvement n'est nécessaire pour la mise en place de l'outil d'évaluation qui pourra donc être utilisé pour tous types de gestes. L'automatisation de la caractérisation du mouvement confère au processus une grande accessibilité : aucune connaissance préalable sur l'action à réaliser n'est requise. Il conviendra également de gérer les différentes morphologies possibles du sujet, ainsi que ses différentes vitesses d'exécution.

1.5 Contributions

Pour répondre à la problématique posée, nous pouvons résumer les apports de cette thèse en trois points principaux :

1. **La modélisation d'un geste à partir d'un jeu de mouvement d'experts**. Pour ce faire, des outils usuels vont être adaptés et généralisés au cas de mouvements humains. C'est le cas du *Dynamic Time Warping* (DTW), méthode d'alignement non linéaire de séries temporelles qui a donné lieu à des techniques de moyennage d'un jeu de séries temporelles telles que le *DTW Barycenter Averaging* (DBA). Ces outils verront leurs contributions accentuées par l'ajout d'un terme de dispersion (la tolérance) en plus du simple moyennage. Certaines de leurs limites seront mises en évidence et une solution sera apportée par l'utilisation de contraintes locales. Ces outils seront ensuite généralisés au cas multidimensionnel afin d'être appliqués à notre cas d'étude de gestes. Pour chaque nouvel apport, une validation sera faite sur une base de données de la littérature afin de conforter les choix adoptés.

2. **L'estimation de la qualité d'un geste novice**. La mesure de qualité du geste sportif va être conduite par comparaison avec le geste expert modélisé précédemment, que l'on appellera "geste nominal". Nous développerons deux niveaux d'alignement par DTW afin de distinguer les recalages locaux et globaux des mouvements experts et novices. Par ce biais, nous déduirons une erreur spatiale et une erreur temporelle, qui à elles deux donneront une estimation de la qualité du geste. L'approche adoptée répondra aux besoins d'indépendance à la morphologie du sujet, à son positionnement dans la zone de capture et à la vitesse d'exécution de son geste. Les résultats obtenus sur des bases de données de service de tennis et de coup de poing de karaté seront confrontés à des annotations d'entaîneurs afin d'être validés.

3. **La mise en place d'un outil d'entraînement temps-réel d'un geste sportif**. Dans un dernier temps, l'ensemble des outils développés contribueront à l'élaboration d'un outil d'entraînement adapté à l'athlète qui souhaite améliorer son geste.

Le rendu sera intelligible, adapté au besoin de l'utilisateur et le plus informatif possible.

À ce jour, ces travaux ont donné lieu à la soumission de trois articles dans des journaux, deux dans des conférences internationales associées à des communications orales, et deux présentations au GDR ISIS :

- M.Morel, C.Achard, R.Kulpa and S.Dubuisson (Soumis en janvier 2017). Time-series Averaging Using Constrained Dynamic Time Warping with Tolerance. In *Pattern Recognition* (Major Revision).
- M.Morel, C.Achard, R.Kulpa and S.Dubuisson. Automatic and Generic Assessment of the Quality of Sport Motions. In *Doctoral Consortium de la conférence Automatic Face and Gesture Recognition (FG)* se déroulant du 30 mai au 3 juin 2017 à Washington.
- M.Morel, C.Achard, R.Kulpa and S.Dubuisson. Automatic Evaluation of Sports Motion : A Generic Computation of Spatial and Temporal Errors. In *Image and Vision Computing*, 2017 (en cours de publication).
- M.Morel, R.Kulpa, A.Sorel, C.Achard and S.Dubuisson (2016). Automatic and Generic Evaluation of Spatial and Temporal Errors in Sport Motions. In *International Conference on Computer Vision Theory and Application*, pages 1-12.
- M.Morel, B.Bideau, J.Lardy and R.Kulpa (2015). Advantages and Limitations of Virtual Reality for Balance Assessment and Rehabilitation. In *Clinical Neurophysiology*, 45 (4-5), pages 315-326.

1.6 Plan du document

Dans un premier temps, le chapitre 2 passera en revue les différents travaux de la littérature relatifs à notre problématique.

Dans un second temps, nous proposons dans le chapitre 3 de nous intéresser à la modélisation de séries temporelles à l'aide d'un algorithme de *Dynamic Time Warping* (DTW). D'abord appliqué à des signaux simples à une dimension afin de valider notre méthode, ce procédé nous permettra à terme de modéliser le mouvement expert à partir de bases de données de gestes. Après une revue de la littérature sur l'alignement et la modélisation de séries temporelles, nous proposerons différentes alternatives aux méthodes actuelles permettant une prise en charge plus adaptée et plus fidèle aux séries temporelles, à des fins de modélisation. Une validation sera faite à la fois sur des signaux 1D mais aussi *via* une classification de gestes.

Ensuite, dans le chapitre 4, nous présenterons les outils développés pour estimer la qualité d'un geste sportif. Pour ce faire, nous expliciterons tout d'abord un geste expert moyen dit *nominal* ainsi qu'une tolérance associée. Nous pourrons alors déduire les erreurs spatiales et temporelles associées à un autre mouvement au cours du temps. Ces résultats seront confrontés à des annotations expertes pour permettre leur validation.

Enfin, dans une cinquième et dernière partie, nous mettrons en place le système d'évaluation à partir des outils créés jusqu'alors.

Une conclusion générale permettra de synthétiser l'ensemble des travaux et d'établir quelques perspectives qui en découlent.

Chapitre 2

Etat de l'Art

Introduction

Dans ce chapitre, nous allons dresser un état de l'art des différentes méthodes s'approchant de près ou de loin à la problématique d'évaluation de gestes sportifs que nous nous sommes fixée. Ainsi, nous établirons le cadre scientifique de cette thèse dans son ensemble. Plus loin dans le manuscrit, nous reviendrons sur des points de la littérature plus spécifiquement liés aux contributions que nous apportons.

2.1 Les grands domaines de l'analyse de gestes

L'analyse du mouvement s'est énormément développée ces dernières années, notamment grâce à l'essor de technologies de captation et de restitution du mouvement. En plus d'une ambition ludique, ce domaine de recherche vise également à simplifier et sécuriser nos actions quotidiennes. Les travaux de détection de mouvement et de reconnaissance de gestes se multiplient et s'appliquent aux jeux vidéo, à la surveillance, à la domotique ou encore au sport.

L'analyse de gestes est en réalité un domaine vaste qui renferme beaucoup de problématiques bien distinctes.

2.1.1 Reconnaissance et analyse du geste

Le premier domaine, probablement le plus répandu, est la reconnaissance de gestes [11, 15, 16, 17, 18]. Dans ce contexte, il s'agit de gommer suffisamment les subtilités des mouvements afin de les résumer à leur essence sans pour autant perdre les saillances. De fait, l'espace est parfois partitionné, les trajectoires résumées ou simplifiées, l'utilisation d'un espace abstrait est largement employé. Reconnaître l'action d'un humain à un moment donné peut trouver des raisons multiples. Dans le domaine militaire par exemple, Dupont *et al.* [19] cherchent à automatiser la reconnaissance d'un certain nombre de signes de la main afin d'entraîner un robot mobile à agir en conséquence. Dans le domaine industriel, détecter le mouvement d'un humain peut permettre une sécurisation

et une meilleure efficacité d'un procédé collaboratif humain-robot comme le montrent Coupeté *et al.* [20]. De façon plus générale, n'importe quelle interaction homme-machine nécessite de savoir quel geste est réalisé par l'humain à tout instant. En domotique, la connaissance d'une action peut permettre le contrôle automatique de différents appareils quotidiens comme le récapitule l'article [21]. D'autres applications de la reconnaissance d'actions peuvent être recensées, telles que la télésurveillance [22], la reconnaissance de la langue des signes [23], la détection de bagarre [24] ou l'aide à la personne [25].

Dans un cadre d'analyse de geste, leur étude permet par exemple de détecter des mouvements complexes tels qu'une feinte au rugby [26]. Dans ce cas particulier, les résultats obtenus mettent en évidence un enchaînement biomécanique complexe qui permettrait à un sportif, après un entraînement approprié, de détecter plus rapidement des feintes au rugby et d'y répondre le plus efficacement possible. Des résultats similaires existent en handball notamment [27, 28]. En sport également, certains chercheurs se focalisent sur la reconnaissance d'une action collective à partir du mouvement de chacun des joueurs sur un terrain de sport, comme en hockey sur gazon par exemple [29]. Dans le domaine médical, analyser un mouvement peut permettre d'objectiver une pathologie [30, 31].

2.1.2 Synthèse de gestes

Une seconde application est la synthèse de gestes. Il s'agit de créer de nouveaux mouvements à partir de mouvements existants. De nombreux travaux ont été menés sur ce sujet, notamment dans l'industrie du jeu vidéo ou en réalité virtuelle. Selon les études, on pourra par exemple synchroniser deux mouvements afin de les fusionner [32, 33], identifier les descripteurs de style du mouvement dans le domaine fréquentiel afin de générer des nouveaux mouvements de styles variables [34, 35], ou encore gérer l'effet d'un environnement variable sur un humanoïde [36]. Cassel *et al.* [37], dans un registre un peu différent mais toujours dans l'objectif de générer un geste, s'appuient sur des règles sémantiques simples du dialogue entre deux humains virtuels (intonation, expression faciale, mouvements de mains et de la tête) pour générer des gestes symboliques appropriés au contexte.

2.1.3 Segmentation de gestes

La segmentation de geste [38, 39, 40, 16] consiste quant à elle à subdiviser une séquence de mouvement en des gestes élémentaires. Ce procédé est souvent rendu complexe lorsque les séquences considérées ne présentent pas de pauses significatives permettant de les subdiviser. La segmentation est souvent un pré-traitement utile à la reconnaissance ou à la synthèse. Certaines méthodes [15] permettent de réaliser en même temps la segmentation et la reconnaissance.

2.1.4 Évaluation de gestes

Plusieurs travaux ont déjà tenté d'évaluer un geste, qu'il soit sportif [17, 41, 42, 43] ou chirurgical [44, 45]. Il existe aussi quelques travaux à application plus artistique, qui

évaluent la performance d'un potier [46] ou analysent le mouvement d'un violoniste au cours de son geste [47].

Nous reviendrons sur ces travaux d'évaluation dans le chapitre 4. Dans ces applications dont le contexte peut paraître proche de celui de la reconnaissance, l'élément distinctif est que la sémantique des gestes à évaluer est connue : il ne s'agira donc pas de simplifier le geste, mais de conserver toutes ses spécificités afin d'être capable de mesurer sa qualité à tout instant et d'en donner un retour pertinent (non abstrait). Ce retour se veut adapté et compréhensible pour l'athlète, l'entraîneur ou l'apprenti chirurgien auquel il est destiné. Dans le cadre de cette thèse, nous proposons de réaliser un entraîneur virtuel et donc, de s'intéresser au geste sportif.

2.2 Représentation du geste

Pour analyser un geste, il convient tout d'abord d'en extraire une information à traiter, c'est-à-dire un *descripteur du mouvement*.

2.2.1 Descripteurs pour résumer un geste

Ce type de descripteur forme un vecteur qui renseigne sur l'état du système. Selon l'enjeu et la problématique fixés, le codage est soit temporel (le signal est codé à chaque instant, ce qui amène à une chaîne temporelle), soit global (tout le geste est codé par un vecteur de caractéristiques).

Le choix du descripteur est primordial et dépend du traitement qui en est fait. Nous allons distinguer trois grandes familles de descripteurs comme le récapitule la figure 2.1 :

- les descripteurs reposant sur un modèle du corps humain ;
- les descripteurs holistiques, qui utilisent la dynamique globale de l'objet (quelconque) en mouvement ;
- les descripteurs locaux, caractérisant les mouvements uniquement à partir de points d'intérêt isolés.

Pour chacune de ces familles, le codage peut être soit temporel, soit global. Dressons un bilan de ces différentes approches.

2.2.1.1 Les descripteurs reposant sur un modèle du corps humain

D'une façon générale, cette catégorie se fonde sur le résultat des études psychophysiques de *Point Light Display* initiées par Johansson en 1973 [48]. Celui-ci atteste que des points lumineux positionnés sur les articulations du corps humain en mouvement vus de profil suffisent au cerveau humain pour reconnaître un geste. À partir de ce postulat, de nombreux travaux exploitent l'évolution de la trajectoire des articulations du corps humain.

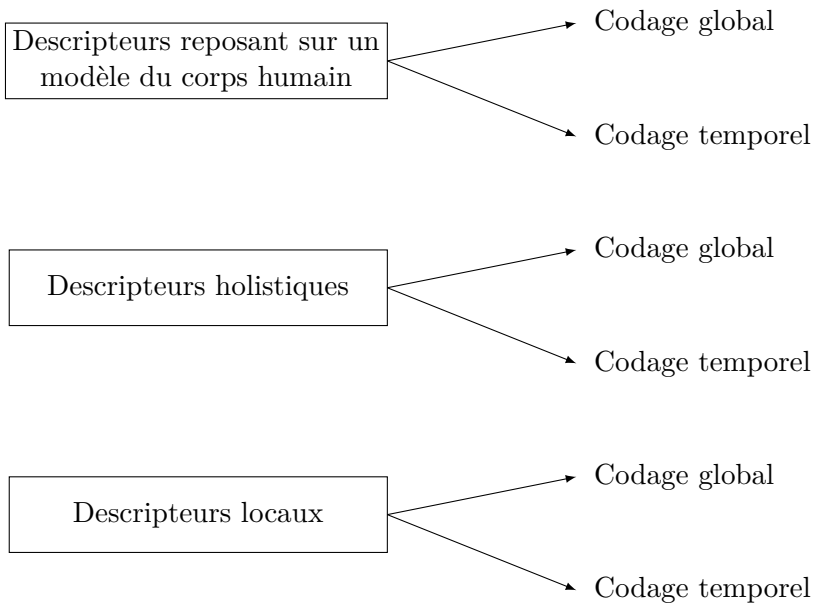


FIGURE 2.1 – Les différents types de codages du geste

Les articulations sont alors extraites d'une image $2D$, d'une carte de profondeur (issue d'une Kinect® par exemple), ou directement d'un système de capture $3D$. Selon le dispositif de capture choisi, la procédure est plus ou moins complexe et le positionnement des articulations plus ou moins précis.

Parmi les descripteurs reposant sur le corps humain, les descripteurs cinématiques et dynamiques sont dits de bas niveau, utilisant la trajectoire, la vitesse et l'accélération cartésiennes ou angulaires des différentes articulations formant le mouvement du corps humain. Ils sont rapidement très volumineux dès lors que beaucoup d'articulations sont prises en compte. Plusieurs travaux synthétisent ces données tridimensionnelles en considérant la courbure plutôt que la trajectoire [39, 49, 44]. Dans un cadre de codage local, la courbure d'une articulation est alors donnée par :

$$\gamma(t) = \frac{\|r'(t) \wedge r''(t)\|}{\|r'(t)\|^3}$$

où $r(t)$ décrit la position de l'articulation considérée et $r'(t)$ et $r''(t)$ respectivement ses dérivées première et seconde.

D'autres travaux considèrent la trajectoire formée par chaque articulation dans l'espace. Ils voient alors le geste comme une extension à la $3D$ d'une écriture manuscrite et appliquent à la reconnaissance de geste un jeu de descripteurs $2D$ cinématiques dédié à la reconnaissance d'écriture [50] (appelé `HBF49`), qu'ils étendent dans l'espace [51].

D'autres travaux considèrent le positionnement relatif des articulations ou des membres. C'est le cas par exemple des travaux décrits dans [52], où est créé à chaque instant un descripteur contenant les positionnements relatifs des articulations. Cette approche lo-

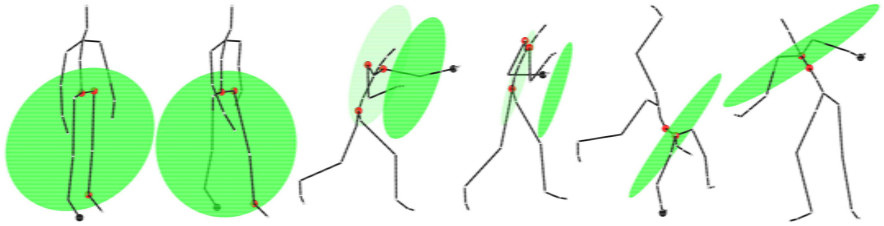


FIGURE 2.2 – Figure extraite de [2]. Les trois articulations en rouge forment un plan symbolisé par un disque vert, le positionnement de l'articulation bleue est évalué relativement à ce plan donnant un résultat binaire. L'association de plusieurs évaluations de positionnement permet d'extraire un vecteur de variables binaires simple et invariant à la morphologie du sujet.

cale est intéressante mais également très lourde avec un jeu de descripteurs d'autant plus important qu'il y a d'instants et d'articulations. Ici, l'auteur applique sa méthode à de la reconnaissance d'actions capturées par Kinect® (donc à relativement basse fréquence et contenant un jeu d'articulations assez restreint). Il applique également une analyse en composantes principales pour réduire la taille du descripteur.

D'autres méthodes discrétisent l'espace et construisent des histogrammes de position d'articulations en cumulant toutes les articulations au cours du temps [53]. Par cette méthode de comptage, le codage devient global et non plus temporel. Cette idée de partionnement de l'espace a également été adoptée par Xia *et al.* en 2012 [54]. Elle est très adaptée à la reconnaissance de geste, puisqu'elle permet de résumer et d'homogénéiser l'information. En revanche, ce qui est bénéfique à la reconnaissance ne l'est plus dans un cadre de l'évaluation puisque ce sont les spécificités qui permettent de discriminer les erreurs d'exécution d'un geste.

D'autres auteurs créent des relations plus globales entre une articulation et un plan approprié formé de trois autres articulations [2, 55, 56], comme l'illustre la figure 2.2. Dans le cas d'un mouvement de marche, périodique, symétrique et très régulier, cette approche est très pertinente. Le descripteur est relativement simple à mettre en place et invariant à la morphologie du sujet. De façon générale, cette méthode convient tout à fait à de la reconnaissance d'action simple, mais souffre elle aussi d'une perte d'information dès lors qu'il s'agit d'évaluer la qualité d'un geste. De plus, si le mouvement est complexe, comme c'est le cas généralement d'un geste sportif, le simple positionnement relatif point-plan risque de varier selon le style d'exécution du mouvement employé. Enfin, les plans à considérer risquent de varier d'un geste sportif à l'autre, faisant perdre toute généricité au modèle.

Des descripteurs géométriques ont également été développés. Il s'agit par exemple de considérer une aire, une boîte englobante, une ellipsoïde englobante du squelette... Une étude considère l'aire formée par le pentagone des effecteurs terminaux (tête, pieds, mains) du corps [3] comme le montre la figure 2.3. Un autre article met en place un contourage itératif du mouvement total à la manière des contours actifs [13]. Une énergie

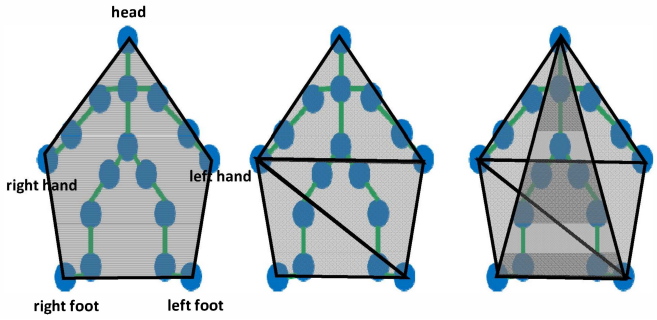


FIGURE 2.3 – Figure extraite de [3]. Les effecteurs terminaux (tête, pieds, mains) sont utilisés pour former des polygones représentatifs de la pose du squelette.

est alors associée au mouvement, dépendant des longueurs des segments et des angles entre les segments composant le contour. On ne considère tout d'abord que les sommets de début et de fin du mouvement, puis on ajoute un sommet minimisant une certaine erreur totale. De cette manière, on extrait d'autant plus de points que la courbure du sommet est élevée, ce qui parait très pertinent.

Enfin, les descripteurs d'effort reposent sur des calculs de plus haut niveau d'énergie [52]. C'est notamment le cas dans les travaux de Komura *et al.* [57] qui évaluent un terme de mouvement d'un coup de poing. Le codage est alors global puisque le descripteur est extrait au terme du mouvement et résume complètement le geste. Onuma *et al.* [58] évaluent quant à eux l'énergie cinétique de chaque articulation, qu'ils utilisent ensuite pour mettre en place un descripteur reposant sur les énergies du haut et du bas du corps.

Encore une fois, ces approches géométriques et énergétiques ne sont pas adaptées à l'évaluation ; d'abord parce qu'elles perdent l'information permettant de discriminer un bon d'un mauvais mouvement, mais aussi parce qu'en se basant sur des données plus haut niveau que les simples positionnements et vitesses des membres, elles ne permettent pas un retour d'information simple *a posteriori*.

2.2.1.2 Les descripteurs holistiques

Les approches holistiques utilisent la dynamique du mouvement. Elles sont plus simples puisqu'elles considèrent l'objet en mouvement dans sa globalité, c'est-à-dire sans *a priori* sur l'objet en mouvement. Elles utilisent, soit la zone en mouvement, soit le flux optique ou bien le gradient des images.

C'est le cas par exemple dans les travaux de Bobick et Davis [4] qui introduisent la notion de motifs spatio-temporels (*Spatio-Temporal Template*). Les auteurs utilisent les différences seuillées entre les images au cours du temps pour créer une image "énergie du mouvement" qui correspond à l'union des images de différence seuillées (notée MEI). En définitive, l'image créée résume le mouvement par un codage global, mais perd la temporalité de celui-ci. De fait l'image "historique du mouvement" (MHI) est également introduite. Afin de conserver l'ordonnancement temporel du geste, les images de diffé-

rence seuillées sont pondérées selon leur indice temporel. Deux exemples d'images MEI et MHI résultant d'un mouvement de bras sont données en figure 2.4.

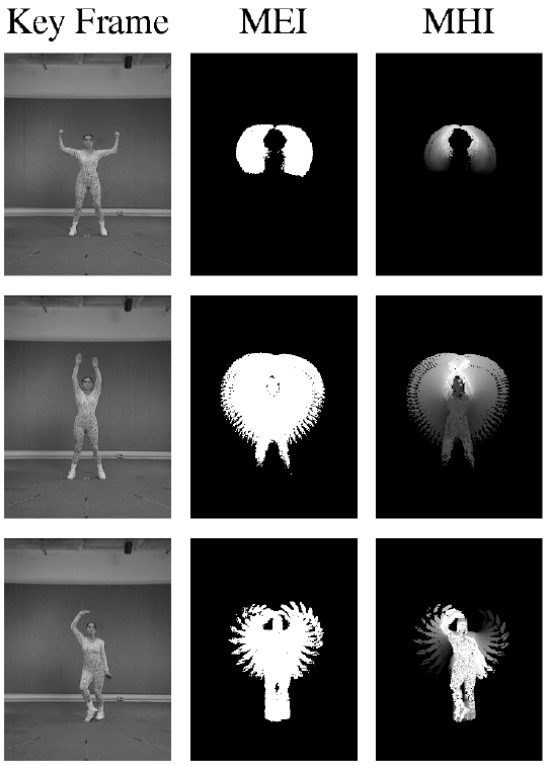


FIGURE 2.4 – Figure extraite de [4]. Énergies du mouvement (MEI) et Historiques du mouvement (MHI) résultant d'un mouvement de bras. Alors que les MEI sont relativement similaires pour les deux derniers mouvements, les MHI, conservant l'information temporelle, permettent de les distinguer.

Un inconvénient majeur de cette méthode intervient lorsque le mouvement s'auto-occulte, c'est-à-dire qu'une posture se superpose à une autre qui la précède dans le mouvement. De plus, son utilisation est plus efficace si l'on considère des gestes pré-segmentés.

Pour pallier à cette limitation, une alternative consiste à conserver l'information temporelle en créant un volume spatio-temporel plutôt qu'une image $2D$ [59]. Il a également été testé de généraliser le MHI au cas d'une captation $3D$ du mouvement avec des Kinects®. Le volume d'historique du mouvement (MHV) est ainsi introduit dans [60].

Non plus reposant sur l'extraction d'une silhouette mais sur l'utilisation du flot (ou flux) optique (*i.e.* le champ des vitesses mesuré à partir des variations de la luminance), des méthodes similaires peuvent être utilisées [61, 62]. En 2003, Efros *et al.* proposent par exemple un codage local par quatre descripteurs contenant les composantes horizontales et verticales du flot optique en x positif, x négatif, y positif et y négatif pour un instant donné [63].

On suppose alors qu'une différence d'image correspond à un mouvement. Dès lors, les moindres changements d'éclairage ou modification de l'arrière-plan (par exemple, des arbres qui bougeraient avec le vent) doivent être évités.

2.2.1.3 Les descripteurs locaux

Les approches dites locales se focalisent sur des points d'intérêts présentant une singularité. Les premières approches sont motivées par les très bons résultats de reconnaissance d'objets par points d'intérêt. Dès lors, des techniques de détection de points d'intérêt $2D$ sont étendues au cas spatio-temporel, comme les descripteurs de Harris par exemple [64] (on parle alors de *Harris* $3D$). Pour caractériser les points d'intérêt obtenus, certains auteurs définissent une zone spatio-temporelle autour du point considéré dont ils extraient un descripteur, par exemple un histogramme d'orientation du gradient ou de flux optique [65].

Scovanner *et al.* extrapolent quant à eux l'utilisation des SIFT (*Scale Invariant Feature Transform*) à la $3D$ [66] pour la reconnaissance d'action.

Une revue de la littérature concernant l'ensemble des descripteurs du mouvement a été écrite par Larboulette *et al.* [67].

En définitive, nous avons vu que l'étude du mouvement humain pouvait reposer simplement sur les positionnements des articulations du mouvement. Selon la manière dont l'extraction des données est réalisée, cette méthode peine parfois à suivre le squelette du sujet, notamment à cause d'éventuelles occultations ou lorsque plusieurs personnes sont présentes dans la scène. Des alternatives existent, par exemple les approches holistiques qui ne nécessitent pas de reconnaître un corps humain dans la scène. Les méthodes utilisées s'en trouvent parfois plus efficaces, mais souffrent quant à elles d'éventuels mouvements de caméra ou de changement de luminosité qui nuisent à la reconnaissance du geste. Une dernière alternative existe, celle d'utiliser des points d'intérêt.

Dans le cadre de cette thèse, les mouvements seront capturés en $3D$ par un système de capture optique précis. Le positionnement des articulations sera donc considéré comme très fiable, c'est pourquoi les descripteurs utilisés reposeront sur ce positionnement. Néanmoins, une fois l'ensemble des articulations considérées, il convient de déterminer lesquelles sont pertinentes.

2.2.2 Discrimination de l'informativité de différents descripteurs

Analyser un geste nécessite souvent de se focaliser sur une partie du corps : celle réalisant le geste. Ainsi, l'évaluation des gestes chirurgicaux se focalisera sur les mains, tandis que, par exemple, l'étude du tir au football sera plus focalisée sur les pieds. Au delà même des membres pris en compte, certaines données, *a priori*, peuvent grandement aider à l'évaluation. C'est sur cela que reposent les travaux décrits dans [12] qui consistent à mettre en place un outil d'entraînement au karaté. En utilisant des descripteurs cinématiques inhérents aux mouvements considérés - et fournis par des entraîneurs de karaté -,

Burns *et al.* analysent les mouvements de plusieurs novices et leur fournissent un outil d'entraînement interactif en réalité virtuelle permettant une progression des sujets.

Il convient de noter que l'outil obtenu n'est en aucun cas générique aux différents sports individuels, mais restreint à un mouvement particulier connu. C'est également en ajoutant des informations relatives au sport que Komura *et al.* [57] proposent un outil d'entraînement aux arts martiaux. Leur outil se focalise sur la minimisation énergétique du mouvement total du défenseur lors d'une attaque, sa prévisibilité d'attaque et sa vitesse de coup de poing, puisqu'on sait ces données être particulièrement liées à la bonne performance du sport considéré. Utilisant le même principe d'ajout de connaissance inhérente au geste, Ward analyse dans sa thèse des gestes de danse de ballet [42]. À nouveau, les paramètres considérés sont très précis et localisés sur les informations les plus pertinentes relativement au mouvement : l'extension du genou, la rotation de la hanche, le décalage thoracique antéropostérieur, *etc.*

Au lieu d'ajouter une connaissance *a priori* sur des indicateurs de performance du mouvement (qu'on ne connaît pas forcément sans l'aide d'un entraîneur), il apparaît judicieux d'effectuer un apprentissage de ces indicateurs. De fait, certains auteurs ajoutent plutôt à leur analyse un prétraitement qui consiste à modéliser le mouvement pour en extraire les articulations les plus pertinentes à analyser. C'est le cas de Ofli *et al.* [68] et de Pazhmouand *et al.* [69] qui discriminent les articulations les plus "informatives" d'une action à partir de leurs variabilités spatiales moyennes sur un sous-segment composant l'action, puis appliquent leurs résultats à une reconnaissance d'action. Finalement, ils choisissent de définir comme articulations informatives les articulations les plus mobiles. Les articulations informatives sont-elles les plus mobiles ? Si l'on considère un mouvement de coup de poing, il semblerait que oui puisque c'est le bras qui renferme l'information. Après avoir discuté avec des entraîneurs de karaté, il apparaît néanmoins que certains gestes requièrent un bon gainage, une stabilité d'un membre qui est au cœur même de l'exécution. La mobilité n'est donc pas toujours la donnée la plus pertinente.

Une autre approche permettant de réduire la dimensionnalité pour ne conserver que les éléments discriminants est reprise par certains auteurs lorsqu'ils appliquent une Analyse en Composantes Principales (ACP) sur les gestes. Raptis *et al.* et Ramakrishnan *et al.* [41, 38] représentent les mouvements obtenus *via* une Kinect® et un système de capture de mouvement optoélectronique (respectivement) de manière angulaire et appliquent une ACP sur ces données pour ne conserver que les premières composantes. De façon similaire, Jiang *et al.* [70] effectuent une décomposition en valeurs singulières de plusieurs mouvements similaires, y concatènent le mouvement à reconnaître et estiment à quel point la décomposition s'en trouve modifiée. Dans un cadre différent de récupération de mouvement (*motion retrieval*), Chao *et al.* [71] représentent également des actions par un ensemble restreint d'harmoniques sphériques orthonormales afin de résumer l'information. Enfin, Barbic *et al.* [40] formulent que la transition d'un geste correspond à l'augmentation rapide de l'erreur de projection d'une décomposition en valeurs singulières sur un nombre restreint de vecteurs propres.
Une autre alternative consiste à gérer la grande dimension en utilisant un espace mul-

tidimensionnel : c'est tout l'intérêt d'utiliser la variété Riemannienne [18, 72, 73]. Typiquement, Devanne *et al.* [73] concatènent les coordonnées tridimensionnelles de l'ensemble des articulations du sujet dans un vecteur et considèrent la trajectoire que décrit ce vecteur dans l'espace multidimensionnel modélisant la dynamique de l'ensemble du mouvement. Les trajectoires sont alors interprétées comme une variété Riemannienne et comparer deux mouvements revient alors à comparer deux formes dans l'espace des formes engendré. Ces techniques sont très performantes lorsqu'il s'agit de quantifier la ressemblance de données. Elles seraient cependant inefficaces dans un contexte d'évaluation puisque toute la description de la dissimilarité des données est perdue dans la projection des données dans un nouvel espace propre physiquement abstrait.
D'autres travaux [16, 14] tentent de discriminer la saillance du mouvement par simplification des courbes le composant. C'est également une idée proposée par Boukir *et al.* [13] lorsqu'ils développent les contours actifs déjà évoqués.
Une alternative à ces approches serait de lier le caractère informatif d'une partie du corps, non pas à sa vitesse, mais à sa variabilité au sein d'une base de données de "bons" mouvements. En effet, en prenant en compte la variabilité d'une articulation parmi une base de données d'experts à un instant donné, on assure qu'un membre particulièrement stable est nécessaire au mouvement. En considérant par exemple un mouvement simple de coup de poing droit libre, on s'attend à mettre en évidence un bras droit très contraint par le mouvement - et donc avec une variabilité faible -, et un bras gauche très libre - de variabilité forte -, nous permettant de distinguer proprement un membre discriminant au mouvement d'un autre. C'est cette méthode de tolérancement que nous proposons de développer dans cette thèse.

2.3 Techniques d'apprentissage statistique

Une fois le geste codé, des outils statistiques doivent être mis en place afin de pouvoir les reconnaître ou les évaluer. Plusieurs outils peuvent être utilisés selon le codage considéré.

Comme cela a été évoqué précédemment, il existe deux types de codage : le codage global et le codage temporel. Dans le premier cas, l'apprentissage consiste à comparer des vecteurs caractérisant des gestes dans leur globalité. Dans le second cas, chaque geste est caractérisé par une chaîne temporelle de descripteurs. À partir d'un codage temporel, plusieurs approches peuvent être considérées : (i) la mise en place d'une mesure de similarité entre les chaînes temporelles, permettant ensuite une discrimination des données par plus proches voisins par exemple ; (ii) la mise en place d'un modèle markovien.

Dans le cadre plus simple d'un codage global, d'autres méthodes peuvent être mises en place telles que les réseaux de neurones, les forêts d'arbres décisionnels ou encore les SVM.

Dans les sections qui suivent, nous revenons rapidement sur le principe de ces différentes techniques.

2.3.1 Méthodes reposant sur un codage temporel

2.3.1.1 Recherche des plus proches voisins

La méthode de recherche du plus proche voisin (*NN* pour *Nearest Neighbor*) est très populaire en reconnaissance de mouvements [64, 74]. Chaque exemple est codé par un vecteur de dimension n et constitue un point dans cet espace. La classification d'un nouveau geste consiste à rechercher le geste le plus proche dans l'espace des descripteurs et lui affecter la classe correspondante. La principale difficulté de cette méthode réside dans le choix d'une métrique de similarité permettant de quantifier les écarts dans l'espace des descripteurs. Par exemple, Mokhber *et al.* [75] se servent de la distance de Mahalanobis pour estimer les écarts entre des moments géométriques $3D$, le tout afin de classifier des comportements humains.

La variante des K plus proches voisins ($K-NN$ pour $K-NearestNeighbor$) consiste à attribuer à un exemple inconnu la classe majoritaire de ses K plus proches voisins.

Nous dressons dans la partie suivante un état de l'art des différentes métriques de similarité entre séries temporelles utilisées.

2.3.1.2 Mesures de similarité

Pléthores de mesures de similarité (ou métriques) entre des séries temporelles ont été utilisées. Que ce soit dans un objectif d'indexation, de reconnaissance ou de classification, une mesure de similarité est toujours requise. Selon le contexte, une simple distance euclidienne (ou autre norme L_p) terme à terme peut être efficace [76], mais restreinte à la comparaison de séries temporelles de même taille et sans déformation temporelle. D'autres travaux ont donc développé des méthodes plus flexibles.

Une première amélioration consiste à considérer l'intercorrélation (ou corrélation croisée) entre les signaux, qui gère alors un décalage fixe entre les données comparées. Pour comparer deux signaux de décalage variable dans le temps, il faut utiliser des méthodes non linéaires, telles que la Déformation Temporelle Dynamique (ou DTW pour *Dynamic Time Warping* en anglais) qui permet un recalage et une mesure de similarité entre des signaux par déformation temporelle élastique [77], comme l'illustre la figure 2.5 qui la compare à la distance euclidienne.

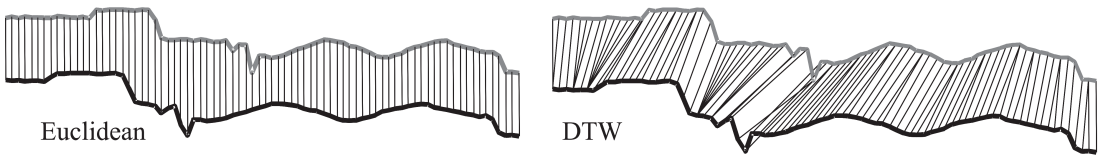


FIGURE 2.5 – Figure extraite de [5]. Comparaison de deux métriques de similarité entre deux séries temporelles. À gauche, la distance euclidienne mesure la distance instant après instant ; de fait, elle reflète mal la distance entre deux signaux de temporalités différentes. À droite, l'alignement par DTW permet un recalage non linéaire des deux signaux. Plus intuitive, elle permet d'apparier les événements similaires.

Le DTW trouve son origine dans l'analyse audio vers les années 1970 [77, 78]. C'est plus tard que son utilisation a été élargie à l'analyse de séries temporelles quelconques telles que la parole [79, 80], la musique [81], l'écriture [82] ou les signaux biologiques [83]. Nombres d'études ont tenté d'améliorer l'efficacité du DTW ces dernières années, dans de nombreux contextes [84, 85, 86, 87, 88].

Dans le contexte du geste, plusieurs auteurs se sont servis du DTW pour aligner et comparer des mouvements. Pham *et al.* [44] par exemple ont calculé les courbures d'un outil chirurgical puis les ont comparées à un modèle grâce au DTW. Le DTW a également été utilisé en obstétrique pour comparer le geste de positionnement de la lame du forceps chez un médecin expérimenté et chez un novice. Sakurai *et al.* [3] ont proposé une méthode permettant de retrouver un geste similaire à un geste donné parmi une base de données. Le système alors mis en place effectue d'abord une normalisation du nouveau mouvement capturé par une Kinect® et calcule l'aire engendrée par le squelette. Cette aire est utilisée comme descripteur du mouvement et un DTW est appliqué pour comparer chaque mouvement de la base avec celui étudié.

Dans tous ces différents travaux, le mouvement est résumé à une seule série temporelle, de sorte qu'aucune information relative à la coordination des membres ne peut être extraite. De fait, cet encodage ne permet pas une analyse fine du mouvement et de ses éventuelles erreurs.

Utilisant également le DTW ou les *Transported Square-Root Vector Fields* (TSRVFs), Veeraraghavan *et al.* [72] et Ben Amor *et al.* [18] utilisent l'espace des formes de Kendall et introduisent l'espace des fonctions de déformations temporelles qui modélisent et permettent d'apprendre la variabilité due aux différences de vitesses d'exécution. Ces travaux sont très pertinents pour la reconnaissance d'action mais ne permettent pas de gérer une asynchronie entre des membres.

Une des limitations de la distance par DTW, soulignée par plusieurs auteurs, est qu'elle ne gère par les éventuels *outliers* présents dans les signaux, au contraire par exemple de la distance de Levenshtein [89]. Cette distance, appliquée aux chaînes de caractères, consiste à compter le nombre d'opérations d'insertions, suppressions ou substitutions de caractères permettant de transformer une première chaîne en une autre. Reposant sur le paradigme de programmation dynamique permettant de déterminer la "sous-séquence commune la plus longue", la distance LCSS fournit également une distance élastique entre deux séries temporelles [90]. Dans ce cas, étant donné que les séries temporelles ne sont pas quantifiées comme les chaînes de caractères, les signaux doivent être adaptés pour pouvoir appliquer la distance de Levenshtein : un critère de correspondance entre deux points doit alors être introduit. Parmi les variantes de mesure de similarité notables reposant sur la distance de Levenshtein, on retrouve la distance EDR (*Edit Distance on Real Sequence*) [91] ou son analogue, la distance EPR [92]. De manière similaire à la distance LCSS, la distance EDR recherche le nombre minimal d'opérations élémentaires permettant de transformer une série temporelle en une autre, mais cette fois-ci on donne également des pénalités aux points non assignés, pénalités d'autant plus importantes que les intervalles de points non assignés sont grands. De façon générale, ces mesures de similarité sont très limitées par le choix d'un paramètre établissant la

condition de correspondance entre deux points. Cette approche binaire (correspondance ou non de deux points) est discutable dans un cadre de comparaison puisqu'elle donne à ce paramètre une importance considérable, et perd nécessairement de l'information par quantification de la distance entre points. Elle a en revanche le mérite d'être robuste au bruit.

À notre connaissance, nous sommes les seuls à avoir comparé les alignements globaux et locaux de deux personnes effectuant un même mouvement. Dans l'approche que nous avons proposée dans [93], un alignement global, *i.e.* reposant sur l'ensemble du squelette de la personne, est d'abord mis en place entre les deux mouvements, permettant d'ajuster temporellement les mouvements. Ensuite, des alignements locaux, c'est-à-dire ajustant les positionnements d'une articulation isolée des deux mouvements, permet d'obtenir de nouveaux alignements potentiellement différents du premier. Ces différents alignements permettent ensuite notamment d'extraire une erreur de synchronie entre les membres d'un sujet durant l'exécution de son mouvement.

2.3.1.3 Modèles Markoviens

Si les méthodes de reconnaissance à partir des plus proches voisins sont performantes, les temps de calcul augmentent rapidement quand le nombre d'exemples de la base de référence augmente. Elles sont donc difficilement implémentables en temps réel alors que la plupart des méthodes de reconnaissance de gestes exigent un fonctionnement "en ligne". D'autres méthodes plus complexes et faisant appel à de l'apprentissage ont alors vu le jour. Les modèles de Markov cachés (HMM pour *Hidden Markov Models*) développés par Lawrence R. Rabiner en 1989 [94] reposent sur l'étude du caractère séquentiel du mouvement en tenant compte de l'incertitude sur les observations du vecteur descripteur considéré. C'est la méthode la plus utilisée en reconnaissance de mouvement [11, 95, 96].

Plus précisément, les données sont modélisées en utilisant une chaîne de Markov à états non observables. Chaque état génère des valeurs observables distribuées selon une certaine distribution de probabilité, par exemple une gaussienne multivariée pour des valeurs continues comme l'illustre la figure 2.6.

Trois types de données paramètrent le modèle :

1. les probabilités initiales de chaque état, souvent notées π_i.

2. la probabilité de transition entre deux états i et j, notée $a_{i,j}$.

3. la probabilité d'une observation k étant dans un état i, notée $b_{i,k}$.

L'hypothèse de Markov présume qu'un état ne dépend que de l'état précédent (hypothèse de Markov de niveau 1) ou bien de n états précédents (hypothèse de Markov de niveau n). De plus, les observations ne sont supposées ne dépendre que de l'état courant.

Les paramètres du HMM peuvent être appris par l'algorithme de Baum-Welch [94], suite auquel la procédure forward-backward estime la probabilité qu'une observation provienne d'un HMM avec le jeu de paramètres connus ou appris. Les HMM et assimilés

sont très puissants pour résoudre des problématiques de reconnaissance de séries temporelles [97, 98, 99, 100, 101, 102]. Reconnaître une action revient alors à trouver la chaîne de Markov qui génère la séquence observée avec la plus grande probabilité.

