Analysis of locomotion and human activity in Free-Living Environments

Sylvain Jung

UNIVERSITE PARIS XIII –SORBONNE PARIS NORD

École doctorale Sciences, Technologies, Santé Galilée

Analyse de la marche et de l'activité humaine dans des environnements libres

Analysis of locomotion and human activity in Free-Living Environments

---

THÈSE DE DOCTORAT
présentée par

Sylvain JUNG

L2TI Laboratoire de Traitement et Transport de l'Information

pour l'obtention du grade de
DOCTEUR en SCIENCES POUR L'INGENIEUR

soutenue le 22 Mars 2023 devant le jury d'examen constitué de :

MOKRAOUI Anissa, Laboratoire de Traitement et Transport de l'Information L2TI Université Sorbonne Paris Nord, Présidente du jury

LE BOUQUIN JEANNES Régine, Université de Rennes 1, Ecole Supérieure d'ingénieurs de Rennes, Rapporteuse

FLEURY Anthony, IMT Nord Europe, Rapporteur

VIENNET Emmanuel, Laboratoire de Traitement et Transport de l'Information L2TI Université Sorbonne Paris Nord, Examinateur

OUDRE Laurent, Centre Borelli ENS Paris Saclay, Directeur de thèse

RICARD Damien, Centre Borelli ENS Paris Saclay, Hôpital d'Instruction des Armées HIA Percy, Co-directeur de thèse

DORVEAUX Eric, Abilycare, Membre invité

GORINTIN Louis, Novakamp, Membre invité

Contents

Remerciements

Mes premiers remerciements s'adressent à Laurent qui a permis d'initier cette thèse, a été un soutien essentiel au cours de ces 3 dernières années et qui m'a donné le goût de la recherche en me montrant à quel point elle pouvait être enrichissante et gratifiante ! Merci encore Laurent, je n'aurais pas pu demander un directeur de thèse plus professionnel, humain et attentif que toi ! Un grand Merci aussi à Damien pour sa bienveillance, sa disponibilité dans les moments charnières et pour ses conseils toujours parfaitement avisés. Ce fut un plaisir immense de travailler avec vous deux. Merci aux membres de mon jury de thèse pour leur relecture rigoureuse et précise, leurs remarques pertinentes, le déplacement pas évident vers le lieu de soutenance et pour les conversations captivantes sur le sujet de la thèse.

Je remercie Eric et Louis pour leur aide et leur soutien sur cette thèse. Merci de m'avoir montré à quoi pouvait ressembler l'écosystème d'un département de recherche.

Merci aux équipes du Centre Borelli avec lesquels j'ai passé de très bons moments et qui m'ont considérablement aidé au cours de ma thèse. Je vous remercie de m'avoir intégré dans votre équipe dynamique, passionnée et talentueuse ! Je souhaite à tout doctorant de faire partie d'un groupe qui pousse autant vers le haut. Merci aux doctorants Sylvain, Alexandre, Sam, Thibaut (j'en oublie évidemment j'en suis désolé) et à tous les autres : Charles, Antoine…

Merci tout particulièrement à l'équipe de Percy qui m'a été d'une aide immense lors de mes prises de mesure et pour tout le reste. Vous avez été formidables et vous formez une équipe motivante et géniale ! Merci Nicolas, Cyril, Damien, et Mona (un grand Merci à toi Mona, tu as été impressionnante et je pense que la suite te réserve pleins de belles surprises !).

Merci au L2TI pour son accueil chaleureux et particulièrement à Anissa pour sa gentillesse bienveillante !

Merci enfin à tous mes amis qui m'ont été d'un soutien indéfectible et qui m'ont offert des moments de décompression salvateurs : Keke, Sab, Marion, Thom, la team du Valois... Vous êtes les meilleurs. Un gigantesque Merci à ma famille (et j'inclue évidemment la belle famille EVEN et plus spécialement Jean-Félix et Michèle) qui continue à être cette bulle d'amour dont j'ai désespérément besoin. Vous avez été merveilleux, comme d'hab. En ma qualité de surémotif contenu, je me suis appuyé sur vous lourdement pour passer les moments compliqués, je vous aime ! Un grand Merci à toi ma Lana adorée : si tu devais un jour lire ses lignes, sache que ton arrivée fut un précieux cadeau et une douce folie motivante pour moi. Sur ces 3 ans, je n'ai eu de cesse de m'émerveiller de la puissance de l'apport affectif que vous m'avez tous offert.

Merci enfin à la plus belle découverte de ma vie : Meryl. Merci pour ces soirées tranquilles, ces voyages intenses vers lesquels tu m'as heureusement poussé, merci d'avoir

contenu mes ronchonnades, d'avoir supporté toutes mes divagations trop longues, merci pour les tapis rouges, merci pour nos craquages de rire, merci pour les séances cinéma, pour les discussions passionnantes, pour tous ces moments sans prétention et pourtant essentiels, merci d'avoir été si parfaite.

Résumé

Ce travail présente une nouvelle approche permettant de produire un récapitulatif visuel de l'activité physique d'un sujet à l'issue d'un protocole utilisant des capteurs inertiels que nous avons implémenté dans des environnements libres. Grâce à cette nouvelle méthode de visualisation, le comportement humain, en particulier la marche, peut maintenant être présenté de manière condensée sous une forme facile à lire et à interpréter. Les séries temporelles collectées lors du suivi des patients dans des environnements de vie libre étant souvent longues et complexes, notre contribution s'appuie sur une chaîne de traitement innovante composée de méthodes de traitement du signal et d'algorithmes d'apprentissage automatique. Une fois établie, la représentation graphique finale permet de résumer les d'activité présentes dans les données et peut être rapidement appliquée à des séries temporelles nouvellement acquises. En résumé, les objectifs de cette thèse sont : (i) la segmentation des données brutes issues des centrales inertielles en régimes homogènes avec une procédure de détection de ruptures adaptative, (ii) la classification de chaque régime selon ses caractéristiques, (iii) l'extraction de paramètres de chaque régime de marche, (iv) le calcul d'un score à partir de ces caractéristiques et (v) la construction du résumé visuel final, à partir des scores des activités et de leurs comparaisons à des modèles sains. Ce retour graphique est une visualisation détaillée, adaptative et structurée qui aide à mieux comprendre les événements marquants d'un protocole de marche complexe.

La première contribution de cette thèse est une revue sélective des études en environnement libre utilisant des capteurs inertiels pour évaluer l'activité physique. La deuxième contribution consiste en la mise en œuvre de trois protocoles (deux dans des environnements semi-libres et un dans un environnement libre) basés sur la revue de littérature précédente. La troisième contribution est une méthode adaptative de détection de ruptures qui s'applique à des signaux transformés dans le domaine temps-fréquence et qui est basée sur l'apprentissage de pénalité. La quatrième contribution de cette thèse est une méthode de caractérisation qui classe les régimes segmentés en des régimes de marche ou en des régimes associés à des activités sédentaires ou non sédentaires. La dernière contribution est le rendu du retour visuel final.

Abstract

This work presents a novel approach to create a graphical summary of a subject's physical activity during a protocol in free-living environments that relies on the use of inertial sensors. Thanks to this new visualization, human behaviour, in particular locomotion, can now be condensed in a easy-to-read and user-friendly output. As time series collected while monitoring patients in free-living environments are often long and complex, our contribution relies on an innovative pipeline of signal processing methods and machine learning algorithms. Once learned, the graphical representation is able to sum up all activities present in the data, and can quickly be applied on newly acquired time series. In a nutshell, the objectives of this thesis are : (i) the segmentation of raw data from Inertial Measurement Units into homogeneous regimes with an adaptive change-point detection procedure, (ii) the classification of each regime according to its characteristics, (iii) the extraction of features from each walking regime, (v) the computation of a score using these features and (iv) the construction of the final visual summary from the scores of the activities and their comparisons to healthy models. This graphical output is a detailed, adaptive and structured visualization which helps better understand the salient events in a complex gait protocol.

The first contribution of this thesis is a selective review of studies in free environments using inertial sensors to asses physical activity. The second contribution consists in the implementation of three protocols (two in Semi-Free Living Environments and one in a Free-Living Environment) based on the previous literature review. The third contribution is an adaptive change point detection method which applies signals transformed into the time-frequency domain and which is based on penalty learning. The fourth contribution of this thesis is a characterisation method that classifies segmented regimes into walking regimes or into regimes associated to sedentary or non-sedentary activities. The last contribution is the rendering of the final visual feedback.

Acronyms

6MWT 6 Min Walking Test. 69

COPD Chronic Obstructive Pulmonary Disease. 45, 46, 49, 62

DT Decision Trees. 109

EE Energy Expenditure. 45, 47, 49, 64

FLE Free-Living Environment. 41, 43, 69, 150

FLEs Free-Living Environments. 18, 37, 43, 45, 47, 50–52, 54–62, 64–66, 69, 70, 103, 106, 114, 130, 148, 150, 151

HAR Human Activity Recognition. 39, 45–47, 49, 55, 60, 63–65, 106–108, 114, 126, 129, 151

IMU Inertial Measurement Unit. 41, 70, 87, 114, 115

IMUs Inertial Measurement Units. 35, 46, 47, 50–54, 57, 65, 71, 107, 130, 150

IOU Intersection Over Union. 85, 93, 96, 98

KNN K-Nearest Neighbor. 109

MET Metabolic Equivalent of Task. 49, 64

MLP Multi-Layer Perceptron. 109

MS Multiple Sclerosis. 45, 46

NB Naive Bayes. 109

NN Neural Network. 109

OPT Optimal Partitioning Method. 83, 84, 87

PA Physical Activity. 45–47, 50, 53, 54, 64, 65, 69, 70, 106, 107

PCA Principal Component Analysis. 108, 114, 126, 151

PD Parkinson's Disease. 45, 46, 49, 62

PELT Pruned Exact Linear Time. 83, 84, 87, 91, 102

RF Random Forest. 109

semi-FLE Semi-Free-Living Environment. 18, 38, 69

semi-FLEs Semi-Free-Living Environments. 18, 37, 38, 41, 43, 45, 47, 50, 52, 55–62, 64, 65, 69, 70, 89, 103, 150

SMA Signal Magnitude Area. 108

STFT Short Term Fourier Transform. 87

VM Vector Magnitude. 108

WB Walking Bout. 105, 106

WBs Walking Bouts. 105

List of Figures

List of Tables

Résumé en français

Contexte de la thèse

Contexte général

Au cours des dernières décennies, l'emploi de capteurs inertiels embarqués (centrales inertielles, smartphones ou accéléromètres simples) pour quantifier l'activité physique et notamment la locomotion des individus s'est généralisé [116, 157]. En effet, plusieurs solutions commerciales ont vu le jour afin de fournir aux utilisateurs la possibilité de visualiser l'évolution de leur activité physique notamment via des applications d'auto surveillance [158]. Dans le contexte de la recherche, l'utilisation de ces capteurs inertiels légers et peu onéreux s'est elle aussi intensifiée : plusieurs études cliniques s'appuient sur les mesures d'accélération et de vitesse angulaire sur différentes positions du corps humain fournies par ces capteurs afin par exemple d'entraîner des algorithmes de classification qui permettront d'évaluer plus précisément les dépenses énergétiques des individus sur une période donnée [74]. Ces capteurs permettent par ailleurs d'identifier des marqueurs quantifiés de la chute [133] ainsi que de fournir un suivi innovant de l'influence des traitements d'un sujet sur ses mouvements par exemple [23]. En outre, l'utilisation de ces capteurs permet de proposer des diagnostics précoces, de rendre plus abordables les coûts des suivis de santé, de définir des marqueurs de fragilité [1] etc.. Les conditions d'utilisation de ces capteurs (récupération des données, filtres, extraction de paramètres) et leurs limites dans le cadre des études en milieu clinique - c'est-à-dire dans des environnements contrôlés (hôpital, laboratoire…) - sont à présent largement documentées dans la littérature.

Ces analyses peuvent par ailleurs s'appliquer en ambulatoire (dans un environnement libre non supervisé) : une telle approche permet de quantifier la marche d'un individu de manière plus représentative puisqu'elle évite le syndrome d'Hawthorne ou de blouse blanche [132]. Ce syndrome intervient lorsque les sujets mesurés réalisent qu'ils sont observés et mesurés au cours d'une acquisition. Leur approche d'un protocole au cours d'un test clinique peut ainsi être altérée et la fiabilité de l'interprétation des paramètres caractérisant leur activité physique peut être réduite. Une analyse en environnement libre permet donc d'atténuer cet effet clinique. Néanmoins, l'analyse en ambulatoire (dans des environnements libres sans supervision) fait face à de nombreux défis : ce type d'étude s'est longtemps basé sur des questionnaires déclaratifs où le sujet répertoriait ses activités quotidiennes, leur durée et éventuellement quelques données qualitatives (ressenti, niveau de stress) [45]. Quant aux analyses quantitatives existantes, des compromis entre la précision des méthodes, le coût de calcul imposés par l'utilisation de longs signaux, l'encombrement des capteurs sur de longues périodes de mesure, la complexité des retours quantifiés des analyses de signaux leur sont imposés.

Avant de pouvoir utiliser les capteurs inertiels décrits précédemment pour des études en ambulatoire, il convient de concevoir une approche itérative et cohérente permettant

de récupérer les signaux complexes et non stationnaires produits par ces études et de les traiter pour aboutir à un retour quantitatif et lisible de l'activité physique d'un sujet. Cette transition des données brutes vers des données exploitables et interprétables est un enjeu crucial de la médecine actuelle.

Deux cohortes de sujets sont particulièrement étudiées dans ce manuscrit : les patients cérébro-lésés et les patients ayant subi des chirurgies orthopédiques. En effet, les lésions cérébrales et médullaires ont des conséquences graves sur la marche d'un patient et provoquent des incidents imprévus qui se déroulent dans son environnement naturel. Par ailleurs, au cours de leur rééducation, les patients ayant subi des accidents vasculaires cérébraux, par exemple, ont une forte propension à chuter [121]. L'étude de l'activité physique d'un patient atteint d'une telle pathologie par l'enregistrement de plusieurs signaux physiologiques et leur analyse permettrait une surveillance contrôlée de l'état des patients, de compléter et de préciser les suivis déjà effectués en tests cliniques. En outre, les séjours en chirurgie traumatologique et orthopédique ont représenté une dépense inédite de plusieurs milliards d'euros sur ces dernières années . La chirurgie orthopédique et traumatologique fait d'ailleurs partie des premiers postes de dépense parmi les services de chirurgie [54] . La possibilité de suivre longitudinalement des patients ayant subi une chirurgie ostéoarticulaire revêt donc une importance conséquente et pourrait limiter les coûts décrits auparavant en permettant d'anticiper des réhospitalisations chez certains patients.

Collaboration avec ENGIE Lab CRIGEN et ABILYCARE

Au cours de ma thèse CIFRE, j'ai travaillé avec le laboratoire Nanotechnologies NANO et Capteurs de l'ENGIE Lab CRIGEN (centre de recherche, de développement et d'expertise du Groupe ENGIE dédié aux nouvelles utilisations de l'énergie dans les villes). Dans un contexte où l'hôpital s'ouvre à la ville (ambulatoire, suivi postopératoire ...) ENGIE souhaite faire émerger des solutions innovantes dans le secteur de la e-santé et proposer de nouveaux services pour le patient, les professionnels et les exploitants. En effet, le système de soins est confronté aujourd'hui à plusieurs difficultés majeures parmi lesquelles la gestion de la dépendance, l'accès universel à une prise en charge de qualité ou l'accroissement significatif des dépenses. Ces difficultés peuvent être en partie relevées grâce aux technologies numériques qu'ENGIE souhaite mettre au service de la surveillance en ambulatoire et donc notamment aux défis que cette surveillance occasionne comme expliqué précédemment. En amont de cette thèse, le laboratoire NANO avait donc déjà développé en collaboration avec le centre Borelli et l'entreprise ABILYCARE un projet d'évaluation dynamique et statique de l'activité physique des individus : le projet SMARTCHECK. Ce projet est d'ailleurs aujourd'hui commercialisé par ABILYCARE. Cette évaluation a pour but de fournir un retour quantifié et visuel à des cliniciens qui souhaiteraient évaluer l'activité physique de leurs patients dans un environnement contrôlé. Elle s'appuie sur une mesure de 10 mètres aller retour à l'aide de plusieurs centrales inertielles pour l'aspect dynamique et sur un test d'équilibre sur une plateforme de force pour l'aspect statique. Ce projet s'inscrit dans une démarche d'expansion des activités de l'ENGIE Lab CRIGEN aux enjeux de santé généraux notamment aux aspects de médecine du travail. Cette thèse est le prolongement de cette démarche et doit permettre à l'ENGIE Lab de CRIGEN de s'inscrire plus encore comme un acteur des méthodes innovantes d'e-santé. ENGIE s'attache particulièrement à rendre ce

suivi possible aux praticiens opérant dans des hôpitaux en leur proposant d'identifier par exemple des risques de réhospitalisation chez leurs patients à la suite de leur séjour dans un établissement de santé.

Collaboration avec le Centre BORELLI

Au cours de ma thèse, j'ai collaboré avec le Centre BORELLI, un laboratoire incluant des chercheurs en médecine, des chercheurs en mathématiques appliquées ainsi que des professionnels de santé tels que des neurologues, des chirurgiens ou des anesthésistes. La mission fondamentale de ce laboratoire est d'associer ces différentes expertises afin de quantifier le comportement humain et animal à l'aide de capteurs légers, et d'algorithmes de traitement du signal et d'apprentissage statistique. Pour ce faire, les chercheurs du Centre BORELLI conçoivent et implémentent plusieurs protocoles dans des cadres cliniques en utilisant notamment des capteurs inertiels ainsi que des capteurs d'oculométrie. Ces protocoles permettent d'enregistrer des signaux physiologiques (univariés ou multivariés) qui sont au cœur des travaux du Centre BORELLI. Un premier défi pour le Centre est d'extraire des informations permettant une analyse de ces signaux physiologiques. Le second est de mettre en place une automatisation du traitement de ces données afin de pouvoir fournir aux cliniciens des méthodes innovantes de suivi longitudinal des sujets enregistrés ainsi que la possibilité d'effectuer des comparaisons interindividuelles entre les sujets.

Collaboration avec le Laboratoire du Traitement et du Transport de l'Information L2TI

Au cours de ma thèse, j'ai aussi collaboré avec les équipes du Laboratoire de Traitement et de Transport de l'Information L2TI. Ce laboratoire réunit des équipes de recherche sur deux aspects principaux donc : le traitement des données et le transport de l'information notamment au sein des réseaux. Un des enjeux de l'équipe de recherche est d'étendre son savoir faire à plusieurs domaines existants pour le traitement des données notamment.

Motivations et enjeux : questions scientifiques

Comment mettre en place un protocole adapté à un environnement libre ?

Les études en environnement libre référencées dans la littérature sont variées et proposent des objectifs nombreux. Les études peuvent par exemple différer dans les instructions données aux participants (liberté complète dans la réalisation des activités ou production d'une liste et d'une chronologie détaillée), dans les moyens d'annotations (présence ou non d'un examinateur, utilisation d'une caméra externe ou portable, utilisation d'un agenda que le sujet doit remplir lui-même), dans les spécificités des environnements où les prises de mesure ont lieu (environnement libre complet, campus d'université, hôpital) etc.. Deux types d'environnement couramment implémentés dans les travaux ambulatoires que l'on retrouve dans la littérature se dégagent notamment : les environnements semi libres où certaines conditions de la prise de mesure sont contrôlées par un examinateur et les environnement entièrement libres qui sont exclusivement les environnements naturels des sujets mesurés. Dans cette thèse, l'un des premiers défis était de regrouper ces différentes

observations effectuées dans la littérature dans une revue étayée afin de notamment mieux définir les environnements semi libres et les environnements entièrement libres. Cette étude bibliographique a permis de définir les enjeux de mise en place des dispositifs de mesure dans ces environnements afin de pouvoir construire des protocoles de mesure de l'activité physique qui soient valides dans un environnement libre. Cette étude bibliographique a également permis de cibler au mieux le type de capteurs à utiliser en milieu libre (capacité de stockage des données et autonomie importantes).

Comment analyser les signaux récupérés en sortie des protocoles en environnement libre ?

Dans le contexte des études en environnements libres, les signaux physiologiques récupérés sont complexes : ils sont multivariés, plus longs que des signaux physiologiques récupérés dans des cadres cliniques et non stationnaires. la figure 0.1 montre des signaux d'accélération mesurés dans des environnements semi libres : ces signaux sont composés d'une succession de phases homogènes, correspondant aux différentes phases du protocole. La segmentation de signal ou la détection de ruptures intervient alors comme une étape importante afin de pouvoir dégager des régimes homogènes dans les signaux physiologiques qui permettront une meilleure analyse en sortie. Ces algorithmes permettent, à partir d'une définition mathématique de la notion de changement, de repérer automatiquement ces changements dans des signaux. Un des défis de cette thèse est d'utiliser ces approches pour le traitement des signaux récupérés lors des protocoles implémentés en environnement libre. Ces méthodes permettent de produire une analyse pertinente des informations transportées par ces signaux. La segmentation pourrait être effectuée par un expert : cela est notamment souvent le cas lors d'études effectuées dans des cadres cliniques. Néanmoins, cette annotation manuelle induit des tâches longues et fastidieuses qui doivent être répétées sur chaque enregistrement ce qui constitue un frein au déploiement des méthodes notamment lorsque les annotations s'effectuent sur des signaux récupérés dans un environnement libre. Une méthode de segmentation automatique apparaît donc nécessaire afin d'éviter cette complexité importante dans le contexte de cette thèse.