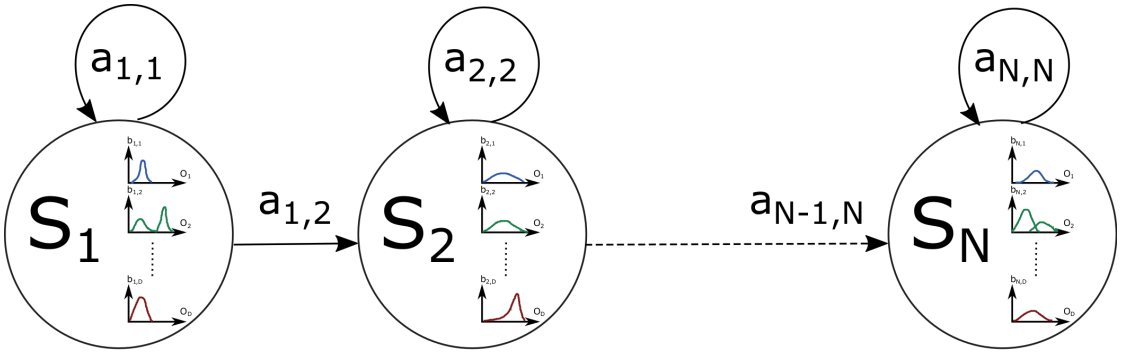


FIGURE 2.6 – Exemple de représentation graphique d'un HMM à N états. Pour chaque état S_i, les densités de probabilité $b_{i,k}$ de chaque descripteur o_k sont tracées. $a_{i,j}$ représente la probabilité de transiter d'un état S_i à un état S_j.

2.3.2 Méthodes reposant sur un codage global

2.3.2.1 Machines à vecteurs de support (SVM)

Les machines à vecteurs de support (*SVM* pour *Support Vector Machines*), proposées par Vapnik en 1995 [103], sont des classifieurs binaires qui reposent sur l'hypothèse d'une séparation linéaire entre les données. Elles consistent à rechercher l'hyperplan "optimal" qui maximise la distance minimale aux exemples d'apprentissage.

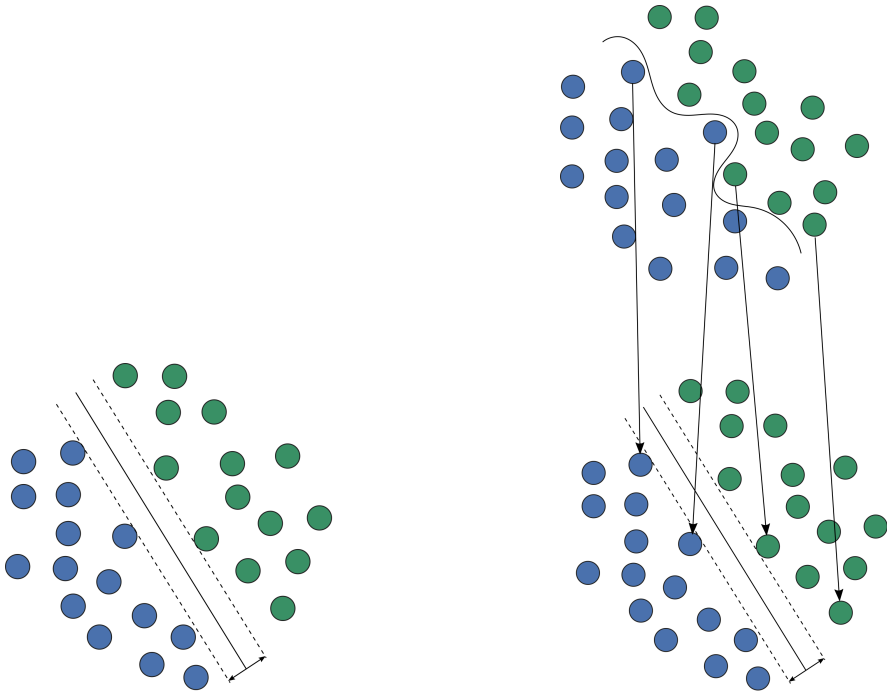
Les descripteurs n'étant pas toujours linéairement séparables, ils peuvent être projetés dans un nouvel espace, souvent de plus grande dimension, dans lequel ils redeviennent séparables. Comme seul le produit scalaire entre les descripteurs est requis dans les SVM, l'astuce de noyau (*kernel trick*) peut être utilisée, évitant ainsi une projection explicite des données.

La figure 2.7 illustre ces deux approches de SVM linéaire et non linéaire. Les SVM ont été utilisés pour la reconnaissance d'actions notamment [64, 104, 105].

2.3.2.2 Forêt d'arbres décisionnels

Les forêts d'arbres décisionnels ont été introduites par Breiman en 2001 [106]. C'est un algorithme utilisé en classification très performant. Comme son nom l'indique, la forêt est constituée de multiples arbres de décision entraînés sur des données différentes.

Dans les cas des forêts d'arbres aléatoires (*random forest*), ces arbres décisionnels sont dits aléatoires puisqu'ils sont construits à partir d'un tirage aléatoire d'échantillons d'apprentissage. Chaque arbre produit un vote par partitionnement récursif des données, et chaque vote contribue à la décision de classification finale. L'élaboration d'un arbre



(a) SVM linéaire (b) SVM non linéaire

FIGURE 2.7 – Représentation $2D$ des descripteurs, colorés en vert ou bleu selon la classe à laquelle ils appartiennent. L'hyperplan de séparation est indiqué par un trait plein noir, linéaire dans le premier cas et non linéaire dans le second. Plus particulièrement, la figure 2.7b illustre un cas non linéairement séparable dans lequel le *Kernel Trick* permet de séparer les descripteurs par un hyperplan dans l'espace transformé.

décisionnel aléatoire est illustré en figure 2.8. Les *Random Forests* ont notamment été utilisées pour la reconnaissance de la langue de signes [107] ou encore pour la détection et la reconnaissance de mouvements de la main [108].

2.3.2.3 Réseaux de neurones

Par analogie avec le fonctionnement et l'architecture biologique d'un neurone, on appelle neurone une structure mathématique avec une entrée, un traitement (ici, une pondération et fonction dite "d'activation"), et une sortie, comme l'illustre la figure 2.9a. Par extension, un réseau de neurones artificiels peut être représenté par un graphe de combinaisons de neurones.

La structure la plus classique d'un réseau de neurones est constituée de plusieurs couches de neurones qui communiquent. Cette structure est appelée perceptron multicouches et est représentée en figure 2.9b. Plus il y a de couches, plus il y a de niveaux d'abstraction. Pour un grand nombre de couches, on parle d'apprentissage profond (ou *deep learning*). En classification, la couche d'entrée contient les différents descripteurs

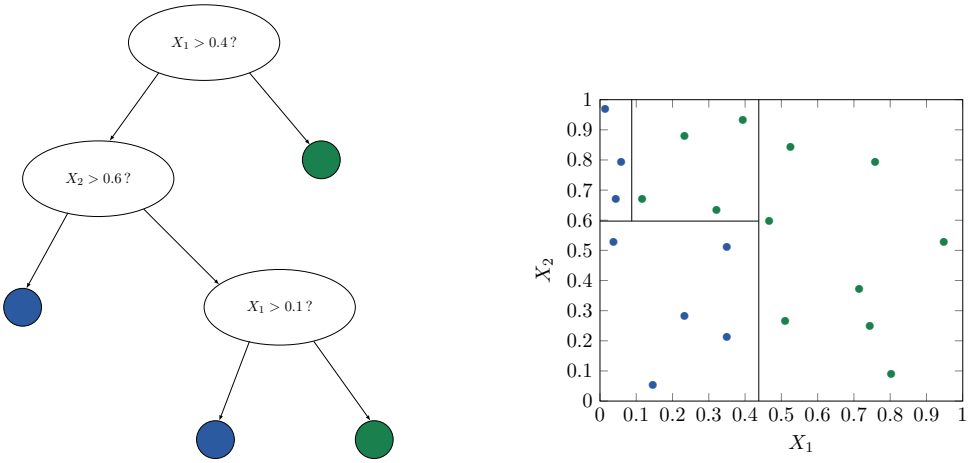
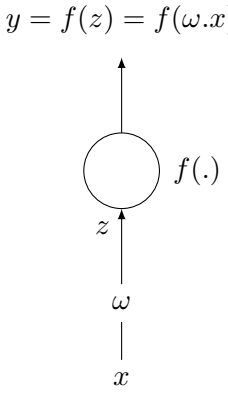
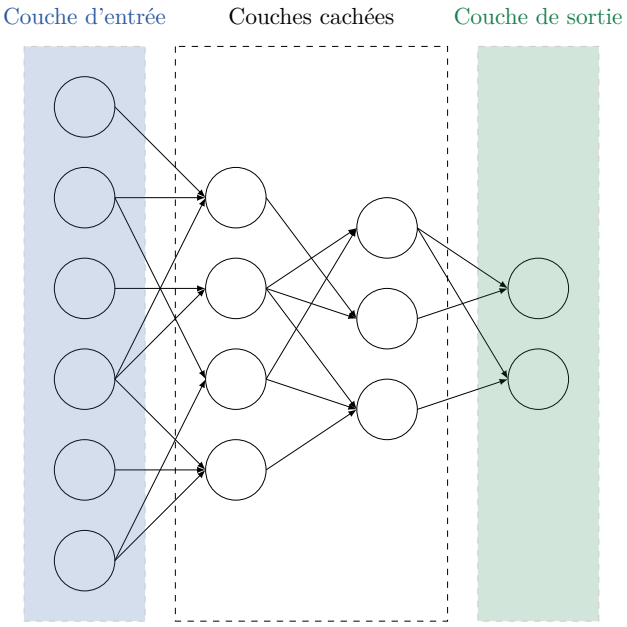


FIGURE 2.8 – Illustration d'un arbre décisionnel aléatoire. A gauche, l'extrémité des branches définit les issues possibles, qui sont atteintes en fonction des décisions prises à chaque étape. Sur la figure de droite est tracé le partitionnement de l'espace (X_1, X_2) qui en résulte.



(a) Représentation mathématique d'un neurone.



(b) Représentation d'un perceptron multicouches.

d'un exemple à classifier. À partir d'une base de données d'apprentissage dont on connaît les classes d'appartenance, par un processus de rétropropagation, les paramètres du réseau de neurones vont être appris afin d'être ensuite restitués sur un exemple test à classifier.

Les réseaux de neurones ont beaucoup été utilisés en reconnaissance de gestes [109,

110]. L'accélération des calculs due à l'amélioration de la technologie actuelle a vu plus récemment le *deep learning* s'étendre, donnant lieu à de multiples publications traitant de la reconnaissance d'actions [111, 112, 113] ou de la reconnaissance de gestes de la main par exemple [114, 115, 116, 117].

Á noter qu'il existe également la classification floue, qui, sous une écriture possibiliste, permet de définir un degré d'appartenance à une classe plutôt qu'une probabilité. Ce formalisme, plus souple, peut palier certains limites des classifieurs listés ci-dessus que nous avons évoquées précédemment. Néanmoins, ce type de classification n'a pas été étudié dans le cadre de cette thèse.

2.4 Bilan

Dans ce chapitre, nous avons tout d'abord dressé un récapitulatif des différents domaines de l'analyse du geste. Il apparaît que l'évaluation se distingue notamment par la nécessité de conserver des informations concrètes afin de procurer à l'utilisateur un rendu pertinent. De plus, la mesure du geste se veut localisée à la fois spatialement et temporellement : le produit fini doit pouvoir informer l'utilisateur sur le membre dont le positionnement est erroné, et à quel instant du geste. Enfin, en plus d'un score, l'appréciation du geste ne doit pas se restreindre à une simple note, mais doit informer l'utilisateur sur l'erreur qu'il commet.

Pour toutes ces raisons, de nombreux outils statistiques présentés ci-dessus ne sont pas appropriés. Ainsi, les SVM, les réseaux de neurones ou les forêts d'arbres décisionnels, qui sont des méthodes discriminatives, ne permettent que de connaître la frontière entre plusieurs classes, sans les caractériser. Ces deux méthodes qui ne passent pas par la création d'un modèle ne peuvent pas être utilisées pour déterminer la qualité d'un geste. Les chaînes de Markov cachées sont des méthodes génératives et la classification passe par la modélisation de chaque classe. Ainsi, un HMM peut être créé pour une seule classe de gestes et la probabilité qu'une réalisation d'un geste soit issue d'un HMM peut-être estimée, donnant ainsi une mesure de la qualité du geste. Le problème est que cette mesure est globale et ne renseigne en rien de l'erreur qui a été commise et de l'instant où elle a été commise. Grâce à l'algorithme de Viterbi, on peut attribuer chaque instant du geste aux états du HMM pour essayer de remonter au temps des erreurs. Le faible nombre d'états des HMM (comparativement au nombre d'instants de réalisation du geste) amènerait cependant à une localisation grossière des erreurs. Ainsi, il semble difficile, à partir des HMMs, d'amener un retour concret et pertinent à l'athlète. Pour la même raison, de nombreux descripteurs trop abstraits ne pourront être utilisés. Par exemple, utiliser uniquement le pentagone formé par les effecteurs terminaux du corps mènerait à une perte très importante de l'information, et ne permettrait pas à l'athlète un retour qui lui soit utile pour une amélioration de son geste.

Pour toutes ces raisons, nous optons dans le cadre de cette thèse pour une évaluation bas niveau reposant sur le positionnement de l'ensemble des articulations du corps à tout instant. L'algorithme DTW a été retenu afin de rendre compte d'une erreur à tout

instant du mouvement.

Dans le chapitre suivant, nous allons mettre en place, grâce à l'algorithme de DTW, une modélisation de séries temporelles, dans la perspective de modéliser le mouvement d'un expert à partir d'un jeu de mouvements exécutés par des experts.

Chapitre 3

Modélisation de séries temporelles à l'aide d'un Dynamic Time Warping (DTW)

Introduction et contexte

Dans ce chapitre, nous allons établir un certain nombre d'outils permettant de modéliser des séries temporelles. À cette fin, les outils seront d'abord développés dans un cadre plus simpliste unidimensionnel. Chaque geste étant un ensemble de séries temporelles, les résultats *1D* seront étendus aux séries multi-dimensionnelles afin de pouvoir modéliser des gestes.

Dans un premier temps, un état de l'art des méthodes de modélisation existantes sera dressé. Afin de modéliser des séries temporelles, un outil par excellence très utilisé est la déformation temporelle dynamique (DTW). De fait, un premier paragraphe nous permettra de formaliser son fonctionnement afin d'appréhender son application dans le cadre de la modélisation.

Dans un second temps, les limitations des outils actuels nous pousseront à proposer des modifications propices à notre objectif de modélisation d'un geste expert.

Enfin, ces outils seront validés *via* deux bases de données de séries temporelles, à la fois en $1D$ mais aussi pour des gestes, le tout anticipant notre objectif final de mesure de qualité d'un geste et de mise en place d'un outil d'entraînement autonome.

3.1 État de l'art

3.1.1 Alignement de séries temporelles par DTW

Avant de se focaliser plus en détail sur l'utilisation du DTW pour modéliser un jeu de séries temporelles, revenons rapidement à sa définition première.

Le DTW est un outil permettant d'aligner deux séries temporelles de façon non

linéaire au moyen d'un chemin de déformation qui fait correspondre les indices temporels des deux signaux. Sa mise en place est relativement simple. En contrepartie, plusieurs auteurs soulignent son temps de calcul très élevé par rapport à une simple distance euclidienne.

Dans un contexte d'indexation, alors que les données sont très nombreuses, la diminution du temps de calcul du DTW peut être utile. Des limites inférieures bornant le temps de calcul du DTW ont alors été introduites dans la littérature afin de rendre le processus d'indexation de séries temporelles par DTW plus rapide [118].

La première étape dans la détermination d'un alignement par DTW consiste à estimer la carte de distance $\mathbf{d}$ de composantes $d_{i,j}$ entre les signaux $(x(i))_{1\leq i\leq M}$ et $(y(j))_{1\leq j\leq N}$ à aligner :

$$d_{i,j} = (x(i) - y(j))^2 \tag{3.1}$$

À partir de cette carte de distance, on calcule la carte de distance cumulée $\mathbf{D}$ de composantes $D_{i,j}$ qui représentent la distance minimale cumulée pour atteindre le point (i, j) en partant de l'origine $(1, 1)$. Elle est donc calculée de la façon suivante :

$$D_{i,j} = d_{i,j} + \min \begin{cases} D_{i,j-1} \\ D_{i-1,j} \\ D_{i-1,j-1} \end{cases} \quad i = 2, ..., M \quad j = 2, ..., N \tag{3.2}$$

et vérifie les conditions initiales :

$$D_{1,1} = d_{1,1} \tag{3.3}$$

$$D_{1,j} = \sum_{p=1}^{j} d_{1,p} \quad j = 1, ..., N \tag{3.4}$$

$$D_{i,1} = \sum_{q=1}^{i} d_{q,1} \quad i = 1, ..., M \tag{3.5}$$

De cette distance cumulée est extrait un chemin de déformation noté ϕ_{xy} qui fait correspondre les indices des signaux $x(i)$ et $y(j)$.

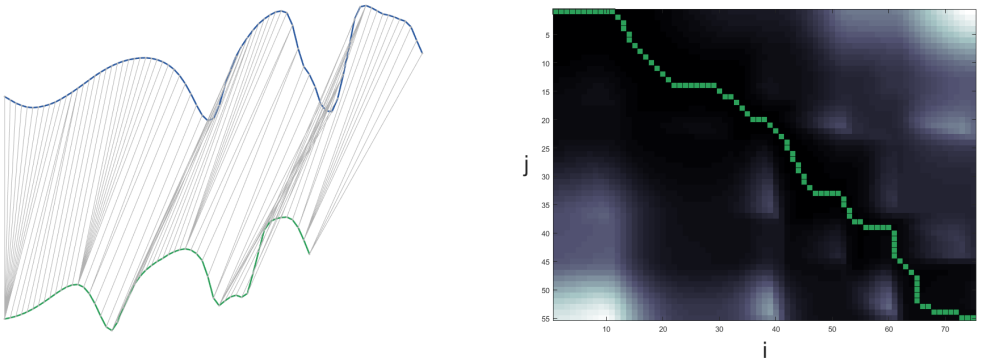
$$\phi_{xy} : \begin{cases} [\![1; K]\!] & \longrightarrow & [\![1; M]\!] \times [\![1; N]\!] \\ k & \longmapsto & \phi_{xy}(k) = (\phi^x_{xy}(k), \phi^y_{xy}(k)) \end{cases} \tag{3.6}$$

ϕ_{xy} vérifie trois conditions :

- La contrainte de monotonicité garantit la conservation de l'ordre temporel.
- Les contraintes aux frontières : $\phi_{xy}(1) = (1, 1)$ et $\phi_{xy}(K) = (M, N)$
- Les contraintes de pas élémentaires : $0 \leq \phi^x_{xy}(k) - \phi^x_{xy}(k-1) \leq 1$ et $0 \leq \phi^y_{xy}(k) - \phi^y_{xy}(k-1) \leq 1 \quad \forall k \in 2...K$, voir équation 3.2.

Le chemin de déformation minimise la distance cumulée finale $D_{M,N}$. Sa longueur K dépend des signaux à aligner, elle est déterminée durant la mise en place du DTW.

La figure 3.1 présente un exemple d'alignement par DTW entre deux signaux $x(i)$ et $y(j)$. Deux représentations de l'alignement par DTW sont visibles : la première par des segments de correspondance entre les indices des signaux (figure 3.1a) et la seconde par le chemin de déformation ϕ_{xy} (en vert sur la figure 3.1b).



(a) (b)

FIGURE 3.1 – (a) Les deux signaux (bleu et vert) sont alignés par DTW. La correspondance d'indices est indiquée par des segments gris. (b) La figure de droite superpose la carte de distance cumulée **D** et le chemin de déformation (en vert). Le blanc correspond à une grande valeur de distance cumulée $D_{i,j}$ et le noir à une faible valeur.

3.1.2 Modélisation de séries temporelles

3.1.2.1 Moyennage de deux séries temporelles

Considérons tout d'abord le cas simplifié de deux séries temporelles $x(i)$ et $y(j)$ à moyenner. Comme nous l'avons exposé en section 3.1.1, le DTW permet d'extraire un chemin d'alignement ϕ_{xy} de taille K contenant les indices d'appariement non linéaire des deux signaux $x(i)$ et $y(j)$ à aligner. Ce chemin de déformation nous permet de créer deux nouveaux signaux déformés $\tilde{x}(k)$ et $\tilde{y}(k)$ de même longueur K tels que :

$$\tilde{x}(k) = x(\phi^x_{xy}(k)) \quad k = 1, ..., K \tag{3.7}$$

$$\tilde{y}(k) = y(\phi^y_{xy}(k)) \quad k = 1, ..., K \tag{3.8}$$

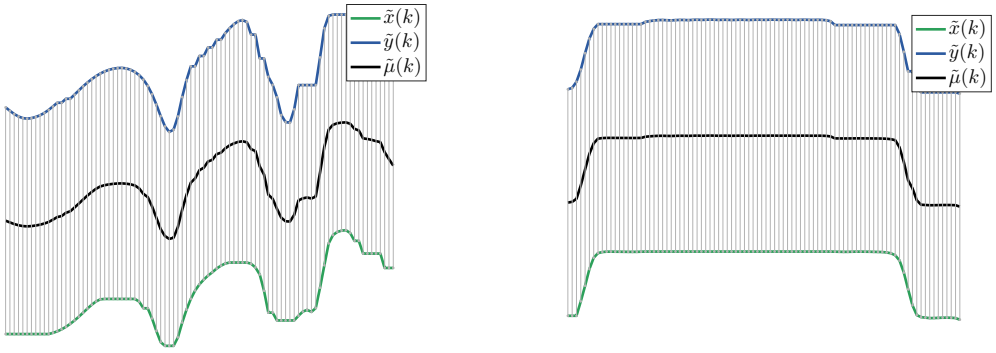
où ϕ^x_{xy} et ϕ^y_{xy} sont définis par l'équation 3.6. La moyenne des deux signaux $x(i)$ et $y(j)$ peut alors être estimée simplement comme la moyenne des signaux $\tilde{x}(k)$ et $\tilde{y}(k)$ pour chaque instant :

$$\mu(k) = \frac{\tilde{x}(k) + \tilde{y}(k)}{2} \quad k = 1, ..., K \tag{3.9}$$

De par sa construction, la taille du signal moyen $\mu(k)$ est alors $K \geq \max(M, N)$ où M et N sont les tailles respectives de $x(i)$ et $y(j)$.

La figure 3.2 illustre la construction du signal moyen $\mu(k)$ sur les deux exemples des figures 3.1 et 3.5 à partir de ϕ_{xy}. $\tilde{x}(k)$ et $\tilde{y}(k)$ (en vert et bleu respectivement) sont moyennés pour obtenir $\mu(k)$ (en noir). On remarque dans le premier cas que la taille de $\mu(k)$ est $K = 91$ tandis que celles de $x(i)$ et $y(j)$ sont $M = 75$ et $N = 55$; et que dans le second cas $K = 92$, $M = 59$ et $N = 49$. Dans les deux cas, on a bien $K \geq \max(M, N)$.

Voyons maintenant comment étendre le moyennage de signaux à une base de données plus grande.



(a) (b)

FIGURE 3.2 – Illustrations de la méthode de moyennage des deux signaux. Les deux figures correspondent aux deux signaux présentés en figures 3.1 et 3.5. Dans les deux cas, $\tilde{x}(k)$ (en vert) et $\tilde{y}(k)$ (en bleu) de même taille K on été obtenus par rééchantillonnage de $x(i)$ et $y(j)$ relativement à ϕ_{xy}. Le signal moyen résultant est $\mu(k)$ (en noir).

3.1.2.2 Extension au moyennage d'un jeu de séries temporelles

Il est donc relativement simple de moyenner deux signaux. De façon générale, le DTW est très répandu en traitement du signal, mais uniquement dans le contexte d'alignement de deux signaux. De fait, il ne permet pas directement d'extraire un signal moyen d'un jeu de plus de deux séries temporelles par simple moyennage de signaux déformés, comme on pourrait naïvement l'envisager. Plusieurs auteurs proposent donc des solutions alternatives au problème posé.

Le processus le plus intuitif est probablement celui proposé par Gupta *et al.* en 1996 : le NLAAF (pour *Nonlinear Alignment and Averaging Filters*). À partir d'un jeu de L séries temporelles $\{x_1(k), \ldots, x_L(k)\}$, les auteurs proposent d'aligner récursivement chaque paire de signaux jusqu'à les avoir tous alignés. Pour commencer, $x_1(k)$ et $x_2(k)$ sont alignés et moyennés pour donner un premier signal moyen $\mu_1(k)$, comme expliqué en section 3.1.2.1. Ensuite, $\mu_1(k)$ est aligné sur $x_3(k)$ et permet d'extraire un nouveau

signal moyen $\mu_2(k)$. Le processus continue jusqu'à avoir aligné tous les signaux. Une alternative également répandue consiste à appliquer une pondération à la moyenne de manière à donner la même influence à chacun des signaux. Le principal inconvénient de cette méthode est la succession de moyennage qu'elle engendre. Comme on l'a vu précédemment, un signal moyen est toujours plus long que les deux signaux qu'il représente. Ainsi, le signal moyen final est d'autant plus long que la base de données l'est (*i.e.* que L est grand).

Le moyennage de forme prioritaire (PSA, pour *Prioritized Shape Averaging*) est une extension du NLAAF proposée par Niennattrakul *et al.* [119]. Les signaux sont d'abord triés selon la ressemblance de leurs formes avant d'être moyennés. Étant donné un jeu de séries temporelles, les signaux les plus similaires sont d'abord détectés, alignés puis moyennés et ainsi de suite de la même manière que pour le NLAAF, jusqu'à obtenir un signal moyen final. Cette fois encore, le signal moyen est d'autant plus long que la base est importante.

Ceci implique tout d'abord un temps de calcul important, la compilation du DTW étant d'autant plus longue que les signaux à aligner sont longs. De plus, le signal ainsi obtenu est bien plus long que les signaux qu'il est censé représenter, ce qui est assez paradoxal et potentiellement problématique selon le contexte de l'étude.

Une approche plus globale, le moyennage barycentrique par DTW (DBA pour *DTW Barycenter Averaging*) a été introduite par Petitjean *et al.* en 2011 [120], et a vite fait figure de référence dans le moyennage de séries temporelles au sein de la communauté.

Contrairement aux travaux précédemment cités, les auteurs dans [120] proposent un algorithme rapide qui assure au signal moyen une taille raisonnable. Les principales étapes de cet algorithme sont les suivantes :

1. Choisir arbitrairement un signal $x_0(k)$ du jeu de séries temporelles. Ce signal initialise le signal moyen : $\mu(k) = x_0(k), \quad k = 1...M_0$ où M_0 est la taille de $x_0(k)$.

2. Itérer IT fois les étapes suivantes :

   (a) Aligner tous les signaux $(x_l(k))_{1 \leq l \leq L}$ sur $\mu(k)$ et extraire les chemins de déformation $\phi_{\mu x_l}$.

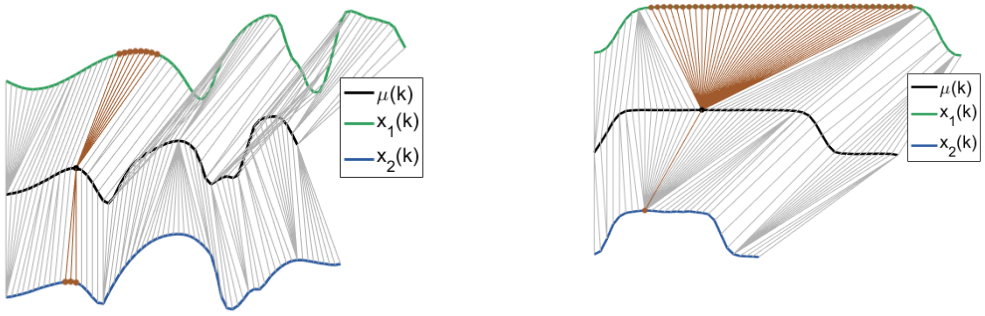
   (b) Mettre à jour tous les points du signal moyen $\mu(k), 1 \leq K \leq M_0$ comme le barycentre des points provenant de toutes les séries temporelles qui lui ont été appariés lors de l'étape (2a).

L'algorithme 1 présente le DBA, et les figures 3.3a et 3.3b l'illustrent avec les deux

signaux $x_1(k)$ et $x_2(k)$ en vert et bleu qui sont simultanément alignés sur $\mu(k)$, en noir.

Données : $x_0(k)$ de taille M_0, $(x_l(k))_{l=1...L}$ de tailles M_l, IT
Résultat : $\mu(k), \quad k = 1, ..., K$
$K = M_0$, $\mu(k) \leftarrow x_0(k), \quad k = 1, ..., K$
pour $it \in 1...IT$ **faire**
  $assoc[k] = \emptyset, k = 1...K$
  pour $l \in 1...L$ **faire**
    $\phi_{\mu x_l} \leftarrow DTW(\mu, x_l)$
    $p \leftarrow length(\phi_{\mu x_l})$
    tant que $p \geq 1$ **faire**
      $(k, n) \leftarrow \phi_{\mu x_l}(p)$
      $assoc[k] \leftarrow assoc[k] \cup \{x_l(n)\}$
      $p \leftarrow p - 1$
    fin
  fin
  pour $k \in 1...K$ **faire**
    $\mu(k) \leftarrow mean(assoc[k])$
  fin
fin

Algorithme 1 : Algorithme de *DTW Barycenter Averaging* (DBA)

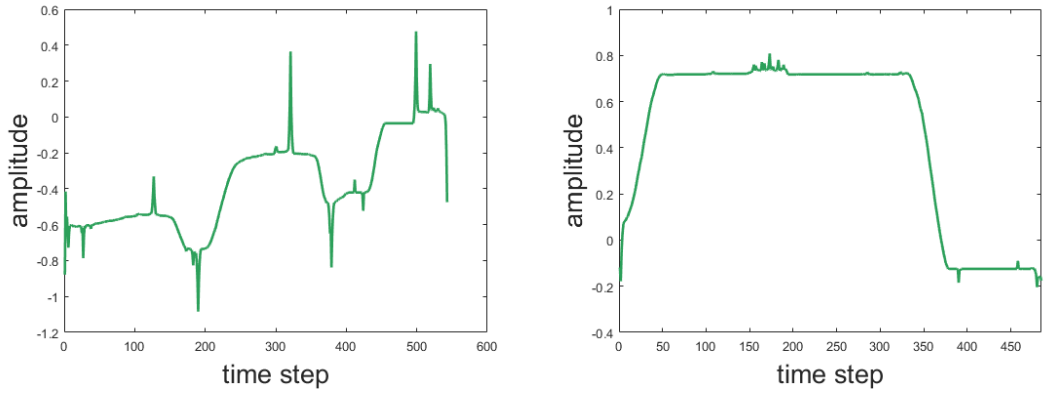


(a) (b)

FIGURE 3.3 – Une itération du DBA sur deux jeux de signaux différents. Le signal noir est le signal moyen $\mu(k)$ résultant de l'itération précédente. Les signaux bleu et vert sont alignés sur le signal noir. Chaque point de $\mu(k)$ est mis à jour comme la moyenne des points qui lui sont appariés (par exemple les points rouges pour l'indice correspondant).

Comme on l'aperçoit sur la figure 3.3b dans un cas d'alignement de signaux à paliers, un point peut être aligné à une multitude d'autres points. De fait, la contribution à la moyenne pour cet instant du signal vert sera beaucoup plus importante (beaucoup de

points) que celle du signal bleu. À l'inverse, pour les indices adjacents, c'est le signal bleu qui contribuera le plus à la création du signal moyen. En définitive, le signal moyen produit se trouve entaché d'un alignement intuitivement malvenu (on aurait instinctivement préféré que le signal vert s'aligne le plus linéairement possible avec le noir, mais ce n'est pas le cas), comme le montrent les figures 3.4a et 3.4b qui représentent les séries temporelles moyennes obtenues par DBA après 4 itérations sur un jeu de plus de 100 signaux dans les deux cas. Il s'en suit une déformation de l'allure des signaux.



(a) (b)

FIGURE 3.4 – Moyennage de deux jeux de séries par DBA après 4 itérations. Les séries temporelles moyennées sont similaires à celles présentées précédemment. Dans le premier cas (a), le jeu est constitué de 319 exemples, et le second (b) de 101 exemples.

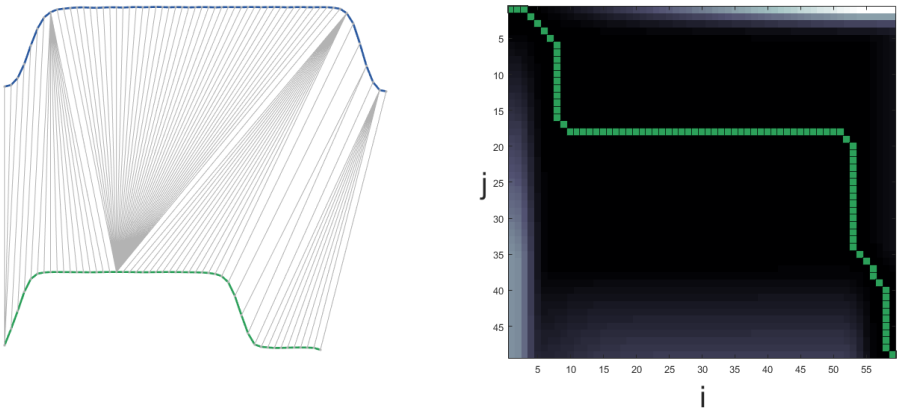
On résume ce phénomène d'alignement inopportun en l'ensemble des "chemins pathologiques", dont le paragraphe suivant va tenter de donner une explication.

3.1.3 Mise en évidence des chemins pathologiques

Une des principales limitations du DBA intervient lorsque le chemin de déformation provenant du *Dynamic Time Warping* "stagne", c'est-à-dire qu'il apparie l'indice d'une série temporelle à plusieurs indices de l'autre série temporelle. On appelle alors le chemin obtenu un "chemin de déformation pathologique". C'est le cas par exemple de l'alignement de la figure 3.5.

Comme on le voit tout d'abord sur la figure 3.5a qui matérialise en gris les appariements d'indices par le chemin de déformation, un indice d'une série temporelle est fréquemment relié à un grand nombre d'indices de l'autre. Ce phénomène se traduit dans l'aspect du chemin de déformation par de longues zones purement verticales ou horizontales (figure 3.5b). Ce phénomène est néfaste, puisqu'il mène à un mauvais recalage qui ne conserve pas les formes des signaux. Il peut avoir plusieurs causes.

De manière générale, l'alignement de deux signaux admettant des amplitudes différentes résulte en l'appariement de tous les points d'amplitudes trop élevées du premier signal avec l'unique point le plus élevé de l'autre, afin d'annihiler au maximum l'impact



(a) (b)

FIGURE 3.5 – Alignement par chemin de déformation pathologique. (a) Les deux signaux (bleu et vert) sont alignés par DTW. Notez les correspondances pathologiques de plusieurs indices sur un seul. (b) La figure de droite superpose la distance cumulée **D** et le chemin de déformation (en vert). La pathologie du chemin de déformation se traduit par de longues zones verticales ou horizontales de stagnation.

des différences d'amplitude. En effet, comme le DTW se base sur une minimisation de distance, la correspondance de points d'amplitudes différentes n'est pas encouragée.

Il en découle un alignement médiocre de séquences de signaux en plateau ou peu variables. C'est le cas de la figure 3.5b. Les deux signaux comportant des plateaux d'amplitudes légèrement différentes, tous les points du plateau du signal bleu, d'amplitudes supérieures à l'amplitude maximale du signal vert, vont s'aligner avec l'amplitude maximale du signal vert, réduite à un point. D'où l'apparition d'un chemin de déformation dit pathologique. Pour surmonter ce problème, certains travaux ont proposé d'aligner, non pas les signaux directement, mais les dérivées des signaux, de manière à s'affranchir du problème dû aux différences d'amplitudes [84, 121]. Cette méthode est cependant très dangereuse car très peu robuste au bruit, menant souvent à des recalages médiocres. Une autre idée serait de normaliser les signaux en amont par un centrage et une réduction des données. Le problème, là encore, est que cette normalisation n'est pas pertinente dès lors que les signaux sont de tailles différentes ou de temporalités variables. Prenons par exemple deux signaux en plateau, dont la durée du plateau diffère largement. La normalisation par centrage réduction mènera alors à des amplitudes très différentes, ne résolvant en rien notre problème. Une normalisation par l'amplitude crête-à-crête ne serait pas non plus efficace puisque bien trop peu robuste au bruit.

Pour limiter l'apparition de ces chemins inopportuns, une méthode que nous proposons de développer est l'utilisation de contraintes dans la détermination de chemin de déformation. Plusieurs formes de contraintes peuvent être mises en place : les contraintes

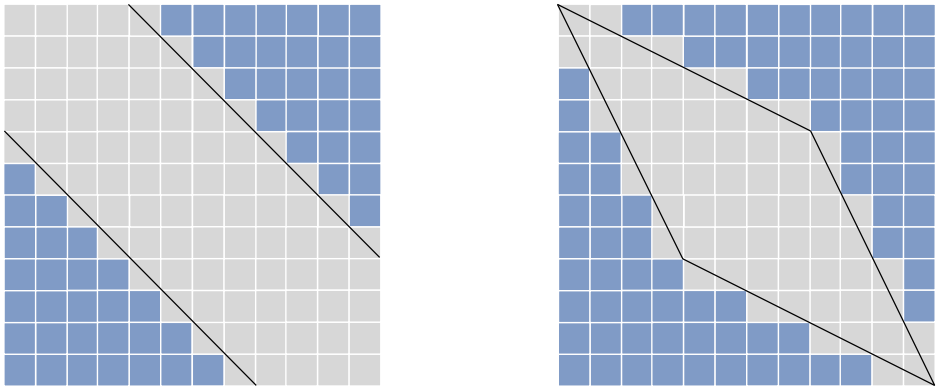
globales et les contraintes locales. L'objectif de la prochaine section est de clarifier ces notions.

3.1.4 Le DTW contraint (CDTW)

3.1.4.1 Contraintes globales

L'ajout d'une contrainte globale permet de limiter le chemin de déformation à une certaine zone. Cette démarche permet de réduire le temps de calcul du DTW, mais aussi de limiter les chemins pathologiques. En effet, si le chemin stagne à un certain indice pendant un certain temps, il se confronte aux frontières de la bande introduite par la contrainte globale et se voit obligé "d'avancer".

Cette notion de contrainte globale a été largement étudiée dans la littérature [122, 123, 124, 125]. Deux contraintes globales se démarquent cependant : la bande de Sakoe Chiba (cf. figure 3.6a) et le parallélogramme d'Itakura (figure 3.6b).



(a) Bande de Sakoe-Chiba (b) Parallélogramme d'Itakura

FIGURE 3.6 – Contraintes globales du chemin de déformation couramment rencontrées dans la littérature. Lors de l'alignement de deux signaux, l'espace de recherche du chemin de déformation est restreinte à la zone grise.

3.1.4.2 Contraintes locales

Une alternative à la contrainte globale est la contrainte locale. Dans ce cas, le chemin de déformation n'est plus simplement contraint globalement par une zone accessible de la carte de distance cumulée, mais localement par un ensemble de déplacements élémentaires admissibles. Les déplacements conflictuels horizontaux et verticaux sont proscrits. De fait, des alternatives au déplacement diagonal doivent être proposées. Une première solution, présentée en figure 3.7a, consiste à considérer trois déplacements locaux obliques comme le proposent Sakoe et Chiba [77]. D'autres contraintes locales ont également été envisagées dans la littérature, comme le résume l'article de Rabiner et Juang [126]. En

réalité, cette contrainte locale contraint globalement le chemin de déformation à appartenir au parallélogramme d'Itakura (figure 3.6b), comme le soulignent Keogh *et al.* [5].

Dans ces travaux, les auteurs proposent le calcul d'une "limite inférieure" au DTW, permettant, si elle est bien utilisée, de réduire considérablement le temps de calcul du processus d'indexation de données. Cette mesure de similarité alternative, bien plus rapide que le DTW, repose justement sur l'utilisation de contraintes locales. Keogh *et al.* notent qu'une contrainte locale consiste finalement à réduire l'espace de recherche de correspondance de l'indice i d'une série temporelle avec l'indice j d'une autre, avec $i - r \leq j \leq i + r$, où r est fixée par la contrainte (et dépend *a priori* de i). De fait, ils introduisent deux signaux annexes U et L qui enveloppent la série temporelle de référence relativement à la contrainte fixée. Enfin, la similarité de limite inférieure (*lower bound*) entre une série temporelle C et la référence Q est simplifiée en une simple distance euclidienne (plus rapide) :

$$LB_{Keogh}(Q,C) = \sqrt{\sum_{i=1}^{n} \begin{cases} (C_i - U_i)^2 & \text{si} \quad C_i > U_i \\ (C_i - L_i)^2 & \text{si} \quad C_i < L_i \\ 0 & \text{sinon} \end{cases}} \tag{3.10}$$

Cette démarche est astucieuse et permet un premier tri des données afin de restreindre le calcul du DTW seulement aux séries les plus probables. Malgré tout, l'approche est réduite au cas de séries temporelles de même taille. Dans tous les cas, elle reste bien sûr moins performante que le DTW et doit s'utiliser en amont dans un but de réduction du temps de calcul.

Notons que l'implication locale-globale n'est pas réciproque, puisque la contrainte globale d'Itakura autorise *a priori* (à condition de ne pas être proche d'une frontière établie par le parallélogramme) dans le chemin de déformation des déplacements localement verticaux ou horizontaux.

Le procédé de contraintes locales amène à davantage de déplacements locaux autour de la diagonale, comme le schématise la figure 3.7b avec K_p le nombre de déplacements élémentaires du plus long déplacement local (par exemple, $K_p = 2$ sur la figure 3.7a). On notera K_{pm} la valeur minimale de K_p permettant d'aligner les signaux (pour deux signaux de tailles M et N à aligner et $K_p < K_{pm}$, il n'est pas possible de relier $(1,1)$ à (M,N) sur la carte de distance cumulée).

La distance cumulée s'en trouve nécessairement modifiée. Pour le cas de la figure 3.7a, elle est donnée par :

$$D_{i,j} = \min \begin{cases} D_{i-1,j-2} + d_{i,j-1} + d_{i,j} \\ D_{i-1,j-1} + d_{i,j} \\ D_{i-2,j-1} + d_{i-1,j} + d_{i,j} \end{cases} \tag{3.11}$$

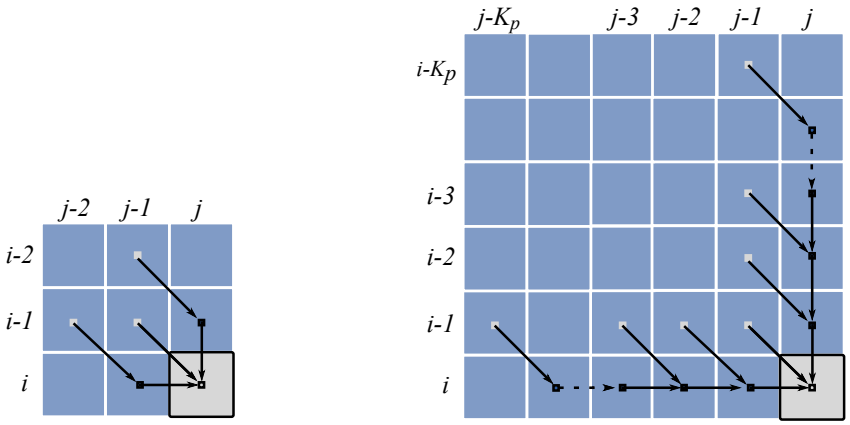


FIGURE 3.7 – Mise en place du CDTW : introduction de nouveaux déplacements élémentaires. Seulement certains déplacements sont autorisés dans le calcul de $D_{i,j}$. (a) Cas particulier de trois déplacements élémentaires admissibles. (b) Cas général de déplacements locaux contraints.

tandis que dans le cas général de la figure 3.7b, elle est donnée par :

$$D_{i,j} = \min \begin{cases} D_{i-1,j-K_p} + \sum_{k=1}^{K_p} d_{i,j-(K_p-k)} \\ D_{i-1,j-(K_p-1)} + \sum_{k=1}^{K_p-1} d_{i,j-(K_p-1-k)} \\ \vdots \\ D_{i-1,j-2} + d_{i,j-1} + d_{i,j} \\ D_{i-1,j-1} + d_{i,j} \\ D_{i-2,j-1} + d_{i-1,j} + d_{i,j} \\ \vdots \\ D_{i-(K_p-1),j-1} + \sum_{k=1}^{K_p-1} d_{i-(K_p-1-k),j} \\ D_{i-K_p,j-1} + \sum_{k=1}^{K_p} d_{i-(K_p-k),j} \end{cases} \tag{3.12}$$

L'état de l'art étant établi, les contributions apportées dans cette thèse concernant le moyennage de séries temporelles est développé et justifié dans les parties qui suivent.

3.2 Moyennage de séries temporelles : le DBA contraint

Nous avons vu en section 3.1.2.2 l'utilité du DBA mais aussi les limites du DTW et les avantages proposés par le DTW contraint (CDTW présenté en section 3.1.4). Nous proposons donc de fusionner ces deux procédés pour introduire le DBA contraint, noté

CDBA (pour *Constrained DBA*), qui est présenté dans l'algorithme 2. Cet innovant apport permet à la fois de moyenner un jeu de séries temporelles tout en assurant un signal moyen de taille raisonnable, mais aussi de s'affranchir des problèmes inhérents au DTW, les chemins pathologiques, qui peuvent détériorer la forme du signal moyen par DBA.

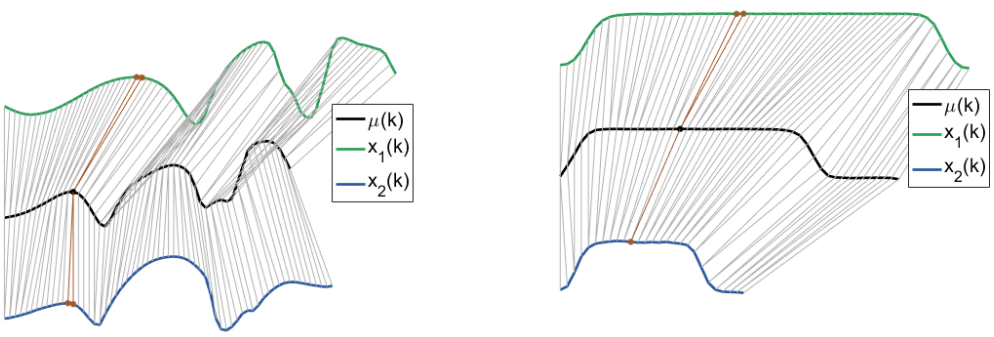
Données : $x_0(k)$ de taille M_0, $(x_l(k))_{l=1...L}$ de tailles M_l, IT, K_p
Résultat : $\mu(k), \quad k = 1...K$
$K = M_0$, $\mu(k) \leftarrow x_0(k), \quad k = 1, ..., K$
pour $it \in 1...IT$ **faire**
  $assoc[k] = \emptyset, k = 1...K$
  pour $l \in 1...L$ **faire**
    $\phi_{\mu x_l} \leftarrow CDTW(\mu, x_l, K_p)$
    $p \leftarrow length(\phi_{\mu x_l})$
    tant que $p \geq 1$ **faire**
      $(k, n) \leftarrow \phi_{\mu x_l}(p)$
      $assoc[k] \leftarrow assoc[k] \cup \{x_l(n)\}$
      $p \leftarrow p - 1$
    fin
  fin
  pour $k \in 1...K$ **faire**
    $\mu(k) \leftarrow mean(assoc[k])$
  fin
fin

Algorithme 2 : Algorithme de DBA Contraint (CDBA)

La figure 3.8 illustre l'alignement par CDTW de deux signaux $x_1(k)$ et $x_2(k)$ sur un signal moyen $\mu(k)$ lors d'une itération du CDBA. Comparativement à la figure 3.3, on remarque que les chemins pathologiques sont nettement réduits (en particulier figure 3.8b).

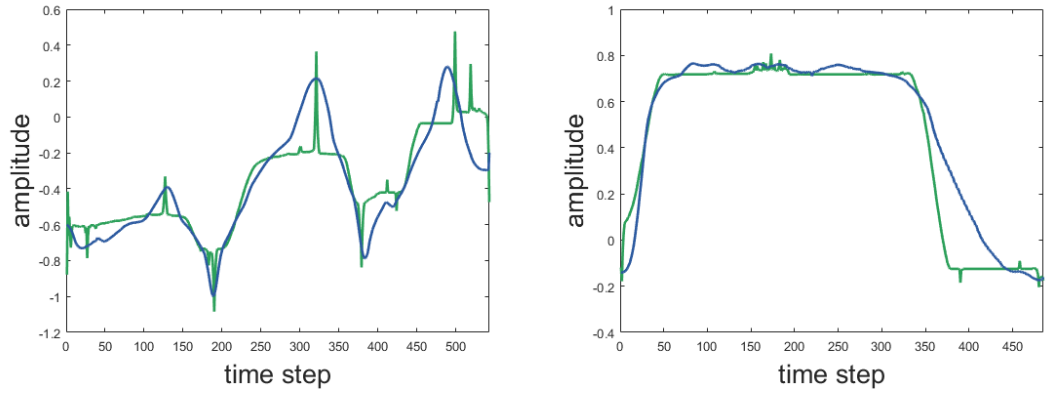
La figure 3.9 compare les signaux moyens des jeux de signaux utilisés précédemment obtenus par DBA (en vert) et par CDBA (en bleu). Le premier jeu, dont deux signaux sont illustrés en figure 3.1a, contient 319 séries temporelles, tandis que le second, présenté en figure 3.5a, en contient 101. D'un premier coup d'œil, on peut remarquer qu'alors que le DBA fait ressortir des discontinuités dues aux appariements multiples sur un même point d'amplitude extrême, le CDBA semble conserver davantage la forme du signal en évitant ces points singuliers. Bien qu'il puisse également apparier plusieurs points à un seul autre, la série engendrée est plus lisse et conforme aux données qu'elle représente étant donnée la contrainte de non stagnation du chemin de déformation.

Cela étant dit, la simple représentation d'un jeu de signaux par sa moyenne est assez pauvre. Elle ne donne notamment aucune information sur la variabilité des signaux au sein du jeu de données. Nous proposons donc d'ajouter autour de cette moyenne une tolérance, image de la variabilité des signaux. Á notre connaissance, c'est la première fois que cette tolérance est introduite dans la modélisation de séries temporelles.



(a) (b)

FIGURE 3.8 – Une itération du CDBA sur deux jeux de signaux différents. Le signal noir est le signal moyen $\mu(k)$ résultant de l'itération précédente. Les signaux bleu et vert sont alignés sur le signal noir. Chaque point de $\mu(k)$ est mis à jour comme la moyenne des points qui lui sont appariés (par exemple les points rouges pour l'indice correspondant).



(a) (b)

FIGURE 3.9 – Signaux moyens obtenus par DBA (en vert) et CDBA (en bleu) après 4 itérations pour les 2 jeux de signaux des figures 3.1a et 3.5a. A l'inverse de ceux obtenus par DBA, les signaux moyens obtenus par CDBA n'engendrent pas de discontinutés.

3.3 Modélisation de la variabilité intraclasse : la tolérance

Afin de mieux modéliser le jeu de séries temporelles, nous proposons d'adjoindre à la moyenne une tolérance qui représente les valeurs admissibles autour de cette moyenne. Cette tolérance est calculée à chaque instant du signal moyen. Elle correspond à l'écart-type du jeu de signaux alignés sur la série temporelle moyenne. L'algorithme 3 résume le processus de modélisation d'un jeu de signaux par CDBA. Notez l'ajout de tolérance

(en rouge) par rapport à l'algorithme précédent ne générant que le signal moyen.

Données : $x_0(k)$ de taille M_0, $(x_l(k))_{l=1...L}$ de tailles M_l, IT, K_p
Résultat : $\mu(k), \sigma(k), \quad k = 1...K$
$K = M_0$, $\mu(k) \leftarrow x_0(k), \quad k = 1, ..., K$
pour $it \in 1...IT$ **faire**
  $assoc[k] = \emptyset, k = 1...K$
  pour $l \in 1...L$ **faire**
    $\phi_{\mu x_l} \leftarrow CDTW(\mu, x_l, K_p)$
    $p \leftarrow length(\phi_{\mu x_l})$
    tant que $p \geq 1$ **faire**
      $(k, n) \leftarrow \phi_{\mu x_l}(p)$
      $assoc[k] \leftarrow assoc[k] \cup \{x_l(n)\}$
      $p \leftarrow p - 1$
    fin
  fin
  pour $k \in 1...K$ **faire**
    $\mu(k) \leftarrow mean(assoc[k])$
    $\sigma(k) \leftarrow std(assoc[k])$
  fin
fin

Algorithme 3 : Algorithme de CDBA avec tolérance

La figure 3.10 illustre la série temporelle moyenne ainsi que la tolérance pour les deux jeux de signaux précédemment étudiés. La tolérance est représentée comme étant $\pm(1 \times \sigma)$, où σ est donc l'écart-type des données à chaque instant.

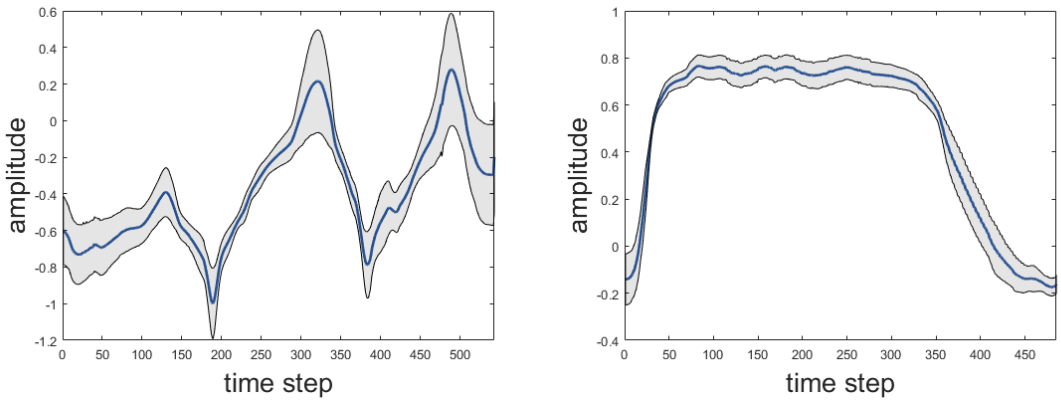
On remarque qu'ajouter la tolérance au DBA est possible mais non pertinent puisque les chemins pathologiques réduiraient la variabilité à seulement quelques points isolés.

Dans la partie suivante, nous allons valider cette nouvelle façon de modéliser les séries temporelles avec la tolérance sur une tâche de classification.

3.4 Bases de données utilisées pour la validation

3.4.1 Séries temporelles 1D : UCRTSArchive

Afin de valider nos contributions d'alignement et de modélisation de signaux de tailles variables, nous mettons en place un protocole de classification sur une archive de différentes bases de données à une dimension *UCR time-series Classification Archive* [7], notamment utilisées dans [119], [127], [128] et [120]. Les caractéristiques de cette archive contenant 20 bases de séries temporelles différentes, telles que des images (Adiac, FaceAll, FaceFour, Fish, OSULeaf, SwedishLeaf, Yoga), des mouvements (GunPoint), des données ECG (ECG200) ou des données simulées (CBF, TwoPatterns), sont données dans le tableau 3.1. Selon la base, le nombre de classes varie de 2 à 50. La taille des signaux est fixe au sein d'une base mais elle varie d'une base à l'autre, de 60 à 637 pas



(a) (b)

FIGURE 3.10 – Séries temporelles moyennes obtenues par CDBA pour deux jeux de signaux distincts. La tolérance, calculée comme $\pm(1 \times \sigma)$, est indiquée par la zone grisée autour de ce signal moyen.

de temps. Chacune des 20 bases de données est séparée en un groupe d'apprentissage et un groupe test distincts.

Base de données	**nb classes**	**apprentissage/test**	**longueur signaux**
50words	50	450-455	270
Beef	5	30-30	470
CBF	3	30-895	128
FaceAll	14	880-1690	131
FaceFour	4	24-88	350
Fish	7	175-175	463
Lighting2	2	60-61	637
Lighting7	7	70-73	319
OSULeaf	6	200-242	427
SwedishLeaf	15	500-625	128
syntheticControl	6	300-300	60
Trace	4	100-100	275
TwoPatterns	4	1000-4000	128
Wafer	2	1000-6174	152
Yoga	2	300-3000	426

TABLE 3.1 – Caractéristiques de l'archive *UCR Time-Series Classification Archives* [7]

3.4.2 Gestes : ArmGesturesM2S

La deuxième base de données utilisée, *ArmGestureM2S* contient 859 gestes, répartis en 15 classes et effectués par 10 sujets différents [6]. Un sujet effectue donc en moyenne 5.73 fois le même geste. Les caractéristiques plus spécifiques de cette base sont indiquées dans le tableau 3.2, il s'agit de gestes de bras classiques comme des applaudissements, des gifles, des saluts, des lancers, des croisements de bras, *etc.* La figure 3.11 illustre ces 15 différents mouvements.

Gestes	**classe**	**nb d'ex.**	**long. moy./classe**
Uppercut	1	58	227
Punch	2	57	235
Claque paume	3	59	241
Claque revers	4	57	237
Salut tête	5	58	356
Salut haut	6	56	377
Gratter menton	7	56	465
Mains hanches	8	57	542
Mains poche	9	55	562
Prendre au niveau des hanches	10	57	330
Prendre au niveau de la poitrine	11	57	324
Prendre en haut	12	58	329
Applaudir	13	58	438
Croiser les bras	14	58	559
Lancer	15	58	295

TABLE 3.2 – Caractéristiques de la base de données *ArmGesturesM2S*, introduite dans [6].

3.5 Moyennage de séries temporelles

3.5.1 Procédure de classification

La procédure de test est la même pour toutes les bases de l'archive de signaux $1D$. Chaque classe notée $\mathcal{C}_c$ est représentée par son signal moyen $\mu_c(k)$ et sa tolérance $\sigma_c(k)$ obtenue à partir des signaux d'apprentissage.

La classification d'un exemple test $x(k)$ se fait selon les 2 procédés suivants :

- DBA : les 2 signaux $x(k)$ et $\mu_c(k)$ sont alignés par DBA pour arriver aux signaux déformés $\tilde{x}(k)$ et $\tilde{\mu_c}(k)$ de même longueur K (éq. 3.7 et 3.8) :

$$\tilde{x}(k) = x(\phi^{x}_{\mu_c x}(k)) \quad k = 1, ..., K \tag{3.13}$$

$$\tilde{\mu_c}(k) = \mu_c(\phi^{\mu_c}_{\mu_c x}(k)) \quad k = 1, ..., K \tag{3.14}$$

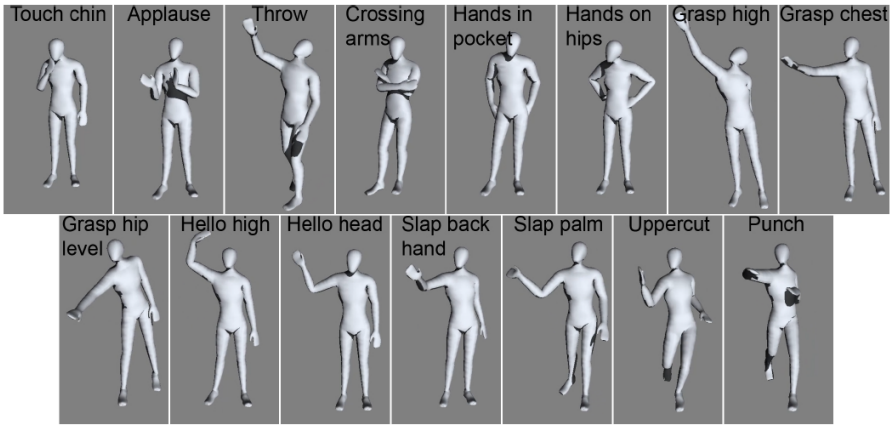


FIGURE 3.11 – Les 15 gestes du haut du corps de la base de données *ArmGesturesM2S*. Certains mouvements ont des propriétés géométriques très similaires. Illustration issue de [6].

La distance entre les signaux est alors la distance cumulée des erreurs à chaque instant :

$$d_c = \frac{1}{K} \sum_{k=1}^{K} \|\tilde{x}(k) - \tilde{\mu_c}(k)\| \tag{3.15}$$

et la classification est réalisée en recherchant la classe qui minimise cette distance.

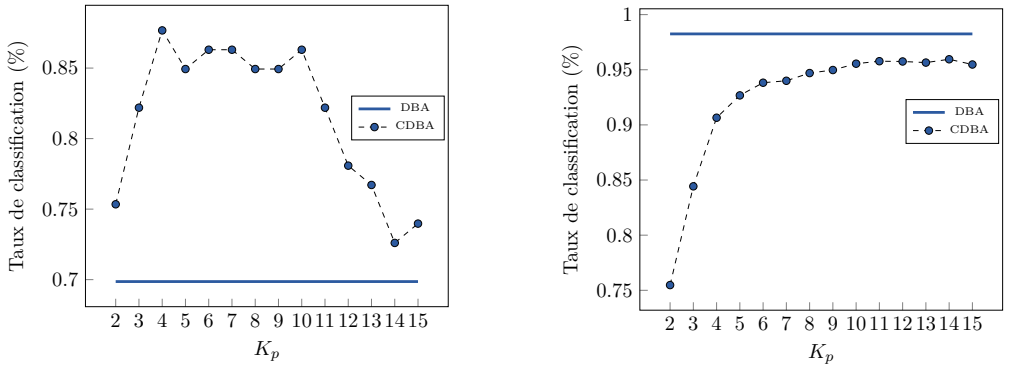
- CDBA : le même principe est utilisé mais les signaux sont maintenant alignés en utilisant le CDBA.

3.5.2 Résultats

Les résultats sans tolérance sont exposés dans le tableau 3.3. Le nombre d'itérations du DBA et du CDBA est fixé à $IT = 10$ (fixé expérimentalement, l'algorithme convergeant très rapidement, nous aurions pu prendre n'importe quelle valeur supérieure à 5). L'alignement par CDTW est paramétré par la contrainte de pente variant de $K_p = 2$ à $K_p = 11$.

Comme on peut le voir dans le tableau 3.3, les meilleurs taux du CDBA (dernière colonne) sont toujours supérieurs ou égaux à ceux du DBA, mis à part pour 2 bases : *CBF* et *TwoPatterns*. En étudiant ces deux bases, elles ont la caractéristique commune d'admettre des séries-temporelles très décalées temporellement ; décalage que le CDBA ne peut rattraper à cause de la limitation de la pente du chemin de déformation, même avec $K_p = 11$. Cependant, on remarque que plus K_p augmente, meilleur est le taux. La figure 3.12b montre notamment que pour l'une de ces deux bases, plus K_p augmente, plus le taux de classification du CDBA tend vers celui du DBA, puisqu'alors la pente du CDBA est de moins en moins contrainte, s'apparentant au DBA non contraint. On

remarque que cette propriété se généralise à de nombreuses bases de données du tableau 3.3. Enfin, il convient de signaler que le choix optimal de la pente du chemin de déformation, K_p, dépend du décalage des séries temporelles intraclasse. Une faible valeur est adaptée aux signaux peu décalés (par exemple *Lighting7*, figure 3.12a), tandis que des signaux très décalés seront d'autant mieux classifiés que K_p sera élevé (par exemple *TwoPatterns*, figure 3.12b).



(a) *Lighting7* (b) *TwoPatterns*

FIGURE 3.12 – Taux de classification en fonction de K_p pour deux bases de données : (a) *Lighting7* qui contient des signaux peu décalés et (b) *TwoPatterns* qui contient des signaux très décalés. Il convient de noter que la valeur optimale de K_p dépend des décalages des signaux intraclasses.

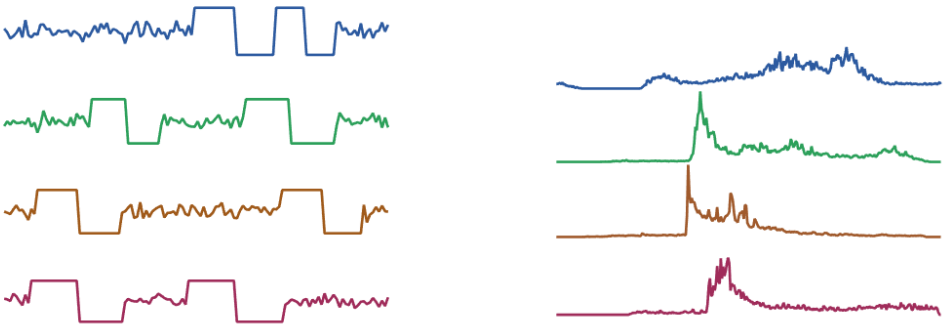
Pour mieux comprendre l'influence de K_p sur les taux de reconnaissance, comparons son effet sur deux bases de données : l'une où il est d'autant plus bénéfique qu'il est faible (*Lighting7*) et l'autre, à l'inverse, où il entraîne de meilleurs taux lorsqu'il est élevé (*TwoPatterns*).

La figure 3.13a illustre quatre exemples de séries temporelles appartenant à la même classe $\mathcal{C}_1$ de la base *TwoPatterns*. Comme on le voit sur cette figure, les signaux peuvent être très décalés au sein de cette classe (par exemple le premier et le second signal).

Leurs alignements obtenus pas DTW ou CDTW avec $K_p = 2$ sont présentés en figures 3.14a et 3.14b respectivement. De par sa limitation de pente, le CDTW ne parvient pas à ajuster les signaux. De fait, l'alignement est incorrect et les formes des signaux ne correspondent pas (figure 3.14b). À l'inverse, le chemin de déformation du DTW n'étant pas contraint, il parvient par des déplacements purement horizontaux et verticaux à faire "rattraper le retard" d'un des signaux sur l'autre et donc à aligner les signaux proprement (figure 3.14a). Ce phénomène se produit à plusieurs reprises lors du calcul du signal moyen. Aussi, tandis que le DTW parvient correctement à aligner l'ensemble des 271 séries temporelles et donc à extraire un signal moyen cohérent relativement aux données qu'il représente (figure 3.14c), le CDBA, reposant sur le CDTW, est fortement impacté par ces mauvais alignements, entraînant un signal moyen dont la forme est très différente de l'ensemble des données qu'il est censé représenter, comme le montre

Database	DBA	CDBA with K_p between 2 to 11 (%)										
	(%)	2	3	4	5	6	7	8	9	10	11	K_p opt
50words	59,24	70,60	66,59	66,15	64,14	63,03	61,92	61,92	61,47	61,69	60,58	**70,60**
Adiac	46,29	49,36	47,32	46,80	47,06	47,57	46,80	47,06	47,32	47,06	47,06	**49,36**
Beef	53,33	43,33	46,67	50,00	46,67	46,67	53,33	53,33	46,67	46,67	46,67	**53,33**
CBF	97,22	92,00	95,89	92,67	94,22	96,78	96,89	96,89	97,11	96,56	96,67	**97,11**
Coffee	96,43	96,43	96,43	96,43	96,43	96,43	96,43	96,43	96,43	96,43	96,43	**96,43**
ECG200	69,00	70,00	73,00	73,00	72,00	73,00	74,00	73,00	74,00	73,00	72,00	**74,00**
FaceAll	78,23	84,68	83,55	81,30	80,18	79,53	79,58	79,35	79,17	79,35	79,05	**84,68**
Facefour	84,09	78,41	78,41	79,55	85,23	84,09	84,09	84,09	84,09	84,09	84,09	**85,23**
Fish	70,29	78,29	73,71	73,14	74,29	75,43	74,29	69,71	69,14	67,43	67,43	**78,29**
GunPoint	72,67	68,00	69,33	72,00	72,67	72,67	70,00	70,67	71,33	70,00	70,00	**72,67**
Lighting2	63,93	68,85	63,93	62,30	65,57	59,02	60,66	63,93	63,93	60,66	62,30	**68,85**
Lighting7	69,86	75,34	82,19	87,67	84,93	86,30	86,30	84,93	84,93	86,30	82,19	**87,67**
OliveOil	83,33	83,33	83,33	83,33	83,33	83,33	83,33	83,33	83,33	83,33	83,33	**83,33**
OSULeaf	39,26	44,63	41,74	40,91	37,19	37,60	35,54	34,71	36,36	34,71	35,12	**44,63**
SwedishLeaf	67,75	70,72	68,00	68,00	68,64	68,80	69,12	68,16	69,28	68,80	68,32	**70,72**
SyntheticC.	99,00	97,00	98,67	98,67	99,00	98,67	99,00	99,00	98,67	99,00	99,00	**99,00**
Trace	100,00	95,00	94,00	95,00	97,00	97,00	98,00	100,00	100,00	100,00	100,00	**100,00**
TwoPatterns	98,25	75,48	84,43	90,65	92,68	93,83	94,00	94,70	94,98	95,55	95,78	**95,78**
Wafer	49,34	65,33	66,14	70,43	70,41	70,07	70,52	71,03	70,44	69,13	69,45	**71,03**
Yoga	54,77	53,47	54,83	52,87	53,13	53,67	54,13	51,27	51,57	51,50	52,93	**54,83**
AVERAGE	**72.61**	**73.01**	**73.41**	**74.04**	**74.24**	**74.17**	**74.40**	**74.18**	**74.01**	**73.56**	**73.42**	**76.88**

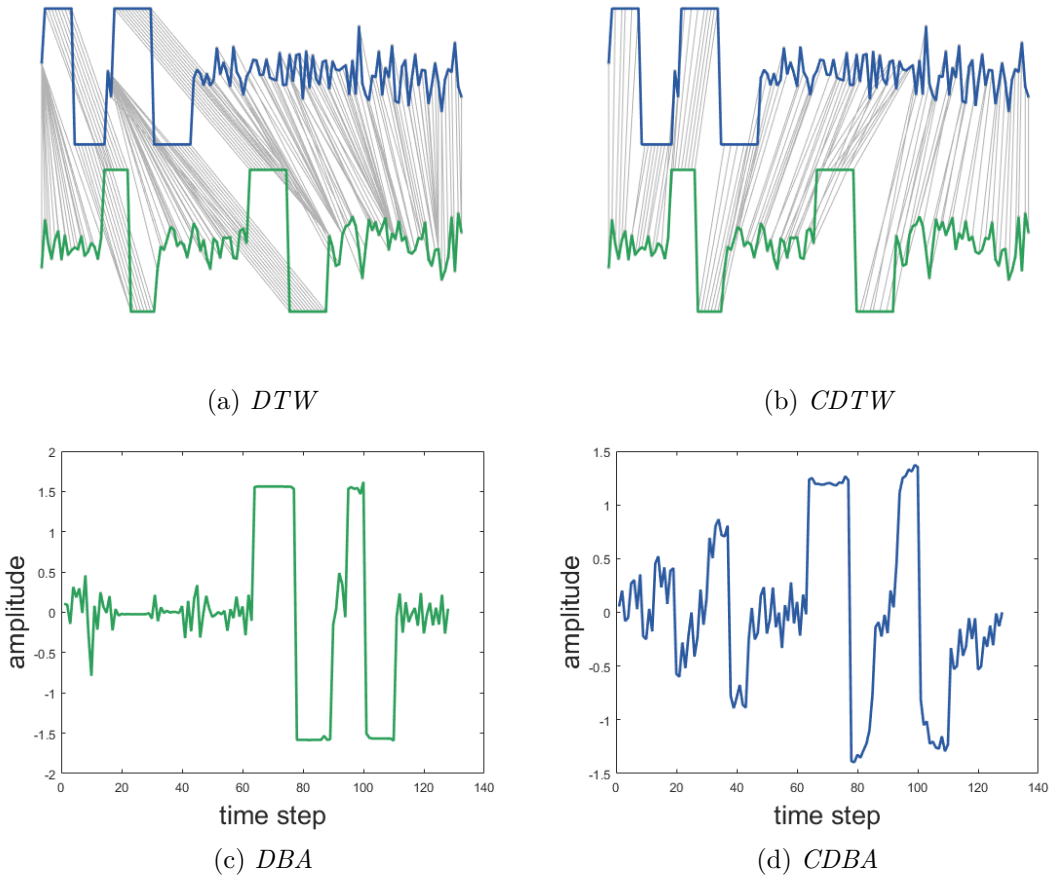
TABLE 3.3 – Taux de classification obtenus par DBA et CDBA avec une contrainte de pente K_p variant de 2 à 11. La dernière colonne reporte les meilleurs taux du CDBA pour chaque base.



(a) *TwoPatterns* (b) *Lighting7*

FIGURE 3.13 – Séries temporelles utilisées pour comparer les comportements du CDTW et du DTW. (a) Quatre signaux appartenant à la même classe de la base *TwoPatterns*. (b) Un signal appartenant à la classe $\mathcal{C}_1$ (en bleu) et trois autres appartenant à la classe $\mathcal{C}_2$ (en vert, marron et rouge) de la base *Lighting7*.

la figure 3.14d. Considérons maintenant un cas très différent pour mettre en évidence l'intérêt du CDBA. La figure 3.13b illustre cette fois une série temporelle appartenant à une classe $\mathcal{C}_1$ (en bleu) et trois autres séries temporelles appartenant à la même autre

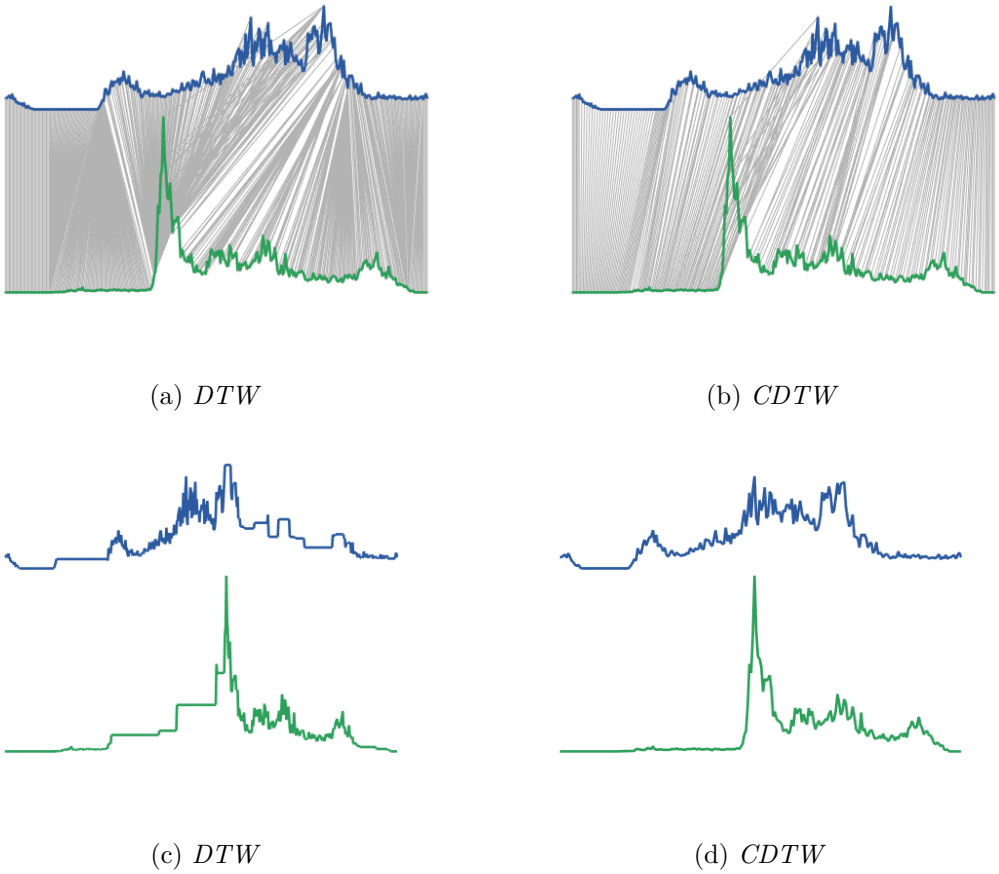


(a) *DTW* (b) *CDTW*

(c) *DBA* (d) *CDBA*

FIGURE 3.14 – Cas particulier de deux signaux très décalés de la base *TwoPatterns*. (a) Les signaux sont alignés par DTW et (b) les signaux sont alignés par CDTW. Dans les deux cas, l'appariement des indices est matérialisé par des segments gris. La seconde ligne présente les séries temporelles moyennes obtenues par (c) DBA et (d) CDBA.

classe $\mathcal{C}_2$ (en vert, marron et rouge) de la base *Lighting7*. Intéressons-nous cette fois à l'alignement de deux signaux appartenant à deux classes différentes. Étant donné que le chemin de déformation du DTW n'est pas limité en pente, il peut admettre autant de déplacements verticaux et horizontaux qu'il le souhaite pour apparier les indices des deux signaux qu'il recale.



(a) *DTW* (b) *CDTW*

(c) *DTW* (d) *CDTW*

FIGURE 3.15 – Illustration des alignements par DTW et CDTW pour deux signaux appartenant à des classes différentes. Première ligne : alignement par (a) DTW et (b) CDTW, avec $K_p = 2$. Deuxième ligne : séries temporelles déformées résultant de l'appariement par le chemin de déformation obtenu par (c) DTW et (d) par CDTW.

La figure 3.15a montre notamment que plusieurs indices d'un signal peuvent être appariés au même indice de l'autre (voir les lignes grises symbolisant les appariements du chemin de déformation) avec le DTW. Les signaux déformés par cet appariement sont représentés en figure 3.15c. Les deux signaux, bien qu'ayant des formes initialement très distinctes, admettent après déformation des formes très similaires qui pourraient mener à une mauvaise classification en aval. Au contraire, de par sa contrainte, le CDTW ne rencontre pas le même problème et les signaux déformés restent bien distincts (figures 3.15b

et 3.15d). Ceci explique le meilleur taux de classification du CDBA par rapport à celui du DBA pour la base *Lighting7* et plus généralement, pour la majorité des bases de données de l'archive.

3.6 Modélisation de la variabilité intraclasse

3.6.1 Procédure de classification

Cette fois-ci, nous allons comparer les taux de classification obtenus par DBA, CDBA et CDBA avec tolérance. Le protocole utilisé pour le DBA et le CDBA est le même que celui explicité en section 3.5.1. Cette fois-ci, afin de rendre compte de la tolérance, nous allons utiliser le chemin d'alignement déduit du CDTW. Celui-ci conduit en effet au chemin $\phi_{\mu_c x} = (\phi^{\mu_c}_{\mu_c x}, \phi^{x}_{\mu_c x})$ (voir équation 3.6) et aux signaux $\tilde{x}(k)$, $\tilde{\mu_c}(k)$ et $\tilde{\sigma_c}(k)$ de même longueur K :

$$\tilde{x}(k) = x(\phi^{x}_{\mu_c x}(k)) \quad k = 1...K \tag{3.16}$$

$$\tilde{\mu_c}(k) = \mu_c(\phi^{\mu_c}_{\mu_c x}(k)) \quad k = 1...K \tag{3.17}$$

$$\tilde{\sigma_c}(k) = \sigma_c(\phi^{\mu_c}_{\mu_c x}(k)) \quad k = 1...K \tag{3.18}$$

On détermine alors la log-probabilité pour le signal $x(k)$ d'appartenir à la classe $\mathcal{C}_c$ par :

$$P(x \in \mathcal{C}_c) = \frac{1}{K} \sum_{k=1}^{K} \ln \left(\frac{1}{\sqrt{2\pi}\tilde{\sigma_c}(k)} \exp \left\{ -\frac{(\tilde{x}(k) - \tilde{\mu_c}(k))^2}{2\tilde{\sigma_c}(k)^2} \right\} \right) \tag{3.19}$$

La classe de $x(k)$ est enfin estimée par maximisation de la log-probabilité $P(x \in \mathcal{C}_c)$.

Le CDBA a été optimisé relativement à K_p. Idéalement, cette optimisation devrait être faite en amont de la classification par un préprocessus d'apprentissage de la valeur de K_p optimale, à partir d'une base de validation. Ce n'est pas ce que nous avons fait, à cause du trop faible nombre d'exemples dans la base d'apprentissage pour pouvoir en extraire une base de validation et conserver un apprentissage du signal moyen et de la tolérance fiable.

3.6.2 Résultats

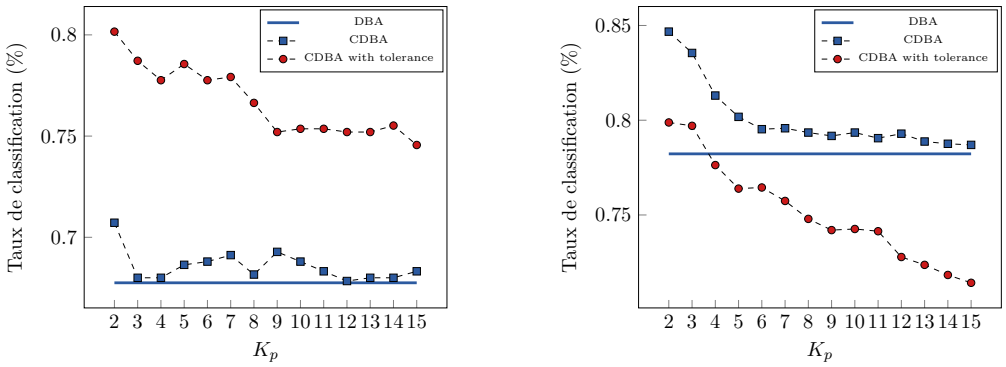
Les résultats sont présentés dans le tableau 3.4.

Une première conclusion est que le CDBA avec tolérance donne en moyenne le meilleur taux de classification (79.97% avec tolérance et 76.88% sans). L'ajout de la contrainte et de la tolérance améliore de plus de 7% les résultats du DBA.

Par exemple, la figure 3.16a montre pour la base *SwedishLeaf* que pour n'importe quelle valeur de K_p, le CDBA est meilleur avec que sans tolérance. Pour certains cas isolés, comme la base *FaceAll* (figure 3.16b), l'ajout de la tolérance détériore légèrement les résultats. Néanmoins, en analysant plus précisément la matrice de confusion de cette

Database	DBA (%)	CDBA (%)	CDBA+tol (%)
50words	59,24	70,60	65,26
Adiac	46,29	49,36	53,70
Beef	53,33	53,33	56,67
CBF	97,22	97,11	97,67
Coffee	96,43	96,43	100,00
ECG200	69,00	74,00	74,00
FaceAll	78,23	84,68	79,88
Facefour	84,09	85,23	87,50
Fish	70,29	78,29	84,57
GunPoint	72,67	72,67	90,00
Lighting2	63,93	68,85	70,49
Lighting7	69,86	87,67	82,19
OliveOil	83,33	83,33	90,00
OSULeaf	39,26	44,63	50,41
SwedishLeaf	67,75	70,72	80,16
SyntheticC.	99,00	99,00	98,33
Trace	100,00	100,00	99,00
TwoPatterns	98,25	95,78	96,13
Wafer	49,34	71,03	81,54
Yoga	54,77	54,83	61,97
AVERAGE	**72,61**	**76,88**	**79,97**

TABLE 3.4 – Taux de classification obtenus par DBA et CDBA avec et sans tolérance. Les résultats par CDBA affichés ont été obtenus après optimisation du coefficient K_p sur chaque base. Le CBDA avec tolérance donne le meilleur taux de classification moyen.



(a) *SwedishLeaf* (b) *FaceAll*

FIGURE 3.16 – Taux de classification par DBA et CDBA avec et sans tolérance en fonction de K_p.

base, on observe qu'une seule classe de cette base est mal reconnue. Nous supposons donc que ce mauvais résultat est dû à la distribution des données, qui ne doit pas être gaussienne, et implique donc une mauvaise modélisation.