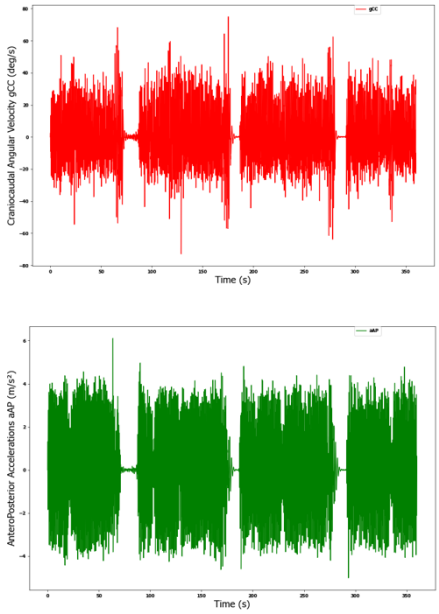
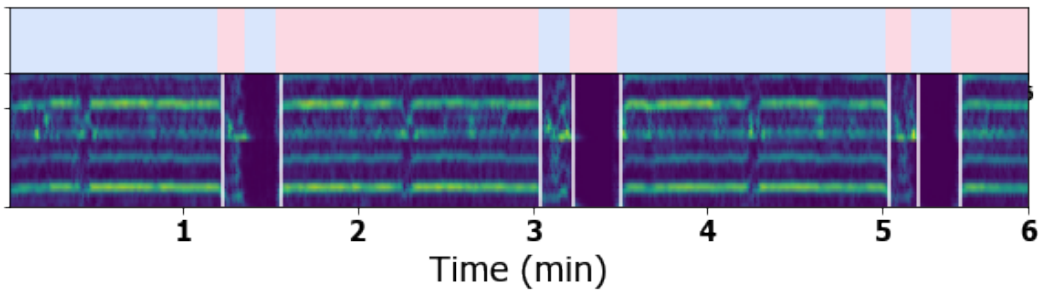


Figure 0.1: Signaux filtrés récupérés dans le bas du dos de participants enregistrés dans un environnement semi libre : vitesse angulaire craniocaudale et accélération antéro-postérieure.

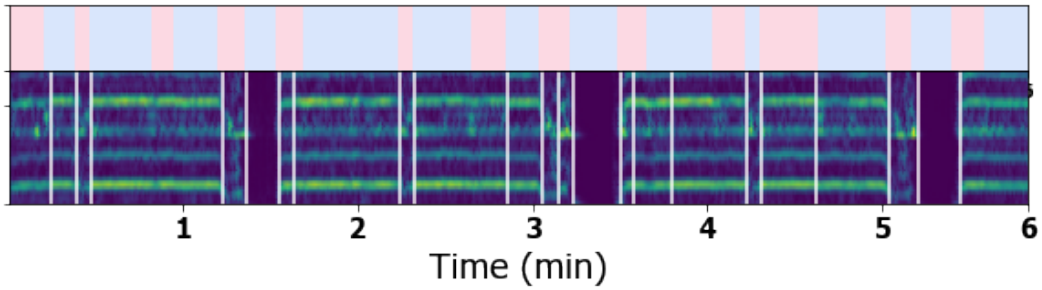
La conception d'une méthode automatisée de détection de ruptures passe notamment par la spécification du type de ruptures que l'on cherche, et du niveau de granularité attendu. Un défi de cette thèse est donc de définir les types de ruptures qui doivent être détectés en mettant en place une méthode supervisée. L'approche utilisée dans cette étude permet en effet, à partir de quelques données annotées, d'isoler des régimes particuliers sur les signaux. Des méthodes non supervisées de détection de ruptures sont parfois utilisées. Néanmoins, dans le contexte de cette étude effectuée dans un environnement libre plus chaotique qu'un environnement clinique, une définition au préalable des types de rupture permet de cibler plus efficacement des types spécifiques de régimes homogènes. Il s'agit donc de pouvoir développer une méthode adaptative de segmentation. Le niveau de granularité de cette détection est contrôlé pour que seules les ruptures annotées soient détectées. Une telle méthode d'apprentissage qui permet de répliquer une stratégie d'annotation sur de nouvelles données est développée dans cette thèse.

Comment caractériser des régimes homogènes segmentés et notamment les régimes de marche ?

Après avoir appliqué notre méthode de segmentation adaptative sur nos signaux, nous récupérons des régimes homogènes segmentés auxquels nous allons donner des labels afin de les caractériser (régimes de marche, d'activité sédentaire, d'activité non sédentaire). Dans la littérature, les portions de signaux physiologiques peuvent être classées selon certains labels prédéfinis (portions de marche, d'activité sédentaire, de course...). Cette caractérisation des signaux permet plusieurs interprétations cliniques. En effet, cela permet d'évaluer par exemple le taux d'énergie dépensée par un sujet dans une journée, l'évolution de l'activité physique d'un patient après l'administration d'un traitement etc...La grande



((a)) Stratégie d'annotation 1



((b)) Stratégie d'annotation 2

Figure 0.2: Évolution des résultats de la détection de ruptures pour un participant témoin en fonction des annotations données en entrée. Les lignes verticales correspondent aux moments où la méthode détecte un point d'arrêt. La partie inférieure des visuels correspond au spectrogramme construit à partir des signaux physiologiques : les différences des régimes dans leur signature spectrale sont visibles. La succession de couleur dans la partie supérieure correspond aux régimes annotés.

majorité des autres études sont des méthodes de classification d'activité qui permettent de caractériser tous les régimes segmentés et non pas uniquement les régimes de marche. Ces approches s'appliquent à des fenêtres glissantes ce qui peut induire un coût computationnel élevé lorsqu'elles sont utilisées sur des longs signaux comme ceux récupérés dans un environnement libre. Un défi de la thèse est de proposer une approche de caractérisation des régimes homogènes isolés sur des signaux longs récupérés dans des environnements libres en évitant les écueils décrits auparavant. L'originalité de notre méthode de classification repose ainsi sur sa condition d'application aux signaux : elle s'applique à des régimes pré-segmentés et non à des portions de signaux sélectionnés par des fenêtres glissantes comme cela est utilisé couramment par les études de la littérature. Cette méthode classe tout d'abord les régimes segmentés en des régimes de marche ou de non marche avant de classer les régimes qui ne sont pas des régimes de marche en des régimes d'activité sédentaire ou d'activité non sédentaire.

Comment construire un outil graphique innovant pour fournir un suivi quantifié de l'activité physique ?

Dans presque tous les travaux cités ci-dessus, les mesures en sortie fournies aux cliniciens sont souvent généralisées/agrégées grâce à des caractéristiques simples adaptées aux objectifs des études (temps passé dans diverses activités ciblées, évaluation de la dépense énergétique à partir du temps passé dans des activités plus ou moins énergivores, nombre de chutes, temps de chute...). Ces mesures de sortie agrégées et moyennées peuvent cacher certains phénomènes pertinents d'intérêt. Une autre approche serait d'utiliser des caractéristiques d'une granularité plus fine telles que celles utilisées en milieu clinique. Cependant, cela impliquerait un plus grand coût de calcul (par exemple sur la détection de tous les pas, des foulées...) et un flux d'informations excessif qui submergerait les cliniciens. Cela les empêcherait d'obtenir une évaluation claire et rapidement compréhensible de l'activité physique de leurs patients. Partant de ce constat, un des défis de cette thèse est d'apporter une solution alternative et intermédiaire permettant une macro-analyse précise, à un faible coût de calcul, ergonomique pour les cliniciens et qui conserve la structure temporelle. Un objectif de cette étude est donc d'extraire pour chaque régime de marche identifié des paramètres cohérents permettant cette macro-analyse. Ainsi, il serait possible de construire des modèles de comparaison sur des sujets sains afin d'aboutir à un retour graphique didactif pour un praticien. Un exemple de retour visuel calculé grâce à notre chaîne de traitement est présenté dans la figure 0.4.

Contributions

Une chaîne de traitement est mise en place dans cette thèse. Elle permet de traiter les signaux récupérés dans un environnement libre pour aboutir à un outil graphique innovant d'évaluation de l'activité physique. Cette approche est donc composée de plusieurs étapes : une analyse de la littérature existante sur les études utilisant des capteurs inertiels afin de quantifier l'activité physique dans des environnements libres, la mise en place de protocoles pour la récupération de signaux en environnement libre, la segmentation des données, la classification des régimes segmentés, l'extraction des paramètres associés à des caractéristiques de la marche et la comparaison de ces paramètres avec le modèle sain. la figure 0.3 résume les étapes successives depuis les données brutes jusqu'au retour graphique final.

Les contributions de cette thèse sont donc résumées comme suit.

- **Revue de littérature** : Nous effectuons une revue complète des études utilisant des capteurs inertiels pour évaluer l'activité physique des sujets dans des environnements libres. Au sein de cette revue, nous nous attachons à recenser les différents objectifs de ces études, les activités effectuées par leurs sujets, les types de capteurs utilisés, les méthodes d'instructions et d'annotations. Nous avons détaillé une partie de cet état de l'art dans une publication précédente [78].
- **Protocoles dans un environnement libre** : Nous concevons et mettons en place trois protocoles (deux en environnements semi libres incluant 15 participants sains et un en environnement entièrement libre incluant 21 sujets sains, 6 patients avec une pathologie orthopédique et 3 patients cérébro-lésés) en s'appuyant sur les conclusions détaillées dans la revue de littérature décrite auparavant. Nous choisissons

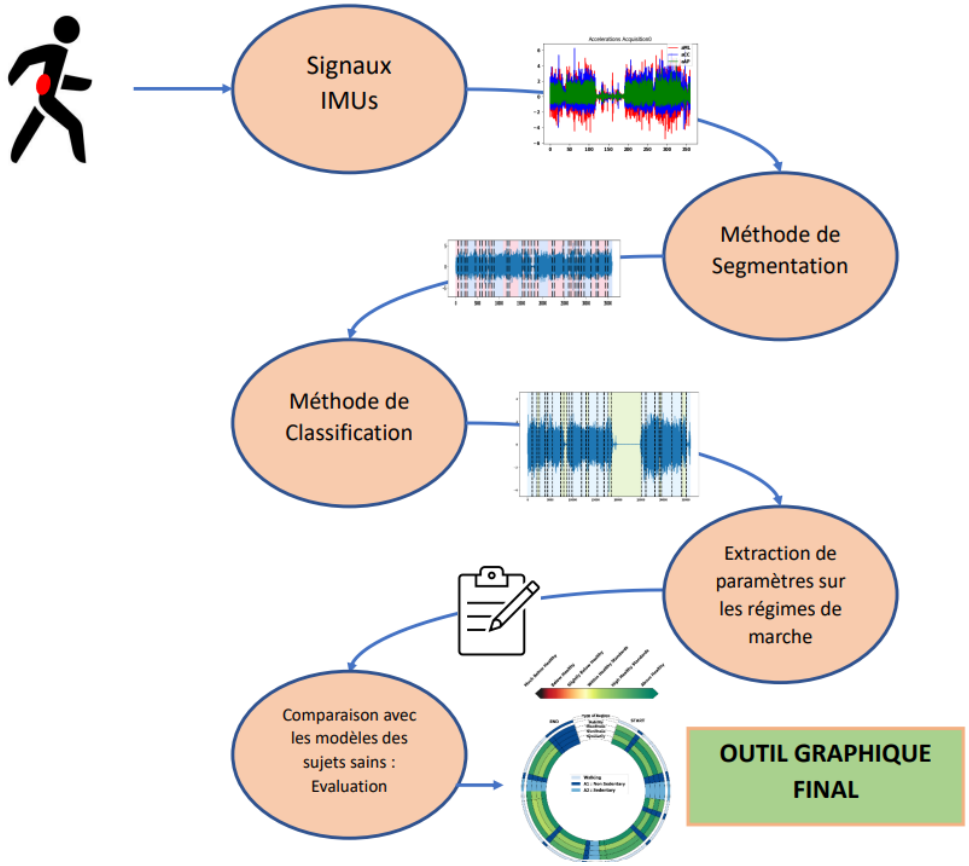


Figure 0.3: Etapes successives de la chaîne de traitement mise en place dans cette thèse.

les capteurs à utiliser, les activités à réaliser au cours des prises de mesure, les transitions entre les régimes d'activité à identifier... Nous incluons et mesurons des sujets sains, des patients cérébrolésés et des patients ayant subi une intervention chirurgicale orthopédique ou sur le point d'en subir une.

- **Segmentation** : Nous implémentons un algorithme adaptatif de détection de ruptures pour traiter les signaux mesurés par les centrales inertielles. La méthode recherche les changements significatifs dans l'espace temps-fréquence à une échelle donnée, c'est-à-dire les instants où le sujet a modifié son comportement/activité. Les signaux sont ainsi segmentés en plusieurs régimes homogènes qui permettront d'extraire des informations des enregistrements. Nous effectuons cette segmentation via un apprentissage de pénalité sur plusieurs signaux annotés. Les résultats de cette méthode sur deux protocoles sont présentés (en utilisant plusieurs métriques d'évaluation) et discutés. Les premiers détails de cette méthode ont été décrits dans une publication précédente [79].

- **Classification** : Une fois les régimes homogènes segmentés, nous les classons comme phases de marche ou phases de non-marche par une procédure de classification supervisée. Un second algorithme identifie, au sein des phases de non marche, les régimes sédentaires et les régimes non sédentaires, fournissant ainsi une labellisa-

tion complète des régimes. Pour chaque tâche, des paramètres sont extraits sur les régimes de marche afin d'entraîner des classifieurs spécifiques (plusieurs classifieurs sont comparés). Les résultats de cette méthode sur deux protocoles sont présentés (Matrices de confusion, précision...) et discutés.

- **Construction de l'outil graphique** : Nous sélectionnons des paramètres à extraire sur les régimes de marche afin d'évaluer quatre critères de marche (Stabilité, Régularité, Vigueur et Symétrie). En outre, nous contrôlons la pertinence et la robustesse du choix de ces paramètres pour caractériser chaque critère d'évaluation de la marche. En utilisant des modèles construits sur ces paramètres et appris à partir de sujets sains, nous associons à chaque régime de marche un score représenté par une couleur distincte, permettant un retour visuel et intuitif. Un exemple de retour visuel calculé grâce à notre chaîne de traitement est présenté dans la figure 0.4. Nous associons ces retours graphiques à des commentaires de praticiens (un neurologue et un chirurgien orthopédique). Les détails de la méthode de classification ainsi que de la construction de l'outil graphique sont décrits dans une publication en cours de soumission.

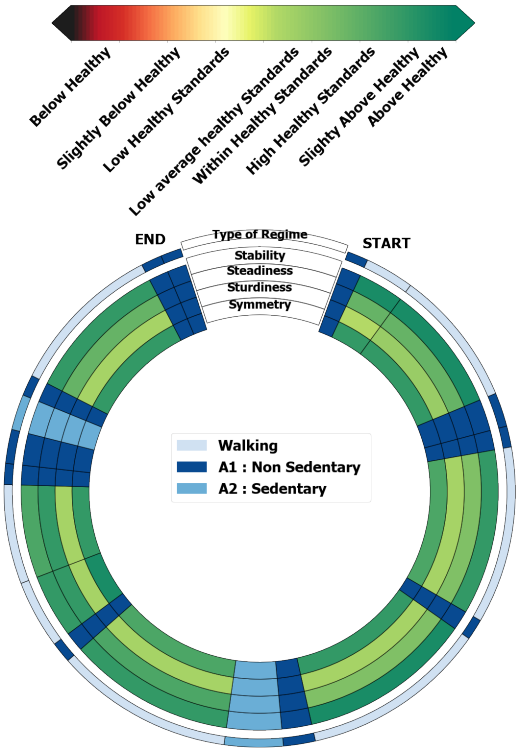


Figure 0.4: Exemple de retour graphique visuel proposé aux praticiens après la mise en place du processus de méthodes appliquées dans cette thèse.

Vue d'ensemble du manuscrit

Le reste du manuscrit est composé de six chapitres :

- Chapitre 2 : Selective Review of FLE/Semi-FLE Studies : Ce chapitre vise à fournir un aperçu sélectif mais complet des études en environnements entièrement libres ou semi libres utilisant des capteurs inertiels pour évaluer l'activité physique, en passant en revue les aspects techniques liés aux capteurs utilisés, les aspects comportementaux tels que les protocoles ou les instructions.
- Chapitre 3 : Protocols' Setups and Contexts : Ce chapitre décrit l'implémentation de trois protocoles (deux en environnements semi libres et un en environnement entièrement libre) basés sur la revue de la littérature précédente.
- Chapitre 4 : Adaptive Changepoint Detection Method : Ce chapitre présente une méthode de segmentation adaptative qui s'applique à des signaux transformés dans le domaine temps-fréquence et qui s'appuie sur un apprentissage de pénalité.
- Chapitre 5 : Classification Method : Il s'agit d'une méthode de caractérisation des régimes homgènes segmentés. Le mécanisme d'extraction des paramètres est détaillé ainsi que celui de l'entraînement des classifieurs utilisés.
- Chapitre 6 : Visual Feedback Rendering : Score Generation, Features' Robustness Les étapes de la construction de l'outil graphique final à partager aux praticiens sont décrites dans ce chapitre. L'extraction de paramètres associés à des critères de marche et l'évaluation de leur pertinence pour caractériser un des critères de marche sont détaillées ici. Des exemples de ces outils graphiques sont par ailleurs présentés et discutés.
- Conclusion : Un rappel des conclusions de chaque chapitre est effectué dans cette section. En outre, des pistes de recherche sont avancées.

1 Introduction

1.1 Context of the thesis

1.1.1 General Context

Over the last few decades, the use of inertial sensors (Inertial Measurement Units (IMUs), smartphones or simple accelerometers) to quantify physical activity and in particular the locomotion of individuals has become widespread [116, 157]. Indeed, several commercial solutions have emerged to provide users with the possibility of visualising the evolution of their physical activity, notably via self-monitoring applications. In the research context, the use of these light and inexpensive inertial sensors has also intensified: several clinical studies rely on the measurements of acceleration and angular velocity on different positions of the human body provided by these sensors in order, for example, to train classification algorithms that will make it possible to evaluate more precisely the energy expenditure of individuals over a given period of time. These sensors also make it possible to identify fall markers [133] as well as to provide innovative monitoring of the influence of a subject's treatments on his movements for instance [23]. In addition, the use of these sensors makes it possible to provide early diagnoses, to reduce the costs of health care, to better define frailty markers, etc. The conditions of use for these sensors (data recovery, filters, extraction of parameters) and their limits in clinical settings - i.e. in controlled environments (hospital, laboratory, etc.) - are now largely documented in the literature.

These analyses can also be applied in ambulatory settings (in free environments): such an approach allows the quantification of an individual's walking in a more representative way since it avoids the Hawthorne or white coat syndrome [132]. This syndrome occurs when the recorded participants become aware that they are being observed and measured during an acquisition. Their approach to a protocol during a clinical test can thus be altered and the reliability of the interpretation of parameters characterising their physical activity can be reduced. An analysis in a free environment can therefore mitigate this clinical effect. Nevertheless, ambulatory studies face several challenges: this type of study has long been based on declarative questionnaires in which the subject lists his or her daily activities, their duration and possibly some qualitative data (feelings, stress level)[45]. As for the existing quantitative assessments, compromises between the precision of the methods, the cost of calculation imposed by the use of long signals, the encumbrance of the sensors over long periods of measurement, the complexity of the quantified feedback of the analyses of signals are imposed on them.

Before the inertial sensors described above can be used for ambulatory studies, an iterative and coherent approach needs to be designed to retrieve the complex, non-stationary signals produced by these studies and process them into a quantitative and interpretable feedback of a subject's physical activity. This transition from raw data to usable and interpretable data is a crucial issue in medicine today.

Two cohorts of subjects are particularly studied in this manuscript: brain-injured patients and patients who have undergone orthopaedic surgery. Indeed, brain and spinal cord injuries have serious consequences on a patient's walking and cause unforeseen incidents to take place in their natural environment. Furthermore, during their rehabilitation, patients who have suffered strokes, for example, have a strong propensity to fall [121]. Studying the physical activity of a patient suffering from such a pathology by recording several physiological signals and analysing them would allow a controlled monitoring of the patient's condition, complementing and refining the follow-ups already carried out in clinical tests. In addition, stays in trauma and orthopaedic surgery have represented an unprecedented expenditure of several billion euros in recent years. Orthopaedic and traumatological surgery is also one of the largest areas of expenditure among the surgical services [54] . The possibility of longitudinally following patients who have undergone orthopaedic surgery is therefore of considerable importance. It could limit the costs described above by making it possible to anticipate rehospitalisation in certain patients.