3.7 Extension à la classification de gestes

3.7.1 Procédure de classification

Afin de vérifier la pertinence de notre approche dans le cas multidimensionnel, nous appliquons désormais notre méthode sur des signaux de gestes. Les données sont alors codées en dimension plus élevée. De fait, l'algorithme de DTW et ses variantes doivent être adaptés. Un geste, codé par son squelette en $3D$, peut être représenté par la position en trois dimensions de A articulations qui varient dans le temps et l'espace. On note $\boldsymbol{X}(k)$ un mouvement à l'instant k. $\boldsymbol{X}(k)$ est donc un vecteur de taille $3 \times A$. La carte distance entre deux mouvements $\boldsymbol{X}(i)$ et $\boldsymbol{Y}(j)$ est donc mise à jour par :

$$d_{i,j} = \|\boldsymbol{X}(i) - \boldsymbol{Y}(j)\|^2 \tag{3.20}$$

Le procédure est ensuite la même que précédemment ; chaque classe est modélisée par un mouvement moyen $\boldsymbol{X}_c(k)$ et une tolérance $\boldsymbol{\Sigma}_c(k)$ de dimension $(3 \times A, 3 \times A)$, obtenus par moyennage des mouvements $\boldsymbol{X}(k)$. Notons que par simplification et souci de temps de calcul, $\boldsymbol{\Sigma}_c(k)$ est supposée diagonale. De même que précédemment, on introduit les variables recalées suivantes :

$$\tilde{\boldsymbol{X}}_c(k) = \boldsymbol{X}_c(\phi^{X_c}_{\boldsymbol{X_c X}}(k)) \tag{3.21}$$

$$\tilde{\boldsymbol{X}}(k) = \boldsymbol{X}(\phi^{X}_{\boldsymbol{X_c X}}(k)) \tag{3.22}$$

$$\tilde{\boldsymbol{\Sigma}}_c(k) = \boldsymbol{\Sigma}_c(\phi^{X_c}_{\boldsymbol{X_c X}}(k)) \tag{3.23}$$

La classification est effectuée, soit par minimisation de la distance cumulée (sans tolérance), soit par maximisation de la log-probabilité de l'équation 3.19 généralisée aux gestes. La log-probabilité que le geste $\mathbf{X}$ appartienne à la classe $\mathcal{C}_c$ est donnée par :

$$P(\mathbf{X} \in \mathcal{C}_c) = \frac{1}{K} \sum_{k=1}^{K} \ln \left(\frac{1}{|\tilde{\boldsymbol{\Sigma}}_c(k)|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\tilde{\boldsymbol{X}}(k) - \tilde{\boldsymbol{X}}_c(k))^T \tilde{\boldsymbol{\Sigma}}_c(k)^{-1} (\tilde{\boldsymbol{X}}(k) - \tilde{\boldsymbol{X}}_c(k)) \right\} \right) \tag{3.24}$$

Si on applique cette équation directement, les résultats de classification obtenus sont plutôt mauvais. On remarque aussi qu'ils sont surtout dûs à certaines articulations quasi immobiles (de tolérance très faible, donc de gaussienne très haute et de largeur très faible). Pour surmonter ce problème, nous proposons de considérer un seuil minimal S de tolérance, et on impose tous les coefficients diagonaux à être supérieurs à ce seuil. Nous avons testé différentes valeurs de S à des fins d'optimisation.

3.7.2 Résultats

Les résultats de classification sont donnés dans le tableau 3.5. On rappelle que la base utilisée, explicité en section 3.4.2, contient 859 gestes, répartis en 15 classes (applaudissement, coup de poing, salut, croisement de bras, *etc.*) et effectués par 10 sujets différents. Un protocole *Leave-One-Subject-Out* a été mis en place afin d'assurer une fiabilité des résultats, c'est-à-dire que chaque geste d'un sujet donné a été classé à partir d'un apprentissage de tous les mouvements effectués par les autres sujets.

Seuil	Sans tolérance							
	DBA	CDBA						
		K_{pm}	$K_{pm}+1$	$K_{pm}+2$	$K_{pm}+3$	$K_{pm}+4$	$K_{pm}+5$	$K_{pm}+6$
0,04	76,834	76,019	**83,818**	83,003	82,654	82,305	81,257	81,141
0,02	76,834	76,019	83,818	83,003	82,654	82,305	81,257	81,141
0,01	76,834	76,019	83,818	83,003	82,654	82,305	81,257	81,141
0,005	76,834	76,019	83,818	83,003	82,654	82,305	81,257	81,141
0,003	76,834	76,019	83,818	83,003	82,654	82,305	81,257	81,141
0,002	76,834	76,019	**83,818**	83,003	82,654	82,305	81,257	81,141
0,001	76,834	76,019	**83,818**	83,003	82,654	82,305	81,257	81,141

Seuil	Avec tolérance							
	DBA	CDBA						
		K_{pm}	$K_{pm}+1$	$K_{pm}+2$	$K_{pm}+3$	$K_{pm}+4$	$K_{pm}+5$	$K_{pm}+6$
0,04	77,066	78,231	83,702	83,236	82,654	82,421	81,723	81,490
0,02	77,299	78,813	**84,284**	83,702	83,236	83,120	82,305	81,839
0,01	79,744	79,511	84,051	**85,215**	84,400	84,400	84,517	84,051
0,005	82,654	78,696	83,818	84,866	**85,332**	**85,332**	**85,332**	84,983
0,003	83,353	77,416	82,654	83,469	**84,284**	83,818	83,702	84,168
0,002	82,887	76,484	81,024	82,421	83,120	82,887	82,654	82,421
0,001	79,744	75,437	80,326	80,792	81,257	80,675	80,442	80,093

Table 3.5 – Taux de classification de gestes par DBA et CDBA avec et sans tolérance avec le paramètre K_p variable, de $K_p = K_{pm}$ jusqu'à $K_p = K_{pm} + 6$.

On remarque dans un premier temps que les résultats d'un processus sans tolérance ne sont bien sûr pas impactés par la modification du seuil, tout simplement parce que le seuil agit sur la tolérance uniquement.

Nous remarquons ensuite que l'utilisation du CDBA (DBA contraint) améliore grandement les résultats par rapport au DBA (on gagne au mieux 7%, sans tolérance, ce qui est loin d'être négligeable). Le meilleur taux de classification est d'ailleurs obtenu pour $K_p = K_{pm} + 3$, où K_{pm} correspond à la valeur minimale de K_p permettant d'aligner les signaux (on rappelle que la contrainte locale induit une contrainte globale de délimitation de chemin en parallélogramme, en deçà de $K_p = K_{pm}$, la zone admissible du chemin de déformation ne permet pas de parcourir la carte de distance cumulée de $(1, 1)$

à (M, N)). Avec $K_p = K_{pm} + 3$, on atteint un taux de bonne classification de 85.3%.

L'ajout de la tolérance a un effet plus ou moins bénéfique selon le seuil choisi.

- Si le seuil est trop grand, on tend vers une classification à tolérance constante, correspondant à une analyse sans tolérance.
- Si le seuil est trop petit, les articulations quasi-immobiles entraînent des tolérances extrêmement faibles qui dictent à elles seules la classification et dégradent les résultats.

Ces comportements se traduisent dans le tableau 3.5 par une amélioration des résultats seulement pour un seuil moyen (S entre 0.003 et 0.01). Avec $S = 0.005$, on parvient notamment à améliorer le taux de classification d'environ 1.5%, validant ainsi l'intérêt de la tolérance mais aussi de la contrainte dans un contexte d'alignement de gestes.

Conclusion

A partir d'outils de la littérature, ce chapitre a permis de mettre en place des approches nouvelles validées sur plusieurs protocoles de classification.

Plusieurs modèles permettant l'alignement de séries temporelles ont tout d'abord été passés en revue, et nous avons décidé d'utiliser le *Dynamic Time Warping* (DTW). Très répandu depuis de nombreuses années, il fait l'objet de nombreuses variantes. Le principe de fonctionnement du DTW et du CDTW a donc été rappelé ainsi que leur effet dans l'alignement de signaux temporels.

Le DTW ne considérant qu'une paire de séries temporelles à aligner, ne peut pas être utilisé tel quel pour le moyennage d'un jeu de signaux. Dès lors, différentes techniques ont été développées pour répondre à ce besoin, dont le NLAAF et le DBA dont les principes ont également été rappelés.

Dans un second temps, les limites des outils actuels ont été mises en avant. Ces difficultés scientifiques ont été résolues par l'introduction de nouvelles approches reposant sur les outils existants de la littérature, d'abord pour aligner des signaux, puis pour les moyenner.

La modélisation d'un jeu de séries temporelles a été améliorée grâce à l'introduction d'une contrainte dans le processus du DBA, mais aussi par l'utilisation d'une tolérance permettant de modéliser la variabilité intraclasse des signaux au cours du temps. Les apports ont été validés sur une tâche de classification.

Enfin, afin d'anticiper l'utilisation de ces nouvelles notions dans le cadre de l'évaluation de gestes, les outils ont été généralisés au cas multidimensionnel d'une classification de gestes. Là encore, les contributions mises au point permettent de surpasser les méthodes actuelles reposant sur le DTW.

Maintenant que tous ces outils ont été établis, nous allons les utiliser pour mesurer la qualité d'un geste sportif, et en déterminer les principales erreurs.

Chapitre 4

Mesure de qualité d'un geste sportif

Introduction

"Plus vite, plus haut, plus fort", telle est la devise des Jeux Olympiques proposée par Pierre de Coubertin à la création du comité olympique en 1894. Pour amener à la performance, l'entraîneur sportif est déterminant : il est celui qui observe, critique, juge le mouvement de son athlète pour lui permettre une meilleure efficacité gestuelle. La tâche n'est pas sans difficulté, l'entraîneur se sert d'ailleurs bien souvent d'enregistrements vidéo lui permettant une meilleure analyse d'un mouvement souvent très rapide et complexe. De fait, la technologie apporte une objectivité très conséquente pour le sportif qui veut s'améliorer. Dans ce chapitre, nous proposons de mettre au service du sportif et de l'entraîneur un outil d'évaluation qui mesure la qualité de son geste à chaque instant de son mouvement. Pour ce faire, son geste est comparé à un geste expert, modélisé à partir d'un jeu de gestes effectués par des experts et obtenu grâce aux outils développés dans le chapitre précédent. La méthode proposée doit répondre à une triple problématique :

- mesurer la qualité d'un geste novice et être capable d'en déterminer les principaux points à améliorer.
- tenir compte automatiquement de la morphologie du sujet.
- s'adapter automatiquement à tous types de gestes et de sports dès lors qu'un jeu de gestes experts est disponible.

Après avoir dressé un état de l'art des différentes approches d'évaluation de geste, nous présenterons les bases de données acquises au laboratoire M2S, puis établirons le processus d'évaluation de la qualité d'un geste novice, en subdivisant l'évaluation selon deux critères : un critère spatial et un critère temporel. Les résultats obtenus seront comparés à des annotations d'entraîneurs sportifs afin d'être validés de manière rigoureuse.

4.1 État de l'art

Si beaucoup de travaux ont été réalisés sur la reconnaissance de gestes ou la synthèse de gestes (section 2.1.4), très peu d'entre eux se sont intéressés à l'estimation de la qualité d'un geste et encore moins à la recherche de la performance dans la réalisation des gestes. Les rares travaux se situent tous dans le domaine chirurgical ou sportif.

4.1.1 Évaluation de gestes chirurgicaux

L'évaluation de gestes chirurgicaux présente un objectif similaire à celui du geste sportif [129]. Bien que les compétences demandées au chirurgien soient différentes de celles requises par le sportif, l'objectif est le même : exécuter un geste optimal pour améliorer la tâche chirurgicale.

L'étude du geste chirurgical s'est étendue ces dernières années, notamment grâce à l'utilisation d'outils robotisés collaboratifs ou de télémanipulation, avec ou sans retour d'effort, permettant l'acquisition simple et précise de données cinématiques du geste. Une base de données très complète de gestes chirurgicaux utilisant le robot de télémanipulation Da Vinci a été développée récemment par Gao *et al.*, comprenant divers gestes tels que des sutures, des nœuds et des manipulations diverses d'aiguilles [130]. Ces gestes ont été exécutés par huit chirurgiens de tous niveaux. Ils sont en outre segmentés en 15 unités d'actions et un jeu de 76 variables permet une analyse fine des données récoltées.

Les approches se focalisent souvent sur l'outil utilisé par le chirurgien. Par exemple en laparoscopie, Hofstad *et al.* [131] évaluent le geste par la douceur du mouvement de l'objet, sa vitesse moyenne et sa trajectoire globale. C'est aussi le cas des travaux de Despinoy *et al.* [45] et Pham *et al.* [44] qui se basent sur la courbure de l'outil chirurgical au cours du temps pour évaluer un geste chirurgical par comparaison avec le geste d'un ou de plusieurs experts. C'est entre autres une des idées reprises par Ahmidi *et al.* [132], qui confrontent cette fois une analyse cinématique bas niveau (accélération, vitesse, profil spectral du mouvement de l'outil) avec une analyse plus haut niveau utilisant notamment la courbure mais aussi la durée d'action, la longueur d'arc, *etc.* Les résultats obtenus se restreignent à une classification de niveau du chirurgien par SVM. Ils sont meilleurs lorsque les descripteurs utilisés sont haut niveau. *A contrario*, cette approche optimisée pour une classe de gestes donnée ne peut *a priori* pas se généraliser à d'autres gestes puisque son paramétrage dépend de la finalité de la tâche choisie.

D'autres travaux s'appuient plus simplement sur l'image du geste chirurgical. Sharma *et al.* [133] tentent de prédire le critère OSATS. Ce critère est un système graduel de notation du mouvement chirurgical reposant sur 7 observations : le respect du tissu, la durée et la quantité de mouvement, la tenue de l'outil, la manipulation de la suture, la fluidité du déroulement des étapes, la bonne connaissance de la procédure et la performance globale. Habituellement évalué manuellement par des médecins, le but de l'étude est d'estimer ce score automatiquement à partir de la vidéo du geste. Les résultats suggèrent une corrélation entre les scores annotés et calculés. Reposant uniquement sur les images et l'extraction de points d'intérêt, cette méthode semble être efficace mais ne permet qu'une évaluation globale du geste sans localiser les éventuelles erreurs commises.

4.1.2 Évaluation de gestes sportifs

Relativement peu de travaux tentent d'évaluer un geste sportif, geste multidimensionnel par nature. Contrairement au geste chirurgical focalisé sur l'outil, l'évaluation automatique du geste sportif nécessite de savoir sur quelle partie du corps se focaliser et à quel moment. Pour cela, plusieurs approches sont possibles : l'apprentissage du mouvement ou l'ajout de connaissances *a priori*. C'est sur cette dernière que reposent les travaux décrits dans [12] qui consistent à mettre en place un outil d'entraînement au karaté. En utilisant des descripteurs cinématiques inhérents aux mouvements considérés - et fournis par des entraîneurs de karaté -, Burns *et al.* analysent les mouvements de plusieurs novices et fournissent un outil d'entraînement interactif en réalité virtuelle permettant une progression des sujets.

Il convient de noter que cette façon de gérer les données conduit à un outil qui n'est pas générique aux différents sports individuels, mais restreint à un mouvement particulier connu. C'est également en ajoutant des informations relatives au sport que Komura *et al.* [57] proposent un outil d'entraînement aux arts martiaux. Leur système se focalise sur la minimisation du mouvement total du défenseur lors d'une attaque, la prévisibilité d'attaque et la vitesse de coup de poing. En effet, ces données sont connues pour être particulièrement liées à la bonne performance du sport considéré. Utilisant le même principe d'ajout de connaissance inhérente au geste, Ward analyse dans sa thèse des gestes de danse de ballet [42]. À nouveau, les paramètres considérés sont très précis et localisés sur les informations les plus pertinentes relativement au mouvement : l'extension du genou, la rotation de la hanche, le décalage thoracique antéropostérieur, *etc.*

Maes *et al.* ont mis en place une plate-forme pour l'apprentissage autonome de la danse [43]. Ici, le défi se trouve simplifié par l'utilisation du tempo de la musique qui impose une synchronie naturelle des mouvements et une vitesse d'exécution constante. La méthode, utilisant une simple corrélation entre le mouvement expert et le mouvement novice, est discutable puisque la variabilité au sein du jeu de mouvement d'experts n'est pas prise en compte et l'analyse, qui est faite globalement, ne permet pas de résultats précis par parties du corps. Kyan *et al.* proposent eux aussi un système d'entraînement de danse [17], mais la temporalité du geste considéré n'est pas appréhendée. Un mouvement est projeté dans un espace sphérique de posture (SSOM) qui permet, à l'aide d'un unique descripteur, d'identifier les catégories de mouvements exécutés par l'athlète, puis de les comparer aux mouvements du professeur. Le score donné à l'utilisateur à chaque instant est global, reposant sur une simple distance entre des descripteurs angulaires des deux poses. L'estimation de l'évaluation de chaque membre est laissée à la charge de l'utilisateur par superposition visuelle de son mouvement et de celui du professeur.

À ce jour et à notre connaissance, seuls Pirsiavash *et al.* se sont intéressés à la question de l'évaluation de gestes multidimensionnels sans connaissance *a priori* du mouvement réalisé [134]. Contrairement à nos travaux, l'acquisition du geste est faite par une caméra. Les poses des sportifs sont ensuite extraites des images, de sorte que la problématique devient tout à fait similaire à la nôtre, à ceci près que les auteurs ont accès à un nombre très important d'annotations. En effet, les vidéos utilisées sont celles de plongeons ou

de chorégraphies de patinage artistique exécutés lors de grands championnats, et pour lesquels les notes des juges sont également disponibles. De fait, grâce à cette grande ressource, un apprentissage automatique par SVM est mis en place pour l'évaluation du mouvement. Notons au passage que cette approche, en plus de nécessiter un nombre très important de données annotées, ne permet pas une mesure de la qualité au cours du mouvement mais uniquement un score global. Enfin, la temporalité des actions n'est pas considérée.

Nous proposons donc dans la partie qui suit une évaluation qui réponde à tous les besoins qui ont été fixés (indépendance à la morphologie, aucune connaissance technique *a priori* du mouvement étudié, localisation temporelle et spatiale des erreurs, estimation quantitative de celles-ci et détermination des points à améliorer dans le geste). Cette approche sera testée sur deux bases de données que nous présentons dans la section qui suit.

4.2 Bases de données et codage du mouvement

4.2.1 Notations et codage du mouvement

Cette partie a pour but d'évaluer la qualité d'un geste sportif avec comme unique prérequis, un jeu de gestes experts effectuant le même mouvement sportif. En effet, il est nécessaire de disposer d'une base de données afin de caractériser le geste et ses variabilités : les sujets ont différentes morphologies, exécutent le geste à différentes vitesses, peuvent avoir des styles différents. Pour autant, deux gestes peuvent être très bons mais aussi très distincts. Il paraît donc primordial d'estimer cette variabilité. Dès lors, plus la taille de la base de données sera grande, plus les résultats seront fiables puisque s'appuyant sur davantage d'exemples.

Par la suite, nous considérons les notations suivantes :

- A : nombre d'articulations.
- M : nombre de membres.
- N_E et N_N : nombre d'experts et nombre de novices respectivement.
- $\boldsymbol{X}_i(k) = \{\boldsymbol{x}_i^a(k), a = 1...A\}$ avec $\boldsymbol{x}_i^a(k) = (x_i^a(k), y_i^a(k), z_i^a(k))$ et $k = 1...M_i$ la trajectoire $3D$ de l'articulation a du $i^{\text{ème}}$ mouvement, de durée M_i. Ainsi, $\boldsymbol{X}_i(k)$, de taille $3 \times A$, représente la pose globale du sujet à l'instant k pour le $i^{\text{ème}}$ mouvement, tandis que $\boldsymbol{x}_i^a(k), a = 1...A$ décrit uniquement la position $3D$ de l'articulation a à l'instant k du mouvement i.
- $\boldsymbol{X}_i^m(k) = \{\boldsymbol{x}_i^a(k), a \in \mathcal{S}_m\}$ où $\mathcal{S}_m = \{a \in 1...A | a$ appartient au membre $m\}$ représente la position de l'ensemble des articulations composant le membre m à l'instant k.

Un geste $X_i(k)$ est donc représenté comme la série temporelle des coordonnées $3D$ des 25 articulations ($A = 25$) composant le squelette du sportif.

Pour créer la base de données nécessaire, les gestes ont été capturés avec un système opto-électronique `Vicon MX-40` (Oxford Metrics Inc., Oxford, UK). Les sujets étaient équipés de 43 marqueurs réfléchissants positionnés sur des repères anatomiques (figure 4.1a). Une douzaine de caméras `Vicon MX-40` positionnées autour du sujet équipé effectuant son mouvement permettent d'extraire grâce à la norme ISB [1] un squelette composé de 25 centres articulaires, qui sont illustrés à la figure 4.1b.

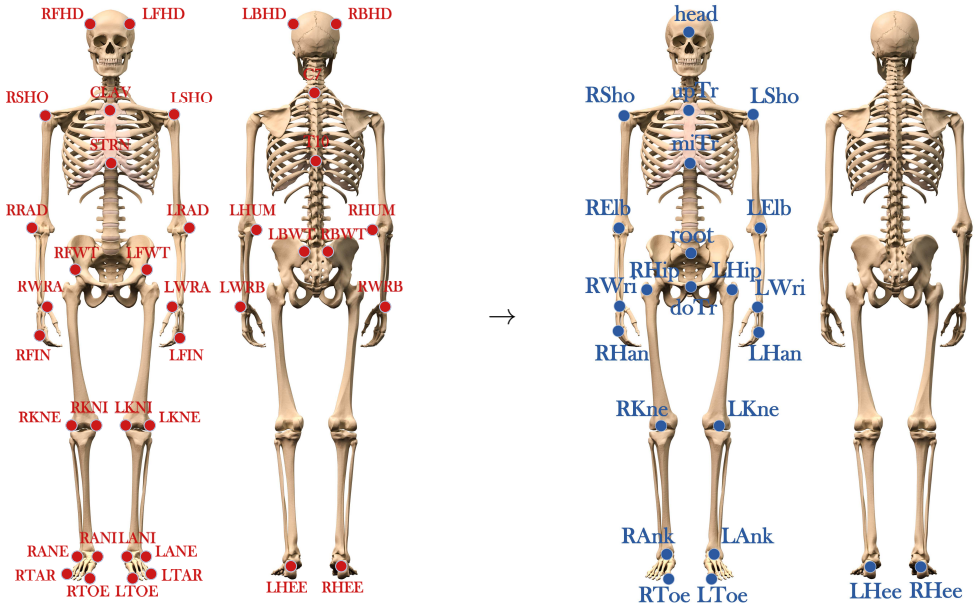


FIGURE 4.1 – (a) Positionnement des marqueurs sur une personne et (b) positionnement des articulations du squelette. On note particulièrement la position de la racine (*root*). Les noms abrègent les articulations correspondantes (par exemple "RElb" abrège *Right Elbow*, *i.e.* coude droit en anglais).

Comme le rappelle la figure 4.2, ces centres articulaires définissent 5 membres ($M = 5$) de la façon suivante :

1. le tronc est composé des épaules, des hanches, de la tête et des articulations ventrales. Les articulations concernées sont "head", "upTr","miTr", "root", "doTr", "RSho", "LSho", "RHip" et "LHip".

2. le bras droit est composé des articulations "RElb","RWri" et "RHan".

3. le bras gauche est composé des articulations "LElb","LWri" et "LHan".

4. la jambe droite est composée des articulations "RKne","RAnk", "RHee" et "RToe".

1. Société Internationale de Biomécanique

5. la jambe gauche est composée des articulations “LKne”,“LAnk”, “LHee” et “LToe”.

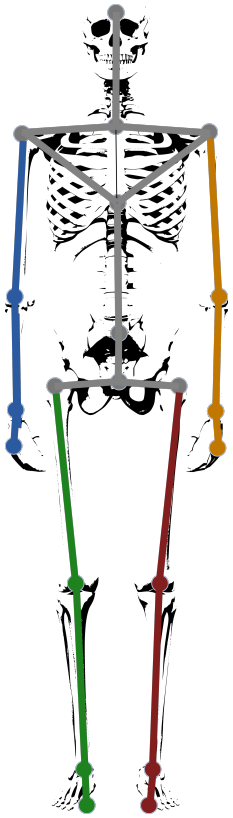


FIGURE 4.2 – Définition des membres considérés : en gris, les articulations du tronc ; en orange, celles du bras gauche ; en bleu, celles du bras droit ; en rouge, celles de la jambe gauche et en vert, celles de la jambe droite.

L'acquisition de deux gestes exécutés par des personnes distinctes amène de façon générale à des variabilités qu'il convient de gérer intelligemment pour appréhender leur comparaison dans les meilleurs conditions possibles. Tout d'abord, le mouvement étant enregistré en $3D$, le positionnement initial du sujet varie. Le problème a été étudié par Rao *et al.* et Parameswaran *et al.* dans le cas de séquences vidéo $2D$, gérant les problèmes d'extraction, représentation et interprétation d'informations visuelles dans le but d'apprendre les mouvements indépendamment du point de vue [39, 135]. Dans le cadre de nos travaux de thèse, le problème est moins complexe puisque l'utilisation de la *motion capture* nous permet de gérer directement les positionnements $3D$ des articulations. En effet, le changement de point de vue peut être résolu par l'utilisation d'un repère mobile fixé sur le sujet.

La variabilité morphologique est un problème plus complexe puisque la comparaison brutale du mouvement d'un adulte d'un mètre quatre vingt avec celui d'un enfant d'un mètre quarante serait bien sûre vaine. Selon le descripteur utilisé (par exemple la repré-

sentation angulaire des articulations), la retranscription du mouvement d'un adulte sur un avatar enfant n'engendrerait pas le même mouvement, par exemple. C'est ce type de problématique qu'ont étudié Kulpa *et al.* pour l'animation d'avatars [36], repris quelques années plus tard par Sorel [11]. D'autres travaux, comme ceux de Sie *et al.* [136], ont mis en place une normalisation très simple des squelettes, comme nous l'avons repris dans nos travaux.
Deux étapes composent le codage des données :

1. Dans un premier temps, l'ensemble des articulations est exprimé dans un repère mobile centré sur l'articulation de la racine ("root") et orienté par les hanches ("RHip" et "LHip'").

2. Dans un second temps, l'ensemble des coordonnées des articulations est divisé par la longueur du torse, elle-même définie comme la distance moyenne entre la tête ("head") et la racine ("root").

Chaque mouvement enregistré a été ensuite segmenté manuellement de manière à conserver uniquement l'information utile du mouvement. Les critères de segmentation ont été conservé pour tous les mouvements de la base (par exemple au tennis, le début du service était enclenché lorsque le bras lanceur débutait son mouvement de lancer).

Enfin, le style est une notion complexe. Le style est défini comme la variabilité observée entre deux réalisations d'un même geste. Les variabilités entre deux mêmes mouvements sont en réalité dues à de multiples facteurs : la technique, le positionnement $3D$ initial, la morphologie du sujet ou encore la vitesse d'exécution. Dans le cadre de cette thèse, on ne tiendra pas compte dans la notion de style des variabilités anthropométriques, du positionnement $3D$ initial ou de la vitesse d'exécution du mouvement. Le style, par exemple, sera le fait qu'un joueur de tennis ramène plus ou moins sa jambe, lève son bras frappeur plus ou moins tôt, lors du service de tennis, sans pour autant que cela n'influe sur sa performance sportive. Cette notion a été étudié très tôt dans la littérature [34, 35].

Dans le cadre de cette thèse, le style d'exécution n'a pas été étudié. Un des buts fixés par cette thèse est de trouver l'essence du geste des experts quel que soit leur style. Le service de tennis, notamment celui des experts, est souvent enclin à différentes variantes de style d'exécution. *A contrario*, le geste de karaté considéré est trop codifié pour admettre des styles. Ces deux gestes sont ceux utilisés dans cette thèse, les bases de données correspondantes sont décrites dans le paragraphe suivant.

4.2.2 Conditions expérimentales : bases de données

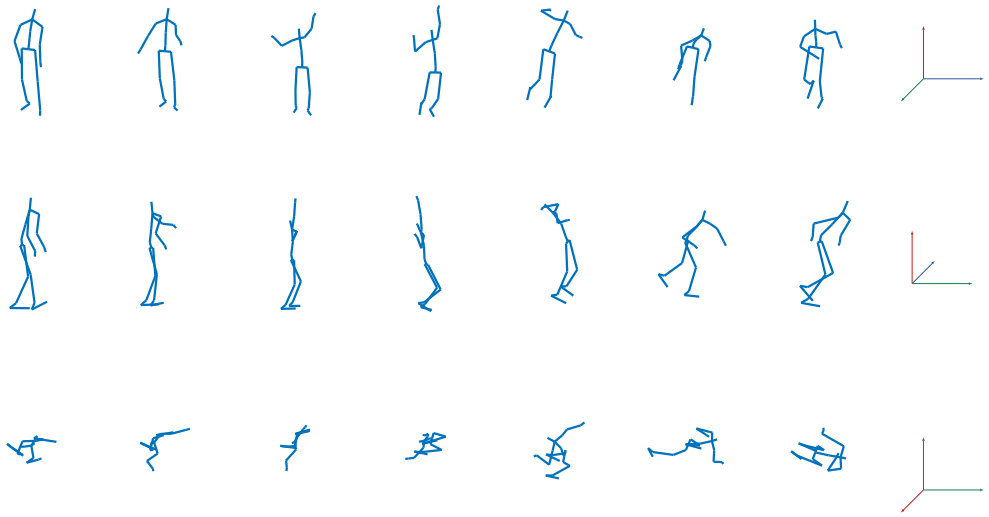
4.2.2.1 Le service de tennis

Un service débute un échange de tennis. Il consiste à lancer la balle à la verticale, puis à la frapper. Au tennis, le service est un coup essentiel. Il permet souvent aux bons joueurs de prendre le dessus sur l'adversaire. C'est le seul coup qui est entièrement maîtrisé par le joueur puisqu'il se donne la balle à lui-même au lieu de la recevoir de l'adversaire.

Le service de tennis est très contraint à la fois spatialement et temporellement. Dans un premier temps, la balle doit être lancée de la manière la plus stable possible. En parallèle, le sujet arme son geste. Il lui reste alors à déclencher la frappe pour être parfaitement synchronisé avec la balle. Un enchainement segmentaire adapté, des appuis des jambes jusqu'aux bras permet de limiter les contraintes articulaires que subit le joueur.

La figure 4.3 illustre en parallèle un service de tennis exécuté par un expert et enregistré image par image, puis un service expert enregistré et extrait par *motion capture* et vu sous plusieurs angles.

(a) Une séquence d'images d'un service de tennis (Flickr, 2012).



(b) Service de tennis squelettisé

FIGURE 4.3 – Kinogramme d'un service de tennis.

247 mouvements d'experts et 72 mouvements de novices ont été utilisés, exécutés par 23 experts (classés) et 8 débutants. Ces données ont été enregistrées au laboratoire M2S de Rennes. Ce geste a été choisi pour sa grande complexité, mais aussi parce qu'il a été très étudié dans ce laboratoire et permet donc d'accéder à un nombre très important de données.

4.2.2.2 Le *Zuki* au karaté

Le *Choku Zuki*, qu'on abrégera en *Zuki* par la suite, est le coup de poest le coup de poing fondamentaling fondamental du karaté. C'est d'ailleurs généralement la première technique apprise sur un *dojo*. Beaucoup de gestes de karaté découlent de ce mouvement fondamental, de sorte qu'un karatéka ne cesse de perfectionner son *Zuki* au cours de sa vie de sportif.

Le *Zuki* requiert une forte synchronisation des deux bras. Alors que le premier se retire en position armée contre la hanche, le second frappe simultanément tout en suivant une trajectoire linéaire qui se termine face au sternum. Les deux bras doivent être très relâchés. En théorie, toute la force provient de la rotation des hanches et de la contraction des muscles abdominaux, synchronisés avec la respiration du karatéka. Les deux poignets doivent tourner de manière homogène durant le mouvement de telle sorte qu'ils aient pivoté de 180° à la fin du geste. Durant tout le mouvement les poings doivent rester fermés.

La figure 4.4 illustre en parallèle un mouvement de *Zuki* exécuté par un expert et enregistré par image, puis un *Zuki* expert enregistré et extrait par *motion capture* et vu sous plusieurs angles.

30 mouvements d'experts et 65 de novices ont été enregistrés et extraits par *motion capture* au laboratoire M2S de Rennes. Ils ont été exécutés par 6 experts et 9 novices. Ce mouvement a été notamment choisi pour la grande rigueur de synchronisme entre membres qu'il requiert.

4.3 Modélisation du mouvement expert

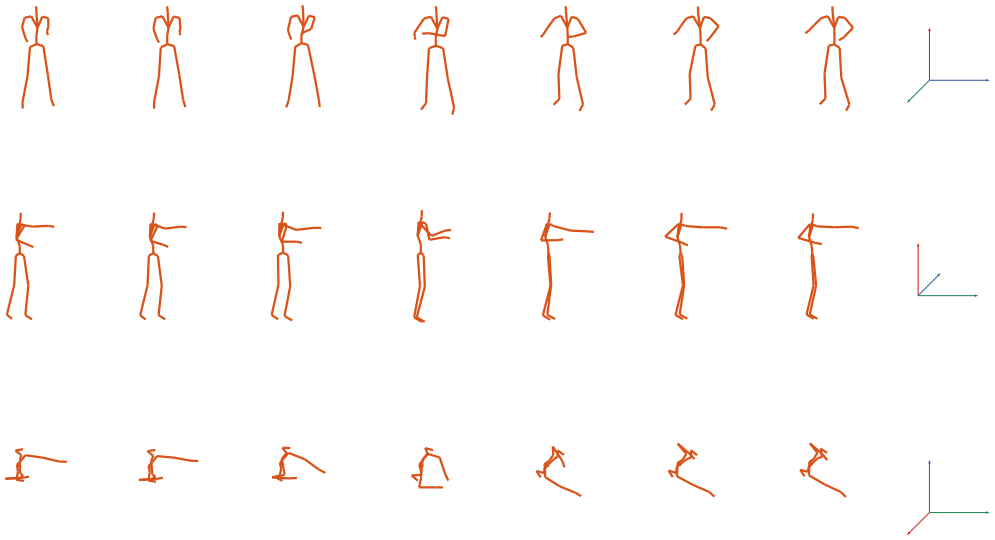
Tous les experts exécutent globalement le même geste. Pour autant, la vitesse d'exécution peut varier d'un sujet à l'autre, de même que le positionnement des membres. Toutes ces variabilités nécessitent une modélisation du geste expert afin de mener à bien la tâche d'évaluation d'un sujet novice. Dans un premier temps, nous allons donc modéliser un mouvement expert moyen, puis allons lui adjoindre une tolérance traduisant la variabilité admissible autour de ce mouvement dit "mouvement nominal" (terme repris de [72]).

4.3.1 Mouvement nominal

Afin de gommer les différences de vitesses d'exécution des mouvements experts, une première étape consiste à les recaler. Dans la mesure où ces mouvements sont considérés comme optimaux, le synchronisme entre les membres n'est pas remis en cause et le recalage des gestes est fait globalement, en considérant tous les membres du corps. Deux gestes $\boldsymbol{X}_0(k)$ et $\boldsymbol{X}_1(k)$ de durées M_0 et M_1 sont alignés grâce à une généralisation du DTW $1D$ à l'alignement multidimensionnel. La carte de distance $\mathbf{d}$ de composantes $d_{k,k'}$, utilisée dans le DTW (équation 3.2), est donnée par :

$$d_{k,k'} = \sum_{a=1}^{A} \|\boldsymbol{x}_0^a(k) - \boldsymbol{x}_1^a(k')\|^2, \qquad k \in \{1...M_0\}, k' \in \{1...M_1\} \tag{4.1}$$

(a) Une séquence d'images d'un mouvement de *Zuki* (Youtube, 2015).



(b) Mouvement de *Zuki* squelettisé.

FIGURE 4.4 – Kinogramme d'un mouvement de *Zuki*.

où $\boldsymbol{x}_i^a(k)$ représente la position $3D$ de l'articulation a, à l'instant k du geste i, comme défini page 74.

Un chemin de déformation $\phi_{\boldsymbol{X}_0\boldsymbol{X}_1}(k), k \in \{1...K\}$ est extrait, conformément à l'équation 3.6 :

$$\phi_{\boldsymbol{X}_0\boldsymbol{X}_1}(k) = (\phi^{\boldsymbol{X}_0}_{\boldsymbol{X}_0\boldsymbol{X}_1}(k), \phi^{\boldsymbol{X}_1}_{\boldsymbol{X}_0\boldsymbol{X}_1}(k)) \tag{4.2}$$

La figure 4.5 illustre la déformation de deux services $\boldsymbol{X}_0(k)$ et $\boldsymbol{X}_1(k)$ relativement à leur chemin de déformation $\phi_{\boldsymbol{X}_0\boldsymbol{X}_1}$ obtenu par DTW. Bien que les mouvements ne soient pas rigoureusement les mêmes, le DTW trouve un compromis qui minimise la distance cumulée totale entre les deux squelettes lors de leurs mouvements.

Pour rendre compte de tous les signaux et comme le DTW ne permet que l'alignement d'une paire de signaux, nous utilisons le DBA. Plus précisément, comme explicité dans la section 3.2, nous proposons d'utiliser le CDBA pour moyenner l'ensemble des gestes experts, afin de limiter les problèmes de chemins pathologiques évoqués dans le chapitre

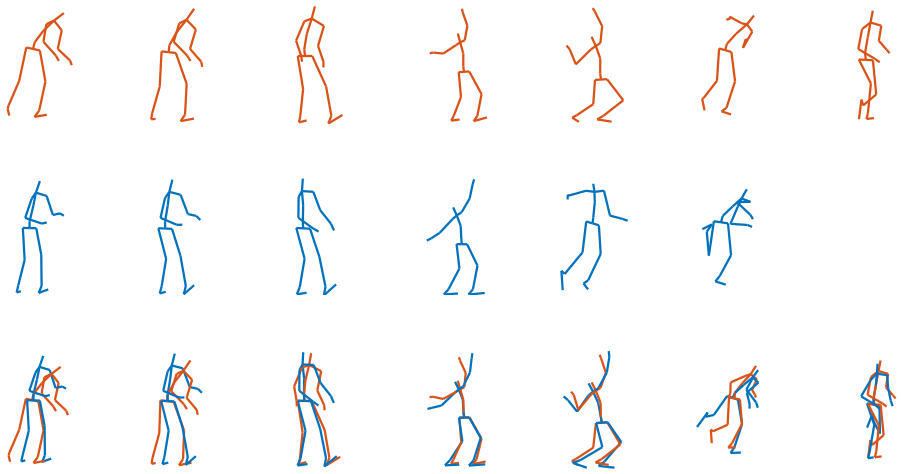


FIGURE 4.5 – Deux services de tennis alignés par le DTW. Les deux premières lignes représentent sous forme de kinogrammes deux services de tennis de durées différentes, la troisième ligne superpose ces deux services une fois alignés par DTW.

précédent. Du CDBA (algorithme 2) résulte alors le geste nominal, nommé $\boldsymbol{X}_n(k) = \{\boldsymbol{x}_n^a(k)\}_{a=1...A}$ par la suite. Ce geste nominal peut être vu comme le geste moyen du jeu de gestes experts alignés les uns avec les autres.

En aucun cas il ne fait office de geste parfait, il doit simplement être vu comme un outil mathématique et non comme un mouvement réel ; dans un premier temps parce que, de par sa mise en place, il ne conserve pas la structure morphologique du corps, et dans un second temps parce que le geste parfait n'existe pas et surtout pas en tant que moyenne de gestes corrects. À l'instar d'une gaussienne $1D$ modélisée par une moyenne et un écart-type, nous allons maintenant affiner la modélisation du geste expert par l'introduction d'une tolérance, image de la variabilité de chaque articulation autour de la position de la même articulation du mouvement nominal.

4.3.2 Tolérance articulaire

Pour tenir compte des variabilités d'exécution des gestes par les experts, on introduit la tolérance. La qualité d'un mouvement dépend donc de la tolérance accordée à chaque articulation autour de sa position nominale. Cette tolérance, bien sûr, dépend du mouvement étudié. Typiquement, le positionnement du pied lors d'un coup de pied de karaté demande bien plus de précision spatiale que le positionnement de ce même pied lors d'un lancer au basket-ball. Cette tolérance dépend bien évidemment aussi de l'articulation considérée et du temps. Les entraîneurs sportifs parlent d'ailleurs bien souvent de positionnement instantané (posture) plutôt que global : on dira par exemple souvent que le tronc de l'athlète de saut de haies doit être le plus bas possible lorsque la jambe est à l'horizontale au-dessus de la haie, plutôt que de parler de la trajectoire totale de

la jambe lors du franchissement.

Reprenant les travaux du chapitre précédent, nous ajoutons au moyennage par CDBA un calcul de la tolérance, comme le fait l'algorithme 3, généralisé au moyennage de séries temporelles à plus haute dimension. De fait, la tolérance $\boldsymbol{\sigma}_n^a(k)$ est calculée pour chaque articulation a et à chaque instant k comme la matrice de covariance des positions de cette même articulation parmi toutes les données expert recalées sur le geste nominal. On notera $\boldsymbol{\Sigma}_n(k) = \{\boldsymbol{\sigma}_n^a(k)\}_{a=1...A}$. Là encore, afin de réduire la complexité en temps de calcul, les termes non diagonaux des matrices de covariance 3×3 sont négligés.

La figure 4.6 montre deux tolérances spatiales du poignet et de la hanche à un instant donné d'un mouvement. Notez la grande différence de tolérance accordée à chacune de ces deux articulations.

FIGURE 4.6 – Tolérance spatiale du poignet gauche (en jaune-vert) et de la hanche droite (en bleu) à un instant k donné. Le mouvement nominal à cet instant est illustré en noir opaque, et correspond à la moyenne des gestes alignés, en gris sur la figure. Pour plus de lisibilité, les tolérances affichées correspondent à $3 \times \boldsymbol{\sigma}_n^a(k)$ à titre illustratif.

4.4 Evaluation du mouvement d'un novice

Un geste sportif peut être mauvais pour de multiples raisons : un bras trop en avant, une jambe trop lente, un manque de rotation du bassin, un bras en retard sur l'autre... Compte tenu de toutes ces possibilités, il apparaît complexe pour le sportif de comprendre et de corriger un tel flux d'informations.

Comme cela a déjà été évoqué, une composante cinématique est à la fois due au spatial et au temporel. Un mouvement est complexe parce que les aspects spatiaux et temporels évoluent simultanément, comme l'illustre un mouvement du jonglage par

exemple. Les gestes sont codifiés par des positions, mais aussi par des synchronisations entre des parties du corps, qui sont souvent primordiales pour mener à bien une tâche. À des fins pédagogiques d'apprentissage, nous proposons donc de distinguer deux erreurs principales : les erreurs spatiales et les erreurs temporelles. En effet, même si le positionnement de chacune des articulations est correct, il se peut que la coordination ne le soit pas. Dans ce cas, une erreur temporelle devrait être détectée. Par exemple, un service de tennis est composé d'un enchainement segmentaire bien précis de la jambe au bras frappeur. Dès lors, notre analyse est relativement bas niveau puisqu'on travaille sur la cinématique du geste uniquement. Toutes ces informations cinématiques seront disponibles à l'entraîneur pour orienter plus objectivement les consignes qu'il donne à son athlète.

Nous allons donc expliciter le calcul de l'erreur spatiale et de l'erreur temporelle. Chacune de ces erreurs est estimée pour chaque membre et à chaque instant du mouvement. Nous estimons que l'analyse par membre est plus pertinente que l'analyse par articulation en anticipant sur l'outil final et son retour visuel. En effet, une analyse par articulation semble être trop ciblée pour l'œil néophyte du novice qui souhaite s'améliorer. Néanmoins, ces informations sont calculées et restent disponibles pour plus de précision, au besoin.

4.4.1 Erreurs spatiales

Soit un geste novice $\boldsymbol{X}_l(k)$ à évaluer. On rappelle que $\boldsymbol{X}_l$ contient l'ensemble des trajectoires des A articulations notées $\{\boldsymbol{x}_l^a(k)\}_{a=1...A}$. Ses erreurs spatiales sont donc estimées à chaque instant k et pour chaque membre m, relativement à la tolérance articulaire autour du mouvement nominal.

Afin de rendre l'erreur spatiale indépendante de l'erreur temporelle, l'estimation de l'erreur est faite à partir d'un alignement membre à membre entre le mouvement du novice et le nominal. On remarque que ce n'est pas ce qui a été fait dans la section 4.3.2 pour évaluer les tolérances articulaires à partir des mouvements experts, tout simplement parce que les mouvements experts sont supposés sans erreur temporelle, c'est-à-dire que le recalage membre à membre mène sensiblement au même recalage que le recalage global.

L'alignement du membre m repose donc sur le calcul de la carte de distance $\mathbf{d}^m$ suivante :

$$d^m_{k,k'} = \sum_{a \in \mathcal{S}_m} \|\boldsymbol{x}_l^a(k) - \boldsymbol{x}_n^a(k')\|^2 \tag{4.3}$$

où $\mathcal{S}_m$ contient l'ensemble des articulations appartenant au membre m. Elle conduit au chemin de déformation $\phi_{\boldsymbol{X}_l^m \boldsymbol{X}_n^m}(k) = \left(\phi^{\boldsymbol{X}_l^m}_{\boldsymbol{X}_l^m \boldsymbol{X}_n^m}(k), \phi^{\boldsymbol{X}_n^m}_{\boldsymbol{X}_l^m \boldsymbol{X}_n^m}(k)\right)$ avec $k = 1...K$ qui aligne le membre m du mouvement novice $\boldsymbol{X}_l$ au membre m du mouvement nominal $\boldsymbol{X}_n$.

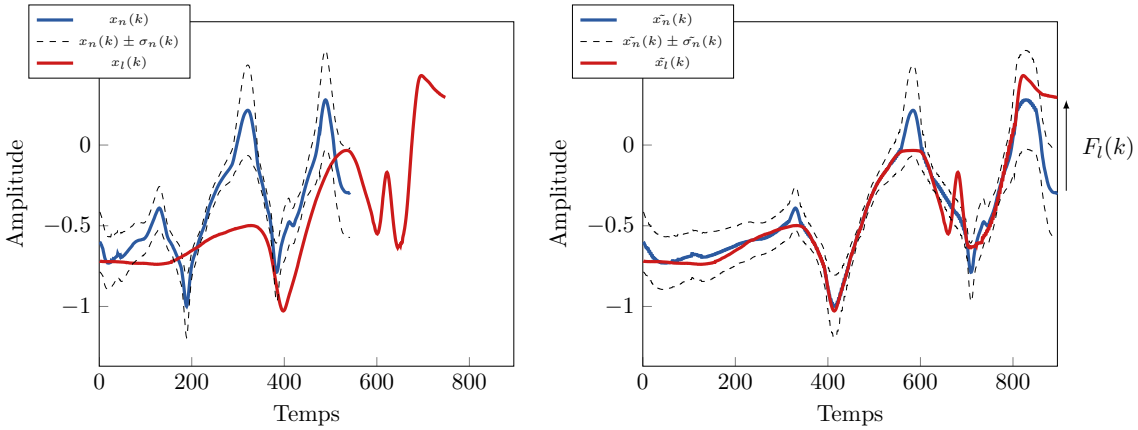


FIGURE 4.7 – Illustration de l'erreur spatiale de la série temporelle $x_l(k)$ dans le cas unidimensionnel, avec $x_n(k)$ et $\sigma_n(k)$ la série temporelle nominale et sa tolérance. À gauche sont représentés les signaux avant recalage. À droite, les signaux ont été déformés selon le chemin de déformation $\phi_{x_l x_n}(k)$.

Notons $\tilde{\boldsymbol{X}}_l^m(k) = \{\tilde{\boldsymbol{x}}_l^a(k)\}_{a\in\mathcal{S}_m}$ et $\tilde{\boldsymbol{X}}_n^m(k) = \{\tilde{\boldsymbol{x}}_n^a(k)\}_{a\in\mathcal{S}_m}$ les signaux ainsi recalés :

$$\tilde{\boldsymbol{x}}_l^a(k) = \boldsymbol{x}_l^a(\phi_{\boldsymbol{X}_l^m \boldsymbol{X}_n^m}^{\boldsymbol{X}_l^m}(k)) \quad k = 1...K, a \in \mathcal{S}_m \tag{4.4}$$

$$\tilde{\boldsymbol{x}}_n^a(k) = \boldsymbol{x}_n^a(\phi_{\boldsymbol{X}_l^m \boldsymbol{X}_n^m}^{\boldsymbol{X}_n^m}(k)) \quad k = 1...K, a \in \mathcal{S}_m \tag{4.5}$$

et $\tilde{\boldsymbol{\sigma}}_n^m = \{\tilde{\boldsymbol{\sigma}}_n^a\}_{a\in\mathcal{S}_m}$ les matrices de covariances recalées :

$$\tilde{\boldsymbol{\sigma}}_n^a(k) = \boldsymbol{\sigma}_n^a(\phi_{\boldsymbol{X}_l^m \boldsymbol{X}_n^m}^{\boldsymbol{X}_n^m}(k)) \quad k = 1...K, a \in \mathcal{S}_m \tag{4.6}$$

L'erreur spatiale est alors estimée par la distance de Mahalanobis entre la position du membre du novice et celle du mouvement nominal :

$$E_l^m(k) = \sqrt{\sum_{a\in\mathcal{S}_m} (\tilde{\boldsymbol{x}}_l^a(k) - \tilde{\boldsymbol{x}}_n^a(k))^T (\tilde{\sigma_n^a}(k))^{-1} (\tilde{\boldsymbol{x}}_l^a(k) - \tilde{\boldsymbol{x}}_n^a(k))} \quad k = 1...K \tag{4.7}$$

avec K la taille du chemin de déformation.

La figure 4.7 illustre ce calcul dans un cas unidimensionnel et avec une seule articulation. $x_n(k)$ représente la série temporelle nominale, $x_l(k)$ la série à évaluer et $\sigma_n(k)$ l'écart-type des données du jeu de signaux d'experts. Notons dès à présent que le signal nominal est de longueur 550 tandis que le signal à évaluer a une longueur de 750. Dans un premier temps, ces deux signaux sont recalés pour arriver à des signaux déformés $\tilde{x_n}(k)$, $\tilde{\sigma_n}(k)$ et $\tilde{x_l}(k)$, $k = 1...K$ avec $K = 900$ dans cet exemple.

Maintenant que les signaux sont recalés, la distance entre les signaux est estimée à chaque instant grâce à la distance de Mahalanobis qui peut se réécrire dans ce cas d'étude $1D$ par $E_l(k) = \frac{|\tilde{x_n}(k) - \tilde{x_l}(k)|}{\tilde{\sigma_n}(k)} \quad \forall k \in \{1...K\}$.

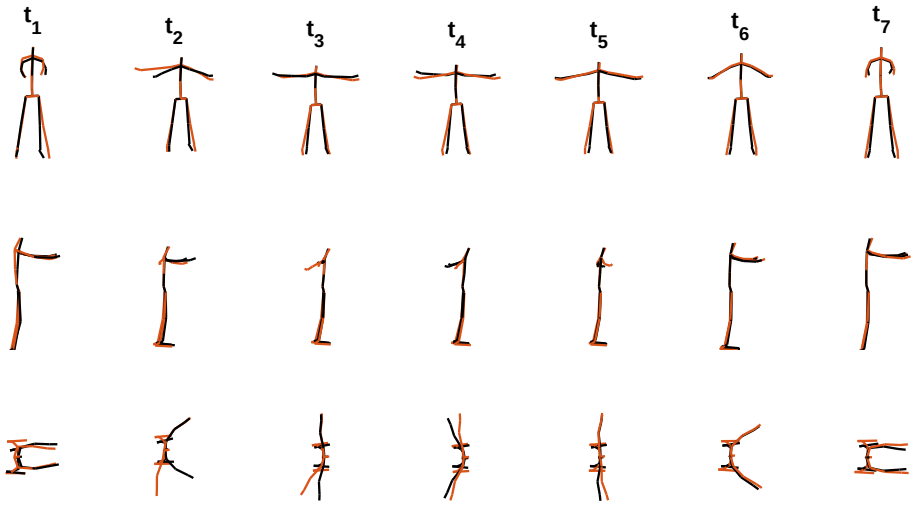
L'erreur étant de la taille du chemin de déformation, elle est potentiellement différente selon le membre considéré. Une simple projection sur l'échelle de temps du mouvement nominal peut cependant permettre d'obtenir des erreurs de même taille. Cela n'a pas été fait ici de manière à conserver un maximum d'information possible.

4.4.2 Erreurs temporelles

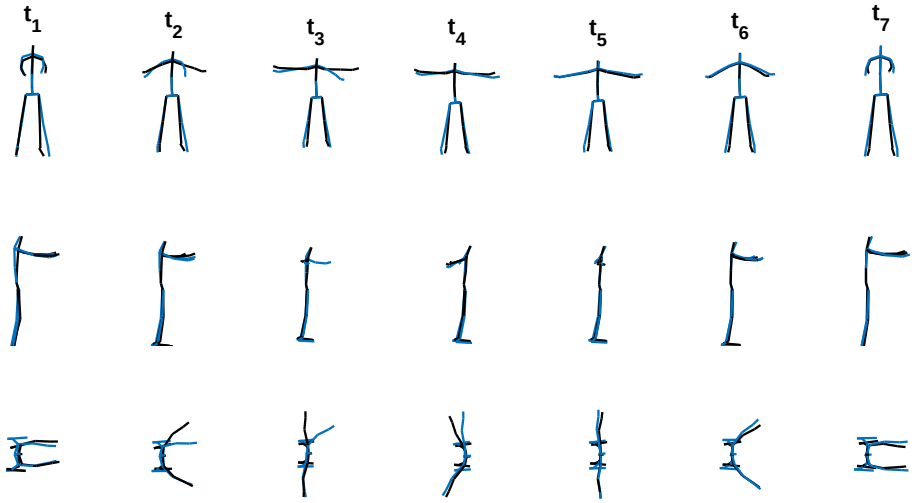
Soient m_1 et m_2 deux membres dont on veut estimer le décalage temporel.

- $\phi_{\boldsymbol{X}_l^{m_1}\boldsymbol{X}_n^{m_1}}(k)$ est le chemin de déformation induit de l'alignement du membre m_1 entre les mouvements novice et nominal.
- $\phi_{\boldsymbol{X}_l^{m_2}\boldsymbol{X}_n^{m_2}}(k)$ est le chemin de déformation induit de l'alignement du membre m_2 entre les mouvements novice et nominal.

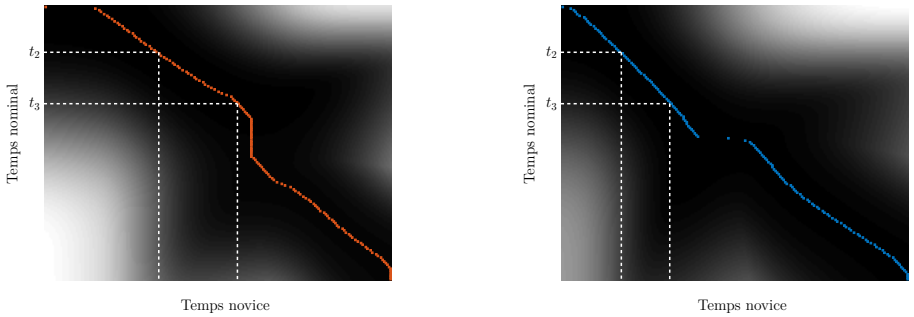
Si m_1 et m_2 sont bien coordonnés, alors le chemin de déformation du membre m_1 est le même que celui du membre m_2 : $\phi_{\boldsymbol{X}_l^{m_1}\boldsymbol{X}_n^{m_1}}(k) = \phi_{\boldsymbol{X}_l^{m_2}\boldsymbol{X}_n^{m_2}}(k) \quad \forall k \in \{1...K\}$. On définit donc le décalage temporel de m_1 relativement à m_2 à l'instant k comme le décalage entre $\phi_{\boldsymbol{X}_l^{m_1}\boldsymbol{X}_n^{m_1}}(k)$ et $\phi_{\boldsymbol{X}_l^{m_2}\boldsymbol{X}_n^{m_2}}(k)$. La figure 4.8 illustre le cas de deux chemins de déformations distincts, pour deux membres différents considérés m_1 et m_2. Afin de bien comprendre le raisonnement, les mouvements considérés sont des mouvements simples d'ouverture et de fermeture de bras. Le squelette noir est considéré comme expert (ses bras sont synchrones), on suppose qu'il s'agit du mouvement nominal. Le squelette coloré correspond au mouvement novice. Dans le premier cas (figure 4.8a), le recalage de tout le corps est basé sur le bras gauche. De fait, le bras gauche est bien recalé à tout instant, mais pas le bras droit, qui est en avance. À l'inverse dans le second cas (figure 4.8b), c'est le bras droit qui impose son recalage, celui-ci est alors bien recalé à tout instant. La dernière figure (4.8c) représente les chemins de déformation ainsi engendrés. En reportant les instants t_2 et t_3 sur cette dernière, on fait apparaître deux temporalités bien distinctes dont découlent l'erreur temporelle.



(a) Recalage d'un mouvement de bras par rapport au bras gauche.



(b) Recalage d'un mouvement de bras par rapport au bras droit.



(c) Distances cumulées et chemins de déformations engendrés par le racalge du bras gauche (à gauche) et par le bras droit (à droite).

FIGURE 4.8 – Mise en place de l'erreur temporelle entre deux membres m_1 et m_2 (ici deux bras) à partir des chemins de déformation qu'ils engendrent. (a) Recalage de deux squelettes basé sur le bras gauche (m_1) uniquement. (b) Recalage de deux squelettes basé sur le bras droit (m_2) uniquement. (c) Chemin de déformations engendrés.

Compte tenu que le processus d'obtention de ces deux chemins ne garantit pas une même longueur K entre $\phi_{\boldsymbol{X}_l^{m_1}\boldsymbol{X}_n^{m_1}}(k)$ et $\phi_{\boldsymbol{X}_l^{m_2}\boldsymbol{X}_n^{m_2}}(k)$, chacun de ces deux chemins est projeté sur l'échelle temporelle du mouvement nominal, de sorte qu'ils deviennent tous deux de la taille du mouvement nominal, M_n. L'erreur temporelle peut alors être estimée par :

$$E_l^{m_1,m_2}(k) = \phi^{\boldsymbol{X}_l^{m_1}}_{\boldsymbol{X}_l^{m_1}\boldsymbol{X}_n^{m_1}}(k) - \phi^{\boldsymbol{X}_l^{m_2}}_{\boldsymbol{X}_l^{m_2}\boldsymbol{X}_n^{m_2}}(k) \quad \forall k \in \{1...M_n\} \tag{4.8}$$

Néanmoins, ce décalage temporel n'est pas significatif dès lors que les membres ne sont pas mobiles. De fait, pour les membres statiques ou quasi statiques, l'alignement issu du DTW n'est pas fiable, les résultats ne sont donc pas significatifs. On introduit donc un coefficient de corrélation $\gamma^{m_1,m_2}(k)$ qui donne d'autant plus d'influence à un décalage temporel que le membre considéré est mobile. Le calcul de l'erreur temporelle se fait alors par :

$$E_l^{m_1,m_2}(k) = \gamma^{m_1,m_2}(k) \times \left(\phi^{\boldsymbol{X}_l^{m_1}}_{\boldsymbol{X}_l^{m_1}\boldsymbol{X}_n^{m_1}}(k) - \phi^{\boldsymbol{X}_l^{m_2}}_{\boldsymbol{X}_l^{m_2}\boldsymbol{X}_n^{m_2}}(k)\right) \quad \forall k \in \{1...M_n\} \tag{4.9}$$

avec :

$$\gamma^{m_1,m_2}(k) = \frac{\max(\|\dot{\boldsymbol{X}}_n^{m_1}(k)\|, \|\dot{\boldsymbol{X}}_n^{m_2}(k)\|)}{\sum_{k=1}^{K} \max(\|\dot{\boldsymbol{X}}_n^{m_1}(k)\|, \|\dot{\boldsymbol{X}}_n^{m_2}(k)\|)} \quad \forall k \in \{1...M_n\} \tag{4.10}$$

où $\|\dot{\boldsymbol{X}}(k)\|$ est l'amplitude de la vitesse moyenne de $\boldsymbol{X}(k)$.

4.5 Méthodolologie

Afin de valider les outils mis en place, nous allons effectuer plusieurs tests à partir d'une vérité terrain issue d'annotations d'entraîneurs. Dans un premier temps, nous validerons l'alignement et la modélisation du jeu de gestes experts *via* une reconnaissance des phases du service au tennis. Par la suite, nous validerons à la fois l'évaluation spatiale et l'évaluation temporelle sur les gestes de tennis et de karaté respectivement.

4.5.1 Annotations

Il est difficile de valider une évaluation puisqu'il n'existe pas de vérité terrain objective permettant de connaître la qualité d'un geste. Afin de valider notre système, nous proposons donc de comparer les résultats de nos évaluations avec les annotations d'entraîneurs. À cette fin, un entraîneur de tennis et un de karaté ont été sollicités afin d'estimer les qualités de certains gestes sportifs.

L'entraîneur de tennis a ainsi annoté 9 mouvements d'experts et 8 de novices, correspondant à 17 personnes distinctes. En amont a été subdivisé chacun des 17 mouvements

de tennis en 4 phases bien connues du service de tennis. Ces phases sont définies à partir d'instants-clé conformément à [137] : à $T_{Ann}(1)$, la tête de raquette est à sa plus haute position, à $T_{Ann}(2)$, la raquette est à sa plus basse position derrière le dos et à $T_{Ann}(3)$ a lieu l'impact entre la raquette et la balle. Ces instants-clé sont illustrés en figure 4.10. On notera $T_{Ann,l}(p)$ l'annotation du $p^{\text{ième}}$ instant clé ($\forall p \in \{1...3\}$) du $l^{\text{ième}}$ mouvement expert annoté.

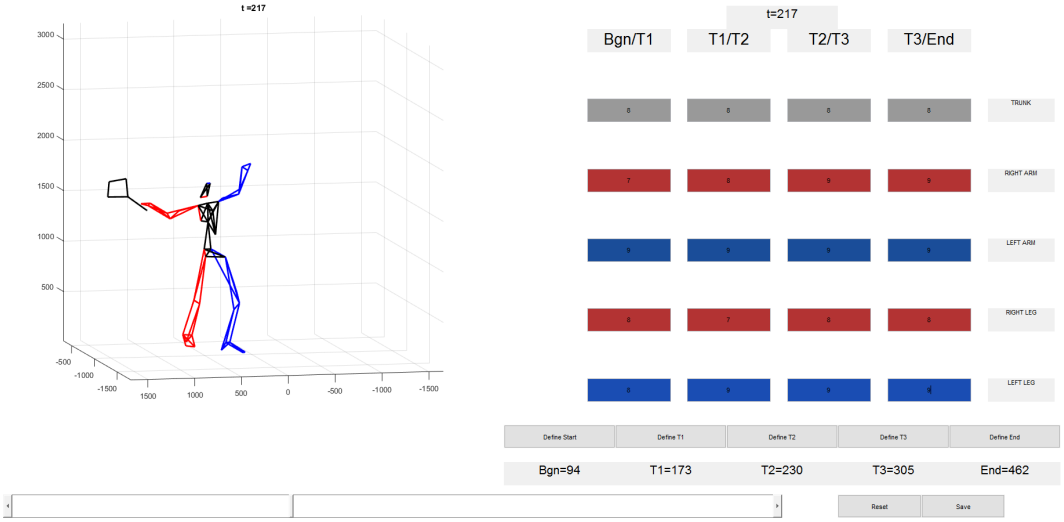


FIGURE 4.9 – Outil d'annotation spatiale pour le service de tennis.

Ensuite, l'entraîneur a estimé 20 scores, un par phase $p \in 1...4$ et par membre $m \in 1...5$ (le tronc, le bras droit, le bras gauche, la jambe droite, la jambe gauche). Chaque score correspond à une note comprise entre 0 et 10. Afin de faciliter la tâche d'annotation, nous avons mis en place un outil, permettant à l'utilisateur de décomposer le mouvement du sujet au cours du temps grâce à un curseur temporel, et ce selon différentes vues. Un code de couleur bleu-rouge permet d'éviter les confusions droite-gauche lors de l'annotation. À la demande de l'entraîneur, une raquette a été ajoutée. La figure 4.9 montre cet outil d'annotation. Il est composé d'une fenêtre de visualisation du mouvement, dont la vue peut changer par simple glissement de la souris sur la figure. La barre de défilement en bas permet d'avancer ou de reculer dans le temps. À droite de l'écran, 20 cases permettent de définir 20 scores : un pour chaque membre et chaque phase du mouvement. Enfin, 5 boutons en bas à droite permettent de définir les 5 instants segmentant les phases du geste sportif considéré (le début du mouvement, $T_{Ann}(1)$, $T_{Ann}(2)$, $T_{Ann}(3)$ et la fin du mouvement).

Comme la synchronisation des membres est nécessaire pour parvenir à frapper la balle, et comme l'entraîneur de tennis a fait part de sa difficulté à estimer la synchronie entre les membres, l'estimation de la qualité du geste de service de tennis se fera uniquement sur les mesures spatiales.

Un entraîneur de karaté a accepté d'annoter les mouvements de *Zuki*, comme ça se

$T_{Ann}(1)$ $T_{Ann}(2)$ $T_{Ann}(3)$

FIGURE 4.10 – Instants-clé subdivisant le service de tennis en quatre phases.

fait lors de compétitions. Cette fois-ci, les annotations ont été uniquement focalisées sur les deux bras du sujet. L'annotation spatiale concerne donc le bon positionnement des deux bras, tandis que l'annotation temporelle vise à vérifier que les deux bras suivent un bon agencement temporel. Comme le mouvement considéré est plus élémentaire que le service de tennis, une seule phase a été considérée. Comme précédemment, le score spatial de chaque membre est compris entre 0 et 10 et un essai par sujet a été noté. Le score temporel entre les 2 bras m_1 et m_2 est estimé entre -10 (retard important de m_1 sur m_2) et 10 (avance importante de m_1 sur m_2) et tous les essais ont été annotés.

Le tableau 4.1 récapitule l'ensemble des annotations recueillies.

	Tennis	**Karaté**
Spatial	17 mvts : 9 experts et 8 novices $m = 1...5$, $p = 1...4$	12 mvts : 4 experts et 8 novices $m = 1...2$, $p = 1$
Temporel	$\emptyset$	95 mvts (tous) $m = 1...2$, $p = 1$

TABLE 4.1 – Récapitulatif des annotations faites par les entraîneurs. m représente les membres évalués et p différentes phases constituant le geste.

4.5.2 Procédure d'évaluation

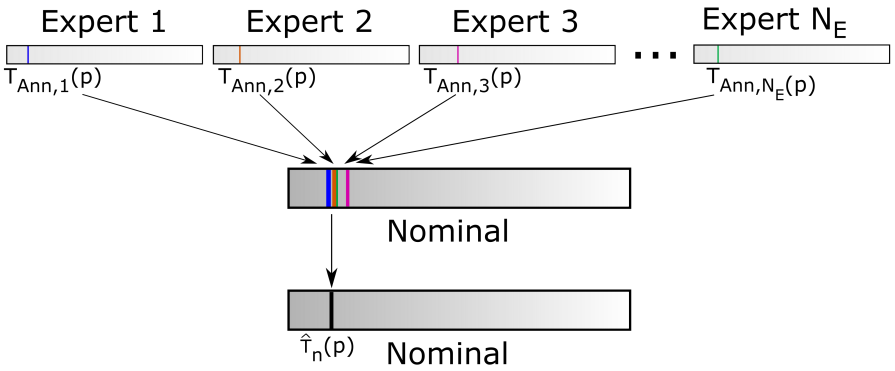
Pour chaque mouvement annoté à classifier, la modélisation des gestes est faite sur l'ensemble des essais réalisés par les experts, à condition que le mouvement à classifier ne soit pas celui d'un expert. Dans le cas contraire, la modélisation est faite sur l'ensemble des gestes experts, excepté ceux réalisés par l'expert dont on considère le mouvement. On parle alors de procédure *leave-one-subject-out*. Ce protocole est utilisé pour le tennis comme pour le karaté.

4.6 Résultats

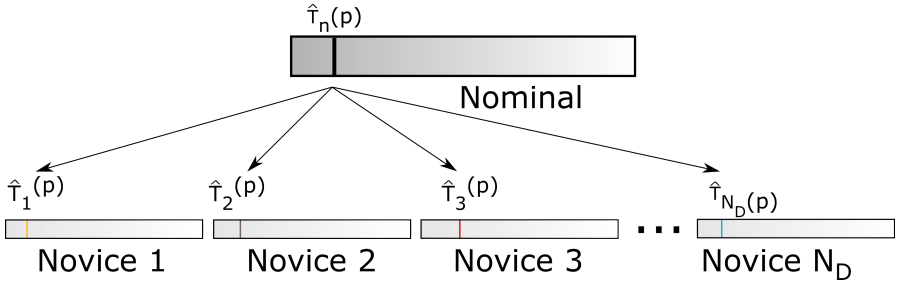
4.6.1 Reconnaissance de phases

L'objectif est d'estimer $\hat{T}_j(p) \quad \forall p \in \{1...3\}$ définissant les phases du mouvement novice j uniquement à partir des essais experts annotés.

Cette procédure permet de valider l'alignement réalisé ainsi que la modélisation du jeu de gestes experts par un mouvement nominal. À cette fin, nous estimons dans un premier temps les phases de l'essai nominal (figure 4.11a) à partir des annotations des experts, puis dans un second temps nous utilisons ces phases (dîtes nominales) pour estimer les phases des mouvements novices (figure 4.11b) avant de les comparer à la vérité terrain.



(a) Première étape : estimation des phases nominales



(b) Deuxième étape : estimation des phases d'un mouvement novice

FIGURE 4.11 – Détection des phases d'un mouvement.

Dans un premier temps, l'alignement par DTW du mouvement expert l sur le mouvement nominal donne un chemin de déformation $\phi_{\boldsymbol{X}_l\boldsymbol{X}_n}(k) = (\phi^{\boldsymbol{X}_l}_{\boldsymbol{X}_l\boldsymbol{X}_n}(k), \phi^{\boldsymbol{X}_n}_{\boldsymbol{X}_l\boldsymbol{X}_n}(k)), k \in \{1...K\}$. On peut alors transférer chacune des annotations de l'expert des phases du geste l sur le geste nominal :

$$\hat{T}_{n,l}(p) = \phi^{\boldsymbol{X}_n}_{\boldsymbol{X}_l\boldsymbol{X}_n}(\tau) \quad \text{t.q.} \quad \phi^{\boldsymbol{X}_l}_{\boldsymbol{X}_l\boldsymbol{X}_n}(\tau) = T_{Ann,l}(p) \tag{4.11}$$

$\hat{T}_{n,l}(p)$ correspond simplement à l'instant du mouvement nominal apparié par DTW à l'instant $T_{Ann,l}(p)$ du mouvement de l'expert annoté. Potentiellement et selon le DTW utilisé, plusieurs points peuvent être reliés à $T_{Ann,l}(p)$. Dans ce cas, la moyenne des instants sera prise en compte :

$$\hat{T}_{n,l}(p) =< \phi^{\boldsymbol{X}_n}_{\boldsymbol{X}_l\boldsymbol{X}_n}(\tau_i) >_{i=1..I} \quad \text{t.q.} \quad \phi^{\boldsymbol{X}_l}_{\boldsymbol{X}_l\boldsymbol{X}_n}(\tau_i) = T_{Ann,l}(p) \quad \forall i \in \{1...I\} \tag{4.12}$$

Les phases nominales sont alors données comme la moyenne des toutes les estimations issues des N_E annotations de mouvements experts :

$$\hat{T}_n(p) = \frac{1}{N_E}\sum_{l=1}^{N_E}\hat{T}_{n,l}(p) \tag{4.13}$$

Les phases des novices peuvent alors être déterminées automatiquement par projection inverse des phases nominales que l'on vient d'estimer, sur les mouvements novices. Pour un geste novice j donné, on aligne son mouvement $\boldsymbol{X}_j$ sur le mouvement nominal. Ce nouvel alignement donne le chemin de déformation $\phi_{\boldsymbol{X}_j\boldsymbol{X}_n}(k) = (\phi^{\boldsymbol{X}_j}_{\boldsymbol{X}_j\boldsymbol{X}_n}(k), \phi^{\boldsymbol{X}_n}_{\boldsymbol{X}_j\boldsymbol{X}_n}(k))$, $k \in \{1...K'\}$. On estimera alors $\hat{T}_j(p)$ par :

$$\hat{T}_j(p) =< \phi^{\boldsymbol{X}_j}_{\boldsymbol{X}_j\boldsymbol{X}_n}(\tau_i) >_{i=1..I} \quad \text{t.q.} \quad \phi^{\boldsymbol{X}_n}_{\boldsymbol{X}_j\boldsymbol{X}_n}(\tau_i) = \hat{T}_n(p) \quad \forall i \in \{1...I\} \tag{4.14}$$

La figure 4.12 montre les résultats obtenus par ce processus et les annotations de l'entraîneur, pour chacun des 8 novices dont les phases ont été annotées. C'est le CDTW avec $K_p = K_{pm} + 1$ qui a été choisi pour la comparaison. Des résultats similaires ont également été obtenus par CDTW avec différents K_p, ainsi que par DTW classique. Les estimations sont très proches de la vérité terrain fournie par les annotations, même lorsque les gestes sont mauvais (comme nous allons le voir dans la section suivante). Le premier instant clé T_1 est le moins bien estimé (comparativement à l'annotation). Cela est dû au fait que c'est un instant fortement identifié par le positionnement de la raquette, information disponible à l'annotateur mais inconnue de l'algorithme (la raquette ne fait pas partie des entrées du système, qui doit rester générique à n'importe quel sport).

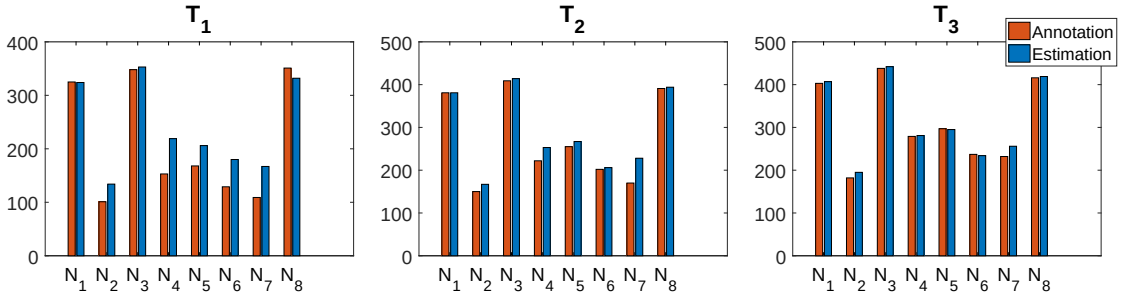


FIGURE 4.12 – Comparaison des estimations instants clés estimés par notre approche avec les annotations de l'entraîneur.

4.6.2 Evaluation spatiale de la qualité d'un geste sportif

Le processus d'évaluation permet d'extraire les erreurs spatiales de chaque membre et ce à chaque instant. Soit $AS_l^m(p)$ l'annotation spatiale donnée par l'entraîneur de la phase p du membre m du mouvement l. Selon les instructions données à l'annotateur, $AS_l^m(p)$ est compris 0 et 10, et est d'autant plus élevé que le geste du membre est correct. Cette annotation peut être comparée à la moyenne de l'erreur spatiale du membre p du mouvement l lors de la phase correspondante :

$$SS_l^m(p) = \frac{1}{|T_p|} \sum_{k \in T_p} E_l^m(k) \tag{4.15}$$

où T_p correspond à la durée totale de la phase p, en nombre d'instants, reportée dans l'échelle temporelle du chemin de déformation et $E_l^m(k)$ est calculée selon l'équation 4.7.

Afin de présenter des résultats plus synthétiques, les scores ont été moyennés sur l'ensemble des M membres annotés ($\frac{1}{M} \sum_{m=1}^{M} AS_l^m(p)$ et $\frac{1}{M} \sum_{m=1}^{M} SS_l^m(p)$). Quatre mesures spatiales sont donc obtenues pour chaque mouvement, une par phase. Les détails des erreurs sont conservés et pourront être utilisés pour davantage de précision.

Les résultats sont illustrés en figure 4.13 (service de tennis) et en figure 4.14 (*Zuki*). La modélisation des mouvements experts choisie est le CDBA (section 3.2) et l'alignement du novice sur le mouvement nominal est fait par CDTW, les deux avec $K_p = K_{pm} + 1$. Les erreurs estimées, qui en théorie peuvent aller de 0 à $+\infty$, sont limitées entre 0 et 5 tandis que les scores annotés varient entre 0 et 10. Les résultats sont concluants puisque le score annoté se révèle d'autant plus grand que l'erreur estimée est faible.

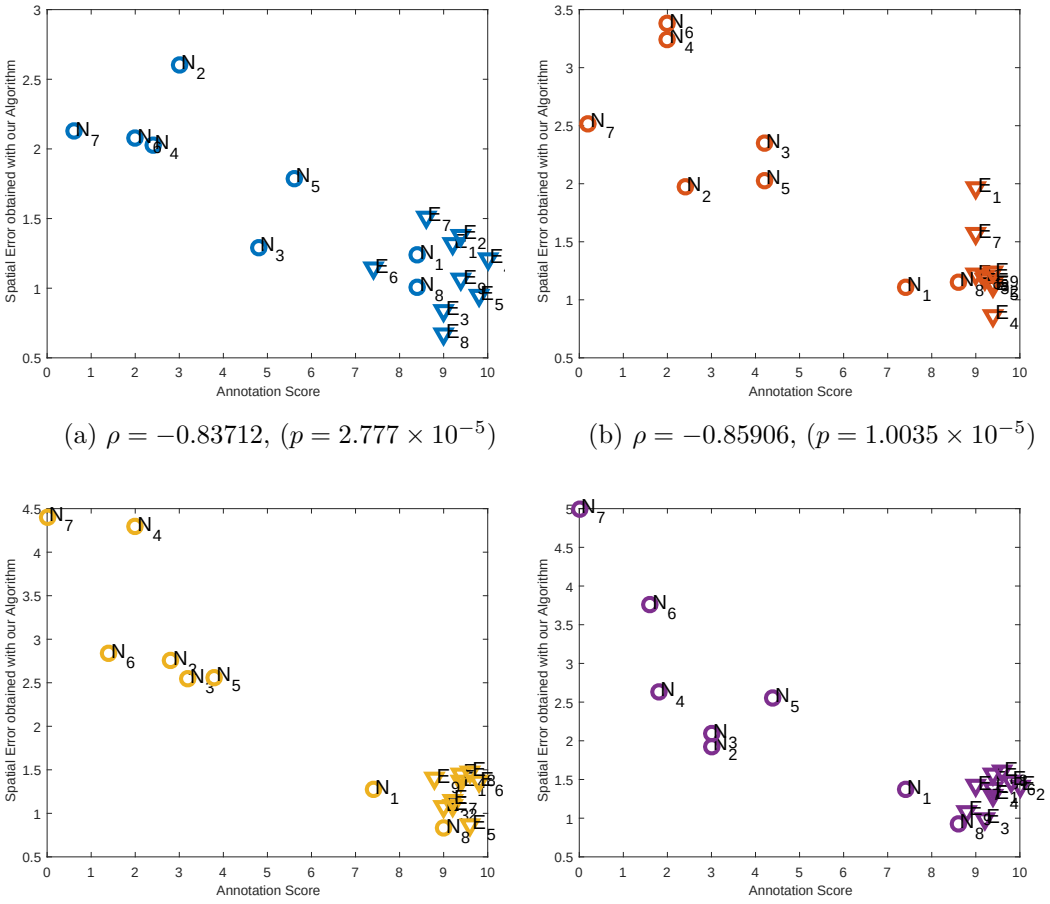
Afin d'obtenir une mesure quantitative de la qualité de l'estimation, la corrélation a été calculée pour chaque phase entre les estimations et les annotations. Les coefficients de corrélation sont très forts puisqu'ils varient entre -0.83 et -0.92.

Afin d'analyser la robustesse et la performance de notre méthode, les résultats sont maintenant établis en utilisant différentes méthodes d'alignement. Afin de synthétiser nos résultats, nous estimons les erreurs globales, c'est-à-dire moyennées à la fois sur les M membres annotés et sur les P phases *i.e.* $\frac{1}{M}\frac{1}{P} \sum_{p=1}^{P} \sum_{m=1}^{M} AS_l^m(p)$ et $\frac{1}{M} \sum_{m=1}^{M} SS_l^m$, avec :

$$SS_l^m = \frac{1}{K} \sum_{k=1}^{K} E_l^m(k) \tag{4.16}$$

Les différentes méthodes d'alignement suivantes sont comparées :

- DTW (et DBA pour la modélisation) sans tolérance
- CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 1$ et sans tolérance
- CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 2$ et sans tolérance



(a) $\rho = -0.83712$, $(p = 2.777 \times 10^{-5})$ (b) $\rho = -0.85906$, $(p = 1.0035 \times 10^{-5})$

(c) $\rho = -0.92573$, $(p = 1.0034 \times 10^{-7})$ (d) $\rho = -0.84995$, $(p = 1.5608 \times 10^{-5})$

FIGURE 4.13 – Comparaison des scores spatiaux avec les annotations de l'entraîneur pour chaque phase du service de tennis. L'alignement choisi est le CDTW avec $K_p = K_{pm} + 1$. Le coefficient de corrélation entre les annotations (axe des x) et les estimations (axe des y) est donné en légende.

- ✳ CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 3$ et sans tolérance
- ✳ CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 4$ et sans tolérance
- ✳ DTW (et DBA pour la modélisation) avec tolérance
- ✳ CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 1$ et avec tolérance
- ✳ CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 2$ et avec tolérance
- ✳ CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 3$ et avec tolérance
- ✳ CDTW (et CDBA pour la modélisation) avec $K_p = K_{pm} + 4$ et avec tolérance

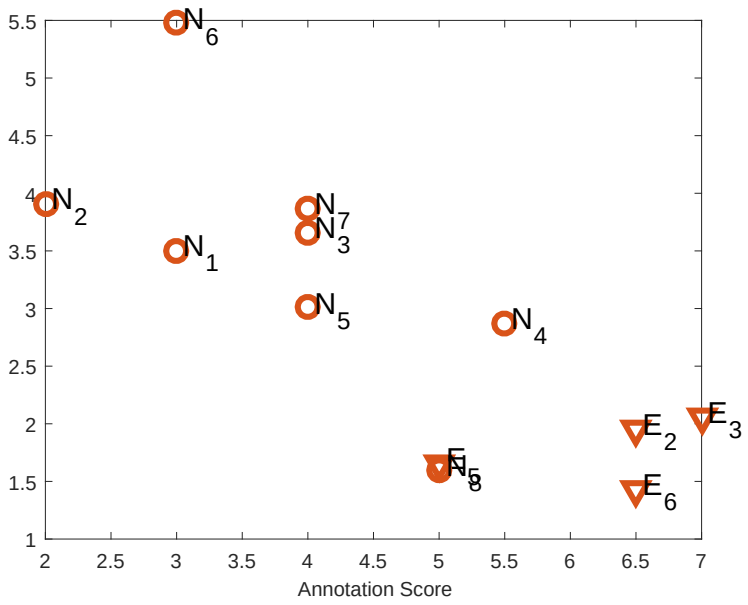


FIGURE 4.14 – Comparaison des évaluations spatiales avec les annotations de l'entraîneur de karaté pour le *Zuki*. L'alignement choisi est le CDTW avec $K_p = K_{pm} + 1$. Le coefficient de corrélation entre les estimations et les annotations est $\rho = -0.77774$ ($p = 0.0029$).

Les résultats de cette comparaison pour le service de tennis sont présentés tableau 4.2, et pour le karaté tableau 4.3. Le CDTW avec tolérance et $K_p = K_{pm} + 1$ semble le plus efficace au regard des annotations. L'utilisation d'un DTW contraint affine la corrélation de l'algorithme avec les annotations (ρ se rapproche de 1 en valeur absolue). On remarque également que quelle que soit la configuration choisie, l'ajout de la tolérance améliore toujours les résultats. L'amélioration est faible pour le tennis mais plus forte pour le karaté, ce qui nous conforte dans l'intérêt de l'utilisation de la tolérance et des contraintes.

4.6.3 Evaluation temporelle de la qualité d'un geste sportif

Étant donné que le service de tennis est très contraint temporellement par l'impact de la balle, mais aussi parce que les entraîneurs de tennis évaluent principalement les postures et non les synchronies entre les membres, l'évaluation temporelle est uniquement faite sur la base de gestes de *Zuki*.

Comme déjà évoqué, le geste de karaté étudié demande une grande rigueur dans la temporalité entre les bras. En fait, il existe deux coordinations principales : celle des bras, et celle de la rotation des poignets. Dans cette étude, nous nous focalisons uniquement sur la coordination des bras étant donné que la rotation angulaire des poignets est une donnée perdue lors de l'extraction du squelette à partir des marqueurs.