1.1.2 Collaboration with ENGIE Lab CRIGEN and ABILYCARE

During my thesis, I collaborated with the Nanotechnologies and Sensors NANO laboratory of ENGIE Lab CRIGEN (research, development and expertise centre of the ENGIE Group dedicated to new uses of energy in cities). In a context where hospitals are opening up to the city (ambulatory care, post-operative follow-up, etc.), ENGIE wishes to develop innovative solutions in the e-health sector and to propose new services for patients, professionals and operators. Indeed, the healthcare system is currently facing several major difficulties, including the management of dependency, universal access to quality care and the significant increase in expenditure. These difficulties can be partly overcome thanks to the digital technologies that ENGIE wishes to put at the service of free-living monitoring and thus notably to the challenges that this monitoring entails as explained above. Prior to this thesis, the NANO laboratory had already developed - in collaboration with the BORELLI centre and the company ABILYCARE - a project for the dynamic and static evaluation of the physical activity of individuals : the SMARTCHECK project. This project is now commercialised by ABILYCARE. The aim of this assessment is to provide quantified and visual feedbacks to clinicians who wish to assess the physical activity of their patients in a controlled environment. It is based on a 10m forward and backward measurement using several inertial units for the dynamic aspect. It uses a balance test on a force platform for the static aspect. This project is part of an approach to expand the activities of ENGIE Lab CRIGEN to general health issues, particularly aspects of workplace medicine. This thesis is an extension of this approach and should enable ENGIE Lab CRIGEN to become even more involved in innovative e-health methods. ENGIE is particularly keen to make this monitoring possible for practitioners operating in hospitals by offering them the possibility of identifying, for example, the risks of re-hospitalisation for their patients following their stay in a health establishment.

1.1.3 Collaboration with the Centre BORELLI

During my thesis, I collaborated with the Centre BORELLI, a laboratory including medical researchers, applied mathematics researchers and health professionals such as neurologists, surgeons or anaesthetists. The fundamental mission of this laboratory is to combine these different expertises in order to quantify human and animal behaviour using light sensors, signal processing and statistical learning algorithms. To this end, Centre BORELLI researchers are developing and implementing protocols in clinical settings using inertial sensors and eye-tracking sensors. These protocols allow the recording of physiological signals (univariate or multivariate) which are at the heart of the Centre BORELLI's work. A first challenge for the Centre is to extract relevant information allowing an analysis of these physiological signals. The second challenge is to automate the processing of these data in order to provide clinicians with innovative methods of longitudinal follow-ups and with the possibility of carrying out inter-individual comparisons between subjects.

1.1.4 Collaboration with the Laboratoire du Traitement et du Transport de l'Information L2TI

During my thesis, I also collaborated with the research teams of the Laboratory of Information Processing and Transport (L2TI). This laboratory brings together research teams on two main aspects: data processing and information transport, particularly within networks. One of the challenges of the research teams is to extend its expertise to several existing fields, particularly in data processing.

1.2 Motivations and issues: scientific questions

1.2.1 How to set up protocols adapted to an open environment ?

Studies in free environments that are referenced in the literature are varied and present a wide range of objectives. For example, the studies may differ in the instructions given to the participants (complete freedom in carrying out the activities or imposed activities), in the means of annotation (presence or absence of an examiner, use of an external or wearable camera, use of a diary that the subject must fill in himself), in the specificities of the environments where the measurements are taken (complete free environment, university campus, hospital) etc. Two types of environments commonly implemented in ambulatory work stand out: the Semi-Free-Living Environments (semi-FLEs) where some measurements' conditions are controlled by an examiner and the Free-Living Environments (FLEs) which are exclusively the natural environments of the measured subjects. In this thesis, one of the first challenges was to bring together these observations made in the literature in a documented review in order to better define the semi-FLEs and the FLEs. This bibliographical study made it possible to define the challenges of setting up measurement mechanisms in free environments in order to be able to construct physical activity assessment protocols that are valid in such conditions. This review also made it possible to better identify the type of sensors to be used in a free environment (high data storage capacity and autonomy).

1.2.2 How to assess the signals retrieved from the protocols setup in a free environment?

In the context of studies carried out in free environments, the recovered physiological signals are complex: they are multivariate, longer than physiological signals recovered in clinical and they are non-stationary. Figure 1.1 shows acceleration signals measured in semi-FLEs: these signals are composed of a succession of homogeneous phases, corresponding to the different phases of the protocol. Signal segmentation or change point detection is then an important step in order to identify homogeneous regimes in the physiological signals that will allow a better analysis of the recorded physical activity. These algorithms allow, based on a mathematical definition of the concept of change, to automatically identify these changes in signals. One of the challenges of this thesis is to use these approaches for the processing of signals recovered from protocols implemented in a free environment. These methods allow to produce a relevant analysis of the information carried by these signals. The segmentation could be performed by an expert: this is often the case in clinical studies. Nevertheless, this manual annotation induces long and tedious tasks that must be repeated on each recording, which is a hindrance to the deployment of the methods. An automatic segmentation method therefore seems necessary to avoid this major complexity in the context of this thesis.

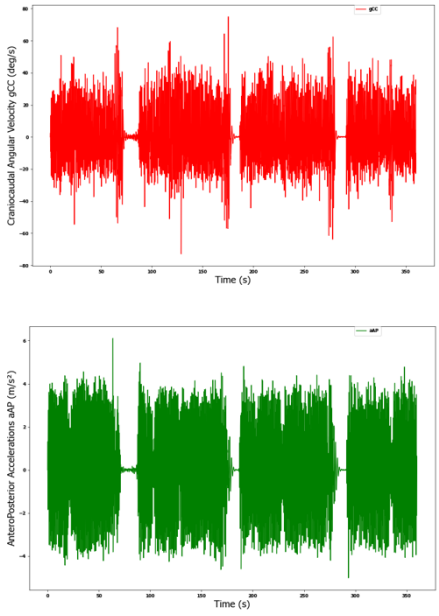
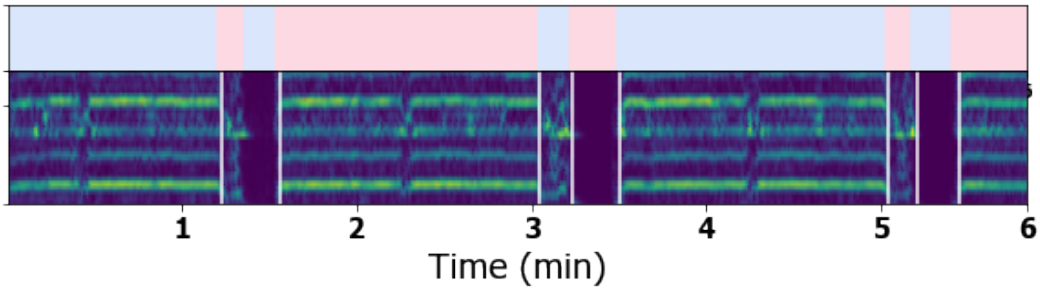


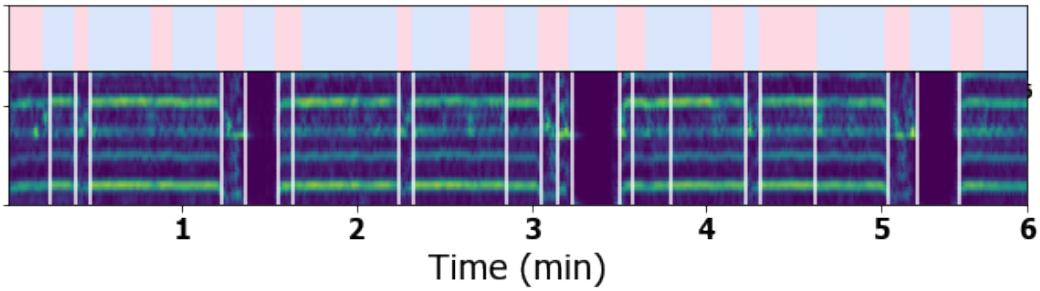
Figure 1.1: Filtered signals recovered from the lower back of participants recorded in a Semi-Free-Living Environment (semi-FLE): craniocaudal angular velocity and anteroposterior acceleration.

The conception of an automated change point detection method requires the specification of the type of change we are looking for, and the expected level of granularity. One of the challenges of this thesis is therefore to define the types of breaks that should be detected by implementing a supervised approach. The approach used in this study makes it possible to isolate specific regimes on the signals from a few annotated data. Unsupervised

methods of detecting breaks are sometimes used. Nevertheless, in the context of this study carried out in a free environment that is more chaotic than a controlled environment, a prior definition of the types of change point makes it possible to target more efficiently specific types of homogeneous regimes. The aim is therefore to develop this kind of adaptive segmentation method. The level of granularity of this detection is controlled so that only annotated breaks are detected. Such a learning method that allows replicating an annotation strategy on new data is developed in this thesis. The evidence of this adaptability to annotations is shown in Figure 1.2.



((a)) Annotation strategy 1



((b)) Annotation strategy 2

Figure 1.2: Evolution of change point detection results for a control participant as a function of the input annotations. The vertical lines correspond to the moments when the method detects a change point. The lower part of the visuals corresponds to the spectrogram constructed from the physiological signals: the differences of the regimes in their spectral signature are visible

1.2.3 How to characterize segmented homogeneous regimes and specifically walking regimes?

After applying our adaptive segmentation method on our signals, we recover segmented homogeneous regimes to which we will give labels in order to characterize them (walking, sedentary activity, non-sedentary activity regimes). In the literature, portions of physiological signals can be classified according to specific predefined labels (portions of walking, sedentary activity, running...). This characterisation allows several clinical interpretations. Indeed, it allows the evaluation of, for instance, the rate of energy expended by a subject in a day, the evolution of a patient's physical activity after the administration of a treatment, etc. Human Activity Recognition (HAR) methods allow the characterisation of all

segmented regimes and not only walking regimes. These approaches apply to sliding windows which can induce a high computational cost when used on long signals such as those recovered in a free environment. A challenge of the thesis is to provide an approach for characterising isolated homogeneous regimes on long signals recovered from free environments while avoiding the pitfalls described above. The originality of our classification method lies in its condition of application to signals: it applies to pre-segmented regimes and not to portions of signals selected by sliding windows as it is commonly used in the literature. The method first classifies segmented regimes into walking or non-walking regimes before classifying non-walking regimes into sedentary or non-sedentary regimes.

1.2.4 How to build an innovative graphical tool to provide quantified monitoring of physical activity?

In almost all of the work cited above, the output measures provided to clinicians are often generalized/aggregated using simple features adapted to the objectives of the studies (time spent in various targeted activities, assessment of energy expenditure from time spent in more or less energy-intensive activities, shadow of falls, fall time...). These aggregated and averaged output measures may hide some relevant phenomena of interest. An alternative approach would be to use features of finer granularity such as those used in clinical settings. However, this would imply a greater computational cost (e.g. on detecting all steps, strides...) and an excessive information flow that would overwhelm clinicians. This would prevent them from obtaining a clear and quickly understandable assessment of the physical activity of their patients. Based on this observation, one of the challenges of this thesis is to provide an alternative and intermediate solution allowing an accurate macro-analysis, at a low computational cost, ergonomic for the clinicians and which preserves the temporal structure. One of the challenges of this thesis is therefore to extract for each identified walking regime, consistent parameters allowing this macro-analysis. Thus, comparison models can be built on healthy subjects in order to provide practitioners with a didactive graphical feedback. An example of such a visual feedback computed with our processing pipeline is presented Figure **??**.

1.3 Contributions

A pipeline approach is implemented in this thesis. It allows to process the signals recovered in a free environment to produce an innovative graphical tool for the evaluation of physical activity. This approach is thus composed of several steps: an analysis of the existing literature on studies using inertial sensors in order to quantify physical activity in free environments, the implementation of protocols for the recovery of signals in free environments, the segmentation of data, the classification of segmented regimes, the extraction of features associated with walking categories (Stability, Steadiness, Sturdiness, Symmetry) and the comparison of these parameters with the healthy models. Figure 1.3 summarizes the successive steps from the raw data to the final graphical output.

The contributions of this thesis are therefore summarized as follows.

- **Literature review** : We conduct a comprehensive review of studies using inertial sensors to assess subjects' physical activity in free environments. Within this review,

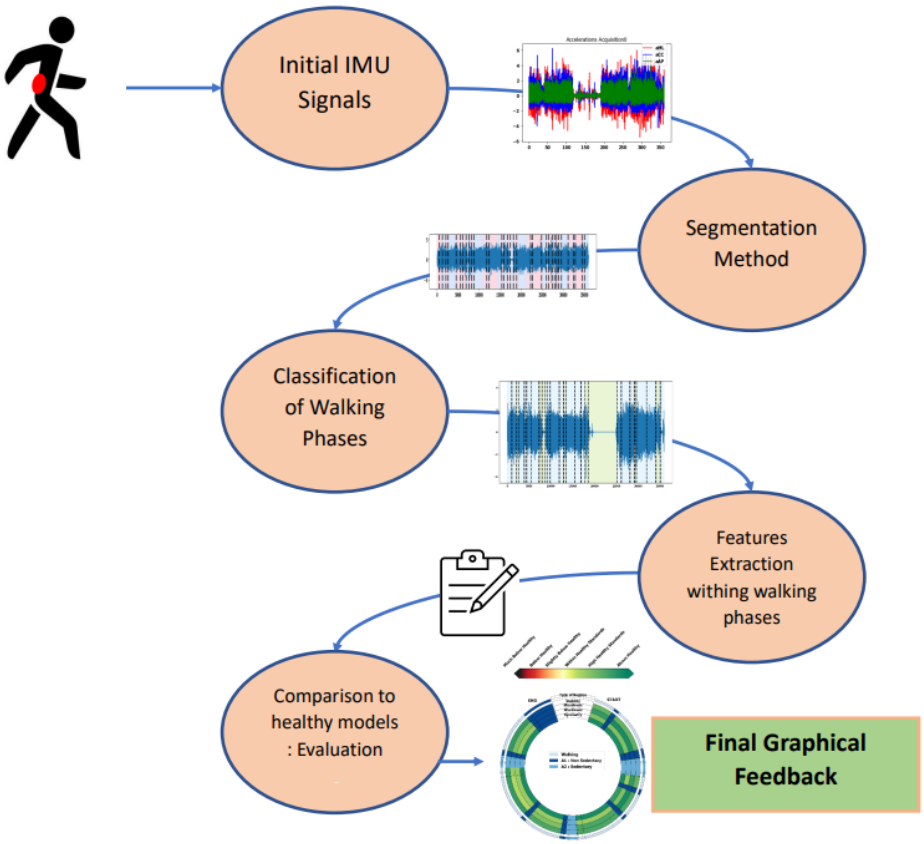


Figure 1.3: Etapes successives de la chaîne de traitement mise en place dans cette thèse.

we focus on identifying the various objectives of these studies, the activities performed by their subjects, the types of sensors used, the instruction and annotation methods. We detailed some section of this state of the art in a previous publication [78].

- **Free-living environments and semi free-living environments protocols** : We conceive and implement three protocols (two in semi-FLEs including 15 healthy participants and one in Free-Living Environment (FLE) including 21 healthy subjects, 6 patients with an orthopedic pathology and 3 cerebro-injured patients) based on the findings detailed in the literature review described before. We choose the sensors to be used, the activities to be performed during the measurements, the transitions between activity regimes to be identified... We include and measure healthy subjects, brain damaged patients and patients who have been or are about to be hospitalized in the orthopedic department.

- **Segmentation** : We implement an adaptive change point detection algorithm to process signals measured by the Inertial Measurement Unit (IMU). This method looks for significant changes in the time-frequency space at a given scale, i.e., the moments when the subject has modified his behavior/activity. Signals are thus segmented into several homogeneous regimes that allow us to extract relevant information from the experiments. We perform this segmentation via penalty learning on

several annotated signals. The results of this method on two protocols are presented (using several evaluation metrics) and discussed. The first details of this method were described in a previous publication [79].

- **Classification** : Once the homogeneous regimes are segmented, we classify them as walking phases or non-walking phases by using a supervised classification procedure. A second algorithm identifies, within the non-walking phases, sedentary and non-sedentary regimes, thus providing a complete classification of the regimes. For each task, parameters are extracted on the walking regimes to train specific classifiers (several classifiers are compared). The results of this method on two protocols are presented (Confusion matrices, accuracy...) and discussed.

- **Visual Feedback rendering** : We select features to be extracted on segmented and classified regimes in order to evaluate four criteria of walking (Stability, Steadiness, Sturdiness and Symmetry). Furthermore, we check the relevance and robustness of the choice of these features to characterize each gait evaluation criterion. Using models built on these features and learned from healthy subjects, we associate to each walking regime a score represented by a distinct color, allowing a visual and intuitive feedback. An example of visual feedback computed with our processing pipeline is shown in figure 1.4. We associate these graphical feedbacks with comments from practitioners (a neurologist and an orthopedic surgeon). The details of the classification method as well as the construction of the graphical tool are described in a publication currently being submitted.

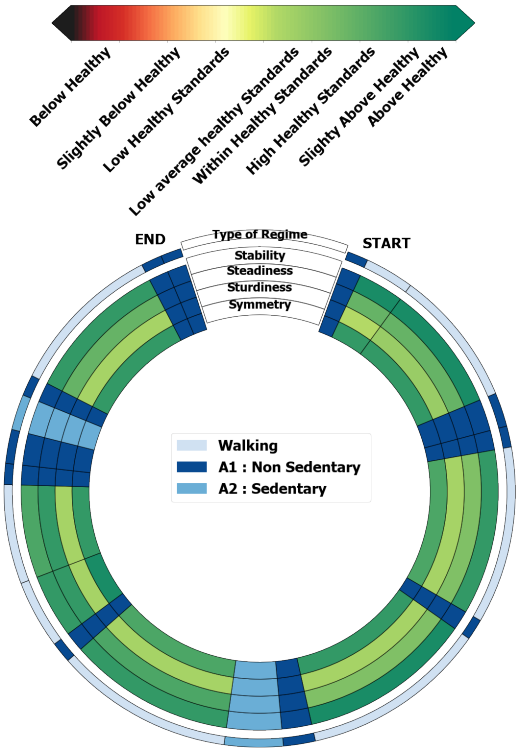


Figure 1.4: Example of visual graphical feedback provided to practitioners after the implementation of the processing pipeline

1.4 Overview of the manuscript :

The rest of the manuscript is composed of six chapters:

- Chapter 2 : Selective Review of FLE/Semi-FLE Studies This chapter aims to provide a selective but comprehensive overview of studies FLEs or semi-FLEs using inertial sensors to assess physical activity, reviewing technical aspects related to the sensors used, behavioral aspects such as protocols or instructions.
- Chapter 3 : Protocols' Setups and Contexts This chapter describes the implementation of three protocols (two in semi-FLEs and one in FLE) based on the previous literature review.
- Chapter 4 : Adaptive Changepoint Detection Method : This chapter presents an adaptive segmentation method that applies to signals tranformed into the time-frequency domain and which is based on penalty learning.
- Chapter 5 : Classification Method This is a method for characterizing segmented homogenous regimes. Features extraction is detailed as well as method to train classifiers.
- Chapter 6 : Visual Feedback Rendering : Score Generation, Features' Robustness The steps involved in building the final graphical tool to be distributed to practitioners are described in this chapter. The extraction of features associated with gait criteria

and the evaluation of their relevance to characterize one of these gait criteria are detailed here. Examples of these graphical tools are also presented and discussed.

- Conclusion : A review of the findings of each chapter is provided in this section. In addition, suggestions for further research are provided.

1.5 Publications

Published international peer-reviewed journal paper :

Sylvain Jung, Mona Michaud, Laurent Oudre, Eric Dorveaux, Louis Gorintin, Nicolas Vayatis, and Damien Ricard. The use of inertial measurement units for the study of free living environment activity assessment: A literature review. *Sensors*, 20(19): 5625, 2020.

Nicolas de l'Escalopier, Cyril Voisard, Mona Michaud, Albane Moreau, Sylvain Jung, Brian Tervil, Nicolas Vayatis, Laurent Oudre, and Damien Ricard. Evaluation methods to assess the efficacy of equinovarus foot surgery on the gait of post-stroke hemiplegic patients: A literature review. *Frontiers in Neurology*, 13, 2022.

Published international peer-reviewed conference paper :

Sylvain Jung, Laurent Oudre, Charles Truong, Eric Dorveaux, Louis Gorintin, Nicolas Vayatis, and Damien Ricard. Adaptive change-point detection for studying human locomotion. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2020–2024. IEEE, 2021.

Submitted to international journal :

Sylvain Jung, Laurent Oudre, Charles Truong, Eric Dorveaux, Louis Gorintin, Nicolas Vayatis, and Damien Ricard. A machine learning pipeline for gait analysis in a semi-free living environment. *Scientific Reports*, 2023

2

Selective Review of FLE/Semi-FLE Studies

2.1 Introduction

In the context detailed in Chapter 1, the goal of this chapter is to identify and describe the uses of wearable inertial sensors in free environments that can be FLEs or semi-FLEs. semi-FLEs studies are studies where some conditions are controlled by the experimenter while FLEs studies are stricly restricted to the subjects' natural environment. This chapter specifically details research works that include phases of movement, particularly walking. It aims at providing a selective yet complete overview on the topic, by reviewing technical aspects linked to the used sensors, behavioral aspects such as protocols or instructions. To that end, the ten most recent years of research on these themes from the early pioneer works to the recent deep learning approaches have been analyzed.

2.1.1 Existing Reviews

Several reviews endeavour to identify the studies using motion sensors to analyse the Physical Activity (PA) of subjects, taking into account at least partially if not exclusively works carried out under free-living conditions. Gorman et al. [71] detail several methods of Energy Expenditure (EE) assessment in free-living settings. De Bruin et al. [48], Byrom and Rowe [30] (Chronic Obstructive Pulmonary Disease (COPD) patients), Tedesco et al. [145], Murphy [108], de Oliveira Gondim et al. [49] and Frechette et al. [66] dwell upon the use of wearable systems (accelerometers or other motion sensors) to monitor activities in specifically targeted cohorts (Parkinson's Disease (PD), Multiple Sclerosis (MS)). Vienne et al. [157], Yang and Hsu [163] and Tedesco et al. [145] also focus on the use of wearable sensors in clinical settings but consider any kind of cohort. Attal et al. [7] and Narayanan et al. [109] analyze articles related to HAR algorithms and classifiers. Schwickert et al. [135] and Henriksen et al. [72] focus on other kinds of studies (respectively studies based on fall detection and studies using one specific brand of sensors). Table 2.1 summarizes all reviews that address some of the pivotal topics of this chapter (motion sensors, free-living settings...). It specifies how the referenced articles address the three main aspects detailed in this chapter: sensors, protocols and algorithms. According to Table 2.1, the majority of the reviews selected for this section generally include articles explaining the various aspects of sensor implementation in a detailed manner, whereas they tend to be less detailed regarding the implementation of protocols (instructions, measurement durations, etc.).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Specific Factors	Focused on FLE/Semi-FLE	√				√		√		√									√
	Considering all cohorts	Elders	√	√	√	√	√	COPD	Elders	√	Elders	√	√	MS	PD	√	√	PD	√
Sensors	Types of Sensors	√	√	√	√	√	√	√	√	√	√	√	√	√	√		√	√	√
	Number of Sensors			√	√			√	√			√			√		√	√	√
	Placements	√	√	√	√	√	√	√	√	√	√	√		√	√		√	√	√
	Technical Characteristics			√	√							√			√			√	√
	Additional Sensors								√		√					√			√
Protocols	Instructions' Details							√									√		√
	Measurement's Durations		√					√				√	√				√	√	√
	Annotations								√			√		√					√
Algorithms	Features			√			√		√	√	√	√		√		√	√		√
	Classifiers						√		√		√	√				√		√	√

Table 2.1: Comparison of our review to other existing reviews dealing with the use of wearable motion sensors in free-living settings. 1 → De Bruin et al. [48] (2008), 2 → Murphy [108] (2009), 3 → Yang and Hsu [163] (2010), 4 → Schwickert et al. [135] (2013), 5 → Gorman et al. [71] (2014), 6 → Attal et al. [7] (2015), 7 → Byrom and Rowe [30] (2016), 8 → Tedesco et al. [145] (2008), 9 → Tedesco et al. [145], 10 → Wang et al. [160] (2017), 11 → Narayanan et al. [109] (2019), 12 → Henriksen et al. [72] (2020),13 → Frechette et al. [66] (2019),14 → de Oliveira Gondim et al. [49] (2020),15 → Prasanth et al. [124] (2021),16 → Benson et al. [21] (2022),17 → Sica et al. [139] (2021),18 → **this chapter**

2.1.2 Scope and limitations of the review

This chapter only includes articles related the assessment of PA of participants in free-living settings and does not address locomotion tests implemented in controlled conditions (such as the 6 MWT, 10m test...) or any other test that can be only performed in controlled settings. Although a significant part of the articles concern pathologies, an in-depth study of the differences between pathologies will not be carried out. In addition, acquisition modes using a specific type of sensor are detailed : inertial motion sensors (gyroscopes, accelerometers, IMUs). Articles focusing mainly on other types of sensors such as GPS, pressure sensors or heart rate sensors are not included, which therefore constitutes a limitation to this study. In addition, this state of the art does not perform a complete analysis of all research detailing HAR processes. Indeed, the willingness to focus on papers that include the use of inertial sensors as well as papers based on free environments does not allow for such a thorough review. Besides, no conditions other than free-living conditions are under consideration in this study, which deliberately limits the scope of this state of the art. Finally, it is to be noted that Google Scholar articles whose duplicates were not found in the PubMed library were not included in this state of the art (further details are

provided in Subsection 2.1.3.)

2.1.3 Methodology

This scoping selective review has been conducted by searching MEDLINE via PubMed, Cochrane and Google Scholar electronic databases to identify articles published from January 1, 2010 to October 20, 2022 whose methods were including the use of wearable sensors such as IMUs, accelerometers or gyroscopes in order to perform an analysis on participants gait and PA in FLEs. Articles that were not directly exploiting output data from these sensors, or that were not performing their protocols in FLEs or at least in semi-FLEs were excluded from the review. According to the scope detailed in section 2.1.2, the following terms were looked after in titles, keywords, abstracts : (((*IMU* OR *accelerometer* OR *gyroscope* OR *inertial sensor*) AND (*free-living* OR *outpatient* OR *real life* OR *out-of-laboratory* OR *unsupervised*)) AND (*walk* OR *gait*).

This selective review was conducted by using Preferred Reporting Items for Systematic Review and Meta-Analyses PRISMA guidelines to select articles as detailed in 2.1. Potentially eligible studies were screened individually by 3 authors MM SJ LO on the basis of abstract and title for FLEs and wearable sensors criterion and of the full text for other criteria. In total, 83 articles meeting the search criteria detailed above were finally selected for this chapter.

2.1.4 Results of screening

As a conclusion, 83 articles were kept for a further analysis. The selection process put in place in regards to the PRISMA flowchart is detailed in 2.1.

2.2 Aims of the Studies

In this chapter, 83 studies linked to the use of IMUs in the context of FLEs were included. However, these studies have various objectives. Out of the 83 articles, six main categories of objectives are put in place and are detailed in Figure 2.2. These categories are obviously subjective and it is to be noted that some of the reviewed studies can have different purposes and thus belong to several categories.

The first main category gathers articles that intend to test a new HAR method by developing an algorithm and/or a measurement device specific to their work [110, 32]. Authors of these studies set up algorithms that allow, from previous annotated observations of participants performing activities (labeled training data), to determine which type of activity is performed during an analysis of a signal section and notably walking bouts which are stationary walking regimes [50, 159]. Some of these studies point towards the study of Machine Learning systems [154, 112, 42] while others focus on other factors such as sensor implementation [46]. This is the group of aims that contains the greatest number of studies (40 papers). Performing HAR allows for providing general metrics about activities durations in order to enable a quantified follow-up of patients' physical activity. This can also be the first step in estimating a patient's EE over long periods.

Some papers use IMUs as a quantification tool for the study of specific cohorts [2, 51, 138] and constitute the second category (32 studies). Some of those studies aim at quantifying patient's pathology ([126, 142, 60, 122, 113], while others quantify the therapeutic effect

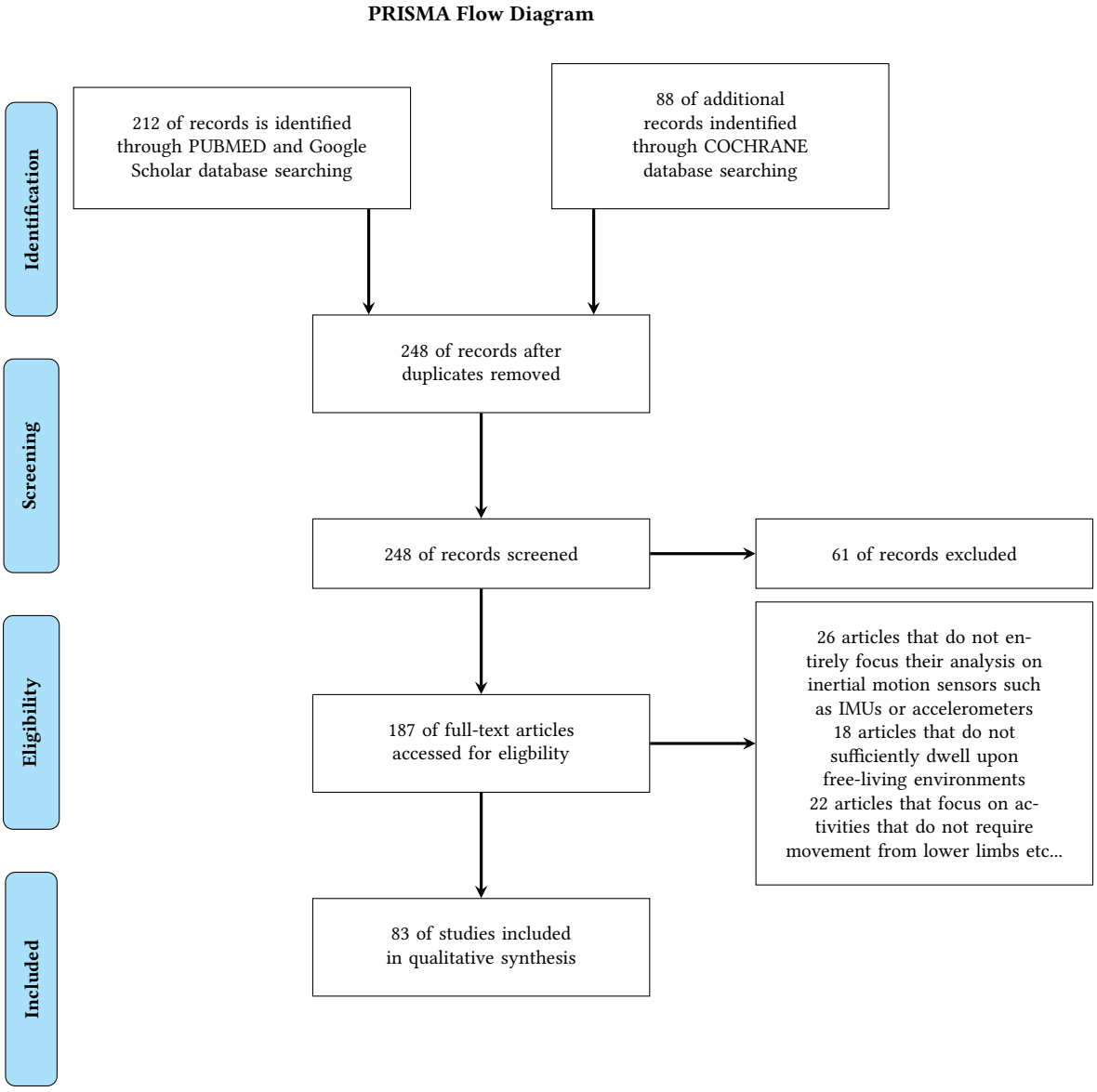


Figure 2.1: PRISMA Flow Chart illustrating the selection process resulting in a list of 83 studies.

on the patient's condition ([77, 138]) or develop algorithms to detect frailty in recorded participants [1]. One large group of these studies intend to detect falls or evaluate falls' risk [81] by relying on methods that have been developed in clinical studies [133, 114]. Some trials focus on healthy participants, others compare the patients included with healthy participants [98, 50, 141, 162], while others dwell specifically upon cohorts with a medical condition. Perriot et al. [122] intended to improve posture detection in COPD participants for instance and Nguyen et al. [112] dwelled upon HAR in patients with PD. In such works, the aim is either to compare the results obtained on certain patients with specific pathologies to control patients or to evaluate the impact that changes in instrumentation (position of sensors, etc.) can have on the results observed in participants with pathologies.

The third group (19 studies) includes works focusing on the analysis of the different characteristics of implanted portable sensors (feasibility, placements, comparisons between types of sensors or between sensors' locations) [136, 99, 61, 55, 38, 164]. Ellis et al. [57] compared the results of HAR depending on the placement of the used sensors (hip or wrist) for instance.
The fourth group of works is dedicated specifically to the evaluation of EE related features (16 studies) [57, 2, 104, 122, 60, 142, 126]. Some of these studies intend to detect the amount of time spent in activities that require a greater or lesser expenditure of energy when carried out (sedentary activities for instance that is to say activities with a Metabolic Equivalent of Task (MET) is below 1.5. MET is the objective measure of the rate a participant expends energy depending on his mass).

Finally, papers comparing the conditions and results in free-living conditions with those obtained in controlled environments [27, 120, 99, 50, 33] and papers dealing with the detection of ancillary parameters (such as wear-time [8, 88], strides or steps [33, 159, 98], ...) constitute the fifth and sixth groups of respectively 14 studies and 6 studies. Wear-time corresponds to the time a participant wear the sensors that he was provided with before the recording of his activities.

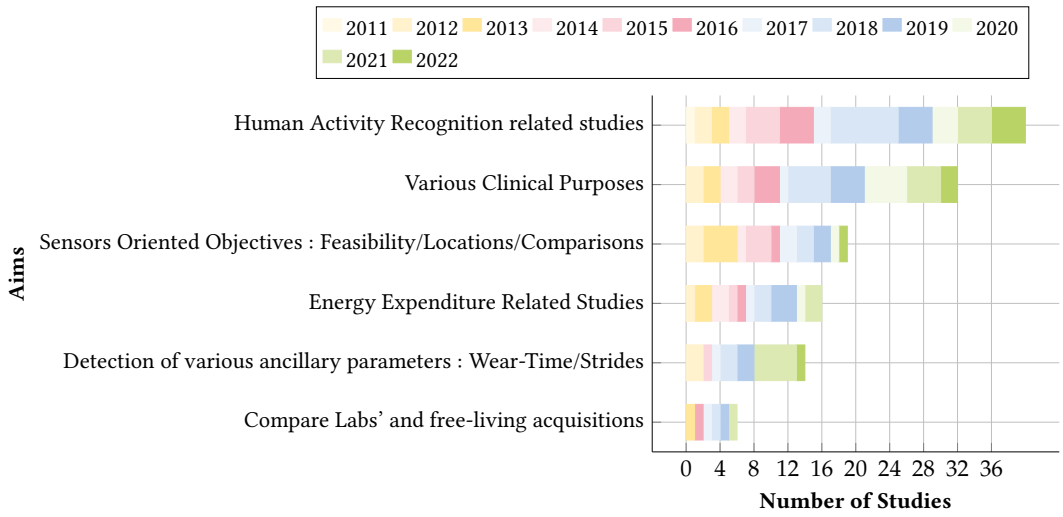


Figure 2.2: Change of the distribution of studies according to their aims over time.

2.3 Sensors

In this section, the first question raised by this chapter will be answered : what are the characteristics in terms of sensors configurations of the main methods for the study of PA in FLEs ? Types, numbers, placements of sensors, their technical characteristics as well as additional sensors' details will be screened in this section.

2.3.1 Type of Sensors

All studies included in this chapter use accelerometers, gyroscope, IMUs [33, 141, 17] or a combination of these three types of sensors. A referencing of the inertial sensors' types among the 83 studies considered in the review is available in Figure 2.3. The devices' brands for sensors that are used more than in one study are shown in Table 2.2.

It appears from Figure 2.3 that the two types of the inertial sensor most used for free-living applications are accelerometers (in particular the Actigraph accelerometers - see Table 2.2) and IMUs : 38.6 % of accelerometers and 39.5 % of IMUs. On the other hand, gyroscopes are rarely used solely (3.5 % of the studies). Table 2.2 also shows that the devices used have smiliar storage and battery capacities. Those technical characteristics are higher for such sensors than XSens sensors' characteristics for example which are often used for gait and locomotion analysis in controlled environments [56]. For the sake of comparison, Shimmer IMUs which are the most used IMUs in the reviewed studies have 8GB of internal storage and between 39h and 69h of battery autonomy while the Actigraph GT3Xs have an autonomy of 25 days and a 4GB intern storage capacity whereas the XSens have a maximum autonomy of 8h and have no internal storage. Some recent studies use smartphones to retrieve details of the same parameters measured by the other types of motion sensors mentioned above. Three of these studies are included in this state of the art [51, 146, 77].

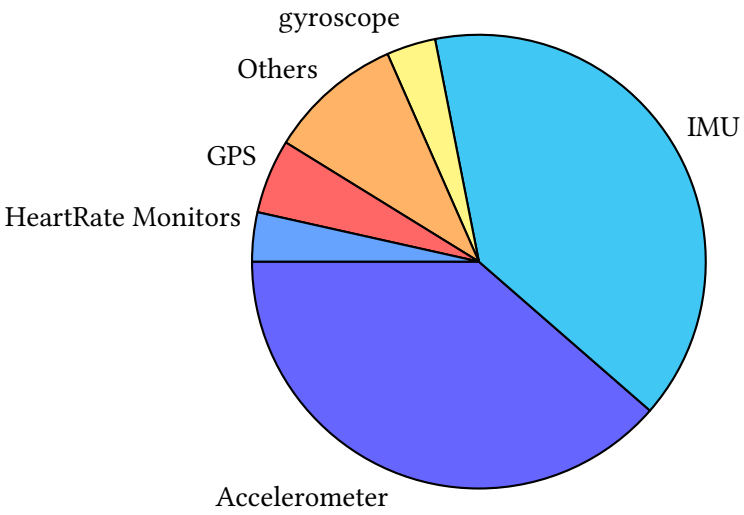


Figure 2.3: Proportions of the use of each kind of sensor in FLEs and semi-FLEs

Device Names	*Type of Sensors*	*Battery Life*	*Storage*	*Nb of Studies*	*Accelerome Range*	*gyroscope Range*	*Sampling Frequency*
Actigraph (*: GT3X) [88]	Accelerometers	25 days	4GB	14	±8 g	NC	30 - 100 HZ
Shimmer (*: Shimmer3)	IMUs	39 hrs - 69 hrs	8GB	7	±2 g (to ±16 g)	±250 dps (to ±2000 dps)	512 Hz
ActivPal (*)	Accelerometers	10+ Days	NC	5	±2 g	NC	20 Hz
Physilog GaitUp (*Physilog 4)	IMUs	23 hrs	8 GB	3	±2 g (to ±16 g)	±250 dps (to ±2000 dps)	1 - 500 Hz
AX3	Accelerometers	14 days at 100Hz	512 MB	3	±2 g (to ±16 g)	NC	12.5 - 3000Hz
GENEActiv	Accelerometers	30 Days	0.5 GB	3	±8 g	NC	10 - 100 Hz
IGS-180 Suit (Xsens)	IMUs	6 hrs	None	3	±16 g	±2000 dps	100 Hz
Dynaport (* MoveMonitor)	Accelerometers	14 Days	1GB	2	±2 g (to ±8 g)	NC	50 - 200Hz
Sensewear [60]	Accelerometers	14 days	20 days	2	±2 g	NC	50 - 60Hz
Hookie AM20	Accelerometers	NC	NC	2	±16 g	NC	100Hz
Actiwatch (Actiwatch 2)	Accelerometers	30 Days	1MB	2	±0.5 – ±2 g	NC	32 Hz
Actical	Accelerometers	194 days	32 MB	2	±0.025 – ±2 g	NC	32 Hz
Empatica E4	Accelerometers	1 Day - 2 Days	60+ hrs	2	±2 g (to ±8 g)	NC	32 Hz

Table 2.2: Details of used inertial sensors. *dps : degrees per second*

2.3.2 Number of Sensors

In some articles, protocols include the use of several sensors positioned on different locations on the body [33, 61, 154]. Figure 2.4 details the distribution of the studies according to the number of inertial sensors associated with them [57, 84, 83, 88, 154, 104, 136, 2, 120]. It shows that the majority of studies in FLEs or semi FLEs - even when using several IMUs, accelerometers or gyroscopes - limit the number of the latter. Of all studies using more

than one sensor, 67% use four or less sensors [142, 158, 32, 46, 109, 99, 69]. Moreover, only three studies use more than ten IMUs - these studies use the IGS-180 suit consisting of seventeen IMUs [113, 112, 12]. Besides, 41 % of the studies use only one sensor for all of their acquisitions [8, 42, 50, 141]. This low number of sensors used in FLEs'/semi-FLEs' studies may appear surprising since the cost of inertial sensors has decreased and their dimensions have reduced in the last ten years. Yet, this could be explained by the fact that reducing the bulkiness due to inertial sensors is helpful in FLEs. With such a reduction, wearing sensors is less likely to serve as a reminder that a measurement is being taken which avoids therefore Hawthorne syndromes [132]. This syndrome affects measurements in clinical/controlled environments and reduces the spontaneous nature of movements observed during these measurements. In particular, although the figures do not allow for a definite conclusion, it seems that the higher the degree of freedom of the environment is, the lower the number of used sensors is. A reasonable explanation is that a too important bulkiness in complete FLEs paradigm is more complicated to manage (installation, charging of the sensors...).

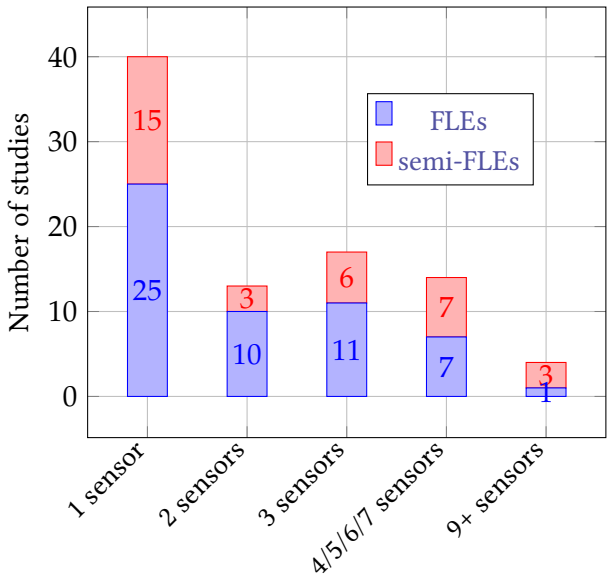


Figure 2.4: Distribution of studies according to the number of sensors associated with them: distinction between FLEs' Studies and semi-FLEs' Studies.

2.3.3 Placement of Sensors

The locations of the sensors is also important as it conditions the types of results that can be obtained in a study. A mapping of the major placements for inertial sensors is available in Figure 2.5.