	Sans tolérance				
	DTW	CDTW			
		$K_{pm}+1$	$K_{pm}+2$	$K_{pm}+3$	$K_{pm}+4$
ρ	-0.94076 ($p = 1.9 \times 10^{-8}$)	-0.9609 ($p = 9.0 \times 10^{-10}$)	-0.95926 ($p = 1.2 \times 10^{-9}$)	-0.95637 ($p = 2.0 \times 10^{-9}$)	-0.95432 ($p = 2.9 \times 10^{-9}$)

	Avec tolérance				
	DTW	CDTW			
		$K_{pm}+1$	$K_{pm}+2$	$K_{pm}+3$	$K_{pm}+4$
ρ	-0.94327 ($p = 1.4 \times 10^{-8}$)	-0.96094 ($p = 9.0 \times 10^{-10}$)	-0.96031 ($p = 1.0 \times 10^{-9}$)	-0.95936 ($p = 1.2 \times 10^{-9}$)	-0.95771 ($p = 1.6 \times 10^{-9}$)

Table 4.2 – Validation de l'évaluation spatiale du service de tennis. Coefficient de corrélation pour les différents protocoles de classification : avec ou sans tolérance, par DTW ou CDTW avec K_p variable.

	Sans tolérance				
	DTW	CDTW			
		$K_{pm}+1$	$K_{pm}+2$	$K_{pm}+3$	$K_{pm}+4$
ρ	-0.71253 ($p = 0.0093$)	-0.74748 ($p = 0.0052$)	-0.7329 ($p = 0.0067$)	-0.72334 ($p = 0.0078$)	-0.72238 ($p = 0.0080$)

	Avec tolérance				
	DTW	CDTW			
		$K_{pm}+1$	$K_{pm}+2$	$K_{pm}+3$	$K_{pm}+4$
ρ	-0.77393 ($p = 0.0031$)	-0.77774 ($p = 0.0029$)	-0.76985 ($p = 0.0034$)	-0.76607 ($p = 0.0037$)	-0.76881 ($p = 0.0035$)

Table 4.3 – Validation de l'évaluation spatiale du *Zuki*. Coefficient de corrélation pour les différents protocoles de classification : avec ou sans tolérance, par DTW ou CDTW avec K_p variable.

Un mouvement expert et un novice sont illustrés en figure 4.15. Une comparaison entre les évaluations temporelles et les annotations est donnée en figure 4.16. Ici, c'est la configuration d'alignement par CDTW avec $K_p = K_{pm}+4$ qui a été choisie puisque c'est celle qui mène au meilleur résultat. Une forte corrélation apparaît entre les estimations et les annotations. Cependant, le sujet N_6 semble entraver la corrélation des données (les essais de N_6 sont représentés en gris sur la figure). En fait, le mouvement du sujet N_6 est très mauvais spatialement (voir figure 4.14). Le novice ne positionne pas ses bras correctement (en position ramenée, ses bras ne sont pas calés le long des hanches mais restent en position haute). De fait, (i) les recalages sont irrémédiablement mauvais, (ii) l'annotation temporelle est complexe et donc peu fiable. En pratique, il parait tout à

fait approprié d'évaluer le décalage temporel entre des membres seulement à condition que ceux-ci effectuent un minimum le bon mouvement. L'évaluation spatiale serait donc faite en amont et l'évaluation temporelle serait considérée uniquement si cette première est "acceptable".

La corrélation des données une fois le sujet N_6 enlevé du jeu de données est tout à fait correcte avec un coefficient $\rho = -0.90792$, proche de 1 en valeur absolue. Rappelons que les annotations sont établies pour tous les mouvements de la base et que le signe de l'annotation dépend du rapport du décalage (avance ou retard du bras droit sur le bras gauche). Cette annotation est d'autant plus grande en amplitude que le décalage est important. De même, l'estimation de l'erreur temporelle amène à un score signé d'autant plus grand que les membres sont fortement retardés.

Remarquons également que la synchronie des bras des experts est considérée comme parfaite puisque toutes les annotations des essais experts sont égales à 0 (pas de retard). À l'inverse, les essais novices sont annotés entre -10 et 10, recouvrant toute la plage de valeurs d'annotations admissibles.

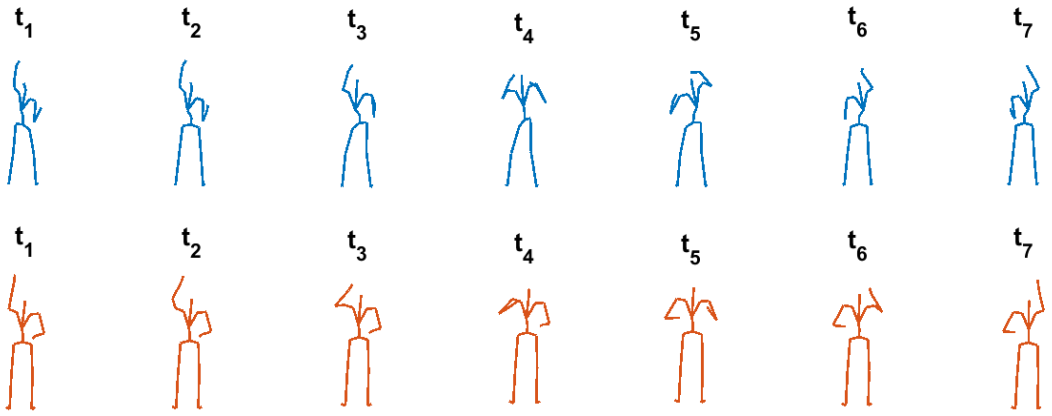


FIGURE 4.15 – Première ligne : un mouvement de *Zuki* exécuté par un expert. Seconde ligne : un mouvement de *Zuki* effectué par un novice. Cet exemple montre une temporalité parfaite des bras de l'expert comparativement au novice qui n'amorce le mouvement de son bras droit que lorsque le bras gauche est pratiquement arrivé à sa position finale (entre t_5 et t_7).

Conclusion

Dans ce chapitre, nous avons adapté les outils développés dans le chapitre précédent à des fins d'évaluation de gestes sportifs, de façon automatique et tout à fait générique. Les erreurs spatiales et temporelles ont été traitées indépendamment du type de sport considéré et de la morphologie du sujet effectuant le geste. À cette fin, les méthodes d'alignement et de modélisation mises en place durant le chapitre précédent ont été

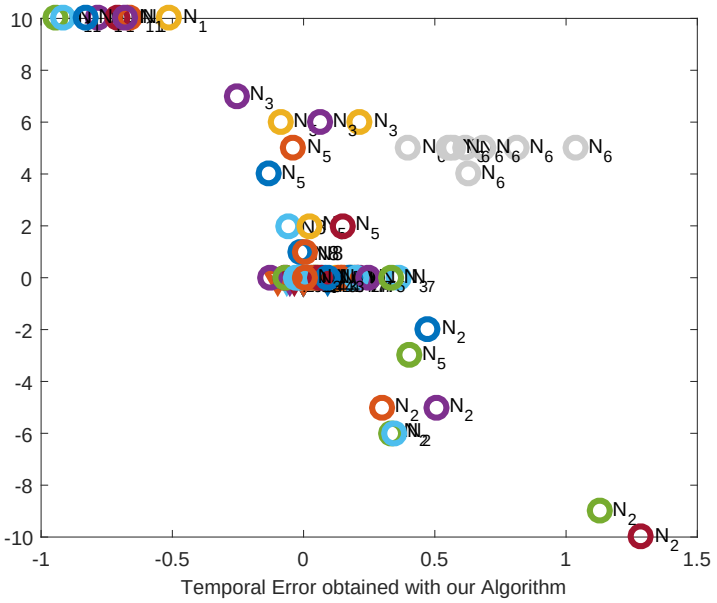


FIGURE 4.16 – Comparaison des évaluations temporelles avec les annotations de l'entraîneur de karaté concernant le geste du *Zuki*. L'alignement choisi est le CDTW avec $K_p = K_{pm} + 4$. Le coefficient de corrélation entre les estimations et les annotations vaut $\rho = -0.90792$ une fois les essais du novice N_6 enlevés (en gris).

généralisées à des signaux de grande dimension.

Les approches explicitées ont ensuite été validées sur deux bases de données acquises au laboratoire M2S, afin de souligner leur généricité : des services de tennis et des coups de poing de karaté (*Zuki*).

Certaines limitations sont tout de même à soulever. Dans un premier temps, la normalisation morphologique est faite de façon succincte en se basant uniquement sur la morphologie globale du sujet (*i.e.* son tronc), négligeant les éventuelles différences morphologiques locales (relatives aux membres). Même si ces différences sont faibles, elles pourraient légèrement modifier nos résultats.

Une autre limitation concerne la modélisation du mouvement expert par moyenne et variance, qui sous-entend de fait une distribution gaussienne des données. En réalité, il est très probable que des patrons de styles puissent être mis en évidence. De fait, un mélange de gaussiennes pourrait être plus approprié à la modélisation du geste expert, selon le sport considéré. Ce travail n'a pas été fait durant cette thèse par manque de données suffisantes. Cette mise à jour pourrait fortement modifier les résultats pour certains sports très disposés à des styles d'exécution bien distincts.

En outre, un choix a été fait de considérer des positions et non des angles. Ce choix a été motivé notamment par Sorel qui a montré que des gestes de bras étaient mieux reconnus à partir de descripteurs de positions (normalisés) plutôt que d'angles [11].

De surcroît, les techniques sportives reposent bien plus souvent sur des enchainements posturaux qu'angulaires. De fait, il paraît plus intuitif de travailler sur des positions. Il serait néanmoins intéressant dans une future étude d'appliquer la méthode développée sur des descripteurs angulaires et de confronter les différents résultats obtenus.

Maintenant que l'évaluation spatiale et temporelle de mouvements sportifs a été testée et validée sur des bases de données, il est temps d'en venir au but même de ce travail de thèse : la mise en place d'un entraîneur virtuel permettant à un sportif de pouvoir se perfectionner de manière autonome, ce en complément de séances d'entraînement classiques.

Chapitre 5

Entraîneur virtuel : vers un outil d'entraînement automatique en ligne

Introduction

L'évaluation d'un geste étant désormais mise en place, il convient maintenant de procurer à l'utilisateur qui souhaite progresser des informations pertinentes qui soient faciles à assimiler. Le but de ce chapitre est donc d'adapter l'outil d'évaluation pour qu'il soit utilisé en ligne et de proposer au sportif un retour d'informations adapté à son niveau et conforme au processus d'apprentissage moteur de l'humain. Dans un premier temps, nous allons donc faire le point sur les différents dispositifs technologiques permettant de capter et d'informer le sujet quant au mouvement qu'il réalise. Les différents choix de retours à transmettre au sportif seront ensuite discutés afin d'adopter une démarche d'entraînement la plus propice possible à la progression sportive. Dans un second temps, nous adapterons notre outil d'évaluation pour le transposer à une utilisation en ligne. Enfin, l'outil d'entraînement sera introduit et testé auprès d'un sujet novice de tennis.

5.1 État de l'art

5.1.1 Les différents systèmes d'entraînement

Un système d'entraînement interactif est un outil communicant avec l'utilisateur afin de lui fournir un retour pertinent à partir d'informations sur son geste. Récemment, l'essor de la réalité virtuelle a permis la mise en place de systèmes communicants où le sujet interagit directement avec son environnement. Les plate-formes immersives (CAVE) ou à moindre coût les casques (Oculus Rift®, HTC Vive®) offrent une expérience d'immersion très stimulante pour l'utilisateur, mais aussi particulièrement adaptée parce que les environnements ainsi créés sont standardisés, reproductibles et contrôlables.

Basé sur cette technologie, un certain nombre de systèmes d'entraînement sont développés. Après des traumatismes importants, la faculté d'équilibre d'une personne peut-être très restreinte. Certains dispositifs immersifs peuvent alors aider à la rééducation. Chez les enfants, on préférera les jeux interactifs immersifs, tels que ceux proposés par Hawkins *et al.* [140] et Barton *et al.* [141] par réalité virtuelle, dans lesquels l'enfant est plongé dans un monde virtuel à thème fantastique, dirigeant un tapis volant ou bien un dragon par simple mouvement du corps, lui-même capté à l'aide de marqueurs placés sur le bassin et le tronc de l'enfant. Chez l'adulte, d'autres dispositifs utilisant le transfert de poids du patient peuvent être mis en place de manière à proposer un outil interactif qui aide le patient à améliorer son contrôle postural. C'est le cas par exemple des travaux de Cikaljo *et al.* [31] qui ont mis en place un dispositif en aluminium et en bois permettant à des patients ayant subi un AVC une rééducation interactive en toute sécurité. Tenus par l'armature, les patients n'ont qu'à transférer leur poids sur la plate-forme afin de faire bouger un personnage virtuel sur un écran et lui faire éviter les obstacles le plus vite possible. Plusieurs tests classiques permettent ensuite d'évaluer la performance et la progression de la personne. De façon générale, tous ces différents dispositifs reposent à la fois sur l'enregistrement d'un mouvement, l'extraction d'informations pertinentes quant au geste réalisé selon l'objectif fixé, mais aussi et surtout à son aspect ludique, prenant et immersif. La revue proposée dans [142] résume tous ces travaux utilisant la réalité virtuelle à des fins de rééducation et d'évaluation de l'équilibre. En sport, la réalité virtuelle est employée pour le handball [28, 27], le baseball [143] ou encore le rugby [26, 144]. Cependant, dans ces différents travaux, la réalité virtuelle a été utilisée pour provoquer un comportement par l'athlète et l'extraire, plutôt que pour le faire s'améliorer, au contraire de Burns *et al.* qui proposent un système d'apprentissage par réalité virtuelle d'un geste de karaté [12].

Il est également possible, en considérant le même principe d'entrées-sorties que celui utilisé par la réalité virtuelle et schématisé en figure 5.1, de mettre en place d'autres systèmes interactifs.

Dans ce cadre, il convient de distinguer les entrées et les sorties de ces systèmes : les entrées permettent de capturer le geste afin qu'il puisse être analysé par notre processus d'évaluation. Les sorties fournissent à l'utilisateur un retour d'information pour lui permettre de s'améliorer.

En premier lieu, comme explicité dans le chapitre 1, il existe aujourd'hui de nombreux systèmes de captation du mouvement qui pourraient être utilisés en entrée du dispositif. Selon la finalité du système d'entrainement conçu, différents choix peuvent être adoptés conformément aux contraintes fixées (coût, encombrement, *etc.*). Dans le cadre de cette thèse, c'est le système Optitrack® qui a été retenu. Néanmoins, les méthodes utilisées resteront applicables à n'importe quel outil choisi, à l'imprécision du système près.

En second lieu, concernant la sortie du système d'entraînement, il existe une multitude d'outils permettant de transmettre un retour d'informations (ou *feedback*) sur le mouvement réalisé par une personne. Le premier retour utilisé, bien sûr, c'est le retour visuel, qui est très simple à mettre en place et très facile d'accès à une personne chez elle, avec un simple écran. C'est notamment ce qu'utilisent les technologies telles

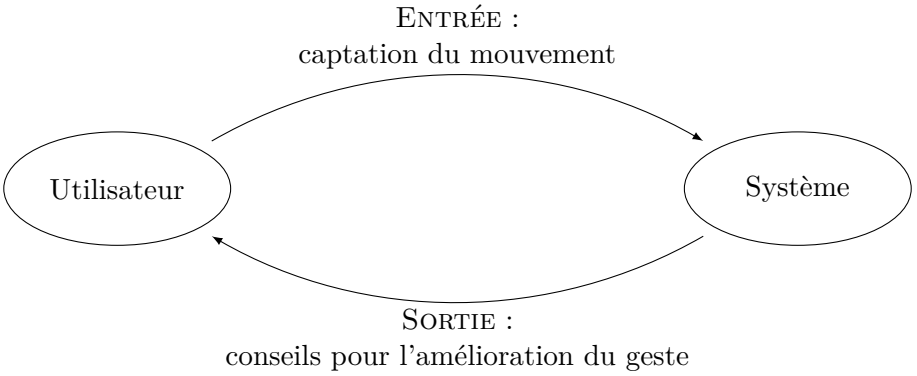


Figure 5.1 – Principe général de communication d'un système d'entrainement

que la Wii®, la XBox® ou encore les applications mobiles, qui captent le mouvement soit à l'aide d'accéléromètres (la Wii® avec la télécommande et beaucoup d'applications mobiles puisque les smartphones actuels contiennent une multitude de capteurs de mouvement), soit par carte de profondeur (la Kinect® utilisée avec la XBox®), soit par plate-forme de pression (la Balance-board de la Wii®), *etc.* Dans la majorité des cas, les retours transmis à l'utilisateur sont directement intégrés dans des jeux à visée simplement ludique mais parfois aussi éducative ou rééducative comme c'est le cas de nombreux jeux d'entraînement (ou *exergames*) dont la revue [138] dresse une vue d'ensemble. Les systèmes à stéréovision étendent cette technique de retour visuel à la vision $3D$ pour plus d'immersion du sujet.

Les retours haptiques peuvent également être choisis, mais ont des contraintes plus fortes de mise en place. De fait, ils sont moins répandus et très spécifiques à la tâche choisie.

Le retour audio est également appréciable, mais est globalement moins bénéfique que les autres dispositifs. Il peut cependant être utilisé de façon additionnelle, et a l'avantage d'être très peu coûteux. La revue proposée dans [139] résume l'ensemble de ces technologies de retour d'information dans le cas de mouvements.

Enfin, des systèmes multimodaux peuvent être appréhendés [145, 146]. Ils permettent d'augmenter le spectre des retours sensoriels, mais peuvent amener de nouveaux problèmes d'incohérence et ou de conflit entre les différentes entrées.

5.1.2 Retours d'information

Apprendre des compétences motrices est un processus complexe rendant compte de nombre de difficultés cognitives. La principale difficulté consiste à extraire les informations pertinentes de l'environnement d'apprentissage. Cette tâche est bien sûr d'autant plus complexe qu'elle implique plusieurs membres, que leurs vitesses sont élevées, que leurs positionnements sont précis, que leurs synchronies sont rigoureuses, *etc.*

La perception de l'espace chez l'humain n'obéit pas aux règles de la géométrie eucli-

dienne. De plus, pour percevoir l'espace et agir en conséquence, le cerveau doit jongler avec de nombreux référentiels liés à chacun de ses capteurs. En fonction de la tâche à exécuter, différents référentiels vont être activés. Lorsqu'il s'agit de donner le retour d'informations d'une de ses actions à un sportif, l'enjeu est donc très complexe puisqu'il s'agit de lui faire assimiler un déplacement de son corps dans l'espace avec des moyens externes réduits : la parole, la visualisation et/ou le retour haptique.

Ce retour d'informations peut être donné pendant l'exécution de la tâche motrice (on parlera de *parallel feedback*), ou après (on parlera de *terminal feedback*). Le *parallel feedback* est controversé, il apparaît que son utilisation mène à sa dépendance ; en d'autres termes, il est bénéfique lorsqu'il est utilisé mais ce bénéfice est perdu lorsqu'il n'y a plus de retour d'informations [147]. Ceci s'explique notamment par le fait qu'un *parallel feedback* peut modifier la tâche : la stratégie de contrôle est perturbée par la concentration du sujet sur le retour donné, de sorte que le contrôle moteur est différent de celui sans retour d'informations [148]. Il apparaît cependant sur certaines études que cet effet néfaste ne le soit plus lorsque les mouvements sont complexes, par exemple les mouvements sportifs, comme le précise la revue de littérature proposée dans [149]. Dans ce cas, c'est l'automatisme de l'action répétée qui prévaut et permet à une personne d'assimiler le geste. D'un autre côté, l'auto-évaluation est également très bénéfique puisqu'elle accroît les capacités de détection d'erreur par soi-même [150]. Alors qu'elle est possible lors du *terminal feedback* (juste avant que le retour d'informations n'apparaisse), elle ne l'est pas s'il est parallèle.

Il a été montré qu'en début d'apprentissage, les indications parallèles ou terminales très fréquentes étaient bénéfiques. À l'inverse, lorsque l'apprenant a une idée déjà bien précise du mouvement (la première phase d'apprentissage a été assimilée), il profitera davantage du retour si celui-ci est plus rare. La fréquence du retour d'informations doit donc décroître avec le niveau de l'apprenant.

Cette diminution de fréquence du retour a été testée relativement au temps et à la performance. Il apparaît que la raréfaction du retour relativement à la performance du sujet est pertinente, cependant les seuils de niveau sont bien sûr difficiles à fixer [151]. Une alternative est de laisser au sujet le contrôle du retour quand il le souhaite.

De part sa facilité de mise en place, c'est le retour visuel qui sera choisi au cours de cette thèse, néanmoins plusieurs types de retours visuels peuvent être mis en place. De façon générale, il conviendra d'opter pour un retour d'informations ni délétère (perturbant ou incompréhensible par l'athlète), ni addictif (si celui-ci est trop intrusif, alors son absence fera perdre à l'apprenant toute la technique qu'il pensait avoir acquis).

Un retour abstrait reposant sur des courbes est souvent préféré lors d'un mouvement simple [149]. Il convient alors de conserver les conventions colorimétriques : le rouge évoque ce qui est faux, le vert ce qui est correct. Lorsque les mouvements sont plus complexes (comme c'est le cas pour l'apprentissage de gestes sportifs), une représentation abstraite ne peut être conservée. On lui préfère alors les visualisations naturelles : soit par affichage côte-à-côte de la référence et du sujet, soit par superposition (on force alors l'apprentissage par imitation).

Les différents travaux de la littérature à ce sujet suggèrent que l'efficacité de la

méthode par superposition dépend de la quantité des membres considérés : trop de membres superposés peuvent surcharger l'apprenant d'informations qui en deviennent alors illisibles. C'est le cas par exemple de Chua *et al.* qui ont obtenu de meilleurs résultats lorsque l'entraîneur virtuel était positionné en face ou à côté de l'apprenant plutôt que superposé à lui, dans le cas du Tai Chi [152].

Dans cette thèse, nous choisissons d'opter pour un retour visuel terminal par superposition. Afin de ne pas surcharger d'informations l'apprenant, chaque erreur spatiale d'un membre sera visualisée indépendamment du reste du corps grâce à un codage colorimétrique. Enfin, il serait pertinent d'adapter le retour au niveau et à la progression de l'athlète. Ceci ne sera pas appliqué dans ce premier prototype, qui sera cependant rendu adapatable et personnalisable en perspective d'un dispositif abouti à destination du sportif.

Avant de mettre en place ce retour d'informations, il convient d'adapter au temps réel le processus d'évaluation mis en place pour qu'il soit utilisable en ligne, afin de permettre à un novice d'avoir un retour immédiat et automatique sur le geste qu'il vient d'exécuter.

5.2 Transposition à un système en ligne

Afin de transférer notre système au temps réel, il convient de l'adapter en conséquence. Rappelons dans un premier temps la construction générale du dispositif. Notre méthode se décompose en deux étapes distinctes : la première de modélisation du geste expert qui est faite en amont de l'évaluation à partir du jeu de mouvements experts ; et la seconde qui fonctionne au fur et à mesure que le novice exécute son geste. C'est donc l'alignement du geste novice sur le geste nominal qui doit être fait en temps-réel. Afin d'obtenir un geste novice segmenté, certains travaux ont recours à des subterfuges simplificateurs. Par exemple, l'utilisation d'un squelette fantôme permet de commencer la captation uniquement lorsque le squelette du novice est superposé au squelette fantôme [121]. D'autres détectent des postures neutres en amont et en aval du geste [6]. Bien sûr, ces méthodes sont contraignantes pour le novice puisqu'elles le forcent à adopter un certain positionnement. Dès lors, nous proposons de nous affranchir de cette contrainte, et de permettre au novice de commencer son geste quand il le souhaite et de le terminer plus ou moins rapidement, sans contrainte avant ou après son mouvement.
Le processus d'évaluation élaboré précédemment reste inchangé : une fois le mouvement segmenté et aligné, les calculs d'erreurs spatiale et temporelle peuvent être appliqués sur le geste, en temps-réel. Un interfaçage adapté doit alors pouvoir rendre compte au novice des erreurs principales de son mouvement, comme nous allons le voir par la suite. La figure 5.2 résume le processus général d'entraînement du novice.

Résumons les différents enjeux du transfert de notre méthode en temps-réel en ces quelques points :

- Le geste novice à acquérir n'est pas segmenté, puisque la personne qui veut s'évaluer doit pouvoir commencer son geste quand elle le souhaite et le terminer plus ou

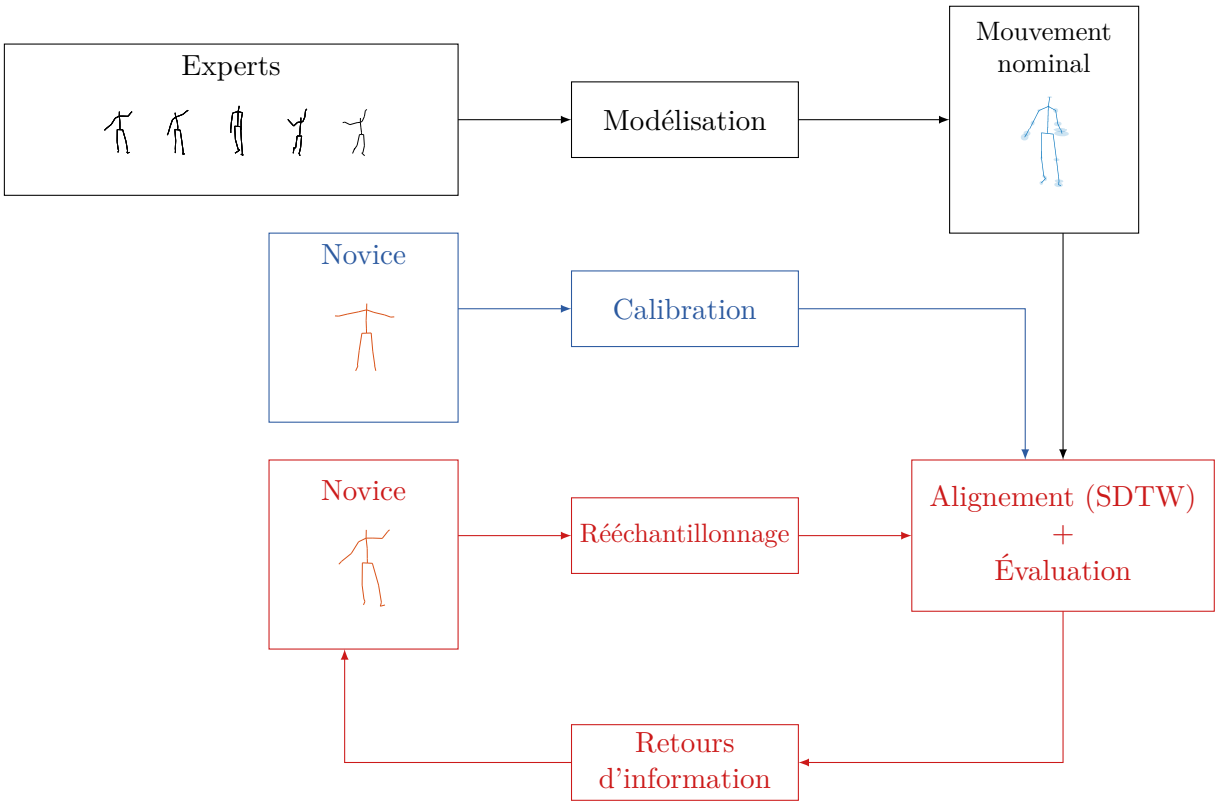


FIGURE 5.2 – Schéma global d'entraînement du novice. Les traits noirs indiquent les processus hors ligne, les rouges les processus en ligne et répétés lors de l'entraînement. En bleu est indiqué le processus de calibration, qui n'a lieu qu'une fois au début de la séance et qui est utilisé lors de l'alignement et l'évaluation du geste novice.

moins rapidement. De fait, un processus d'alignement segmentant va être explicité en partie 5.2.1.

- Les représentations squelettiques obtenues peuvent être différentes selon le procédé d'acquisition, c'est-à-dire que les articulations considérées dans la définition du squelette diffèrent d'un outil d'acquisition à l'autre. Nous devrons alors mettre en place un recalage entre squelettes, comme cela sera explicité en section 5.2.2.

- De surcroît, les fréquences d'acquisition des gestes experts et du geste à évaluer pouvant être très différentes (la Kinect® permet une fréquence d'acquisition de 30 Hz tandis que le système Vicon® utilisé peut fonctionner à 200Hz typiquement), les mouvements à aligner sont de durées très différentes. C'est pourquoi un ré-échantillonnage des données pourrait être mis en place.

- Le système d'acquisition peut présenter une précision très faible qu'il convient de caractériser afin de l'ajouter à la tolérance articulaire du geste nominal. Ici, nous avons utilisé l'Optitrack® et négligé son imprécision qui est certes supérieure à celle du système optotélectronique Vicon® utilisé chez les experts, mais très faible et donc négligeable devant la tolérance articulaire du geste nominal.

5.2.1 Le DTW segmentant (SDTW)

Le DTW segmentant (ou *Subsequence DTW*, aussi abrégé SDTW) répond au besoin de l'alignement de signaux multiples séquentiels non segmentés. À partir d'une série temporelle de référence (ou motif de référence), le SDTW permet de reconnaître et d'aligner les motifs similaires d'une séquence temporelle, comme l'illustre la figure 5.3. Appliqué au contexte d'évaluation de gestes, il permet d'enregistrer plusieurs mouvements du sportif au cours du temps sans avoir recours à une segmentation manuelle. Deux méthodes ont été développées pour parvenir à ce but : l'approche classique, plus naïve, étudiée notamment par Muller et Anguerra [153, 154] et l'approche temps-réel développée par Sakurai *et al.* [155], aussi appelée algorithme "SPRING".

La mise en place de base du SDTW repose sur une modification très simple du DTW. Alors que le DTW classique impose de connaître le début et la fin du chemin de déformation cherché, le SDTW quant à lui cherche à les optimiser afin d'extraire un chemin de déformation flottant. Soit $(x(i))_{1\leq i\leq M}$ le motif de référence, et $(y(j))_{1\leq j\leq N}$ la séquence temporelle à segmenter. Appliqué sur ces signaux, le SDTW consiste à extraire les chemins de déformation qui alignent $(x(i))_{1\leq i\leq M}$ sur $(y(j))_{a_n\leq j\leq b_n}$. a_n et b_n $(1 \leq a_n \leq b_n \leq N)$ représentent le début et la fin des motifs retrouvés dans le signal $y(j)$. L'initialisation de l'algorithme du DTW est modifiée de manière à ne privilégier aucun départ (voir la section 3.1.1).
Au lieu de :

$$D_{1,j} = \sum_{p=1}^{j} d_{1,p} \qquad j = 1, ..., N \tag{5.1}$$

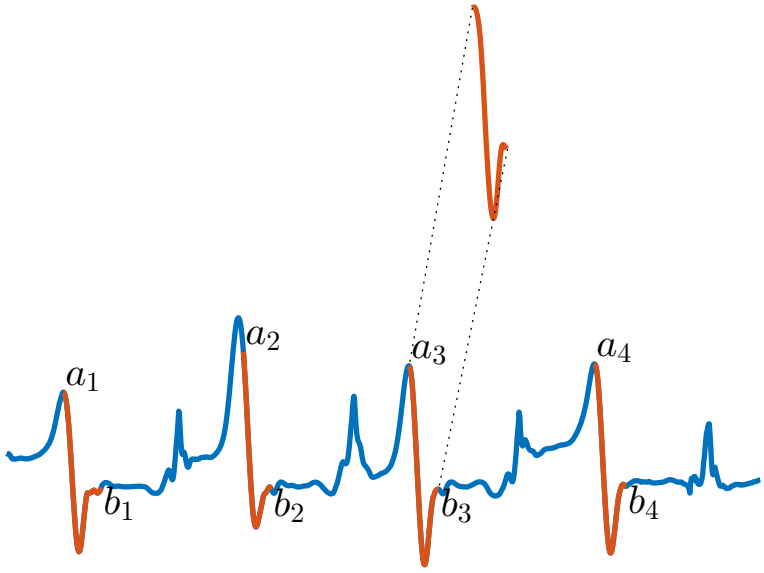
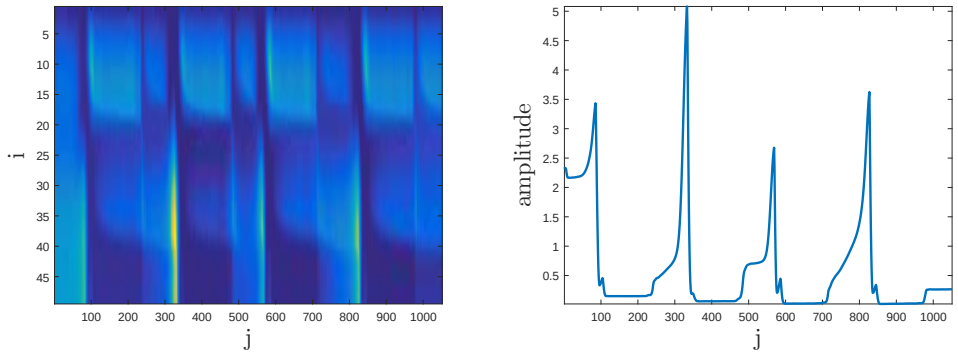


FIGURE 5.3 – Contexte du SDTW : à partir d'une séquence non segmentée, le but est de reconnaître et aligner chacun des motifs similaires à un certain motif (représenté en rouge sur la première ligne)

le SDTW impose :

$$D_{1,j} = d_{1,j} \qquad j = 1, ..., N \tag{5.2}$$

Dans le cas d'études des signaux $1D$ de la figure 5.3, la nouvelle carte de distance cumulée obtenue est représentée figure 5.4a. Cette carte a maintenant une forme particulière du fait de la nouvelle initialisation de la première ligne. Le but du SDTW est de rechercher des chemins de déformation traversant la carte de la première à la dernière ligne et menant à une distance faible entre le motif et la sous-partie de $y(j)$. Pour cela, une idée est de rechercher les minima locaux sur la dernière ligne de la carte correspondant aux valeurs de $D_{M,j}$. Cette ligne est représentée figure 5.4b, où les quatre minima locaux correspondant à la fin des 4 motifs apparaissent clairement.



(a) (b)

FIGURE 5.4 – (a) Carte de distance cumulée initiale issue de l'application du SDTW entre les signaux $x(i)$ et $y(j)$. (b) Dernière ligne de la carte de distance cumulée $D_{M,j}$.

Cette problématique étant posée, deux principaux algorithmes tentent d'y répondre. Passons tout d'abord en revue la démarche de la méthode classique, qui nous permettra ensuite de comprendre la méthode temps-réel que nous utiliserons.

5.2.1.1 Approche classique

La méthode classique consiste tout simplement à rechercher le minimum global de $D_{M,j}$ qui sera affecté à b_1, puis à extraire le chemin de déformation par rétropropagation, jusqu'à atteindre la première ligne de la carte $\boldsymbol{D}$. Une fois la première sous-séquence trouvée, il reste à trouver les potentiels autres chemins de déformation, *i.e.* les autres sous-séquences cachées dans $y(j)$. Pour éviter qu'une sous-séquence correspondant à la première sous-séquence ne soit trouvée à nouveau, on fixe les valeurs de $D_{M,j}$ au voisinage de D_{M,b_1} à l'infini. On peut alors appliquer à nouveau le processus pour obtenir les autres sous-séquences.

L'algorithme est réitéré jusqu'à ce qu'il n'y ait plus de minimum en dessous d'un certain seuil s. Ce seuil doit être choisi de telle sorte que les sous-séquences ne correspondant pas suffisamment au motif ne soient pas détectées. On pourrait aisément

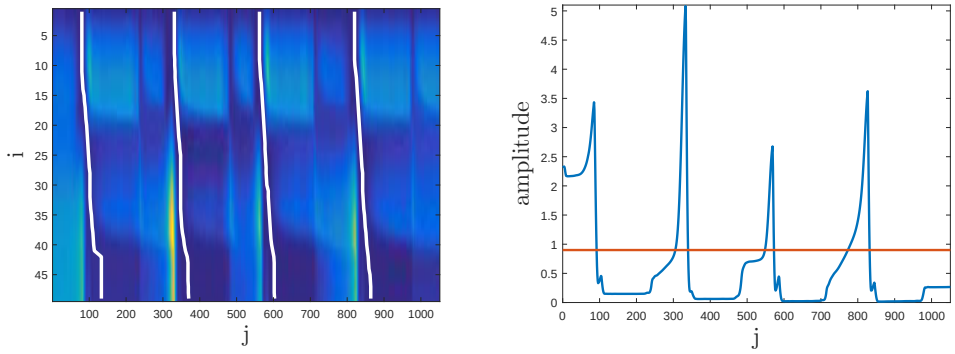
imaginer un apprentissage de ce seuil à partir d'une base d'apprentissage de signaux similaires à ceux à aligner. L'algorithme 4 récapitule les différentes étapes du processus de SDTW classique.

La figure 5.5 présente la distance cumulée obtenue entre les deux signaux de la figure 5.3 sur laquelle sont superposés les chemins de déformation obtenus *via* l'algorithme classique avec un seuil fixé à $s = 0.9$.

Données : $x(i)$ de taille M, $y(j)$ de taille N, seuil s, voisinage δ, $p = 1$
Résultat : $(a_n, b_n), n = 1...p$
Calcul des cartes de distance $\boldsymbol{d}$ et de distance cumulée $\boldsymbol{D}$ entre $x(i)$ et $y(j)$ avec les nouvelles conditions initiales (équation 5.2).
tant que $\min_j D_{M,j} < s$ **faire**
  $b_p = \arg\min_j D_{M,j}$
  a_p est déterminé par rétropropagation de b_p dans $\boldsymbol{D}$
  $D_{M,c} = +\infty \quad \forall c \in \{b_p - \delta, b_p + \delta\}$
  $p = p + 1$
fin

Algorithme 4 : Algorithme de *Subsequence DTW* (SDTW)

Enfin, des post-traitements sur les chemins de déformations peuvent permettre un filtrage d'éventuels chemins caduques, par exemple les chemins trop courts (quasi verticaux) détectés à tort par l'algorithme.



(a) (b)

FIGURE 5.5 – (a) Carte de distance cumulée initiale issue de l'application du SDTW entre les signaux $x(i)$ et $y(j)$. En blanc sont tracés les chemins de déformation obtenus. (b) Dernière ligne de la carte de distance cumulée $D_{M,j}$ (en bleu) et seuil choisi de manière à ce que les sous-séquences ne correspondant pas au motif ne soient pas détectées (en rouge).

5.2.1.2 Approche temps-réel

L'approche classique ne permet pas une analyse en temps-réel ; la recherche de minimum global n'étant pas possible sur des données qui arrivent progressivement. Dès lors, une nouvelle technique a été développée par Sakurai *et al.* [155, 156], aussi appelée algorithme SPRING. Cet algorithme se veut plus efficace et utilisable en temps-réel.

Cette fois, la carte de distance cumulée, dont les conditions initiales n'ont pas changé (équation 5.2), est mise à jour à chaque instant, dès que la trame est reçue. Ainsi, à chaque nouvel échantillon de $y(j)$, une nouvelle colonne est ajoutée à la matrice de composantes $D_{i,j}$. On retient au fur et à mesure la distance cumulée minimale d_{temp} (valeur minimale de la dernière ligne de $\boldsymbol{D}$) et l'indice b_{end} correspondant. La difficulté est alors de savoir si, après avoir détecté un minimum, il n'y en aura pas un plus petit quelques instants plus tard. Il faut donc un critère supplémentaire.

En plus de la matrice de distance cumulée, on propage une matrice $\boldsymbol{S}$ appelée "matrice de départ". $\boldsymbol{S}$ est donnée par : $\forall i \in 2...M$ et $\forall j \in 2...N$,

$$S_{i,j} = \begin{cases} S_{i,j-1} & \text{si } \arg\min_{i,j}\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\} = (i, j-1) \\ S_{i-1,j} & \text{si } \arg\min_{i,j}\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\} = (i-1, j) \\ S_{i-1,j-1} & \text{si } \arg\min_{i,j}\{D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}\} = (i-1, j-1) \end{cases} \tag{5.3}$$

avec les conditions initiales suivantes :

$$S_{1,j} = j \quad \forall j \in 1...N \tag{5.4}$$

La matrice de départ $\boldsymbol{S}$, correspondant à l'alignement par SDTW "temps-réel" des deux mêmes signaux que précédemment, est présentée en figure 5.6. On remarque les bandes de couleur uniforme correspondant à des points admettant le même indice de départ.

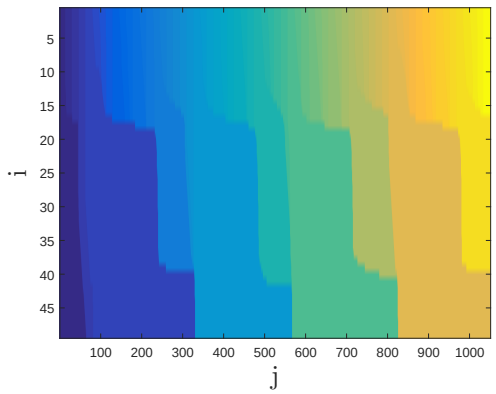


FIGURE 5.6 – Matrice de départ $\boldsymbol{S}$ générée par l'application du SDTW "temps-réel" entre le signal motif et la séquence temporelle de la figure 5.3.

De par sa construction, la dernière ligne de la matrice de départ, de composantes $S_{M,j}$, contient ainsi dans chaque colonne j l'indice k correspondant au début de la sous-séquence la plus similaire à la série temporelle patron et finissant à l'indice j. Le processus global peut donc se résumer de la façon suivante :

1. Initialisation : à la réception de la première trame de la séquence, $d_{temp} = D_{M,1}$ et $b_{end} = 1$.

2. À chaque instant, mise à jour de la matrice de distance cumulée et de la matrice de départ.

3. Si à l'instant j, $D_{M,j} < d_{temp}$, alors actualisation de d_{temp} et b_{end} : $d_{temp} = D_{M,j}$ et $b_{end} = j$.

4. Pour s'assurer que d_{temp} est bien un minimum local, le calcul est prolongé jusqu'à atteindre un instant j tel que $S_{M,j} > b_{end}$. On garantit ainsi la non-présence d'un autre chemin possible dans la séquence en cours.

5. Dès qu'un chemin est trouvé, il est enregistré et d_{temp} et b_{end} sont réinitilisés.

6. L'algorithme s'arrête lorsque toutes les trames sont arrivées.

L'algorithme 5 récapitule les étapes de construction des sous-séquences alignées de $y(j)$ par l'algorithme SPRING.

Données : $x(i)$ de taille M et $y(1)$, $p = 1$, seuil s
Résultat : $(a_n, b_n), n = 1...p$
tant que *Nouvelle trame j : y(j)* **faire**
  Mise à jour de $\boldsymbol{d}$, $\boldsymbol{D}$ et $\boldsymbol{S}$ (eq.5.2)
  si $D_{M,j} < d_{temp}$ **alors**
    $d_{temp} = D_{M,j}$
    $j_{temp} = j$
  fin
  si $d_{temp} < s$ *et* $S_{M,j} > j_{temp}$ **alors**
    $b_p = j_{temp}$ et $a_p = S_{M,b_p}$
    $p = p + 1$
  fin
fin

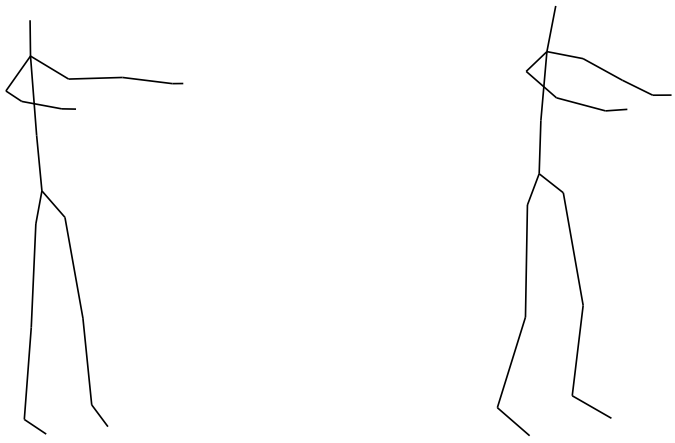
Algorithme 5 : Algorithme de *Subsequence DTW (SPRING)*

C'est cette approche qui sera privilégiée dans nos travaux puisqu'elle permet une utilisation en temps-réel, particulièrement nécessaire pour les besoins de notre application.

Il est enfin à noter que le SDTW (qu'il soit codé en temps-réel ou non), de part sa mise en place, privilégie la diagonalité du chemin de déformation (c'est-à-dire l'alignement linéaire) comme l'a souligné Muller [153]. De fait, lorsque les fréquences des systèmes de capture diffèrent largement, un ré-échantillonnage des données peut être très pertinent.

5.2.2 Recalage de squelettes par transformations locales

À systèmes de capture différents, squelettes potentiellement différents. Plusieurs facteurs expliquent ceci. Dans le cas de systèmes de capture opto-électronique (type Vicon® ou Optitrack®), les centres articulaires sont déterminés à partir du positionnement des marqueurs. Ce calcul est très souvent intégré dans le logiciel de capture (comme c'est le cas de l'Optitrack® par exemple). Bien que contraints à un repérage anatomique rigoureux pour le positionnement des marqueurs, la localisation des centres articulaires varie donc d'un système à l'autre. De fait, les squelettes diffèrent. Nous considérerons cependant, comme c'est le cas la plupart du temps, que le nombre d'articulations et que leur hiérarchie au sein de la structure du squelette est toujours la même quel que soit le système considéré. Ce jeu d'articulation est celui considéré en figure 4.1b. Dans le cas d'une technique différente d'acquisition des données (par exemple, en utilisant la carte de profondeur comme le fait la Kinect®), le squelette engendré est là encore potentiellement distinct. Un exemple est illustré en figure 5.7 avec deux squelettes issus de deux systèmes de capture différents. Il apparaît très clairement des positionnements de centres articulaires distincts.



(a) Squelette extrait *via* la Kinect®.

(b) Squelette extrait *via* le système opto-électronique Vicon®.

FIGURE 5.7 – Illustration de la nécessité du recalage de squelette par transformations locales. Un même mouvement acquis par une Kinect® (a) ou par un système de capture optoélectronique Vicon® (b) positionne différemment les centres articulaires du sujet.

Nous proposons de mettre en place un recalage spatial simple des données par transformation linéaire. Notons que pour plus de rigueur et dans le cas de données davantage déformées, un recalage non linéaire devrait être considéré.

Deux articulations obtenues par deux systèmes de capture indexés 1 et 2 n'ont pas forcément le même positionnement dans un repère global, noté ici $\mathcal{R}_0$. Ceci peut poser

problème lorsque l'acquisition des gestes experts n'est pas effectuée avec le même système que l'acquisition des gestes novices. Nous proposons d'apprendre la transformation entre les positions des articulations à partir d'une étape de calibration. Ce recalage dépendant des conditions de capture (*i.e.* du matériel utilisé), il doit être effectué durant la séance d'entraînement.

Dans un contexte idéal, ce recalage devrait être déterminé à partir d'une capture simultanée d'un même mouvement par les systèmes 1 et 2. En pratique, en supposant que le novice s'entraîne chez lui avec un système à faible coût, c'est impossible. Par conséquent, nous proposons de faire rejouer au novice un mouvement très simple par imitation de l'expert. À partir des données acquises par l'expert et de celles du novice, le recalage peut être calculé.

Soit le schéma de la figure 5.8, où l'on considère le positionnement des articulations du bras droit (RSho, RElb et RWri, cf. figure 4.1) obtenues par les outils de capture 1 (en bleu) et 2 (en vert). Les repères orthogonaux directs $\mathcal{R}_1$ et $\mathcal{R}_2$ sont mis en place de la façon suivante, pour $i = 1$ ou 2 :

- $RElb_i$ correspond au positionnement du coude.
- $\boldsymbol{x}_i$ est dirigé par le vecteur allant du coude ($RElb_i$) au poignet ($RWri_i$)
- $\boldsymbol{z}_i$ est orthogonal au plan contenant les trois articulations du bras ($RSho_i$, $RElb_i$ et $RWri_i$).
- $\boldsymbol{y}_i$ est orthogonal à $\boldsymbol{x}_i$ et $\boldsymbol{z}_i$ et $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i)$ est une base directe.

Chacune de ces bases orthogonales est ensuite normée, ce qui leur confère un statut de bases orthonormées.

Considérons l'outil de capture 2 comme l'outil utilisé par le novice et l'outil 1 celui utilisé par les experts. On recherche la transformation permettant d'exprimer la position de l'articulation centrale acquise grâce à l'outil 1, notée $RElb_1$ (de coordonnées $\begin{pmatrix} x_{O_1} \\ y_{O_1} \\ z_{O_1} \end{pmatrix}$ dans le repère $\mathcal{R}_0$) connaissant sa position $RElb_2$ (de coordonnées $\begin{pmatrix} x_{O_2} \\ y_{O_2} \\ z_{O_2} \end{pmatrix}$ dans le repère $\mathcal{R}_0$). Pour ce faire, nous allons extraire la transformation qui permet de passer du repère 1 au repère 2. Cette transformation sera utilisée durant l'entraînement pour estimer $RElb_1$ à partir uniquement de $RElb_2$ (l'outil 2 étant utilisé par le novice). Soit $\boldsymbol{T_{12}}$ cette transformation. La relation de Chasles sur les transformations linéaires permet tout d'abord d'exprimer $\boldsymbol{T_{12}}$ en fonction de $\boldsymbol{T_{01}}$ et $\boldsymbol{T_{02}}$:

$$\boldsymbol{T_{12}} = \boldsymbol{T_{10}}.\boldsymbol{T_{02}} \tag{5.5}$$

$$\boldsymbol{T_{12}} = \boldsymbol{T_{01}^{-1}}.\boldsymbol{T_{02}} \tag{5.6}$$

avec :

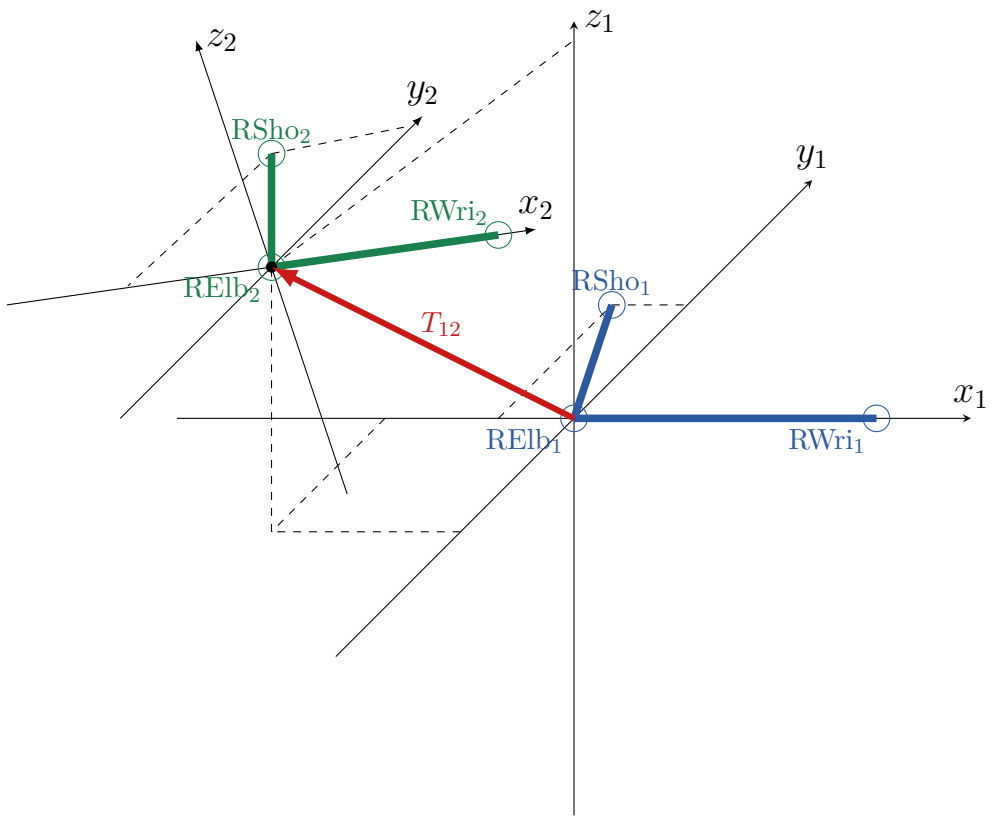


FIGURE 5.8 – Mise en place des bases locales $\mathcal{R}_1 = (RElb_1, \frac{\boldsymbol{x_1}}{\|\boldsymbol{x_1}\|}, \frac{\boldsymbol{y_1}}{\|\boldsymbol{y_1}\|}, \frac{\boldsymbol{z_1}}{\|\boldsymbol{z_1}\|})$ et $\mathcal{R}_2 = (RElb_2, \frac{\boldsymbol{x_2}}{\|\boldsymbol{x_2}\|}, \frac{\boldsymbol{y_2}}{\|\boldsymbol{y_2}\|}, \frac{\boldsymbol{z_2}}{\|\boldsymbol{z_2}\|})$ de l'articulation du coude obtenues *via* deux systèmes de capture différents. En bleu sont représentées les 3 articulations $RSho_1$, $RElb_1$ et $RWri_1$ du bras droit obtenues grâce au premier outil de capture. En vert celles obtenues grâce au deuxième outil de capture. À partir de ces trois points dans l'espace, on introduit un repère orthonormé direct avec $\boldsymbol{x}$ dirigé selon l'avant-bras et $\boldsymbol{z}$ orthogonal au plan formé par le bras et l'avant-bras.

$$\boldsymbol{T_{ij}} = \begin{pmatrix} \begin{pmatrix} & \boldsymbol{R_{ij}} & \end{pmatrix} & \begin{pmatrix} x_{O_j} \\ y_{O_j} \\ z_{O_j} \end{pmatrix}_{\mathcal{R}_i} \\ 0 \quad 0 \quad 0 & 1 \end{pmatrix} \quad \forall (i,j) \in \{0...2\} \tag{5.7}$$

La position de $RElb_1$ dans $\mathcal{R}_0$ s'exprime donc par rapport à $RElb_2$ de la façon suivante :

$$\begin{pmatrix} x_{O_1} \\ y_{O_1} \\ z_{O_1} \end{pmatrix}_{\mathcal{R}_0} = \begin{pmatrix} \begin{pmatrix} & \boldsymbol{R_{02}} & \end{pmatrix} & \begin{pmatrix} x_{O_2} \\ y_{O_2} \\ z_{O_2} \end{pmatrix}_{\mathcal{R}_0} \end{pmatrix} . \begin{pmatrix} x_{O_1} \\ y_{O_1} \\ z_{O_1} \end{pmatrix}_{\mathcal{R}_2} \tag{5.8}$$

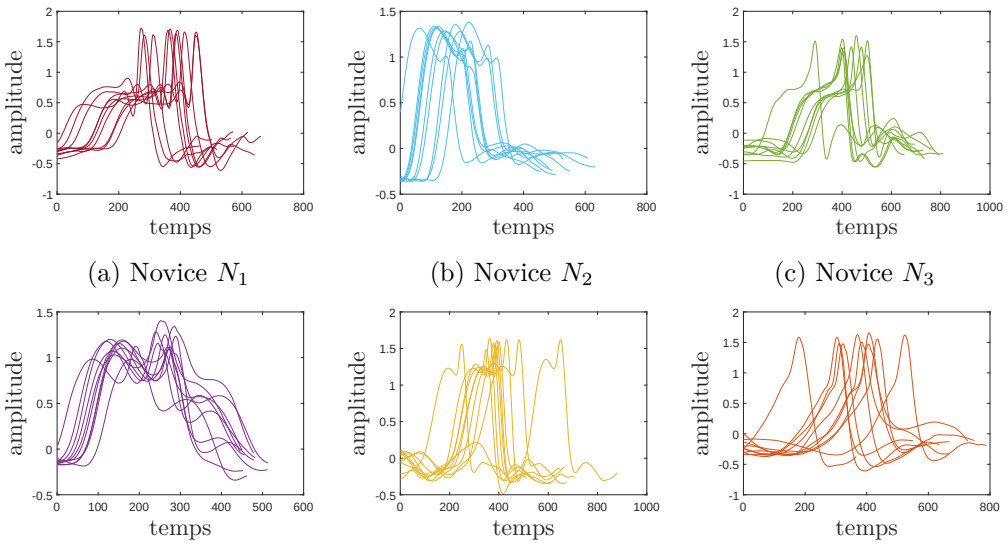
où $\boldsymbol{R_{02}}$ et $\begin{pmatrix} x_{O_1} \\ y_{O_1} \\ z_{O_1} \end{pmatrix}_{\mathcal{R}_2}$ ont été calculés durant la calibration. Le calcul de cette transformation a été fait sur tout le mouvement de calibration. Le positionnement moyen relatif a ensuite été conservé.

Les squelettes étant alors recalés spatialement et les mouvements pouvant être recalés temporellement, notre processus d'évaluation du geste peut alors être appliqué. La partie suivante va traiter de la mise en place du déroulement d'une séance d'entraînement.

5.2.3 Mise en place d'une interface adaptée

5.2.3.1 Étude de la répétabilité des erreurs d'un novice

Afin d'évaluer la fréquence de rendu des informations, nous nous proposons d'étudier la répétabilité des erreurs d'un novice. La figure 5.9 présente les variations en hauteur d'une même articulation (la main droite) pour plusieurs novices différents, dénommés ici de N_1 à N_6. Chacun de ces novices exécute le service 10 fois, les 10 trajectoires sont superposées.



(a) Novice N_1 (b) Novice N_2 (c) Novice N_3

(d) Novice N_4 (e) Novice N_5 (f) Novice N_6

FIGURE 5.9 – Projection en z de la trajectoire de la main droite lors d'un service de tennis exécuté par 6 novices distincts (de N_1 à N_6). Chaque novice réalisant 10 essais à la suite, on obtient à chaque fois 10 trajectoires qui sont superposées.

Comme nous pouvons le remarquer, malgré une temporalité variable, les motifs très similaires reviennent chez un même novice, et ces motifs sont variables d'un novice à l'autre. Indépendamment de sa taille (les échelles diffèrent), on peut très raisonnablement penser que le mouvement du novice reste très similaire d'une réalisation à l'autre dès

lors qu'on ne lui procure pas de retour d'information sur les erreurs commises.

Pour cette raison, nous confortons l'idée suggérée lors de l'état de l'art, consistant à donner à l'utilisateur un rendu terminal après chacune de ses tentatives.

5.2.3.2 Déroulement d'une séance d'évaluation

Maintenant que les différentes adaptations au temps réel ont été faites, la séance d'entraînement peut être planifiée. Elle se compose des étapes suivantes :

1. L'utilisateur choisit le nombre de tentatives qu'il souhaite réaliser.
2. Un processus de calibration débute : il s'agit de calculer les recalages des squelettes par transformations locales. À cette fin, on donne la consigne au sujet de faire suivre à son squelette (qu'il voit en temps réel) le mouvement d'un squelette de référence. Par DTW segmentant, on récupère le mouvement du novice segmenté. Enfin, à partir des deux mouvements, on détermine le recalage (*i.e.* la transformation T_{12}).
3. Le novice effectue son mouvement.
4. Le SDTW ajuste le mouvement du novice sur le geste nominal. Les erreurs spatiales et temporelles sont calculées après recalage du squelette du novice sur celui de l'expert par transformations locales. La plus grande erreur est exposée au novice : une animation de son mouvement superposé au nominal est affichée, un codage colorimétrique lui indique quel membre est à corriger, et à quel moment.
5. Les deux étapes précédentes sont répétées autant de fois que le novice l'a choisi au début.
6. Un récapitulatif général permet à l'athlète de rejouer ses différents mouvements et de lire les erreurs correspondantes. Une jauge de progression permet également de l'informer sur l'évolution de sa performance globale au cours des essais.

5.3 Utilisation de l'outil d'évaluation temps-réel

Afin de tester la faisabilité du système établi, une démonstration d'entrainement est faite sur un novice.

La licence *Body* du logiciel Motive® a été utilisée en association avec 8 caméras infrarouges Optitrack® et 37 marqueurs positionnés sur des repères anatomiques bien spécifiques du sujet. Le logiciel et sa licence permettent d'ajuster un squelette sur les points des marqueurs extraits en $3D$. Ce squelette est envoyé à tout instant *via* une connexion réseau à un second PC sur lequel Matlab effectue les calculs et l'interfaçage. Les instructions et retours sont projetés à l'utilisateur (cf. figure 5.10). L'outil d'entraînement développé a été testé sur des services de tennis. De fait, uniquement l'erreur spatiale a été considérée, pour les mêmes raisons que précédemment (4.5.1). Dans un premier temps, l'utilisateur indique le nombre d'essais qu'il souhaite réaliser comme le

montre la figure 5.11. Ensuite, on lui donne les consignes de suivi de mouvement pour la calibration de son squelette, c'est l'étape illustrée aux figures 5.12, 5.13 et 5.14. Le recalage est alors calculé. S'en suit l'enregistrement du mouvement à proprement parler, que l'utilisateur commence et termine quand il le souhaite (figures 5.15 et 5.16). Le SDTW se charge alors de le segmenter et de l'aligner sur le mouvement nominal. L'erreur principale du mouvement est affichée au novice, comme le témoigne la figure 5.17 (dans ce cas présent, le bras droit est dit trop à droite lors de la frappe). L'enregistrement du mouvement recommence autant de fois que l'a choisi l'utilisateur au début de la séance. Enfin, une fois les essais terminés, un récapitulatif général est proposé à l'athlète.

Ce récapitulatif, illustré en figure 5.19, contient plusieurs champs :

- À droite, la visualisation du mouvement du novice est superposée au mouvement nominal. Une barre de défilement permet d'avancer plus ou moins dans le temps. Plusieurs vues peuvent être choisies, l'utilisateur peut également changer la vue manuellement en glissant sur le squelette avec la souris.

- En haut à droite, l'essai considéré est choisi. Tous les essais réalisés pendant la séance sont disponibles.

- Juste en dessous, un tableau récapitule les 5 erreurs spatiales principales. Ces erreurs sont localisées sur un membre, elles sont ensuite expliquées très simplement : on dira dans le cas d 'une erreur spatiale qu'un membre est "trop haut", "trop bas", "trop à gauche", "trop à droite", "trop en avant", "trop en arrière", ou dans le cas d'une erreur spatiale, qu'un membre est "en retard" ou "en avance" sur un autre. Des scores sont également donnés et les erreurs sont localisées temporellement. Par exemple, la première erreur concerne le bras gauche qui est "trop haut" entre les instants 363 et 544. La valeur maximale atteinte par l'erreur est de 11.34. Rappelons que conformément à l'hypothèse gaussienne des données que nous nous sommes fixée, 99.7% des experts ont une erreur inférieure à 3.

- Par simple clic sur l'un des boutons correspondant aux erreurs temporelles indiquées, le squelette novice est recalé localement par rapport au membre considéré et l'erreur en question est indiquée par coloration rouge à l'instant où le membre est mal positionné.

- En dessous, un deuxième tableau récapitule cette fois les erreurs temporelles. Comme elles n'ont pas été considérées ici, ce tableau est resté vide.

- Enfin, un graphe de progression pourrait être affiché. Ça n'a pas été fait ici puisque seulement 2 essais ont été réalisés, mais le principe est très simple. Il suffit d'établir un intervalle de scores entre un "bon service" et un "très mauvais service" (dont les scores sont à définir selon les attentes de l'entraineur par exemple), puis de transposer le score global de l'essai en pourcentage du score parfait, et d'établir ainsi un graphe de variation au cours des séances. La progression peut être considérée sur tout le corps, ou se focaliser sur un membre en particulier, comme le

proposent les différents boutons à droite du graphe de progression. Ils apporteraient au novice une motivation supplémentaire particulièrement appréciée dans un cadre d'apprentissage.

Notons que ce premier prototype ouvre un vaste champ de possibilités quant à la mise en place d'outil d'entrainement sportif. Ici, différents niveaux d'interprétation des erreurs sont disponibles : le premier, se basant uniquement sur des scores, le second, explicitant brièvement la nature de l'erreur commise. Pour aller plus loin, différentes phases d'apprentissage pourraient être proposées afin d'adapter le retour d'informations et sa fréquence à la progression du sujet.

FIGURE 5.10 – Session d'entraînement d'un novice avec Optitrack®. 37 marqueurs sont positionnés sur le novice, 8 caméras extraient leurs positionnements en $3D$. L'interface d'entraînement est projetée au novice lors de sa session.

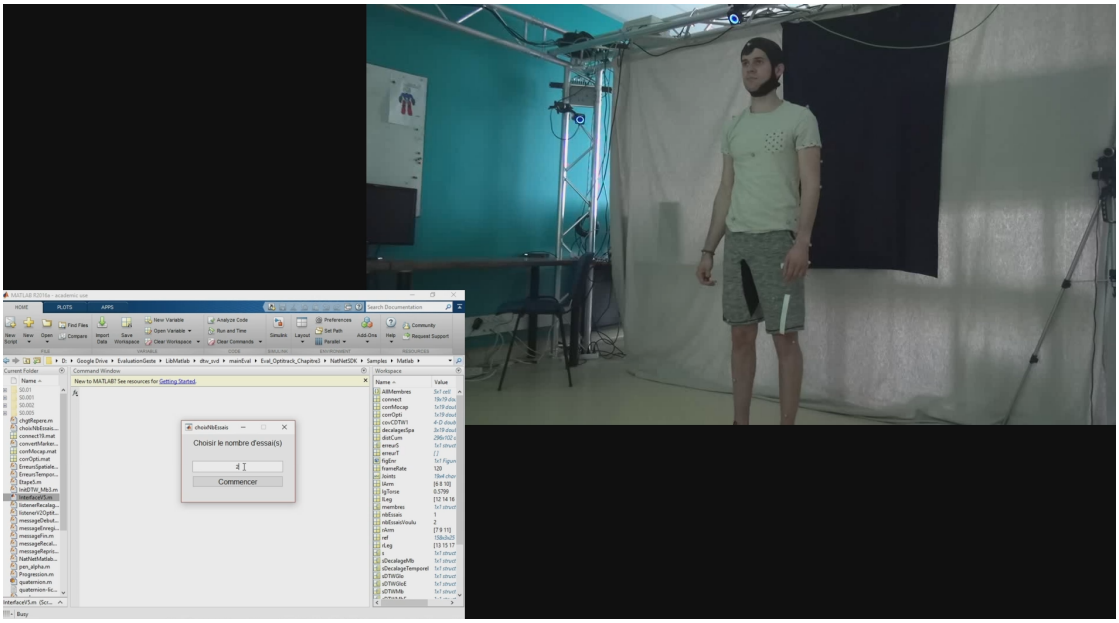


FIGURE 5.11 – L'utilisateur indique le nombre d'essais qu'il souhaite réaliser.

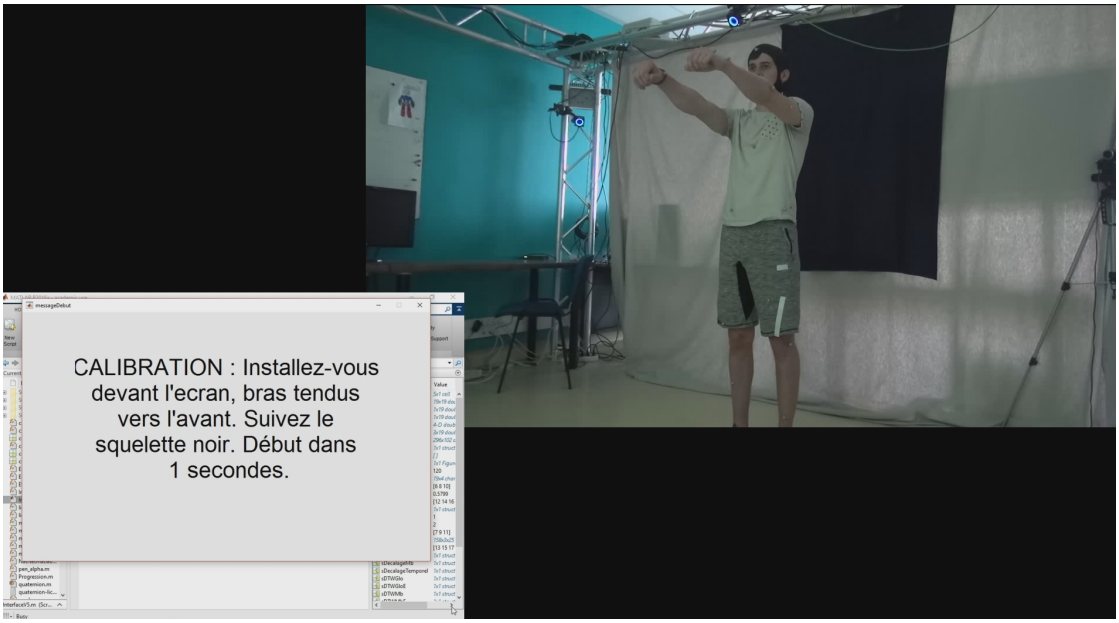


FIGURE 5.12 – Des instructions sont données à l'utilisateur quant à la phase de calibration : il doit faire suivre à son squelette le mouvement d'un squelette de référence.

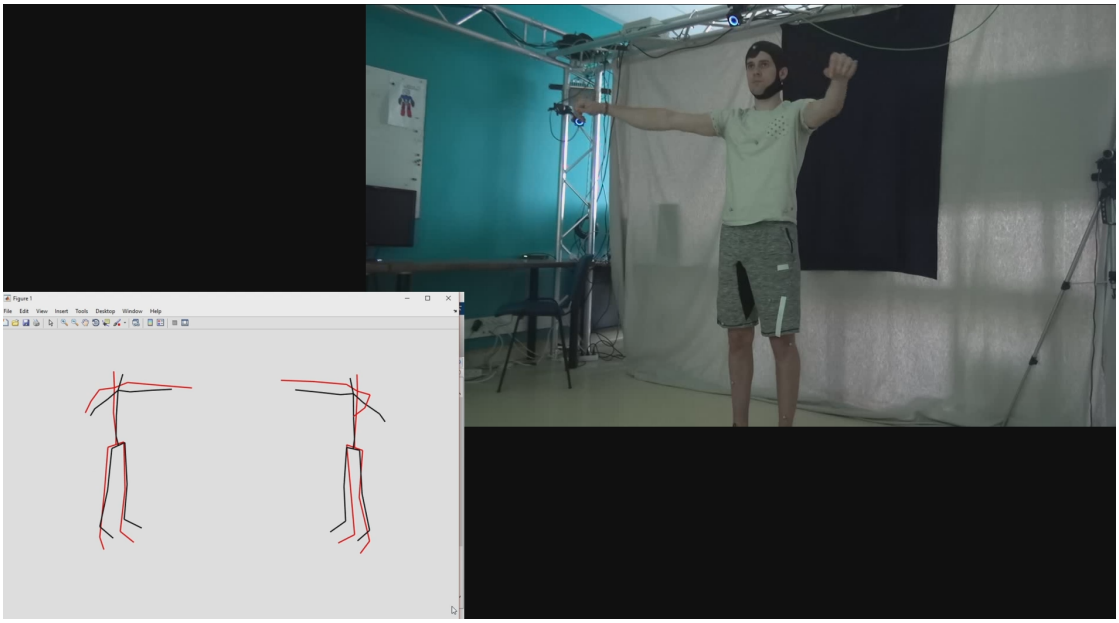


FIGURE 5.13 – Son squelette est modélisé par des traits rouges, celui à suivre est en noir. Deux vues sont affichées.

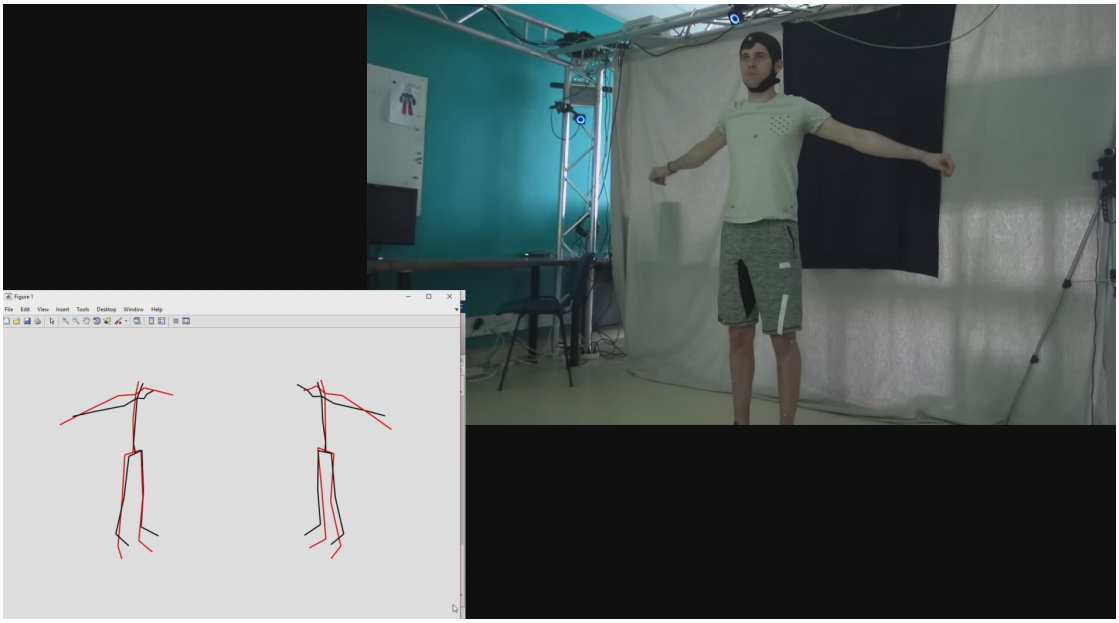


FIGURE 5.14 – Le mouvement attendu est un mouvement très simple d'ouverture et de fermeture des bras.

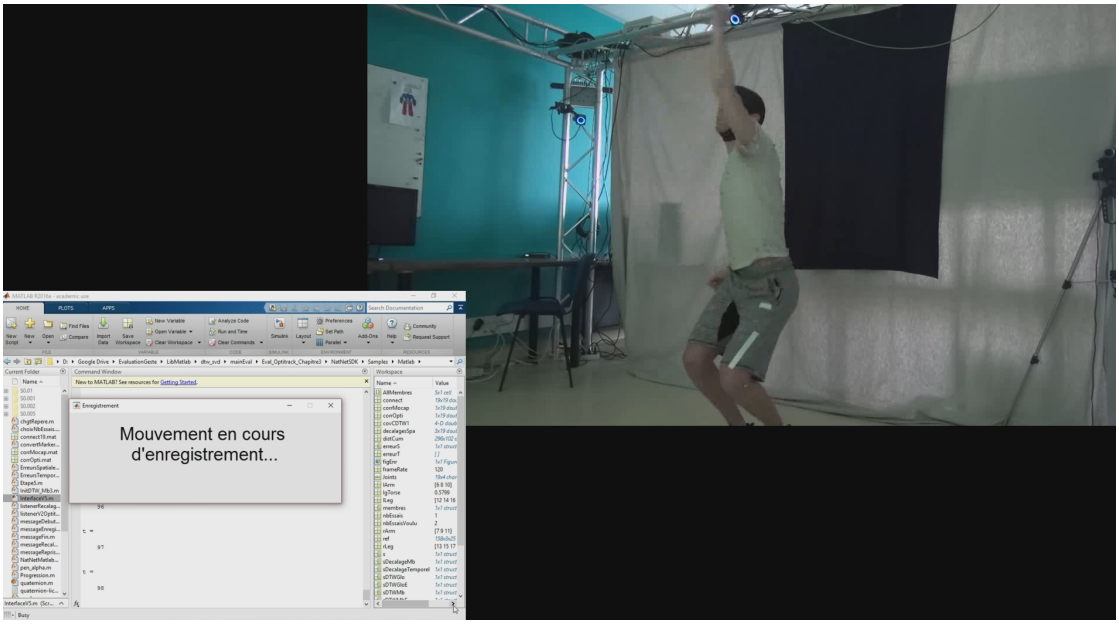


FIGURE 5.15 – Une fois la calibration terminée, l'utilisateur peut exécuter son mouvement qui sera enregistré et évalué.

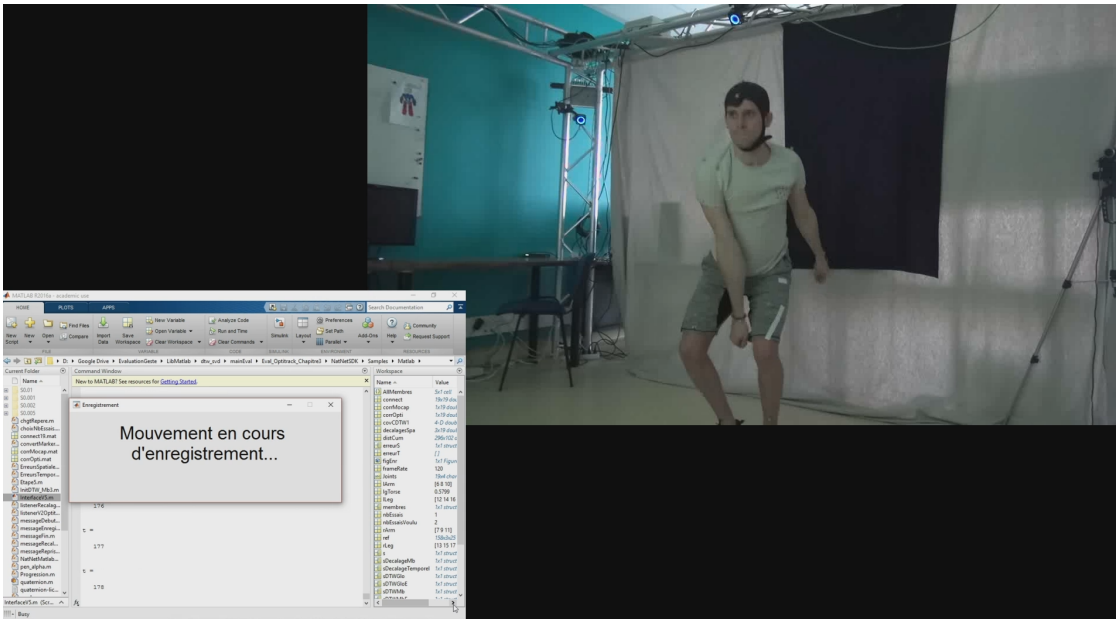


FIGURE 5.16 – Le sujet termine son mouvement de service de tennis.

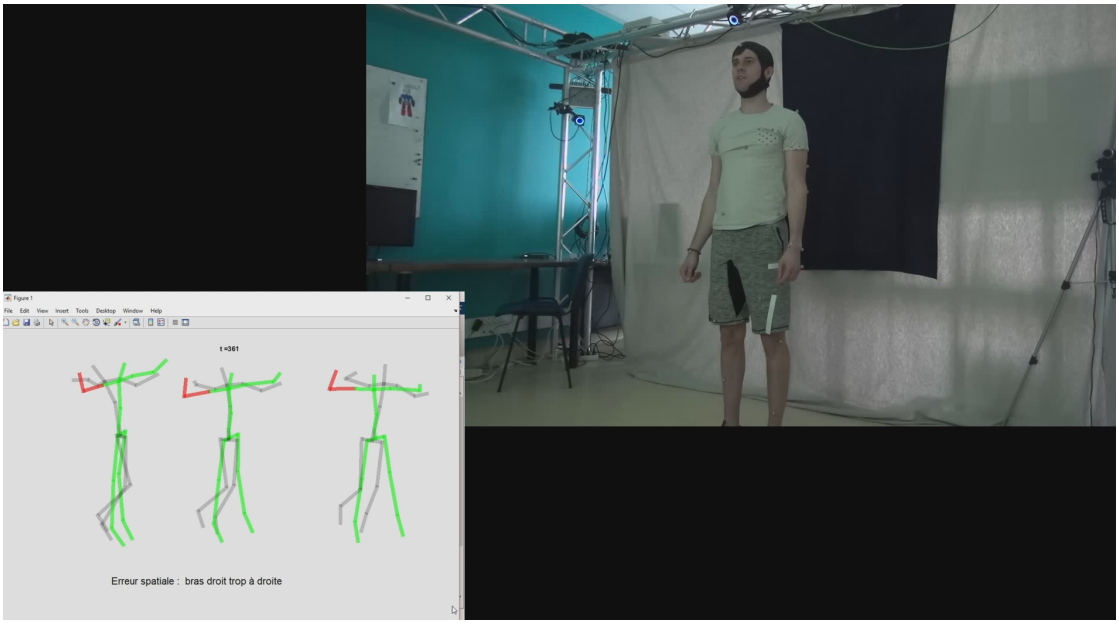


FIGURE 5.17 – L'erreur principale est affichée au joueur avec un code couleur adapté.

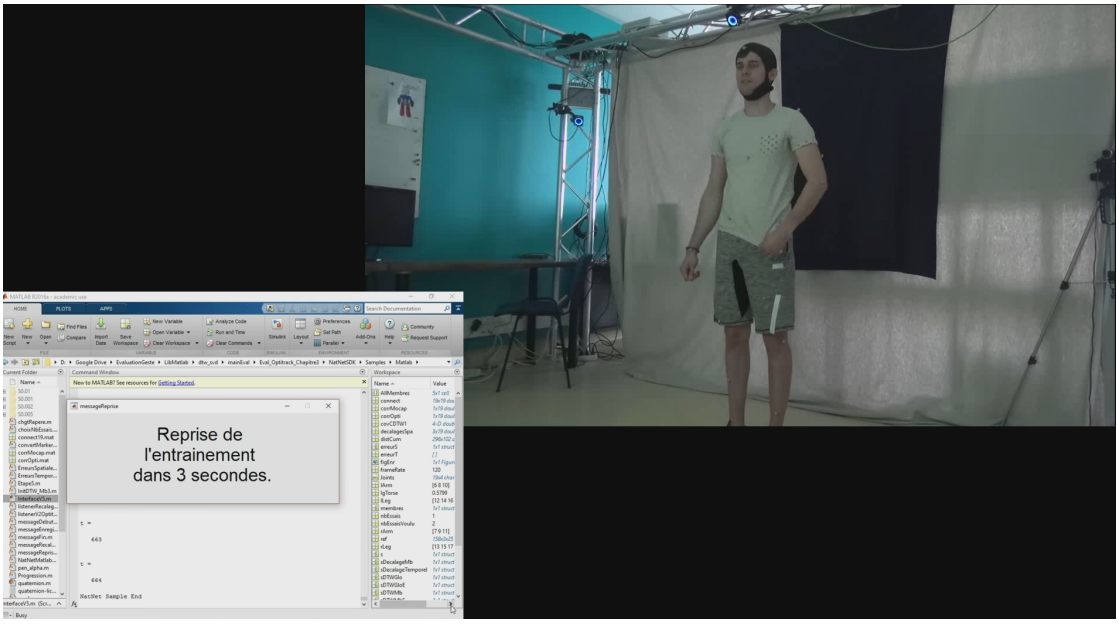


FIGURE 5.18 – Le processus se répète autant de fois que l'a décidé l'utilisateur au départ.

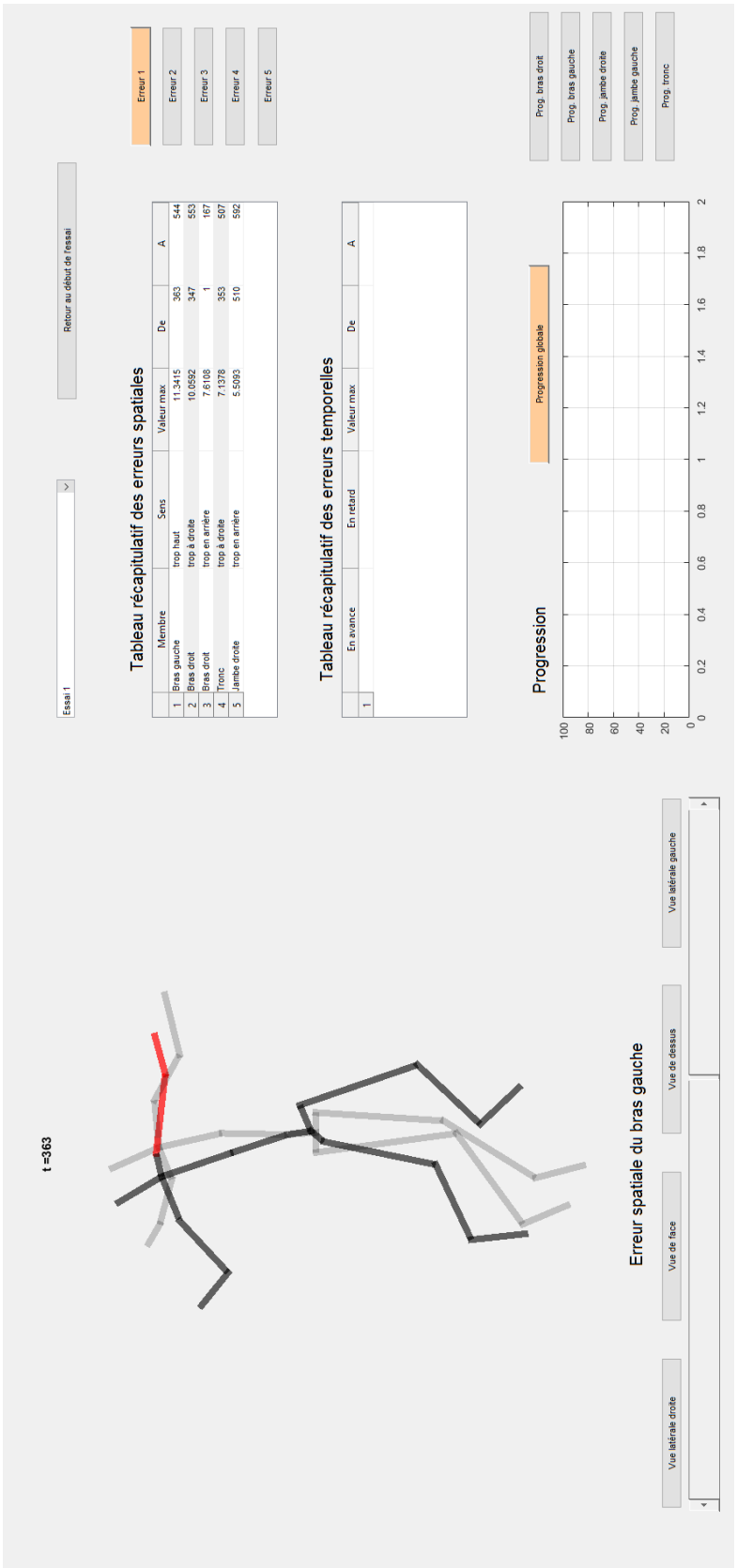


FIGURE 5.19 – À la fin de l'entraînement, un récapitulatif global est donné à l'utilisateur.