It appears that the sensor placements in FLEs are homogeneous, including both lower and upper body parts. A clear trend does not emerge even if the wrist and Lower-Back (L3-L5 vertebrae) positions seem to be predominant (50 iterations in reviewed works [8, 42, 119, 146, 100, 126, 52, 53, 123]). Several inertial sensors are sensors implanted on a watch (ActiWatch, GENEActiv...) hence the high proportion of sensors placed on the wrist. In addition to this, it is reported in one study [136] that the participants tend to prefer this

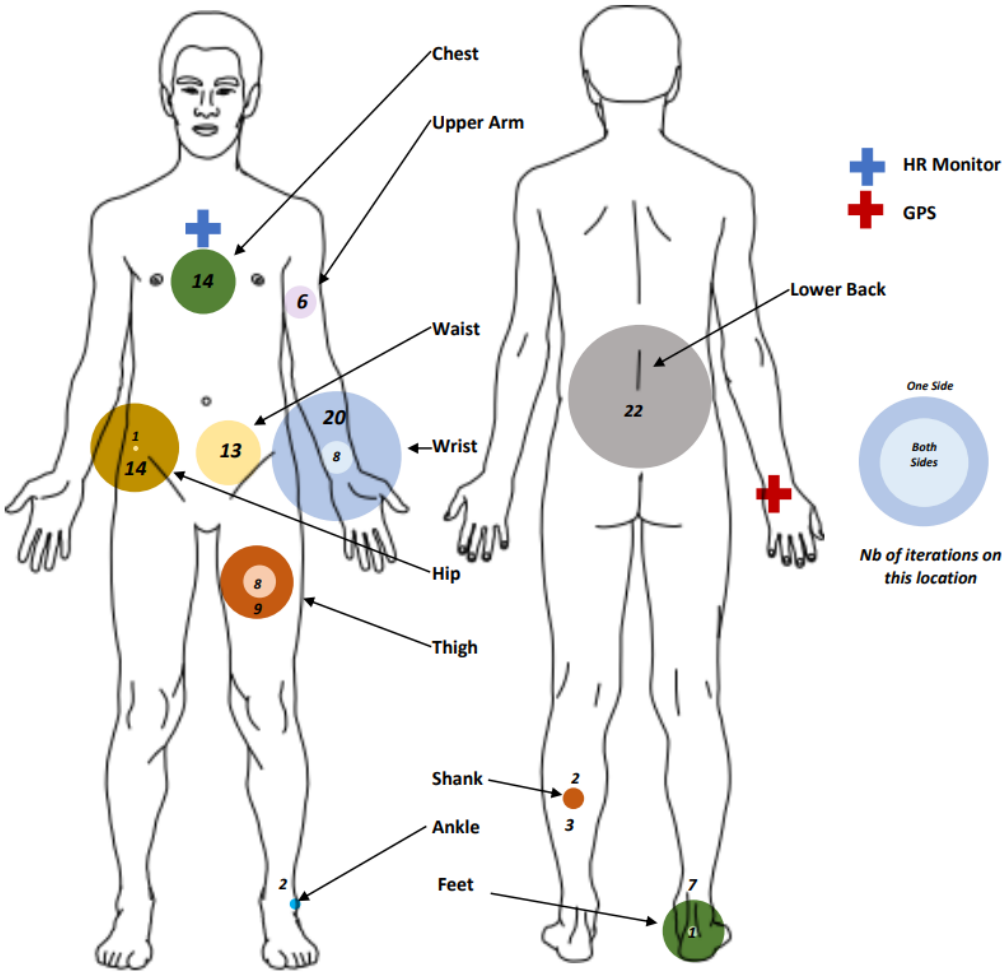


Figure 2.5: Sensors (IMUs, Accelerometers, gyroscopes + other kind of sensors : GPS, HeartRate Monitors) placements. Each circle displayed on the graph shows the proportion of use of the location (its radius is matched with the number of times that location is used in relation to the total number of locations used). Within these circles, another light-coloured circle can be displayed : it corresponds to the proportion of the number of times the sensors placed on the location of the circle are placed on both sides of the participant. The remaining darker part represents the proportion of the number of times the sensors placed on the same location are placed on only one side of the participant. The average position of the other sensors (GPS, HeartRate Monitors) are also indicated by crosses.

configuration since sensors are more comfortable and less cumbersome to wear on the wrist. However, another study [140] has shown that 15.6% of their participants who wore accelerometers on the wrist violated the protocol for one or more days : sensors were worn on the wrong hand during 6.9% of the days. During the periods of discrepancies, the daily PA was miscalculated by more than 20%. It therefore appears that behind the expected sensor locations, the correct placement of the device also has a significant effect on the results. Works using lower-back location usually place sensors thanks to velcro straps or

other mechanisms preventing sensors' undesired displacements. It seems that few studies conducted in open environments use portable sensors attached to the feet of participants. This can be explained by the fact that in unsupervised conditions, the use of a sensor on this location can present a detrimental congestion for the smooth running of the activities to be performed by the observed subject. These results clearly demonstrate the constraints and compromises that any protocol in FLEs relies on: getting the cleanest signals while still achieving a good acceptability for participants.

2.3.4 Technical Characteristics

As displayed on Table 2.2, sensors used in FLEs can be chosen according to several technical characteristics such as storage, battery life or range of measurements. One of these important characteristics is the sampling frequency. This parameter influences both the level of precision of the processing and some practical considerations such as storage, size, or energy consumption. It is therefore important to locate and identify possible trends on the chosen value of this parameter according to the type of study and its associated objectives. The evaluation of daily PA by IMUs requires the selection of an adequate sampling frequency. The choice of this frequency must be based on the acceleration power of the human movement in order to be able to acquire all the data relating to it. In 1997, Bouten et al. [24] studied the acceleration power of human motion by distinguishing the upper body parts from the lower body parts. It was observed that the acceleration power in the upper body varies from 0.5 to 5 Hz. In the lower body, the heel strike can however produce acceleration frequencies up to 60 Hz. Knowing that, and depending on the sensors locations, sampling frequencies ranging from 10 Hz [67, 99] to more than 200 Hz [100, 68] have been used in the literature. As can be seen in 2.6, there is large variety of sampling frequencies used articles. In several articles, the choice of the sampling frequencies is justified by considerations similar to those of Bouten et al. [24], and it is accepted in several publications that a sampling frequency greater than 20 Hz is an acceptable choice to capture most every day activities [101, 146, 110]. Indeed, according to Karantonis et al. [80], Bianchi et al. [22] and Khusainov et al. [86], all human body movements are within the range of 0 to 20Hz hence the importance of having sensors with sampling frequencies at least above 40 Hz (Nyquist criterion).

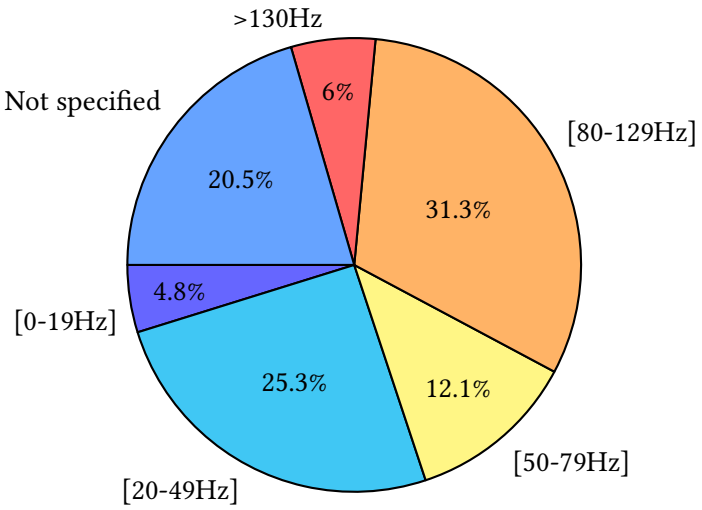


Figure 2.6: Distribution of sampling frequency among all articles. Note that three articles use sampling rates increasing from 10Hz to 200 Hz in 10 Hz increments. They are thus counted in each corresponding slice of the pie chart.

2.3.5 Additional Sensors

Although our review is not focused on non-inertial sensors, several studies use inertial sensors associated with an additional sensor such as a GPS or heart rate monitors [60, 63]. Some of them heavily rely on these sensors (GPS for instance) [107, 138, 63, 44, 76, 4]. As far as GPS are concerned, they can enable trajectory reconstruction [107, 138, 44, 76] hence their usefulness in some cases (to correlate the data measured by the inertial sensors with the mapping of the movement of the participants in their environment). Tedesco et al. [145]'s review identifies some GPS sensors that can be found in studies tracking activity but mixes up free environments (semi-FLEs and FLEs) with controlled environments. GPS tracking is sometimes used in studies in order to visualise walking bouts or walking habits of participants. Nevertheless, accuracy of GPS is greatly degraded indoors, hence the almost unique use of these sensors in outdoor environments. Moreover, GPS do not provide valid data for vertical position. In some cases, GPS can also act as a good reference to correct the absolute positions of inertial sensors during horizontal movement phases. Indeed, they can be used to correct the drift errors of an algorithm named pedestrian-dead-reckoning (with a Kalman smoother filter) intended to reconstruct the trajectory of participants and thus to be able to look for stationary walking phases [159].

2.3.6 Impact of Sensors' Setup

The variation in the types of sensors used, their numbers, their placement on the participants, as well as their technical characteristics are significant. This variation implies differences in signal retrievals, in their processing, and therefore in the accuracy of the HAR calculation. Indeed, several articles detail in particular the impact of different types of sensors by comparing, for example, different models of accelerometers used within the same protocols and on the same locations [61, 99]. Comparisons can be made on the accuracy of the calculation of certain features or on the accuracy of HAR. Furthermore, based on the

analysis of these same types of results, some studies highlight the impact of differences in motion sensors' placements on the retrieval and the use of data [57]. Depending on the way the sensors are set up, different features can also be computed. For instance, one acquisition performed with one sensor located on the lower back does not enable the same features' retrieval than one acquisition with a sensor placed on the wrist. Researchers also have shown that sensor positioning errors could lead to variations (displacement within a body part due to insufficiently reinforced sensor mounting for instance). These variations can lead to a loss of orientation information that significantly affect the raw data [91]. Corrections (such as the use of orientation-robust features) to avoid these changes can be implemented when designing the measurement setup.

2.4 Protocols

In this section, insights on the different protocols used in the studies are provided. Instructions to perform activities, activities' and environments' details as well as inclusion criteria and annotations' trends are presented. As previously mentioned, one difficulty in the analysis of human movement is to reach a satisfactory balance in the experimental conditions. Indeed, although the goal is to record a movement as natural as possible, studies also aim at reaching the greatest possible accuracy. This forms a gradient of experimental condition more or less "natural" to which each experimenter sets the parameters. In the introduction of this chapter, we defined FLEs as environments that are not controlled by the experimenter. The participant thus has no indication of the environment, and complete freedom of movement. On the opposite, environments established by research teams are considered as semi-FLEs. It includes laboratory, indoors or outdoors space, even replica of apartments. As it will be detailed in this section, it should be noted that these definitions are strictly based on the environment, but that there is also a gradient of freedom in the activities according to the instructions given.

2.4.1 Instructions

Communicating the protocol to participants influences the way they perform the activities. For each protocol it is necessary to be consistent with the walking variable to be measured and the environmental conditions. Sustakoski et al. [143] notice a difference in walking speed on various protocols performed by the same participants. Overall, the paper prompts discussion on the notion that slight changes, such as walking on a computerize walkway or on the ground, can influence walking speed. Rehman et al. [128] compare the impact of walking protocols and gait assessment systems on patients with Parkinson's disease, and underline the impact of the protocol on the activities performed. They observe at the level of two different walking protocols a difference in the participants' performance (pace, rhythm, variability, asymmetry). Thus, the format in which instructions are communicated is an important parameter to fix in a protocol. In the literature, different types of instructions are observed, which leave more or less freedom of interpretation as displayed in Figure 2.7. There is a continuum of situations between semi-FLEs' studies that still include precise instructions in the way of laboratory settings and FLEs' studies where the subjects are completely free.

One type of instruction regularly found in semi-FLEs [83, 120, 164, 77, 98, 52, 9, 110, 43, 17, 101] is the presentation of the activities to be performed : how to perform them

and for how long. These instructions are classified as “Imposed Activities with specific Instructions". In this case, participants have rigorous instructions and therefore no freedom of interpretation and completion of the activities. Other ways of instructing often found in semi-FLEs [51, 105, 2, 100] are "Imposed activities without any instructions". In those conditions, participants are told which activities to perform but there is no precision on how to perform them. This type of protocol allows more freedom as to how perform the activities which makes them more natural. A third level of instruction used in articles is the suggestion of activities : some activities are suggested to the participants but without any additional instructions on how to perform them. In addition, the configuration of the environment itself can contribute to the suggestion (letting a pen on the floor which will imply that the participant leans forward to pick it up [134, 33, 44, 27, 63, 67, 112, 12]). In this case, it gets closer to FLEs' situations where participants have complete freedom when performing activities. The last case observed mostly in FLEs' situations is "complete freedom". Participants are thus equipped with one or more IMUs and continue their daily activities at home (including going to work, hobbies, home activities) without any instruction or suggestion [141, 50, 55, 76, 88, 61, 99, 136, 57, 140, 104, 122, 126, 60, 142, 52, 9, 110]. In the latter case participants are in the most natural conditions possible (both environment and activity). A major aspect of the instructions are sometimes modified according to studies : exceptions in which the sensors must be removed (shower [88], sleep [76]...) according to the experts who set up the measurement protocols. This modification of the wear-time (which is a parameter measured as precisely as possible by certain research teams) has a direct impact on the total recording time on a typical day. Participants from one study were even asked to remove their sensors whenever they felt skin irritation around the latter [46]. These requests for sensor removal may exist even when the instructions on the activities to be carried out are free: these two aspects are distinct.

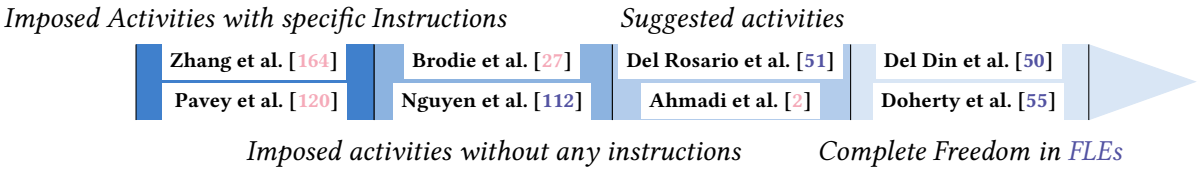


Figure 2.7: Frieze detailing the different types of instruction given to participants during FLEs' and semi-FLEs' studies. *Citations colored in red are citations of semi-FLEs' studies. Citations colored in blue are citations of FLEs' studies.*

2.4.2 Environments

In addition to instructions, environments also play a part in the definition of FLEs/semi-FLEs conditions. The repartition of studies included in this state of the art according to these categories is displayed on Figure 2.8. It appears that the majority of the reviewed acquisitions are performed in FLEs (among all reviewed acquisitions, 60.8% are in FLEs and 39.2% are in semi-FLEs). Acquisitions are specific recordings and several acquisitions can be put in place in one study. Several studies include both FLEs and semi-FLEs acquisitions (15.5% of the reviewed studies).

Within FLEs studies, a distinction is made between certain papers depending on participants' environment. Indeed, there are two types of environments: those that are familiar to the subject and those that are unfamiliar. Concerning articles using familiar environments, some studies in FLEs are limited to the habitats of the participants[140, 57, 88, 136, 50, 104, 142, 60, 126, 84, 146, 4, 10, 8] while others include all the environments daily frequented by the participant like library, gym, university daily commute [83, 104].

semi-FLEs studies can be divided into two categories: indoor and outdoor studies. Indoor movements can be carried out on conveyor belts, on a 10m path (or another defined distance) previously traced in a controlled environment (laboratory). Some studies take place in a reconstruction of an appartement: these environnements are simulated FLEs [63, 12, 105, 113]. Others take place inside the laboratory and, or in infrastructures next to it to be able to perform some activities : inside university campus [17, 101] for instance. In those studies, participants are asked to perform a course that includes several activities that can not be performed only in a laboratory. Some studies imply indoor and outdoor parts [159].

FLEs studies don't seem to substitute for other experimental conditions, but provide access to other data that complement those obtained under semi-controlled conditions. semi-FLEs, FLEs and controlled conditions thus appear to be complementing each other to provide the most complete study of human activity.

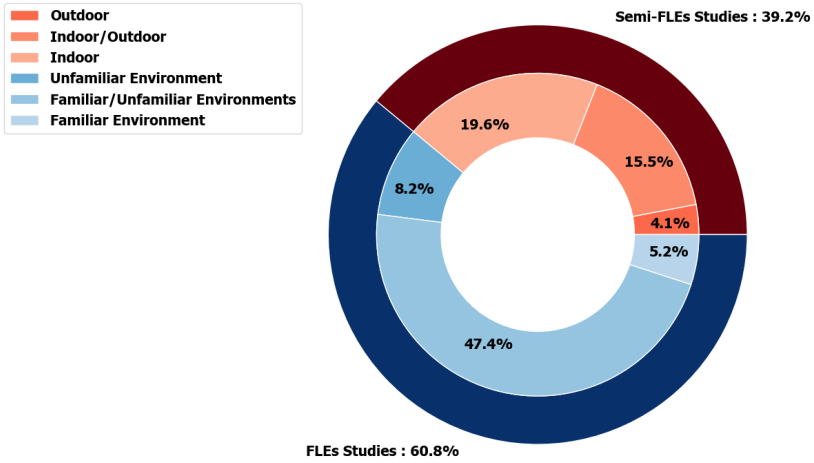


Figure 2.8: Number of studies focusing on Free-Living Environments (FLEs), semi-Free-Living Environments (Semi-FLEs), or both of them. The proportion of FLEs' Studies performed in Familiar, Unfamiliar or both types of environments and the proportion ofsemi-FLEs' Studies performed in outdoor, indoor or both types of locations are also detailed.

Figure 2.9 displays a detailed list of the environmental categories used in FLEs and semi-FLEs studies. This figure also shows the number of times the measures are implemented in these types of environments. It can be noticed that most of the protocols are implemented in environments which are familiar to the participants (homes, offices, etc.) but that some other arrangements are possible and quite frequently encountered (hospitals, rehabilitation centers, etc.).

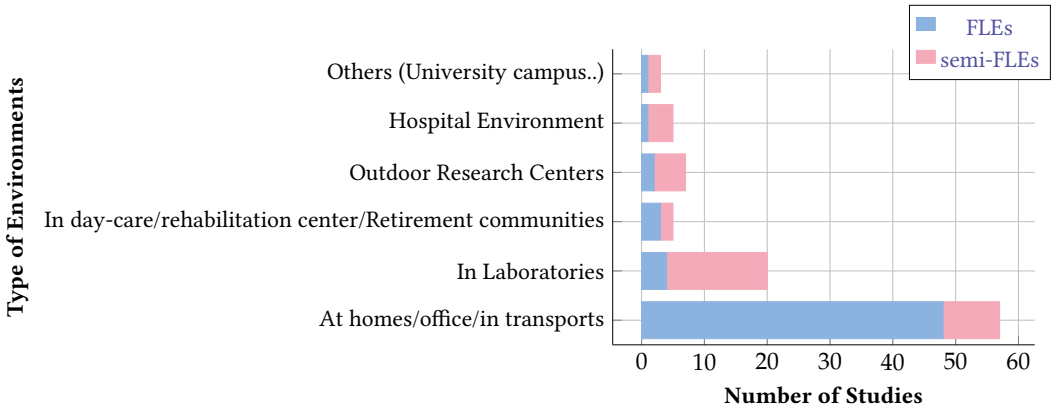


Figure 2.9: Details of the used Environments

2.4.3 Activities

In this section, the most frequently studied activities in free-living settings are presented. This section is organized in two parts, the first one being dedicated to full FLEs and the second one to semi-FLEs. Figure 2.10 shows the distribution of the studied activities for both FLEs and semi-FLEs.