Conclusion

Ce dernier chapitre a permis de mettre en application les outils développés et validés précédemment, dans un contexte d'entraînement en temps réel. À des fins d'entraînement adapté d'un novice qui souhaite améliorer son mouvement de façon autonome, plusieurs considérations ont été prises en compte.

Dans un premier temps, l'outil de capture a dû être reconsidéré. En effet, celui choisi lors de l'enregistrement de mouvements d'experts étant encombrant et coûteux, il n'est pas adapté à l'entraînement d'une quelconque personne chez elle. Par conséquent a dû être réfléchie l'utilisation d'un dispositif autre que celui de la phase d'apprentissage. La fréquence d'acquisition s'en trouve possiblement modifiée, de même que le positionnement des centres articulaires du squelette extrait. Des solutions ont été proposées par recalage spatial et rééchantillonage temporel des données.

Dans un second temps, le contexte d'entraînement n'assure en aucun cas une segmentation automatique des mouvements. Le DTW segmentant (ou *Subsequence DTW*) a donc été proposé pour répondre à ce besoin.

Enfin, le rendu informatif jouant un rôle prépondérant dans le processus d'apprentissage d'un sportif, les différents procédés ont été passés en revue et une procédure construite et justifiée de déroulement de séance a été mise en place. Le rendu informatif visuel terminal avec superposition a été choisi, tout en assurant les conventions colorimétriques adaptées.

Pour conclure ce chapitre, une démonstration a été faite sur des services de tennis. Par conséquent et pour les mêmes raisons qu'explicitées auparavant, uniquement l'erreur spatiale a été prise en compte. Bien que l'Optitrack® soit coûteux et donc peu réaliste dans un cadre d'entraînement d'une personne lambda, c'est ce dispositif d'enregistrement qui a été utilisé pour des raisons pratiques.

Néanmoins, le procédé mis en place pourrait s'adapter à n'importe quel autre moyen technique.

Conclusion...

Cette thèse s'inscrit dans le cadre d'une collaboration entre un laboratoire de biomécanique (le M2S de Rennes) et un laboratoire de traitement de signal et apprentissage statistique (l'ISIR de Paris). Elle vise à concevoir un système d'entraînement autonome de n'importe quel sport individuel à partir uniquement de gestes exécutés par des experts du domaine. Un tel système trouve des débouchés dans l'apprentissage de mouvements techniques auprès de sportifs de tous niveaux. On pourrait également transférer un tel système à l'apprentissage de gestes chirurgicaux. La mise en œuvre de cet outil soulève toutefois plusieurs problèmes scientifiques auxquels nous avons tenté de répondre dans ce manuscrit.

Au cours de cette thèse, nous avons spécifiquement abordé la problématique de la modélisation d'un jeu de séries temporelles multidimensionnelles. En réalité, la littérature se concentrant beaucoup sur la classification de données, il est plus fréquent de rencontrer des outils de différenciation plutôt que de modélisation. Néanmoins, certains outils existent et c'est l'un d'entre eux que nous avons décidé d'utiliser, celui reposant sur l'algorithme de déformation temporelle dynamique (ou DTW), pour la simple raison qu'il était primordial pour la suite de conserver une information temporelle précise des données, de même que de conserver des variables concrètes telles que les positions. Le DTW permet d'aligner deux séries temporelles entre elles [77]. D'abord étudié dans un cas unidimensionnel, nous nous sommes penchés sur ses effets et avons rapidement mis en évidence une limitation majeure qui concerne les chemins de déformation pathologiques. Nous avons donc proposé certaines contributions afin de répondre au besoin fixé, en utilisant non pas un DTW mais sa version contrainte localement, le CDTW. Un défi scientifique majeur s'est alors posé, celui du moyennage d'un jeu de séries temporelles. Partant du CDTW, nous avons proposé l'algorithme du CDBA répondant aux limitations actuelles. Pour parfaire ce premier modèle, nous avons ajouté une information majeure de tolérance au sein du jeu, témoignant de la dispersion de celui-ci à tout instant. Ce premier apport a été validé sur une base de données $1D$ de la littérature, *UCR Time-Series Classification Archive*. Après généralisation au cas multidimensionnel, il a ensuite été validé sur une base de gestes de la littérature, *ArmGesturesM2S*. Dans les deux cas, la validation s'est faite dans un cadre de classification de données.

Dans un second temps, les méthodes ainsi développées ont été appliquées à l'évaluation de gestes sportifs. Pour ce faire, deux bases de données ont été acquises auprès d'experts et de novices : l'une est composée de services de tennis ; l'autre de coups de

poings de karaté (*Zuki*). Ces mouvements étant effectués par des personnes de morphologies différentes, selon des positions et orientations variables, une première étape de codage et normalisation des données a été requise. Ensuite, les outils développés dans la partie précédente ont permis de modéliser le mouvement de tous les experts par un mouvement moyen (qu'on a appelé "mouvement nominal") et une tolérance articulaire. La comparaison d'un mouvement novice à ce modèle permet de définir deux mesures de performance : l'erreur spatiale et l'erreur temporelle. Nous estimons en effet qu'à elles deux, ces erreurs englobent la majorité des fautes commises lors d'un geste quel qu'il soit. La première concerne le positionnement des membres à chaque instant, la seconde la coordination (ou synchronie) de ceux-ci. Afin de valider les procédés ainsi formés, les bases de données de tennis et de karaté ont été annotées. Ceci a permis de confronter les erreurs obtenues avec les annotations d'experts. Dans les deux cas, une forte corrélation nous a permis de confirmer la pertinence de notre approche.

Enfin, dans un troisième et dernier temps, ce processus d'évaluation a été mis à contribution afin d'élaborer un outil d'entrainement au geste sportif. Pour ce faire, les méthodes jusqu'ici applicables hors ligne ont été adaptées pour être applicables en ligne. La prise en compte d'un outil d'acquisition différent de celui utilisé par les expert a également été discuté. Le rendu d'informations à donner au joueur est également sujet à de nombreuses analyses. Ces différentes études nous ont poussé à lui procurer un retour visuel à la fin de chacun de ses mouvements. Finalement, une démonstration a été faite auprès d'un joueur novice dont le mouvement a été enregistré à l'aide de l'outil de capture de mouvement Optitrack®.

... et perpectives

Outre ces conclusions spécifiques à chaque étude, les travaux menés au cours de cette thèse permettent de tirer plusieurs constats généraux et d'esquisser quelques perspectives à plus ou moins brève échéance.

Tout d'abord, la normalisation morphologique des squelettes est faite de façon très succincte. En particulier, les éventuelles différences morphologiques locales (c'est-à-dire relatives aux membres) n'ont pas été prises en compte. Même si ces différences peuvent paraître ici négligeables, elles pourraient très largement pénaliser le procédé dans le cas de l'évaluation du mouvement d'un enfant par exemple. Une normalisation plus performantes des squelettes pourrait être réfléchie dans le cadre de travaux futurs.

L'approche gaussienne que nous faisons des données de positionnement articulaire au sein du jeu de mouvements d'experts peut également être discutée. En effet, cette hypothèse simplifie beaucoup l'étude mais est également assez simpliste. Une première amélioration possible serait de considérer un mélange de gaussiennes par exemple. Cette approche pourrait faire ressortir des styles d'exécution, qui sont pour le moment noyés dans une représentation gaussienne commune à tous les gestes. Malheureusement, le manque de données ne nous permettait pas ici de mener à bien cette approche.

On a fait le choix dans cette thèse de ne pas considérer les articulations peu mobiles dans le calcul de l'erreur temporelle. En effet, celles-ci se révèlent parasites puisque les chemins de déformation qui leur sont associées sont très peu informatifs. Bien que le fait de comparer ces chemins semble très efficace, la pondération et le seuillage des membres mobiles pourraient *a contrario* être re-réfléchis pour s'adapter davantage à l'ensemble des cas possibles.

Concernant le prototype de recherche, une limitation majeure qui fait de lui un produit seulement quasi-fini est la qualité du retour sensoriel donné là l'utilisateur. De futurs travaux pourraient s'atteler à la mise en place d'un interfaçage plus accessible avec une adaptation du système relativement à la progression du sujet, de manière à l'associer à un vrai programme d'entrainement.

Enfin, des choix ont été pris quant au nombre de membres, à leur composition, au seuil de tolérance, *etc.* De futurs travaux pourraient faire appel à l'apprentissage profond (ou *deep learning*) pour résoudre les différents problèmes posés à plus haut niveau d'abstraction.

Bibliographie

[1] B. Mac, "Hurdle clearance," 2015. http ://www.brianmac.co.uk/hurdles/. (document), 1.6

[2] M. Müller, T. Röder, and M. Clausen, "Efficient content-based retrieval of motion capture data," *ACM Transactions on Graphics*, vol. 24, no. 3, p. 677, 2005. (document), 2.2, 2.2.1.1

[3] K. Sakurai, W. Choi, L. Li, and K. Hachimura, "Retrieval of similar behavior data using kinect data," in *14th International Conference on Control, Automation and Systems (ICCAS)*, pp. 1368–1372, IEEE, 2014. (document), 2.2.1.1, 2.3, 2.3.1.2

[4] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001. (document), 2.2.1.2, 2.4

[5] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2004. (document), 2.5, 3.1.4.2

[6] A. Sorel, R. Kulpa, E. Badier, and F. Multon, "Dealing with variability when recognizing user's performance in natural 3d gesture interfaces," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 8, p. 19, 2013. (document), 3.4.2, 3.2, 3.11, 5.2

[7] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The UCR time series classification archive," July 2015. (document), 3.4.1, 3.1

[8] S. Dubuisson and C. Gonzales, "A survey of datasets for visual tracking," *Machine Vision and Applications*, vol. 27, pp. 23–52, 2016. 1.3

[9] P. Plantard, E. Auvinet, A.-S. Pierres, and F. Multon, "Pose estimation with a kinect for ergonomic studies : Evaluation of the accuracy using a virtual mannequin," *Sensors*, vol. 15, no. 1, pp. 1785–1803, 2015. 1.3

[10] S. Noiumkar and S. Tirakoat, "Use of optical motion capture in sports science : A case study of golf swing," in *International Conference on Informatics and Creative Multimedia*, pp. 310–313, IEEE, 2013. 1.3

[11] A. Sorel, *Gestion de la variabilité morphologique pour la reconnaissance de gestes naturels à partir de données 3D.* PhD thesis, Rennes 2, 2012. 1.3, 2.1.1, 2.3.1.3, 4.2.1, 4.6.3

[12] A.-M. Burns, R. Kulpa, A. Durny, B. Spanlang, M. Slater, and F. Multon, "Using virtual humans and computer animations to learn complex motor skills : a case study in karate," *BIO Web of Conferences*, vol. 1, p. 00012, 2011. 1.3, 2.2.2, 4.1.2, 5.1.1

[13] S. Boukir and F. Chenevière, "Compression and recognition of dance gestures using a deformable model," *Pattern Analysis and Applications*, vol. 7, no. 3, pp. 308–316, 2004. 1.3, 2.2.1.1, 2.2.2

[14] C. Halit and T. Capin, "Multiscale motion saliency for keyframe extraction from motion capture sequences," *Computer Animation and Virtual Worlds*, vol. 22, no. 1, pp. 3–14, 2011. 1.3, 2.2.2

[15] G. Vaquette, C. Achard, and L. Lucat, "Information fusion for action recognition with deeply optimized hough transformed paradigm," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 423–430, 2016. 2.1.1, 2.1.3

[16] M. Barnachon, *Reconnaissance d'actions en temps réel à partir d'exemples.* PhD thesis, Lyon I, 2013. 2.1.1, 2.1.3, 2.2.2

[17] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan, "An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, pp. 1–37, Mar. 2015. 2.1.1, 2.1.4, 4.1.2

[18] B. A. Boulbaba, J. Su, and S. Anuj, "Action recognition using rate-invariant analysis of skeletal shape trajectories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2015. 2.1.1, 2.2.2, 2.3.1.2

[19] M. Dupont and P.-F. Marteau, "Gesture control system for mobile robots," in *Conference on l'Interaction Homme-Machine*, pp. 1–6, ACM Press, 2015. 2.1.1

[20] E. Coupeté, F. Moutarde, and S. Manitsaris, "A user-adaptive gesture recognition system applied to human-robot collaboration in factories," in *International Symposium on Movement and Computing*, pp. 1–7, ACM Press, 2016. 2.1.1

[21] A. C. de Carvalho Correia, L. C. de Miranda, and H. Hornung, "Gesture-based interaction in domotic environments : State of the art and hci framework inspired by the diversity," in *Human-Computer Interaction – INTERACT 2013* (D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and

M. Winckler, eds.), vol. 8118, pp. 300–317, Berlin, Heidelberg : Springer Berlin Heidelberg, 2013. 2.1.1

[22] S.-C. Huang, "An advanced motion detection algorithm with video quality analysis for video surveillance systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 1, pp. 1–14, 2011. 2.1.1

[23] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998. 2.1.1

[24] E. Esen, M. A. Arabaci, and M. Soysal, "Fight detection in surveillance videos," in *Content-Based Multimedia Indexing (CBMI)*, pp. 131–135, IEEE, 2013. 2.1.1

[25] P. C. Roy, S. Giroux, B. Bouchard, A. Bouzouane, C. Phua, A. Tolstikov, and J. Biswas, *A Possibilistic Approach for Activity Recognition in Smart Homes for Cognitive Assistance to Alzheimer's Patients.* Paris : Atlantis Press, 2011. 2.1.1

[26] S. Brault, B. Bideau, K. Kulpa, and C. Craig, "Detecting deceptive movement in 1 vs 1 based on global body displacement of a rugby player," *The International Journal of Virtual Reality*, vol. 8, no. 4, pp. 31–36, 2009. 2.1.1, 5.1.1

[27] N. Vignais, B. Bideau, C. Craig, B. S., M. F., and R. Kulpa, "Virtual environments for sport analysis : Perception-action coupling in handball goalkeeping," *The International Journal of Virtual Reality*, vol. 8, no. 4, pp. 43–48, 2009. 2.1.1, 5.1.1

[28] B. Bideau, F. Multon, R. Kulpa, L. Fradet, B. Arnaldi, and P. Delamarche, "Using virtual reality to analyze links between handball thrower kinematics and goalkeeper's reactions," *Neuroscience Letters*, no. 372, pp. 119–122, 2004. 2.1.1, 5.1.1

[29] A. Bialkowski, P. Lucey, P. Carr, S. Denman, I. Matthews, and S. Sridharan, "Recognising team activities from noisy data," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 984–990, IEEE, June 2013. 2.1.1

[30] S. M. Anzalone, E. Tilmont, S. Boucenna, J. Xavier, A.-L. Jouen, N. Bodeau, K. Maharatna, M. Chetouani, and D. Cohen, "How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+time) environment during a joint attention induction task with a robot," *Research in Autism Spectrum Disorders*, vol. 8, no. 7, pp. 814–826, 2014. 2.1.1

[31] I. Cikaljo, M. Rudolf, N. Goljar, and Z. Matjacic, "Virtual reality task for telerehabilitation dynamic balance training in stroke subjects," in *Virtual Rehabilitation International Conference*, pp. 121–125, 2009. 2.1.1, 5.1.1

[32] S. Menardais, R. Kulpa, F. Multon, and B. Arnaldi, "Synchronization for dynamic blending of motions," in *SIGGRAPH Symposium on Computer Animation*, p. 325, ACM Press, 2004. 2.1.2

[33] L. Kovar, M. Gleicher, and F. Pighin, "Motion graphs," in *SIGGRAPH Symposium on Computer Animation*, p. 1, ACM Press, 2008. 2.1.2

[34] A. Bruderlin and L. Williams, "Motion signal processing," in *SIGGRAPH '95 Proceedings of the 22nd annual conference on Computer graphics and iterative techniques*, pp. 97–104, ACM Press, 1995. 2.1.2, 4.2.1

[35] M. Unuma, K. Anjyo, and R. Takeuchi, "Fourier principles for emotion-based human figure animation," in *SIGGRAPH'95 Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 91–96, ACM Press, 1995. 2.1.2, 4.2.1

[36] R. Kulpa, F. Multon, and B. Arnaldi, "Morphology-independent representation of motions for interactive human-like animation," *Computer Graphics Forum, Eurographics 2005 special issue*, vol. 24, no. 3, pp. 343–352, 2005. 2.1.2, 4.2.1

[37] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, B. Douville, S. Prevost, and M. Stone, "Animated conversation : Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents," in *Conference on Computer graphics and interactive techniques*, pp. 413–420, 1994. 2.1.2

[38] A. S. Ramakrishnan and M. Neff, "Segmentation of hand gestures using motion capture data," in *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*, pp. 1249–1250, 2013. 2.1.3, 2.2.2

[39] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *International Journal of Computer Vision*, vol. 50, pp. 203 – 226, Nov. 2002. 2.1.3, 2.2.1.1, 4.2.1

[40] J. Barbic, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard, "Segmenting motion capture data into distinct behaviors," *Graphics Interface*, pp. 185–194, 2004. 2.1.3, 2.2.2

[41] M. Raptis, D. Kirovski, and H. Hoppe, "Real-time classification of dance gestures from skeleton animation," in *SCA '11 Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, p. 147, ACM Press, 2011. 2.1.4, 2.2.2

[42] R. E. Ward, *Biomechanical Perspectives on Classical Ballet Technique and Implications for Teaching Practice*. PhD thesis, University of New South Wales, Sydney, Australia, 2012. 2.1.4, 2.2.2, 4.1.2

[43] P.-J. Maes, D. Amelynck, and M. Leman, "Dance-the-music : an educational platform for the modeling, recognition and audiovisual monitoring of dance steps using spatiotemporal motion templates," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 35, 2012. 2.1.4, 4.1.2

[44] M. T. Pham, R. Moreau, and P. Boulanger, "Three-dimensional gesture comparison using curvature analysis of position and orientation," in *EMBC'10*, pp. 6345–6348, IEEE, 2010. 2.1.4, 2.2.1.1, 2.3.1.2, 4.1.1

[45] F. Despinoy, D. Bouget, G. Forestier, C. Penet, N. Zemiti, P. Poignet, and P. Jannin, "Unsupervised trajectory segmentation for surgical gesture recognition in robotic training," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 6, pp. 1280–1291, 2016. 2.1.4, 4.1.1

[46] S. Manitsaris, A. Glushkova, F. Bevilacqua, and F. Moutarde, "Capture, modeling, and recognition of expert technical gestures in wheel-throwing art of pottery," *Journal on Computing and Cultural Heritage*, vol. 7, no. 2, pp. 1–15, 2014. 2.1.4

[47] N. Rasamimanana and F. Bevilacqua, "Effort-based analysis of bowing movements : Evidence of anticipation effects," *Journal of New Music Research*, vol. 37, no. 4, pp. 339–351, 2008. 2.1.4

[48] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973. 2.2.1.1

[49] W. Ding, K. Liu, F. Cheng, and J. Zhang, "Stfc : Spatio-temporal feature chain for skeleton-based human action recognition," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 329–337, 2015. 2.2.1.1

[50] A. Delaye and E. Anquetil, "Hbf49 feature set : A first unified baseline for online symbol recognition," *Pattern Recognition*, vol. 46, pp. 117–130, Jan. 2013. 2.2.1.1

[51] Y. Boulahia, E. Anquetil, R. Kulpa, and F. Multon, "Hif3d : Handwriting-inspired features for 3d skeleton-based action recognition," in *23rd International Conference on Pattern Recognition*, 2016. 2.2.1.1

[52] X. Yang and Y. Tian, "Effective 3d action recognition using EigenJoints," *Journal of Visual Communication and Image Representation*, vol. 25, pp. 2–11, Jan. 2014. 2.2.1.1, 2.2.1.1

[53] M. Ziaeefard and H. Ebrahimnezhad, "Hierarchical human action recognition by normalized-polar histogram.," in *ICPR*, pp. 3720–3723, IEEE Computer Society, 2010. 2.2.1.1

[54] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in *2nd International Workshop on Human Activity Understanding from 3D Data (HAU3D) in conjunction with IEEE CVPR*, pp. 20–27, IEEE, 2012. 2.2.1.1

[55] M. Müller and T. Röder, "Motion templates for automatic classification and retrieval of motion capture data," *Proceedings of the 2006 ACM SIGGRAPH*, pp. 137–146, 2006. 2.2.1.1

[56] T. Röder, *Similarity, Retrieval, and Classification of Motion Capture Data.* PhD thesis, Rheinischen Friedrich- Wilhelms- Universität, Bonn, 2006. 2.2.1.1

[57] T. Komura, B. Lam, R. W. H. Lau, and H. Leung, "e-learning martial arts," in *Advances in Web Based Learning – ICWL 2006* (W. Liu, Q. Li, and R. W.H. Lau, eds.), vol. 4181, pp. 239–248, Springer Berlin Heidelberg, 2006. 2.2.1.1, 2.2.2, 4.1.2

[58] K. Onuma, C. Faloutsos, and J. K. Hodgins, "Fmdistance : A fast and effectiv distance function for motion capture data," *Eurographics*, 2008. 2.2.1.1

[59] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007. 2.2.1.2

[60] D. Weinland and E. Boyer, "Action recognition using exemplar-based embedding," in *Conference on Computer Vision and Pattern Recognition*, pp. 1–7, IEEE, 2008. 2.2.1.2

[61] M. Jain, H. Jégou, and P. Bouthemy, "Better exploiting motion for better action recognition," in *International Conference on Computer Vision and Pattern Recognition*, pp. 2555–2562, IEEE, 2013. 2.2.1.2

[62] L. Wang and D. Suter, "Recognizing human activities from silhouettes : Motion subspace and factorial discriminative graphical model," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, 2007. 2.2.1.2

[63] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in *International Conference on Computer Vision - Volume 2*, pp. 726–733, IEEE, 2003. 2.2.1.2

[64] I. Laptev, B. Caputo, C. Schüldt, and T. Lindeberg, "Local velocity-adapted motion events for spatio-temporal recognition," *Computer Vision and Image Understanding*, vol. 108, no. 3, pp. 207–229, 2007. 2.2.1.3, 2.3.1.1, 2.3.2.1

[65] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proceedings of the 14th International Conference on Computer Communications and Networks*, pp. 65–72, IEEE, 2005. 2.2.1.3

[66] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM International Conference on Multimedia*, pp. 357–360, 2007. 2.2.1.3

[67] C. Larboulette and S. Gibet, "A review of computable expressive descriptors of human motion," in *Proceedings of the 2nd International Workshop on Movement and Computing*, pp. 21–28, ACM Press, 2015. 2.2.1.3

[68] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Sequence of the most informative joints (smij) : A new representation for human skeletal action recognition," in *Journal of Visual Communication and Image Representation*, pp. 8–13, IEEE, 2012. 2.2.2

[69] H. Pazhoumand-Dar, C.-P. Lam, and M. Masek, "Joint movement similarities for robust 3d action recognition using skeletal data," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 10–21, July 2015. 2.2.2

[70] Y. Jiang, I. Hayashi, M. Hara, and S. Wang, "Three-dimensional motion analysis for gesture recognition using singular value decomposition," in *IEEE International COnference on Information and Automation*, pp. 805–810, IEEE, 2010. 2.2.2

[71] M.-W. Chao, C.-H. Lin, J. Assa, and T.-Y. Lee, "Human motion retrieval from hand-drawn sketch," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 5, pp. 729–740, 2012. 2.2.2

[72] A. Veeraraghavan and A. K. R. Chowdhury, "The function space of an activity," in *Conference on Computer Vision and Pattern Recognition*, pp. 959–968, IEEE Computer Society, 2006. 2.2.2, 2.3.1.2, 4.3

[73] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo, "Reconnaissance d'actions humaines 3d par l'analyse de forme des trajectoires de mouvement," in *Compression et Représentation des Signaux Audiovisuels*, 2014. 2.2.2

[74] R. Yang and S. Sarkar, "Coupled grouping and matching for sign and gesture recognition," *Computer Vision and Image Understanding*, vol. 113, no. 6, pp. 663–681, 2009. 2.3.1.1

[75] A. Mokhber, C. Achard, and M. Milgram, "Recognition of human behavior by space-time silhouette characterization," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 81–89, 2008. 2.3.1.1

[76] B. Yi and C. Faloutsos, "Fast time sequence indexing for arbitrary lp norms," in *Proceedings of the 26th International Conference on Very Large Data Bases*, pp. 385–394, 2000. 2.3.1.2

[77] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978. 2.3.1.2, 2.3.1.2, 3.1.4.2, 5.3

[78] L. R. Rabiner, "Considerations in dynamic time warping algorithms for discrete word recognition," *The Journal of the Acoustical Society of America*, vol. 63, no. 1, pp. 575–582, 1978. 2.3.1.2

[79] G. Kang and S. Guo, "Variable sliding window DTW speech identification algorithm," in *Ninth International Conference on Hybrid Intelligent Systems*, pp. 304–307, IEEE, 2009. 2.3.1.2

[80] W. Abdulla, D. Chow, and G. Sin, "Cross-words reference template for dtw-based speech recognition systems," in *Conference on Convergent Technologies for Asia-Pacific Region*, vol. 4, pp. 1576–1579, Allied Publishers Pvt. Ltd, 2003. 2.3.1.2

[81] Ning Hu, R. Dannenberg, and G. Tzanetakis, "Polyphonic audio matching and alignment for music retrieval," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 185–188, IEEE, 2003. 2.3.1.2

[82] I. Guler and M. Meghdadi, "A different approach to off-line handwritten signature verification using the optimal dynamic time warping algorithm," *Digital Signal Processing*, vol. 18, no. 6, pp. 940–950, 2008. 2.3.1.2

[83] B. S. Raghavendra, D. Bera, A. S. Bopardikar, and R. Narayanan, "Cardiac arrhythmia detection using dynamic time warping of ECG beats in e-healthcare systems," in *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pp. 1–6, IEEE, 2011. 2.3.1.2

[84] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proceedings of the 2001 SIAM International Conference on Data Mining* (V. Kumar and R. Grossman, eds.), pp. 1–11, Society for Industrial and Applied Mathematics, 2001. 2.3.1.2, 3.1.3

[85] F. Zhou and F. D. la Torre Frade, "Canonical time warping for alignment of human behavior," in *Advances in Neural Information Processing Systems Conference (NIPS)*, December 2009. 2.3.1.2

[86] F. Zhou and F. de la Torre, "Generalized canonical time warping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2015. 2.3.1.2

[87] A. Heloir, N. Courty, S. Gibet, and F. Multon, "Temporal alignment of communicative gesture sequences," *Computer Animation and Virtual Worlds*, vol. 17, pp. 347–357, July 2006. 2.3.1.2

[88] D. Gong, G. Medioni, and X. Zhao, "Structured time series analysis for human action segmentation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1414–1427, 2014. 2.3.1.2

[89] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Doklady Akademii Nauk SSSR*, vol. 163, no. 4, pp. 845–848, 1966. 2.3.1.2

[90] M. Vlachos, D. Gunopoulos, and G. Kollios, "Discovering similar multidimensional trajectories," in *Proceedings of the 18th International Conference on Data Engineering*, pp. 673–684, 2002. 2.3.1.2

[91] L. Chen, M. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pp. 491–502, 2005. 2.3.1.2

[92] L. Chen and R. Ng, "On the marriage of lp-norms and edit distance," in *Proceedings of the Thirtieth international conference on Very large data bases*, pp. 792–803, 2004. 2.3.1.2

[93] M. Morel, R. Kulpa, A. Sorel, C. Achard, and S. Dubuisson, "Automatic and generic evaluation of spatial and temporal errors in sport motions," in *International Conférence on Computer Vision Theory and Application*, pp. 1–12, 2016. 2.3.1.2

[94] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989. 2.3.1.3, 2.3.1.3

[95] O. P. Concha, R. Y. Da Xu, Z. Moghaddam, and M. Piccardi, "Hmm-mio : An enhanced hidden Markov model for action recognition," in *CVPR*, pp. 62–69, IEEE, 2011. 2.3.1.3

[96] Y. Jiang, "An HMM based approach for video action recognition using motion trajectories," in *International Conference on Intelligent Control and Information Processing*, pp. 359–364, IEEE, 2010. 2.3.1.3

[97] C. Achard, X. Qu, A. Mokhber, and M. Milgram, "A novel approach for recognition of human actions with semi-global features," *Machine Vision and Applications*, vol. 19, no. 1, pp. 27–34, 2008. 2.3.1.3

[98] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura, "Speech parameter generation algorithms for HMM-based speech synthesis," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 1315–1318, IEEE, 2000. 2.3.1.3

[99] M. Gales and S. Young, "The application of hidden markov models in speech recognition," *Foundations and Trends® in Signal Processing*, vol. 1, no. 3, pp. 195–304, 2007. 2.3.1.3

[100] S. B. Wang, A. Quattoni, L.-P. Morency, and D. Demirdjian, "Hidden conditional random fields for gesture recognition," in *CVPR*, pp. 1521–1527, 2006. 2.3.1.3

[101] S. Zhong and J. Ghosh, "Hmms and coupled hmms for multi-channel eeg classification," in *Proceedings of the 2002 International Joint Conference on Neural Networks*, pp. 1154–1159, IEEE, 2002. 2.3.1.3

[102] K. Kahol, P. Tripathi, and S. Panchanathan, "Computational analysis of mannerism gestures," in *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 946–949 Vol.3, IEEE, 2004. 2.3.1.3

[103] V. N. Vapnik, *The nature of statistical learning theory.* Springer, 1998. 2.3.2.1

[104] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions : A local svm approach," in *International Conference on Pattern Recognition*, pp. 32–36, IEEE, 2004. 2.3.2.1

[105] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," in *International Conference on Computer Vision (ICCV)*, pp. 1–8, IEEE, 2007. 2.3.2.1

[106] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. 2.3.2.2

[107] R. Su, X. Chen, S. Cao, and X. Zhang, "Random forest-based recognition of isolated sign language subwords using data from accelerometers and surface electromyographic sensors," *Sensors*, vol. 16, no. 1, p. 100, 2016. 2.3.2.2

[108] X. Zhao, Z. Song, J. Guo, Y. Zhao, and F. Zheng, "Real-time hand gesture detection and recognition by random forest," in *Communications and Information Processing* (M. Zhao and J. Sha, eds.), vol. 289, pp. 747–755, Springer Berlin Heidelberg, 2012. 2.3.2.2

[109] J. Van Vaerenbergh, R. Vranken, L. Briers, and H. Briers, "A neural network for recognizing movement patterns during repetitive self-paced movements of the fingers in opposition to the thumb," *Journal of Rehabilitation Medicine*, vol. 33, no. 6, pp. 256–259, 2001. 2.3.2.3

[110] E. B. Pizzolato, M. dos Santos Anjo, and G. C. Pedroso, "Automatic recognition of finger spelling for LIBRAS based on a two-layer architecture," in *Symposium on Applied Computing*, p. 969, ACM Press, 2010. 2.3.2.3

[111] D. Wu, L. Pigou, P.-J. Kindermans, N. D.-H. Le, L. Shao, J. Dambre, and J.-M. Odobez, "Deep dynamic neural networks for multimodal gesture segmentation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1583–1597, 2016. 2.3.2.3

[112] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013. 2.3.2.3

[113] P. Barros, G. I. Parisi, D. Jirak, and S. Wermter, "Real-time gesture recognition using a humanoid robot with a deep neural architecture," in *International Conference on Humanoid Robots*, pp. 646–651, IEEE, Nov. 2014. 2.3.2.3

[114] P. Barros, S. Magg, C. Weber, and S. Wermter, "A multichannel convolutional neural network for hand posture recognition," in *Artificial Neural Networks and*

Machine Learning (S. Wermter, C. Weber, W. Duch, T. Honkela, P. Koprinkova-Hristova, S. Magg, G. Palm, and A. E. P. Villa, eds.), vol. 8681, pp. 403–410, Springer International Publishing, 2014. 2.3.2.3

[115] H.-I. Lin, M.-H. Hsu, and W.-K. Chen, "Human hand gesture recognition using a convolution neural network," in *International Conference on Automation Science and Engineering (CASE)*, pp. 1038–1043, IEEE, Aug. 2014. 2.3.2.3

[116] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3d convolutional neural networks," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1–7, IEEE, 2015. 2.3.2.3

[117] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *International Conference on Signal and Image Processing Applications*, pp. 342–347, IEEE, Nov. 2011. 2.3.2.3

[118] E. Keogh, T. Palpanas, V. B. Zordan, D. Gunopulos, and M. Cardle, "Indexing large human-motion databases," in *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*, pp. 780–791, VLDB Endowment, 2004. 3.1.1

[119] V. Niennattrakul and C. A. Ratanamahatana, "Shape averaging under time warping," in *6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 626–629, IEEE, 2009. 3.1.2.2, 3.4.1

[120] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, pp. 678–693, Mar. 2011. 3.1.2.2, 3.4.1

[121] A.-M. Burns, *On the Relevance of Using Virtual Humans for Motor Skills Teaching : a case study on Karate gestures*. PhD thesis, Rennes 2, 2013. 3.1.3, 5.2

[122] C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, pp. 11–22, Society for Industrial and Applied Mathematics, 2004. 3.1.4.1

[123] R. Gaudin and N. Nicoloyannis, "An adaptable time warping distance for time series learning," in *5th International Conference on Machine Learning and Aplications (ICMLA)*, pp. 213–218, IEEE, 2006. 3.1.4.1

[124] D. Yu, X. Yu, Q. Hu, J. Liu, and A. Wu, "Dynamic time warping constraint learning for large margin nearest neighbor classification," *Information Sciences*, vol. 181, no. 13, pp. 2787–2796, 2011. 3.1.4.1

[125] B. Ben Ali, Y. Masmoudi, and S. Dhouib, "Tabu search for dynamic time warping global constraint learning," in *6th International Conference of Soft Computing and Pattern Recognition*, pp. 376–381, IEEE, 2014. 3.1.4.1

[126] L. Rabiner and B. Juang, *Fundamentals of speech recognition.* Prentice-Hall, 1993. 3.1.4.2

[127] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, "Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm," *Knowledge and Information Systems*, vol. 47, no. 1, pp. 1–26, 2016. 3.4.1

[128] F. Petitjean and P. Gançarski, "Summarizing a set of time series by averaging : From Steiner sequence to compact multiple alignment," *Theoretical Computer Science*, vol. 414, pp. 76–91, Jan. 2012. 3.4.1

[129] C. E. Reiley and G. D. Hager, "Task versus subtask surgical skill evaluation of robotic minimally invasive surgery," in *Medical image computing and computer-assisted intervention – MICCAI 2009* (D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), vol. 5761, pp. 435–442, Springer-Verlag, 2009. 4.1.1

[130] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh, C. C. G. Chen, R. Vidal, S. Khudanpur, and G. D. Hager, "Jhu-isi gesture and skill assessment working set (jigsaws) : A surgical activity dataset for human motion modeling," in *Medical Image Computing and Computer-Assisted Intervention*, 2014. 4.1.1

[131] E. F. Hofstad, C. Vapenstad, M. K. Chmarra, T. Lango, E. Kuhry, and R. Marvik, "A study of psychomotor skills in minimally invasive surgery : what differentiates expert and nonexpert performance," *Surgical Endoscopy*, vol. 27, no. 3, pp. 854–863, 2013. 4.1.1

[132] N. Ahmidi, P. Poddar, J. D. Jones, S. S. Vedula, L. Ishii, G. D. Hager, and M. Ishii, "Automated objective surgical skill assessment in the operating room from unstructured tool motion in septoplasty," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, no. 6, pp. 981–991, 2015. 4.1.1

[133] Y. Sharma, T. Plotz, N. Hammerld, S. Mellor, R. McNaney, P. Olivier, S. Deshmukh, A. McCaskie, and I. Essa, "Automated surgical OSATS prediction from videos," in *International Symposium on Biomedical Imaging (ISBI)*, pp. 461–464, 2014. 4.1.1

[134] H. Pirsiavash, C. Vondrick, and A. Torralba, "Assessing the quality of actions," in *European Conference on Computer Vision*, pp. 556–571, Springer, 2014. 4.1.2

[135] V. Parameswaran and R. Chellappa, "View invariants for human action recognition," in *International Journal of Computer Vision*, vol. 66, pp. 83–101, Kluwer Academic Publishers, 2003. 4.2.1

[136] M.-S. Sie, Y.-C. Cheng, and C.-C. Chiang, "Key motion spotting in continuous motion sequences using motion sensing devices," in *IEEE International Conference on Signal Processing*, pp. 326–331, IEEE, 2004. 4.2.1

[137] C. Martin, *Biomechanical analysis of the tennis serve : relationships with performance and upper limb injuries.* Theses, Université Rennes 2, 2013. 4.5.1

[138] A. Matallaoui, J. Koivisto, J. Hamare, and R. Zarnekow, "How effective is "exergamification" ? a systematic review on the effectiveness of gamification features on exergames," in *Proceedings of the 50th Annual Hawaii International Conference on System Sciences (HICSS)*, pp. 3316–3325, 2017. 5.1.1

[139] N. Ukita, D. Kaulen, and C. Röcker, "Towards an automatic motion coaching system - feedback techniques for different types of motion errors," in *Proceedings of the International Conference on Physiological Computing Systems*, pp. 167–172, 2014. 5.1.1

[140] P. Hawkins, M. Hawken, and G. Barton, "Effect of game speed and surface perturbations on a postural control in a virtual environment," in *Proceeding of the 7th ICDVRAT*, pp. 311–318, 2008. 5.1.1

[141] G. J. Barton, M. B. Hawken, R. G. Foster, G. Holmes, and P. B. Butler, "The effects of virtual reality game training on trunk to pelvis coupling in a child with cerebral palsy," *Journal of NeuroEngineering and Rehabilitation*, vol. 10, no. 1, p. 15, 2015. 5.1.1

[142] M. Morel, B. Bideau, J. Lardy, and R. Kulpa, "Advantages and limitations of virtual reality for balance assessment and rehabilitation," *Clinical Neurophysiology*, vol. 45, no. 4-5, pp. 315–326, 2015. 5.1.1

[143] R. Ranganathan and L. Carlton, "Perception-action coupling and anticipatory performance in baseball batting," *Journal of Motor Behavior*, vol. 39, no. 5, pp. 369–380, 2007. 5.1.1

[144] G. Watson, S. Brault, R. Kulpa, B. Bideau, J. Butterfield, and C. Craig, "Judging the 'passability' of dynamic gaps in a virtual rugby environment," *Human Movement Science*, vol. 30, no. 5, pp. 952–956, 2010. 5.1.1

[145] J. Von Zitzewitz, P. Wolf, V. Novaković, M. Wellner, G. Rauter, A. Brunschweiler, and R. Riener, "Real-time rowing simulator with multimodal feedback," *Sports Technology*, vol. 1, no. 6, pp. 257–266, 2008. 5.1.1

[146] A. M. Wing, M. Doumas, and A. E. Welchman, "Combining multisensory temporal information for movement synchronisation.," *Experimental brain research*, vol. 200 3-4, pp. 277–82, 2010. 5.1.1

[147] A. Salmoni, R. Schmidt, and C. Walter, "Knowledge of results and motor learning. a review and critical reappraisal," *Psychological Bulletin*, vol. 95, no. 3, pp. 355–386, 1984. 5.1.2

[148] A. J. Kovacs, J. Boyle, N. Grutmatcher, and C. H. Shea, "Coding of on-line and pre-planned movement sequences," *Acta Psychologica*, vol. 133, no. 2, pp. 119–126, 2010. 5.1.2

[149] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic and mutlimodal feedback in motor learning : A review," *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 21–53, 2013. 5.1.2

[150] S. Swinnen, A. Richard, D. Nicholson, and D. Shapiro, "Information feedback for skill acquisition : Instantaneous knowledge of results degrades learning," *Journal of Experimental Psychology : Learning, Memory and Cognition*, vol. 16, no. 4, pp. 706–716, 1990. 5.1.2

[151] D. C. Ribeiro, G. Sole, J. H. Abbott, and S. Milosavljevic, "Extrinsic feedback and management of low back pain : A critical review of the literature," *Manual Therapy*, vol. 16, no. 3, pp. 231–239, 2011. 5.1.2

[152] P. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, and R. Pausch, "Training for physical tasks in virtual environments : Tai chi," in *Virtual Reality Conference*, pp. 87–94, 2003. 5.1.2

[153] M. Müller, *Information retrieval for music and motion : with 26 tables*. Berlin : Springer, 2007. 5.2.1, 5.2.1.2

[154] X. Anguera and M. Ferrarons, "Memory efficient subsequence dtw for query-by-example spoken term detection," in *International Conference on Multimedia and Expo*, pp. 1–6, IEEE, 2013. 5.2.1

[155] Y. Sakurai, C. Faloutsos, and M. Yamamuro, "Stream monitoring under the time warping distance," in *International Conference on Data Engineering*, pp. 1046–1055, IEEE, 2007. 5.2.1, 5.2.1.2

[156] M. Toyoda and Y. Sakurai, "Subsequence mtching in data streams," *NTT Technical Review*, vol. 11, no. 1, 2013. 5.2.1.2