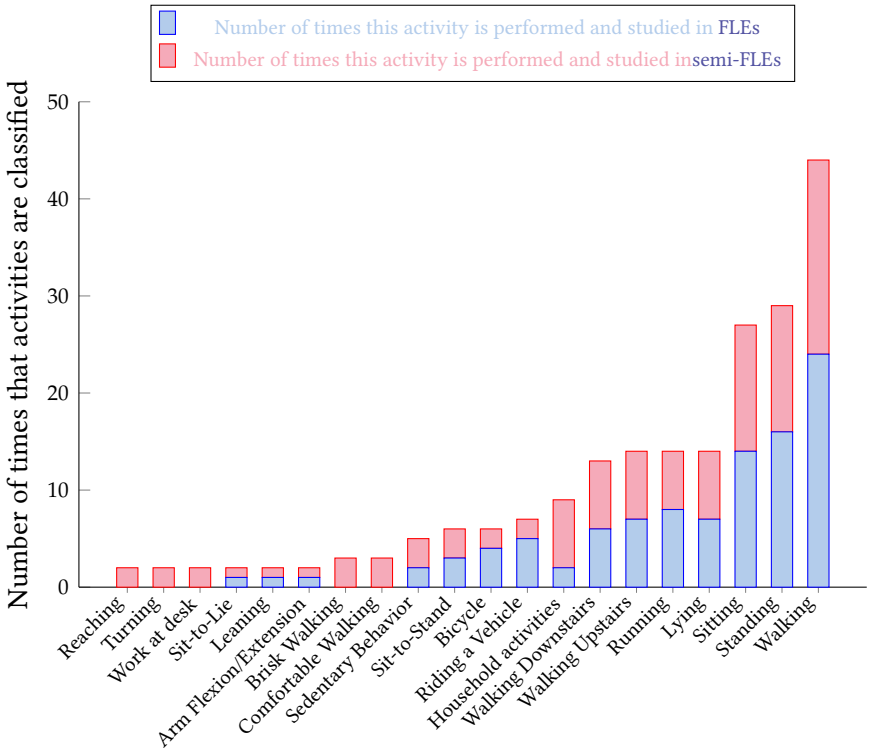


Figure 2.10: Details of the distribution of the studied activity types into FLEs andsemi-FLEs. *Total Number of Activities performed an studied in* ***FLEs = 96****. Total Number of Activities performed and studied in* ***Semi-FLEs = 108.***

It appears that out of the 96 activities in FLEs, three activities stand out as the most

analyzed : standing, sitting and walking. Several articles simultaneously study these three activities [122, 68, 51, 57, 10]. Participants of these studies had to perform all of these three activities. A second slice of activities also seems to emerge and is composed of the following movements: running [57, 120, 67], lying down [154, 109, 51, 68], going up and down a staircase[53, 46]. Several studies are also interested in the detection of postural transitions (sit-to-stand, stand-to-lie ...) [51, 120].

semi-FLEs enable a better control of performed activities and allow for more accurate annotations of these activities. It results in an optimised computation of HAR classifiers' accuracies for instance. As seen in Figure 2.10, the same initial conclusions as for the FLEs studies can be drawn. The majority of the activities measured are: standing, sitting, walking, with a slight increase in the percentage of studies analysing walking and running. Some new activities to be detected are also referenced in the semi-FLEs studies: turning and reaching [113]. semi-FLEs' studies focus more on household activities [164, 100] which are simpler to put in place and to annotate in such an environment. Besides, outdoor activities such as vehicle travels are less frequent (two semi-FLEs' [67] vs. four FLEs' [76, 57]), probably because they are more difficult to setup in this kind of environment.

It should be noted that some studies try to detect groups of activities (sedentary activities [154, 104, 55], stationary activities [67, 120, 69]...) which included several specific activities (standing, sitting, lying down for the group of so-called stationary activities for instance), etc.

2.4.4 Measurements' Durations

Durations of recordings are also an important characteristic of referenced studies. Figure 2.11 displays the distribution of measurements' durations for both FLEs and semi-FLEs studies. Strong differences between the two experimental conditions can be observed.

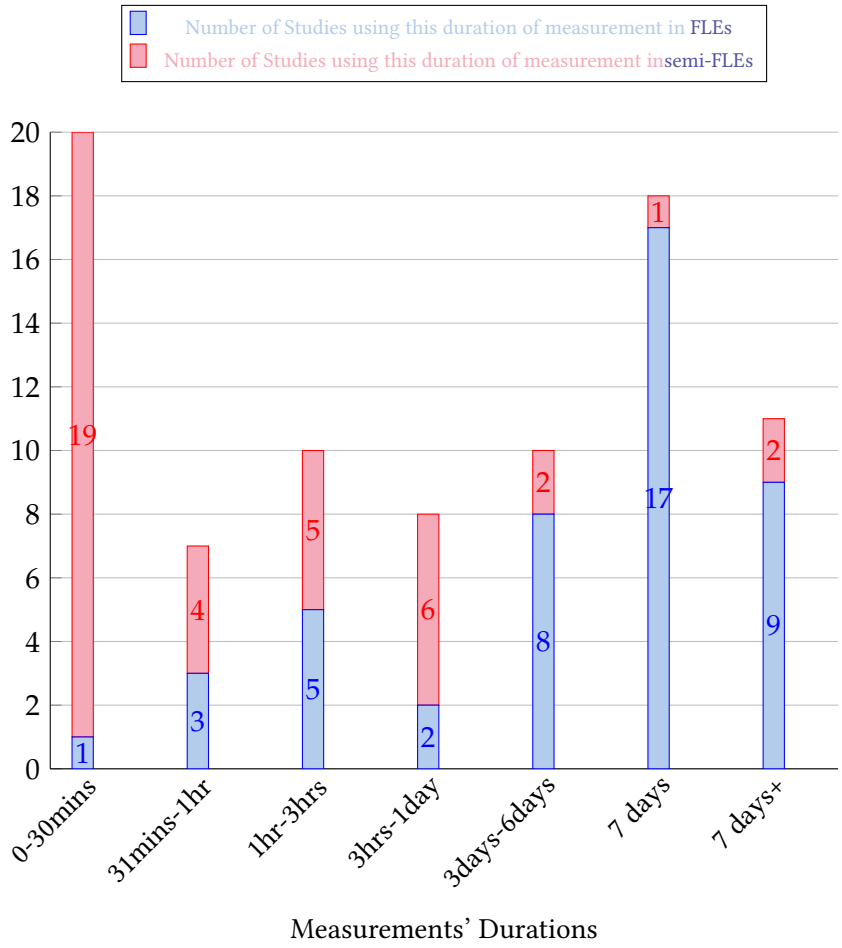


Figure 2.11: Measurements' Durations in in FLEs' andsemi-FLEs' Studies

FLEs studies mostly consider durations greater than 1 hour and even 3 days in some cases. Indeed, of all listed durations (one duration is not available in one article), 57.8% of the studies measure participants over seven days or more [57, 84, 46], 75.6% percent of the studies measure subjects over three days or more [33, 88, 55], 91.1% percent of the studies measure subjects over one hour or more [123]. This corresponds to recurrent objectives in FLEs studies: to work on extensive databases mimicking the participants' physical activity as closely as possible. These long acquisitions allow to observe movements as spontaneous as possible, which is the interest of free-living studies.

The vast majority of recordings in semi-FLEs' studies are completed in less than one day (87.2 %). This is consistent with the observation that in semi-FLEs, deconstruction of motion phases is more easily feasible since the environment is controlled. A clear separation between studies in semi-FLEs is observed around one parameter: the continuity or not in the measurements made. Some research groups measure all movements by having them all performed in one single recording [44, 100, 38, 164], while other teams measure movements one after the other with a separation between each [162, 105, 27]. There are more studies that evaluate their subjects continuously than non-continuously. This corresponds to the desire of these studies to work on databases specific to the field of Free-Living (long and extensive). Besides, semi-FLEs' studies record their participants on short periods

(19 semi-FLEs' studies record less than 30 minutes). It adds up to the control such studies already have by constraining the space in which subjects perform activities. There is a significant number of studies that measure their participants over several days though [154, 42].

2.4.5 Inclusion and Exclusion criteria

One of the advantages of ambulatory studies is that they allow longitudinal follow-up of participants varying from a few hours to several days. It is especially interesting for studying pathological patients without being limited to a one-off test of a few minutes during the hospital visit. This explains why a large proportion of pathological case studies are conducted in FLEs [8, 52, 57, 60, 84, 122, 126, 141, 142, 158]. The definition of inclusion and exclusion criteria determines the part of the population targeted by the study and therefore the precision of the study. Participants' characteristics can sometimes affect their motor skills (taking medication that mimics the symptoms of another disease, co-morbidity of pathologies…) hence the importance to describe the latter. Those characteristics (healthy or with specific diseases) are described in Figure 2.12 and further detailed in Figure 2.13.

In studies including only healthy participants (the majority of studies in this chapter), age is a recurring inclusion or exclusion criteria. Some studies focus on elderly populations, aged 65 and over [27, 4, 9, 12, 38, 107, 119, 51]. Others record younger cohorts, with participants between adolescence and under 30 years of age [67, 76, 88, 99, 120, 136, 104, 63]. The remaining studies use the age, gender and BMI of the participants as a criterion [154, 46, 164, 100, 44], or only the average age of included participants [85, 110, 109, 101]. One study in particular includes only breast cancer survivors [83].

In the group of studies focusing on specific cohorts, a substantial part of studies dwell upon neurological pathologies as PD [77, 98, 17, 113, 112, 50], cerebral palsy [33, 2], accidental brain injuries (strokes) [146, 52, 53, 52, 53] and traumatic brain injuries [141]. There is also a significant part of non neurological pathologies such as COPD [142, 60, 126, 122] and obesity [57] where locomotion is affected. All these pathologies have in common locomotion as biomarker which makes them a coherent choice of study.

For all above-mentioned studies, a mandatory inclusion criterion is the official diagnosis of the disease under study [50, 33, 112, 17, 84, 52, 53, 142, 57, 77, 98, 141]. Disease scaling questionnaires are sometimes necessary to include patients in the study like Parkinson pathology severity degree Hoehn and Yahr scale [50, 77, 98, 17, 113]. Functional tests can also be used as inclusion criteria as found in studies with COPD participants for instance for whom a functional lung test is necessary as a diagnosis [122, 126, 142, 60, 158]. A recurring exclusion criterion is the co-morbidity of the studied pathology with another pathology that might have the same motor symptoms[50, 141] and the treatment that could also interfere with motor symptoms [50].

Finally, some inclusion and exclusion criteria are common to both study groups. Whether participants are pathological or not, general health information such as Body Mass Index are collected [164, 44, 63, 33]. Furthermore, in all studies, physical (orthopedic or muscular) and cognitive impairments often are exclusion criteria [61, 50, 33, 104, 12, 46] which are detected by questionnaires: Montreal Cognitive Assessment (MOCA)[50, 113], Mini Mental State Exam (MMSE)[50, 77], Geriatric Depression Scale (GDS)[4, 50] for mental state. It is thus notable that inclusion and exclusion criteria are grouped around age, health and BMI. These 3 characteristics allow for a precise selection of participants.

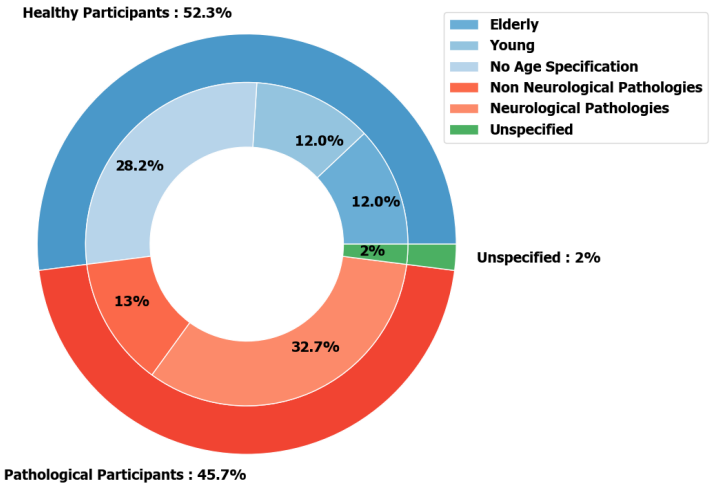


Figure 2.12: Distribution of studies according to participants state of health of participants. In shades of blue are represented the proportion of studies with non-ill participants (in light blue those with young participants, in intermediate blue those with elderly participants). The proportion of studies with ill participants is shown in shades of red. A distinction is made between neurological (dark red) and non-neurological (light red) diseases.

Participant Characteristics	*Nb of Studies*
Healthy participant	
No age specification	26
Elderly participant (>65 years old)	1
Youth participant (<30 years old)	11
Pathological participant	
Parkinson Disease	15
Traumatic Brain Injuria, Stroke, Hemiparetic patients	8
Chronic Obstructive Pulmonary Disease	7
Cancer	2
Cerebral Palsy	2
Obesity	1
Others (Heart Failures, Fallers, Other Neurological Disorders	9
No participant information	2

Figure 2.13: Detail of the number of studies according to the characteristics of included participants. Studies containing healthy participants are divided into three categories.

2.4.6 Annotations and Meta Data

The analysis of referenced articles shows that the "annotation of activities" aspect is essential in several studies, in particular in order to provide verification data and to compute accuracy values for the classifiers used for HAR. Sensors such as cameras, audio recorders,

GPS, force plates, etc. can be used to annotate the timings of activities and compare them with those detected by the classifiers chosen by the investigators. Annotations are not only chosen to give an accuracy score for HAR but also to evaluate the number of Moderate Vigorous Physical Activities MVPA : high energy level activities [55, 136, 99]. Annotations allow to evaluate the number of movements performed with a specific score of MET [104, 142, 158].

As for the cameras, one study investigated how both inertial and vision sensors can simultaneously be used to enable human HAR [35]. According to Figure 2.14, most common devices for annotation are video cameras that are widely used in FLEs [27, 51, 55, 104, 67, 9, 4, 10] and semi-FLEs [41, 105, 17, 101]).

Instead of additional sensors, some studies use direct manual annotations by experimenters[164, 61, 33, 134, 77, 52, 112, 53, 43, 100, 113]. Of course, those types of annotations are only relevant in studies that take place in laboratories (semi-FLEs). In FLEs, participants have to take notes into a diary that is given back to the experimenters at the end of the protocol or when the subjects take off the wearable sensor [88, 76, 98, 122, 77, 85]. In specific studies focused on the evaluation of EE, ground truth annotations can be provided by additional clinical exams. For instance, in [126, 60] the analysis of urine samples is used to track the subject's EE. The patient is asked to collect urine samples at a fixed time which is subsequently analyzed.

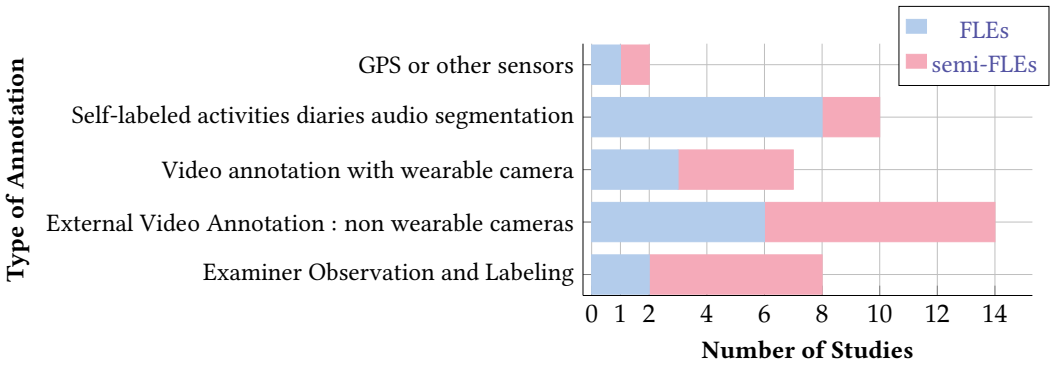


Figure 2.14: Distribution of the different kinds of activities/physical behaviors/stationary phases annotations depending on environments

Observations from examiners to annotate are more used in semi-FLEs than in FLEs where it is more complex to implement. Wearable or fixed cameras and GPS are used at a similar rate between FLEs and semi-FLEs studies. semi-FLEs enable a better control of performed activities and allow for more accurate annotations of these activities. It results in an optimised computation of classifiers' accuracies for HAR. GPS are sometimes used to detect stationary walking phases using trajectory reconstructions (using a standard inertial navigation algorithm termed pedestrian dead-reckoning (PDR) on a study [159]). This allows for the study of the PA of a subject on these phases and to compare the results obtained in FLEs with those obtained in fully controlled environments.

Similarly, some questionnaires are also found to assess fatigue and ability to perform some physical activities and then study the physical capacity of a subject : Multidimensional Fatigue Inventory (MFI)[50], Nottingham Activity of Daily Living Scale[112], Physical Activity Readiness Questionnaire (PAR-Q) [43, 100]. These questionnaires are means

of completing annotations because it enables to assess as good as possible the physical performance of patients.

2.5 Discussion and conclusion

IMUs and accelerometers stand out as the most commonly used sensors in this overview of studies in free environments. One of their main locations are the wrist and the lower-back since this is convenient to use thanks to the development of watch sensors or velcro straps. The characteristics of the free environments in which the studies take place is also a major factor of distinction between the reviewed works. Two types of such environments emerge: FLEs (environments that are not controlled by the experimenter) and semi-FLEs (environments established by research teams : laboratories, indoors out outdoors space...). semi-FLEs' conditions provide a smooth transition from controlled to full ambulatory settings i.e FLEs' conditions. Within those environments, the three most common types of activity studied in the reviewed works are walking, standing and sitting. According to the studies referenced in this state of the art, the instructions given to participants to carry out these activities vary in their degree of explicitness (some impose activities, others suggest them, while others leave the subjects complete freedom in their natural environment). In addition, the recording times for taking measurements via motion sensors vary from one research team to another. The short recording durations (less than five hours) are often conducted in semi-FLEs' studies while the FLEs' studies setup longer recording times (three days or more). In all FLEs studies, annotation has a decisive role in data collection. The quantity, quality and accuracy of the data thus depend on new ways of annotating, hence the diversity of it : diary, video or tape recording. As soon as the implementation of sensors and the setup of the protocols are put in place and enable measurements, the computed data are to be processed. In several contexts, HAR is a crucial step for the study of PA. Finally, the majority of studies in free-living conditions use protocols with a wide variety of implementations. Nevertheless, studies in FLEs conditions are completely standard yet while semi-FLEs conditions are more commonly used. With regard to protocols' characteristics such as instructions, measurement times or annotations, some trends are emerging (in FLEs conditions: longer measurement times, annotations by the participants themselves or by monitoring systems other than examiners are becoming more popular, etc...).

Thus, within studies evaluating the physical activity of participants in free environments, some clear tendencies emerge : the massive use of accelerometers and IMUs compared to gyroscopes, the fact that these sensors are mainly positioned on the wrist, lower back or waist, the predominance of activities such as walking, standing, sitting in the protocols. These few observations, which make it possible to identify choices that are generally shared between studies concerning the implementation of devices and protocols in particular, may prove to be the main recommendations for future studies. Nevertheless, the differences between the studies are still notable and numerous, which prevents from bringing out a standard approach to assess physical activity in free environments.

One recurring question in this chapter is linked to the notion of environment and to the distinction between semi-FLEs and FLEs. As a matter of fact, a shared feature of this continuum of protocols is the psychological dimension of the participants and its impact on their behaviour. The white coat effect (or Hawthorne effect), for example, is caused by

the presence of medical personnel and influences the physiological measurements of the participants by altering their behaviour [156]. Likewise, under experimental conditions, the presence of an experimenter indeed influences the participant's performance [115]. Thus, the "artificial" side of a study is to be taken into account in the analysis of results, even in FLEs environments. Whether it is through social interaction or by knowing that we are part of a test, a psychological effect affects our behaviour. Other psychological tendencies should be taken into account in the analysis of human physical activity. The trend to compare themselves to others illustrated by the "better than average" can also induce differences between participants' physical activities. There is also a current trend towards self-evaluation with the development of self-tracking applications and increasingly connected objects.

Throughout this chapter, the diversity of the studies included is noteworthy. In spite of the search filters, studies are very disparate, at protocols level, where we have shown the diversity of experimental conditions established. There is a splintering of studies according to the sensors brand used which implies a lack of the same sensors configuration (axis definition, accuracy, sampling frequency...). Each brand has its own software and therefore of handling by scientists. This diversity is also found in the dataset used by each study. A large majority of studies record their own data with recruitment and testing of participants. This makes it possible to have customized data very specific to each study. Very few studies included in this state of the art use already existing dataset [42, 160, 84], and even fewer aim at creating an open access dataset. There is no standardization of dataset between the different articles. The diversity and lack of standardization of studies contribute to a compartmentalization of human activity research. This disparity may be a hindrance to research, which justifies the need for unification. One way to unify the studies would be the creation of a public and universally accessible dataset, as in the recent study by Garcia et al [69]. Additional ways to unify the studies could be to work on a common goal such as the challenges (see for instance created by [149] for example), to have a shared aim for the whole community [149]. Although progress has been made in the detection of physical activity using inertial motion sensors, longer, larger and more accurate studies are needed in order to be able to track patients longitudinally. Nevertheless, this state of the art provides further studies intending to develop protocol in free-living environments with clear possible choices concerning sensors' characteristics, placements, number or performed activities, measurements' durations that depend on studies' aims and setup environments.

3

Protocols' Setups and Contexts

3.1 Introduction

The first challenge of this thesis was to set up protocols to record long signals in semi-FLEs and FLEs. These protocols are based on the tools listed in the literature as detailed in Chapter 2 for such environments. Measurements were carried out as follows : the first protocol took place in a semi-FLE, the second protocol in a semi-FLE with freer conditions and finally the last one in a complete FLE.

3.2 Protocols Purposes

Three new semi-FLEs and FLEs protocol were built in collaboration with clinical experts (neurologists, orthopaedic surgeons) in order to perform longitudinal follow-ups, inter-individual comparisons on included cohorts as well as to assess effects of neurological disorders and orthopedic defects on physical activity in free environments.

Protocol 1 aims to be an intermediary between the controlled conditions of clinical protocols and the freer conditions of FLEs protocols. The duration of this protocol has been chosen to be identified with the 6 Min Walking Test (6MWT). In this test, participants are asked to walk as much distance as possible over 6 minutes by walking back and forth along a 30m corridor with instructions from an examiner. In our study, only the 6 minutes duration is retained since it corresponds to average figures used for standard semi-FLEs studies in the literature as presented in Figure 2.11. Indeed, 19 studies in semi-FLEs' conditions (48.7% of studies) were carried out in less than 30 minutes. Our new protocol also includes only a limited number of instructions to preserve the spontaneous behaviors found in ambulatory phases. Participants are not restricted to a 30m corridor and are not needed to walk as far as possible as in the 6MWT for instance. They are nevertheless instructed to perform specific activities : leaning, sitting, standing, climbing up and down stairs, U-turns which are commonly performed by subjects in the literature in FLEs as presented in Chapter 2 as well as in controlled conditions. These activities were also chosen according to the needs of the clinical experts in order to efficiently compare included cohorts.

Protocol 2 was implemented to test whether changes could occur with less supervision and therefore more freedom for participants. Instructions were indeed modified to get closer to FLEs' conditions. The aim was to let participants entirely free to perform their PA by using wearable cameras to annotate transitions. Using a camera so that annotations could be more precise notably between end-of-lap activities. Besides, less activities were

asked to be performed and no specific sequence of activities were instructed as in Protocol 1 in order to add degrees of freedom for participants. By doing so, we gradually come closer to FLEs conditions as depicted in Figure 2.7.

Protocol 3 was built to address identical biomedical purposes but in new FLEs' conditions. As 91% of the studies measure subjects over one hour in FLEs: it was decided to set this protocol's duration to 90 minutes. Besides, almost no instructions were given except to perform a minimum walking duration (10 minutes). Annotations were performed thanks to a wearable camera as in Protocol 2. As such, this protocol is found to be on the far right side from the instructions' frieze presented in 2.7.

3.3 Participants

The ID-RCB number of the committee for the protection of individuals (Comité de Protection des Personnes) in which this study is included is : 2021-A00087-34.

3.3.1 Protocol 1

Fifteen healthy subjects (27.13$\pm$4.35 year-old, 7 men and 8 women) and one pathological patient (suffering from normal pressure hydrocephalus NPH) were measured on a semi-controlled protocol.

3.3.2 Protocol 2

Twenty-one healthy subjects (33.4 $\pm$14.42 year-old, 10 men and 11 women), 6 patients having undergone or about to undergo an orthopedical surgery (43.83$\pm$19.39 year-old, 1 man and 5 women) and 3 patients suffering from a neurological pathology due to a cerebral lesion (72.6 $\pm$4.03 year-old, 1 man and 2 women) were measured on a semi-controlled protocol. Characteristics of participants with neurological or orthopedical affections are presented in Tables 3.1 and 3.2.

3.3.3 Protocol 3

Six healthy subjects (26.17 $\pm$1.46 year-old, 4 men and 2 women) were measured on an unsupervised protocol.

3.4 Sensors

Subjects were equipped with a Shimmer3 IMU (with sampling frequency F_s=100 Hz, battery life = 39-69hrs, storage = 8GB, ±2 g (to ±16 g) for the accelerometer, ±250 dps (to ±2000 dps) for the gyroscope) [78] presented in Figure 3.1 on their lower back L5 as displayed in Figure 3.3 in every protocol. This sensor stands out as one of the most used in unsupervised environments as stated in Chapter 2 (used in 7 studies). Its autonomy, internal storage and high sampling frequency allow for valuable recordings in semi-FLEs and FLEs. Indeed, the sampling frequency can reach values close to values used in clinical studies (>100Hz) which enables precise analysis of the PA. As displayed in Figure 2.6, the Shimmer3 IMU is one of the 6% of sensors referenced in our state of the art studies that

Detailed Affection	Sex	Age	Size	Weight	Walking Speed (m/s)	Observed Walking Difficulty (from + to +++)
immediate preoperative phase of a knee ligamentoplasty	F	25	164	70	1.16	+
immediate preoperative phase of an ACL ligamentoplasty and meniscal tears	M	25	187	92	1.27	+
one day after a total hip replacement	F	70	168	79	1.05	+++
rehabilitation after a knee sprain with ACL rupture	F	51	165	53	1.32	+
multiple right hip surgeries resulting in chronic gluteus medius insufficiency causing Trendelenburg type lameness	F	66	163	73	1.38	++
immediate postoperative phase of an ACL ligamentoplasty and meniscal tears	F	25	164	70	1.25	+

Table 3.1: Characteristics of participants with orthopedical affection in Protocol 2

Detailed Affection	Sex	Age	Size	Weight	Walking Speed (m/s)	Observed Walking Difficulty (from + to +++)
Affection of the cranial pairs and neck muscles after radiotherapy treatment of a Chordoma without locomotion disorder	F	76	157	47	1.22	+
Post-radiation left brachial plexitis that occurred 20 years after radiation treatment for breast cancer. Complete paralysis of the entire left upper limb	F	75	158	69	1.17	+++
Post-radiation leukopathy after treatment of a left temporal glioma	F	67	144	54	1.09	++

Table 3.2: Characteristics of participants with neurological affection in Protocol 2

provides the ability to set the frequency to over 100Hz. It is noteworthy that using Shimmer3 IMUs at high sampling frequencies does not induce a loss in battery capacity which is crucial in free environments. Besides, it can easily be attached to the lower-back thanks to a belt strap.

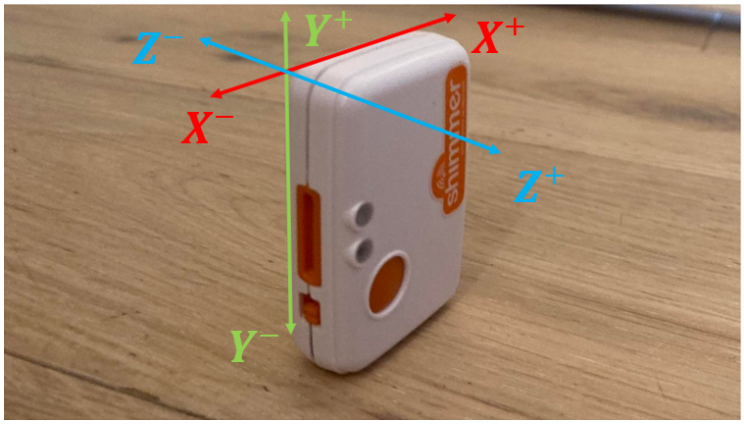


Figure 3.1: Shimmer 3 IMU and associated axis

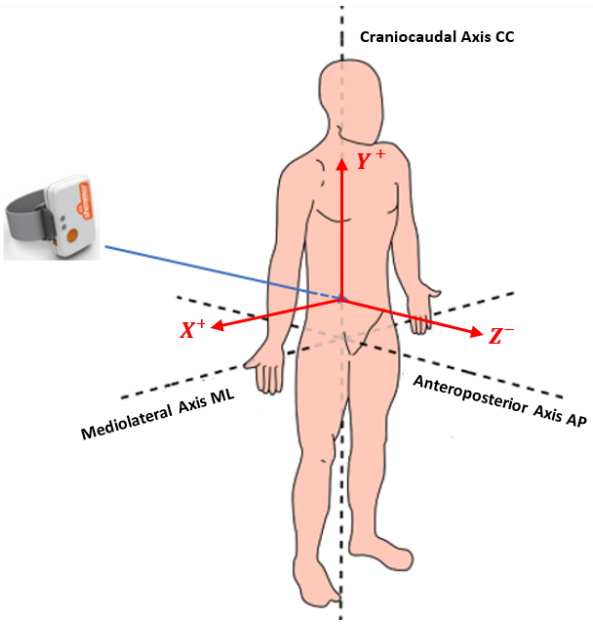


Figure 3.2: Description of axis on which raw signals are recorded from the lower back

Figure 3.3: Inertial sensor location used for implemented protocols

3.5 Pathway, Activities, Transitions, Instructions

3.5.1 Protocol 1

Subjects were asked to complete several laps of the Neurophysiology Department at Percy Hospital (a semi-controlled environment) for a total protocol duration of precisely 6 minutes. All activities were chosen according to the most performed movements in the literature as detailed in Chapter 2. The protocol, illustrated on Figure 3.4, contains several regimes that are either walking phases (denoted **W●**) or activities (denoted **A●**) as displayed in Table 3.3.

Transitions between two walking periods correspond to 90° Turn (as opposed to activities A1 which involves a door opening during the 90° turn) : they will be grouped in the same transition category since they are very similar. At the end of each lap, after having climbed down the stairs, subjects perform three activities in the same order : leaning (A5), standing still (A3) and sitting still (A4) before resuming walking.

Table 3.3: Details of annotated phases.

Activities/Walking Phases	Details
W0	Walk (Start → 1)
W1	Walk (1 → 2)
A1	Door Opening and 90 degrees Turn
W2	Walk (2 → 3)
W3	Walk (3 → 4)
A2	Going up 3 steps stairs U-turn and going down 3 steps stairs
A3	Standing Still
A4	Sitting Still
A5	Leaning
W4	Walk (5 → 6)
W5	Walk (6 → 1)

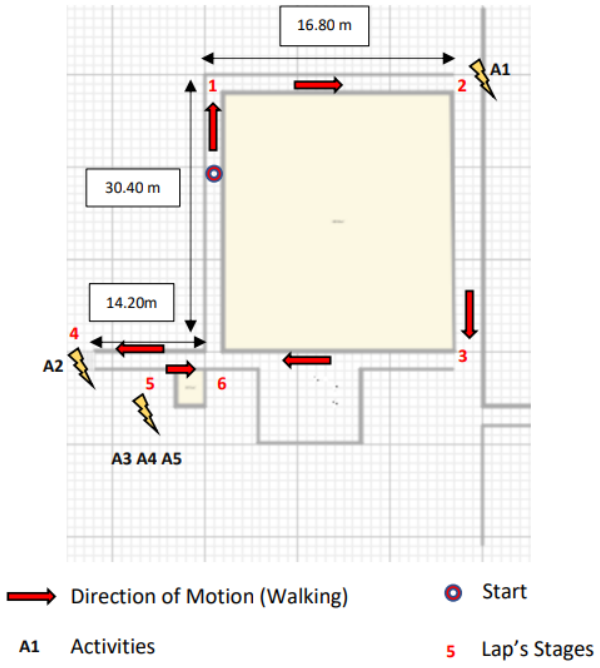


Figure 3.4: Description of the semi-controlled protocol. Numbers displayed indicate the position of the subject during his/her path.

During each experiment, an examiner (an expert able to identify change points) has annotated all change points described above that correspond to transitions between walking phases/activities, activities/walking phases, activities/activities or walking phases/walking phases. These annotations consist in precise timestamps and will be used as ground truth change points' labels for training the supervised segmentation algorithms presented in Chapter 4.

3.5.2 Protocol 2

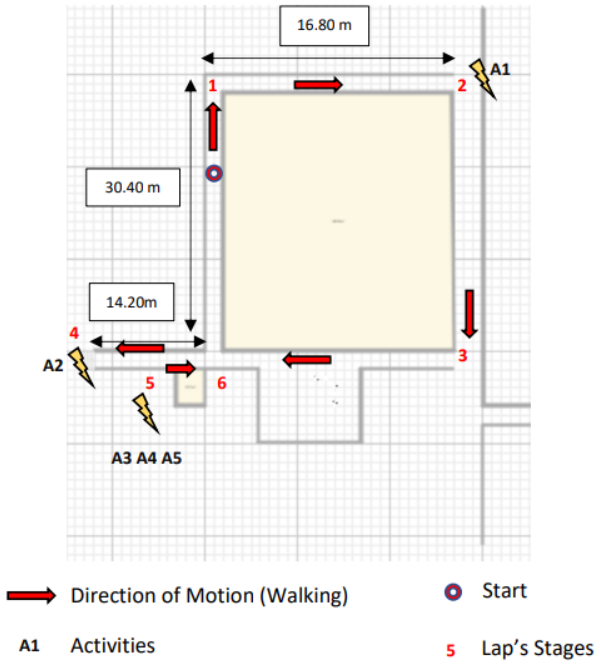


Figure 3.5: Description of the semi-controlled protocol. Numbers displayed indicate the position of the subject during his/her path.

In Protocol 2, subjects were asked to complete several laps of the Neurophysiology Department at Percy Hospital (a semi-controlled environment) and to perform activities at the end of each lap (climbing up and down some stairs, leaning, sitting, standing) for a total protocol duration of precisely 6 minutes.

The protocol, illustrated on Figure 3.5, contains several regimes that are either walking phases (denoted **W●**) or activities (denoted **A●**). During each experiment, change points were better identified than for Protocol 1 thanks to a camera carried by the subject who was let alone to perform the protocol. These annotations were performed in collaboration by one expert and consist in precise timestamps that will be used as ground truth change points' labels. Characteristics of these transitions are displayed in Table 3.3. Participants were only instructed to perform one end-of-lap activity among leaning (A5), standing still (A3) , sitting still (A4) and not to successively perform them all as in Protocol 1 to improve annotations. This change was implemented to improve annotations quality. Illustrations of

transitions recorded by the wearable camera used for annotations are presented in Figure 3.6.

3.5.3 Protocol 3

Subjects were measured in their offices' environment (Université de Paris' offices containing stairs, corridors, elevators...) for 90 mins in a complete unsupervised manner. They were only asked beforehand to perform a short list of activities (leaning, sitting, standing) when they wanted during the experiment. They were also instructed to perform walking for at least 10 minutes whenever they felt to during the experiment. Participants were carrying a wearable camera so that further applied methods' results could be compared to ground truth observations. The short list of activities to perform was suggested as presented in Figure 3.7.

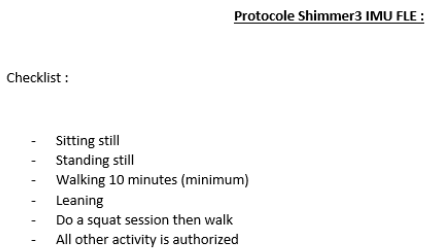


Figure 3.7: Protocol 3 activities' instructions

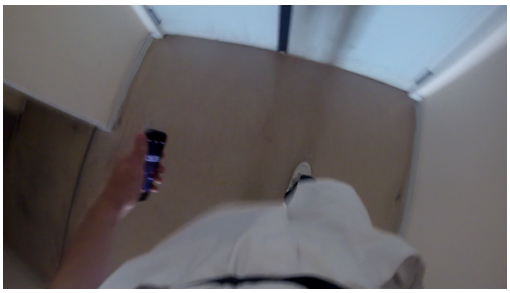
3.6 Recorded signals

Three linear accelerations signals (m/s²) and three angular velocities signals (deg/s) from the lower back (corresponding to three specific axes : craniocaudal CC axis, mediolateral ML axis and anteroposterior AP axis as detailed in Figure 3.2) are recorded in every protocol. Raw signals are filtered between 0.5 and 5 Hz to remove the noise [144, 20, 150] with a Butterworth bandpass filter (4th order).

Two examples from pre-processed signals (craniocaudal angular velocity and anteroposterior acceleration) recorded from one complete experiment performed on Protocol 2 are displayed in Figure 3.8. Those same signals taken from homogeneous walking regimes in an healthy participant from Protocol 2 are presented in Figure 3.9. In this figure, signals seem non-stationary depending on the performed activities. This implies the need for segmentation in order to characterise stationary regimes within the signal. Figure 3.10 displays the same filtered signals during one transition separating the walking phase between corner 1 and corner 2 (W1) and the (A1) activity (opening a fire door and performing a 90° turn) phases. A black line corresponding to the beginning of (A1) activity and to the ending of the walking phase has been added. Both signals appear to differ between (W1) and (A1) portions : anteroposterior acceleration displays higher values on the walking phase and craniocaudal angular velocity displays higher amplitudes in steps' patterns in the activity portion. Steps' patterns also vary in the activity phase compared to the walking phase.

((a)) Start

((b)) W0 → W1



((c)) W1 → A1

((d)) A1 → W2

((e)) W2 → W3

((f)) W3 → A2

((g)) A2 → A4

((h)) A4 → W4

((i)) W4 → W5

((j)) End of the lap

Figure 3.6: Screenshots of performed transitions in Protocol 2 from a wearable camera used for annotations

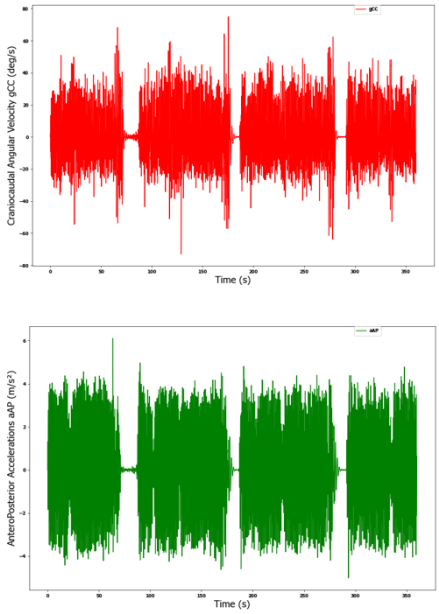


Figure 3.8: Filtered signals retrieved from the lower back : craniocaudal angular velocity and anteroposterior acceleration. A complete healthy participant's recording from Protocol 2 is displayed.

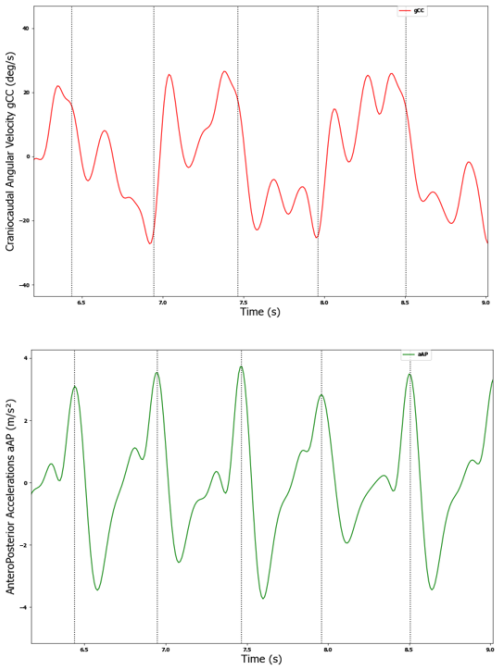


Figure 3.9: Filtered signals retrieved from the lower back : anteroposterior acceleration and craniocaudal angular velocity. Signals are taken from one homogeneous walking regime of one healthy participant from Protocol 2. Two strides (four steps) are included in those examples. Dotted lines correspond to initial feet contacts.

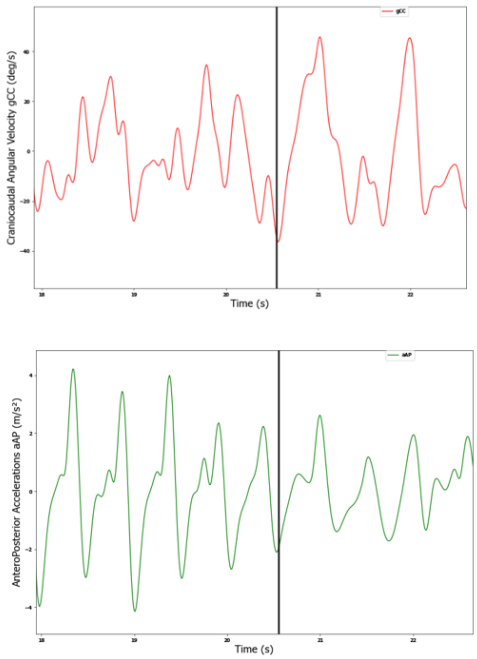


Figure 3.10: Filtered signals retrieved from the lower back of one healthy participant in Protocol 2 : craniocaudal angular velocity and anteroposterior acceleration. Black line corresponds to the beginning of (A1) activity (opening a fire door and performing a 90° turn) and to the ending of the walking phase between corner 1 and corner 2 (W1) for one healthy participant from Protocol 2.

4

Adaptive Change Point Detection Method

4.1 Introduction

In the previous chapter, we described several protocols enabling to respond to identified biomedical issues (longitudinal follow-up, inter-individual comparisons, etc.). These protocols are composed of several homogeneous activities and walking phases that we want to isolate in order to characterize them. One of our first tasks is therefore to segment the signals recovered during the measurements in order to identify these protocols' phases. Once these homogeneous and stationary regimes are segmented, they can be characterized according to their nature, compared with each other and evaluated according to healthy subject standards as presented in Chapter 6.

Intuitively, this segmentation depends on the meaning given to the notion of *change* and *homogeneity* that we seek to detect. In order to adapt the strategy to the signals of interest, we propose to use a supervised approach that learns from annotated data the type of changes that are meaningful in our context and thus the adequate *scale* for the segmentation algorithm. The objective is to find as precisely as possible the instants of change points that are similar to the annotated change points provided by experts.

Several change point detection methods exist in the literature and apply to various study conditions. These methods are developed to meet specific needs: the nature of the signals to be segmented (multimodal or one-dimensional), the annotation strategy, the number of change points. Before being able to choose an approach, a first review of their uses in the literature as well as their associated evaluation metrics is performed. Then, characteristics of our method chosen according to this state of the art, their implementation on our signals, step by step results and a discussion on annotations' strategy are presented.

4.2 State of the art

Notations : In this section, we use the following notations. Let $y = \{y_0, y_1, \ldots, y_T\}$ denote a $\mathbb{R}^d$-valued signal with $d \geq 1$. $\{y_t\}_{t=0}^{t=T}$ is denoted $y_{0..T}$. A $(b-a)$ long sub signal $\{y_t\}_{a+1}^{b}$ between samples a and b $(0 < a < b)$ is denoted $y_{a..b}$. A set of K indexes (or a partition of K indexes) is denoted $\mathcal{T} = \{t_1, t_2, ..., t_K\} \subset \{0, ..., T\}$.

4.2.1 Quantitative criterion and problem statement

In this chapter, we consider a $\mathbb{R}^d$-valued multidimensional non stationary random signal $y = y_0, ...y_T$. This signal is assumed to be piecewise stationary meaning that some of

its characteristics abruptly change at some unknown instants $t_1^*,....t_K^*$. The objective of change point detection methods is to estimate indexes of those instants t_K^*. The number of change points K^* to find may or may not be known.

Our segmentation approach is meant to be offline, meaning that it detects change points after the data has been collected. With regard to these features, a common solution is to define the best partition $\hat{\mathcal{T}}$ (list of detected change point instants) on a signal as the partition that minimizes a previously defined quantitative criterion $\mathcal{V}$. Assumption is made that $\mathcal{V}$ is of sum of costs computed on segmented regimes. The choice of these costs is entirely based on the conditions of application of the change point detection method (number of change points, annotation strategy...).

Let $\mathcal{T} = \{t_1, t_2, \ldots, t_K\}$ be a tested partition with K change points for $\mathcal{T}$ and $c(\bullet)$ the cost function that evaluates the goodness-of-fit of regimes from $\mathcal{T}$:

$$\begin{aligned} \mathcal{V}(\mathcal{T}, y) &= \sum_{k=0}^{K} c(y_{t_k} ... y_{t_{k+1}}), \\ \hat{\mathcal{T}} &:= \arg\min_{\mathcal{T}} \mathcal{V}(\mathcal{T}, y) \end{aligned} \tag{4.1}$$

All segmentation approaches that are to be applied rely then on three specific parameters :

- the definition the cost used for the assessment of the quantitative criterion : the cost function is an evaluation of segmented regimes' homogeneity. Cost values are expected to be low when associated to homogeneous regimes.
- the search method : the procedure that seeks for the optimal partition according to the chosen cost that is to say the lowest possible value of the sum of the costs on all segmented regimes.
- the constraint on the number of change points : a complexity penalty is added if the number of change points to detect is unknown to balance $\mathcal{V}(\mathcal{T}, y)$. The choice of this complexity penalty relies on the amplitude of changes to detect : if the value of the chosen penalty is high, only most significant change points will be detected and inversely if the value of the chosen penalty is low.

4.2.2 Costs

Several types of costs can be used to define $\mathcal{V}$. Let $y_{a..b}$ be values of y within a regime delimited between a and b instants, $\bar{y}_{a..b}$ the empirical mean of y on this regime and $\hat{\Sigma}_{a..b}$ the empirical covariance matrix of this regime. Cost functions are defined as follows :

- The cost related to the mean-shift model $c_{\mathcal{L}_2}$ also called the quadratic error loss [39, 102, 117] which is the most used in the literature. $c_{\mathcal{L}_2}$ is used for a Gaussian distribution with fixed variance. When using this cost, only mean changes within y are detected.. $c_{\mathcal{L}_2}$ can be applied to multidimensional signals.

$$c_{\mathcal{L}_2}(y_a..y_b) = \sum_{t=a+1}^{b} \|y_t..\bar{y}_{a..b})\|_2^2, \tag{4.2}$$

- The c_Σ cost function detects changes in mean and variance. It is used for stock market situations [95] or for biomedical applications [37] for instance :

$$c_\Sigma(y_{a..b}) = (b-a)log(det(\hat{\Sigma}_{a..b})) + \sum_{t=a+1}^{b} (y_t - \bar{y}_{a..b})' \hat{\Sigma}_{a..b}^{-1} * (y_t - \bar{y}_{a..b}), \quad (4.3)$$

- Several other cost functions are used to a lesser extent and will not be further detailed here: the cost function related to a Poisson distribution [40, 89], the cost function adapted to signals with heavy noise tails, the cost function adapted to autoregressive signals (notably ECG/EEG signals) [125] or speech recognition tasks [6].

Instead of using a complex cost function, signals can be transformed so that the changes manifest themselves as changes in the mean for example, in order to use a simpler cost function. It is possible for instance to first transform signals into adequate domains (such as the time-frequency domain) and then detect mean-shifts [151].

4.2.3 Search methods and constraint

There are two types of search methods that can be applied to perform segmentations : optimal methods (Pruned Exact Linear Time (PELT), Optimal Partitioning Method (OPT)) and approximate methods (window sliding [28, 59], binary segmentation [15, 36], bottom-up segmentation [82]) which are not explored further here. Indeed, the considered approach for our study being offline (applied after the data recording) and not requiring fast processing, optimal methods with interesting accuracy results represent the most valuable choice.

These optimal methods are setup to find the optimal partition $\hat{\mathcal{T}}$ whose change points minimize $\mathcal{V}$. They apply to two different conditions: OPT if the number of change points to detect is known [94, 96, 34, 131] and PELT if the number of change points to detect is unknown.

Known number of change points: Optimal Partitioning

This solution based on dynamic programming applies to situations where the number of change points K > 1 to find is known. Optimal partitions are searched as follows :

$$\hat{\mathcal{T}} := \arg\min_{\mathcal{T}} \mathcal{V}(\mathcal{T}, y) \quad (4.4)$$

OPT first determines the last change point prior to T by defining the index t_{K-1} that displays the lowest $c(y_{t_{K-1}}..y_T)$ value. It then relies on following observations to iteratively define $t_1, ..., t_{K-2}$ with $t_0 = 0$ and $t_{K+1} = T$:

$$\begin{aligned} \min_{|\mathcal{T}|=K} \mathcal{V}(\mathcal{T}, y_{0...T}) &= \min_{t_0=0, t_1, ..., t_K, t_{K+1}=T} \sum_{k=0}^{K} c(y_{t_{k...k+1}}) \\ &= \min_{t \leq T-K} \left[c(y_{0..t}) + \min_{t_0=t, t_1, ... t_{K-1}, t_K=T} \sum_{k=0}^{K-1} c(y_{t_{k...k+1}}) \right] \\ &= \min_{t \leq T-K} \left[c(y_{0..t}) + \min_{|\mathcal{T}|=K-1} \mathcal{V}(\mathcal{T}, y_{t...T}) \right] \end{aligned} \quad (4.5)$$

Equation (4.5) shows how to intuitively define the first change point t_1 when optimal partition is already known for previous K-1 change points. OPT displays a complexity of the order $\mathcal{O}(Kn^2)$ and is used for several purposes as DNA sequences [34] or financing [95]...

Unknown number of change points: PELT

PELT method identifies change point instants when the number of change points to find is unknown. Let $|\mathcal{T}|$ be the number of change points, PELT method integrates a linear penalty $\beta|\mathcal{T}|$ which allows to balance out the goodness-of-fit of term $\mathcal{V}(\mathcal{T},y)$. Optimal partitions are searched as follows :

$$\hat{\mathcal{T}} := |\hat{\mathcal{T}}|,\hat{t_1},\ldots,\hat{t_{|\mathcal{T}|}} := \underset{K,t_1,\ldots,t_{|\mathcal{T}|}}{\arg\min} \left(\sum_{k=0}^{|\mathcal{T}|+1} c(y,\{t_k\}_{k=1}^{|\mathcal{T}|}) + \beta|\mathcal{T}| \right) \quad \text{with } t_0 := 0 \text{ and } t_{|\mathcal{T}|+1} := T \text{ by convention.} \tag{4.6}$$

β is the smoothing parameter that must be adapted accordingly to the number of change points finally detected. β allows for a trade-off between a sufficient complexity for a satisfactory accuracy and a low computational cost. This β value is critical, as it controls the sensitivity of the algorithm: large β will only detect strong breaks (change of activity for instance) while low β will also detect small breaks (change within the walking phases).

(Note that $|\hat{\mathcal{T}}|,\hat{t_1},\ldots,\hat{t_{|\mathcal{T}|}}$ depend on β and y.) For a fixed β, this discrete optimization problem is solved efficiently in linear time $\mathcal{O}(n)$ [87] thanks to the PELT method. This pruning method iteratively discards indexes from the list of possible last change points to find the latter. In this approach, each sample is considered separately and iteratively. It can be rejected or not from a list of possible change points according to the following explicit pruning rule for two indexes t and s $(t < s < T)$:

$$\text{if } \left[\min_{\mathcal{T}} \mathcal{V}(\mathcal{T},y_{0\ldots t}) + \beta|\mathcal{T}| \right] + c(y_{t\ldots s}) \geq \left[\min_{\mathcal{T}} \mathcal{V}(\mathcal{T},y_{0\ldots s}) + \beta|\mathcal{T}| \right] \text{ with t<s<T} \quad \text{holds, then t cannot be the last change point prior to T} \tag{4.7}$$

If (4.7) holds, for any $T > s$, the optimal partition with the last change point prior to T being s will be better than any partition with the last change point prior to T being t. As such, a list of optimal change points can be found iteratively by searching change points with indexes that have not been discarded. PELT has been used in several application sectors, notably for oceanographic data [87] and DNA data [73, 103].

4.2.4 Evaluation Metrics

Existing evaluation metrics used to compare predicted segmentations and true segmentations are presented here. Let $\mathcal{T} = t_1,...,t_K$ be a computed partition with K change points and $\mathcal{T}^* = t_1^*,...,t_K^*$ a true partition with K^* change points.

Usual metrics: $F1_{usual}, Recall_{usual}, Precision_{usual}$

A predicted change point is a True Positive (TP) if it is close to a true change point (within a specific positive temporal margin). This margin corresponds to the maximum accepted error for a change point. It must be less equal to the minimum temporal distance between two true change points. Recall which corresponds to the proportion of true change points that are correctly predicted is the ratio of the number of TPs to the number of true change points. Precision corresponds to the proportion of predicted change points that are linked with true change points is the ratio of the number of TPs to the number of predicted change points.

$$\begin{aligned} Precision_{usual} &= \frac{TP}{K} \\ Recall_{usual} &= \frac{TP}{K^*} \\ F1_{usual} &= 2\frac{Precision_{usual} Recall_{usual}}{Precision_{usual} + Recall_{usual}} \end{aligned} \tag{4.8}$$

Best values for Precision, Recall and F1-Scores are 1 while poorest values are 0.

Intersection over Union Metrics: $F1Score_{IoU}$ **,** $Precision_{IoU}$**, and** $Recall_{IoU}$

Another set of metrics similar to Usual Metrics is computed. It is an extension of the basic metrics that allows not to use a fixed tolerance (imposed margin). The segmentation procedure is also assessed with 3 standard evaluation metrics: precision, recall and F1-score. These metrics are computed by using the Intersection Over Union (IOU) index (also known as Jaccard index) [97] between detected and real regimes. IOU quantifies the overlapping degree between ground truth and predicted regimes. It is obtained by dividing the value in samples of the full overlap between two regimes by the total number of samples covered by both evaluated regimes. For each predicted regime, when compared to all annotated regimes, if the IOU score is above a specific threshold for one specific regime, the predicted regime is considered as a True Positive (TP). If there is more than one annotated regime associated to a predicted regime, only one annotated regime (with the highest value of IOU) is definitively associated to this predicted regime. Then, precision, recalls and F1-scores are computed according to Equation (4.9). Small temporal shifts between predicted and real change points that may be due to annotation approximations are not penalized.

$$\begin{aligned} Precision_{IoU} &= \frac{TP}{K} \\ Recall_{IoU} &= \frac{TP}{K^*} \\ F1Score_{IoU} &= 2\frac{Precision_{IoU} Recall_{IoU}}{Precision_{IoU} + Recall_{IoU}} \end{aligned} \tag{4.9}$$

Best values for $Precision_{IoU}$, $Recall_{IoU}$ and $F1Score_{IoU}$ are 1 while poorest values are 0.

Delta Time Error

Delta Time Error Δ_{TE} computes the difference between the instant of a predicted TP and its associated annotation (groundtruth change point). It enables to evaluate the average

offset obtained when trying to predict a specific type of change point for instance. Let t_k^* and t_k respectively denote the sample of an annotated change point and the sample of its associated predicted change point.

$$\Delta_{TE}(\mathcal{T},\mathcal{T}^*) = |t_k^* - t_k| \tag{4.10}$$

Annotation Error

Annotation error AE [152] corresponds to the difference between the number of predicted change points K and the number of true change points K^*. This metric penalizes both oversegmentation and undersegmentation. An annotation error percentage $AE_{\%}$ is also computed in order to compare the value of the annotation error with the total number of true regimes.

$$\begin{aligned} AE(\mathcal{T},\mathcal{T}^*) &= |K^* - K|, \\ AE_{\%}(\mathcal{T},\mathcal{T}^*) &= \frac{AE}{K^*} \end{aligned} \tag{4.11}$$

Haussdorf metric

Haussdorf metric HM [25] evaluates the worst error made when predicted change points over all listed change points. It measures the greatest temporal distance between either one predicted change point from any true change points or between one true change point from any predicted change point. HM is expressed in number of samples :

$$HM(\mathcal{T},\mathcal{T}^*) = max\left[\underset{t^*\in\mathcal{T}^*}{max}\left(\underset{t\in\mathcal{T}}{min}(|t^*-t|)\right) \quad ; \quad \underset{t\in\mathcal{T}}{max}\left(\underset{t^*\in\mathcal{T}^*}{min}(|t^*-t|)\right)\right] \tag{4.12}$$

It also penalizes both oversegmentation and undersegmentation. When using the Haussdorf metric on long signals, one single poorly evaluated change point can lead to poor results.

Rand index metric

Rand index $RI(\mathcal{T},\mathcal{T}^*)$[93] evaluates the total number of agreements between a true partition $\mathcal{T}$ and its associated predicted partition $\mathcal{T}^*$. Each sample from $\mathcal{T}^*$ is compared to all other samples from $\mathcal{T}^*$ to define which couples of samples belong to the same regimes : $gr(\mathcal{T})$ and which couples of samples do not belong the same regime : $ngr(\mathcal{T})$. The same procedure is performed for samples from $\mathcal{T}^*$: $gr(\mathcal{T}^*)$ and $ngr(\mathcal{T}^*)$.

$$\begin{aligned} gr(\mathcal{T}) &= \{(s,t), 1 \leq s \leq t \leq T \text{ s.t s and t belong to the same regime according to } \mathcal{T}\ \}, \\ ngr(\mathcal{T}) &= \{(s,t), 1 \leq s \leq t \leq T \text{s.t s and t belong to different regimes according to } \mathcal{T}\ \}. \end{aligned} \tag{4.13}$$

The number of samples' couples to belong both to $gr(\mathcal{T}^*)$ and $gr(\mathcal{T})$ is computed : $|gr(\mathcal{T}^*) \cup gr(\mathcal{T})|$ as well as the number of samples' couples to belong both to $ngr(\mathcal{T}^*)$ and $ngr(\mathcal{T})$: $|ngr(\mathcal{T}^*) \cup ngr(\mathcal{T})|$.

$RI(\mathcal{T},\mathcal{T}^*)$ can then be calculated :

$$RI(\mathcal{T},\mathcal{T}^*) = \frac{|gr(\mathcal{T}^*) \cup gr(\mathcal{T})| + |ngr(\mathcal{T}^*) \cup ngr(\mathcal{T}^*)|}{T(T-1)} \tag{4.14}$$

$RI(\mathcal{T},\mathcal{T}^*)$ is normalized between 0 (both partitions display a total disagreement) and 1 (both partitions display a total agreement). The Rand Index can provide a detailed evaluation but the associated computational cost when it is computed on long signals is extremely high. Besides, $RI(\mathcal{T},\mathcal{T}^*)$ can display high values on long signals since the $|gr(\mathcal{T}^*) \cup gr(\mathcal{T})| + |ngr(\mathcal{T}^*) \cup ngr(\mathcal{T}^*)|$ term can display higher values when compared to short signals which are not sufficiently attenuated by $T(T-1)$.

4.3 Proposed Method

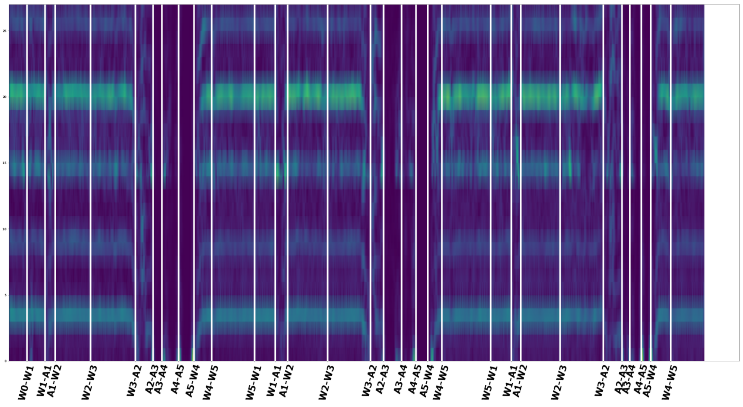
4.3.1 Introduction

Our retrieved signals present several different types of changes, which makes it complex to define an adapted cost function. It was therefore decided to transform our data into time-frequency domain where variations become mean variations. The $c_{\mathcal{L}_2}$ cost is thus chosen to be further used in our method. This simple cost also enables to reduce computational heaviness. PELT method is chosen because of its enhanced computational cost when compared to OPT method. A linear penalty constraint is implemented to use the PELT method.

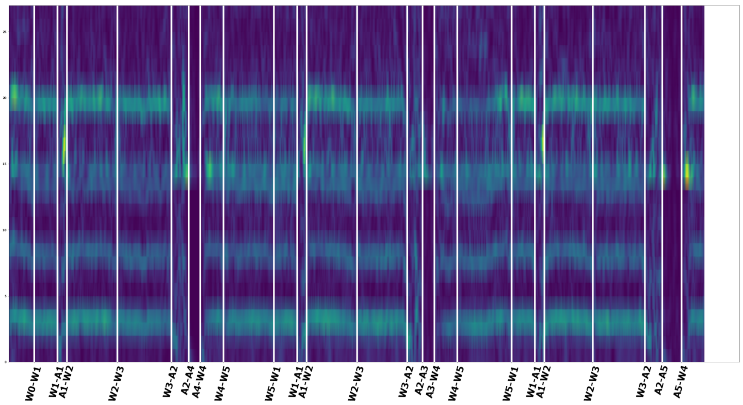
4.3.2 Data Transformation

As detailed in Chapter 3, three linear accelerations signals (m/s²) and three angular velocities signals (deg/s) from the lower back are recorded throughout the protocol. As explained in the introduction, these signals do not have significant mean-shifts and must be modified in order to be able to detect changes. A new domain other than time domain is thus introduced : signals are transformed in the time-frequency domain which enables to better identify these mean-shifts. Indeed, in recorded signals, an alternation of regular periodic regimes with easily noticeable harmonic frequency structures is identifiable which implies obvious mean shifts on investigated spectrograms.

More specifically, two signals of interest are extracted from the raw data (IMU recordings): the craniocaudal angular velocity (gCC) and the anteroposterior acceleration (aAP). These signals are chosen as they are directly influenced by the changes that can be observed during the execution of the protocol (beginning and end of gait, short activities, half turns and turns...). Some studies already use these signals to meet objectives similar to ours (turn detection, detection of activities of daily life...) [129, 111]. These signals are normalized (zero mean and unit variance) before being transformed in the time-frequency domain through Short Term Fourier Transform (STFT) (3 second window length and 0.1s hop length). Only the 0.5 - 5 Hz frequency band, where phenomena of interest are contained, is kept [150]. The norms of the STFT coefficients of each aAP and gCC signals are computed and concatenated, providing $d = 28$ frequency bins per frame (14 per signal). The output data is considered by the segmentation algorithm as a d-dimensional multivariate signal. Output spectrograms for protocols 1 and 2 are displayed in Figure 4.1. In both spectrograms, differences in spectral signatures are visually noticeable between walking



((a)) STFT Spectrogram Protocol 1



((b)) STFT Spectrogram Protocol 2

Figure 4.1: Example of concatenated Spectrograms of STFTs from Craniocaudal Angular Velocity and Anteroposterior Acceleration from protocols 1 and 2. White lines correspond to transitions between activities.

phases and activity phases and notably (A1) (fire door and turn) and (A2) (stairs). Indeed, mean shifts from walking phases and (A1) or (A2) are apparent to the naked eye. Nevertheless, mean shifts between two walking phases are less visible which indicates possible difficulties to detect those transitions.

4.3.3 Annotations and change points

Transitions and Activity/Walking phases

Protocols illustrated in Chapter 3 contain several regimes that are either walking phases (denoted **W●**) or activities (denoted **A●**). During each recording, change points are identified thanks to an examiner that follows participants during recordings in Protocol 1 and thanks to a camera carried by the subject who is let alone to perform his deambulation in Protocol 2. These change points correspond to transitions between walking regimes and activities' regimes, activities' regimes and walking regimes, activities' regimes and other activities' regimes or walking regimes and other walking regimes. Characteristics

of the phases that start and end to these transitions from the two semi-FLEs protocols are displayed in Table 3.3.

Normalized Mean Shift Amplitudes

Intuitively, annotated change points are of various nature, and in order to investigate their properties, we propose to use a metric called normalized mean-shift amplitude $\hat{\Delta}$ computed for each type of change point [147]. This quantity allows to investigate to what extend an annotated change point separating two regimes *left/right* is *hard* to detect. Given a multivariate signal with d dimensions and an annotated change point, the squared normalized mean-shift amplitude $\hat{\Delta}^2$ is defined as :

$$\hat{\Delta}^2 = \frac{1}{d}\sum_{i=1}^{d}\frac{(\mu_{left,i}-\mu_{right,i})^2}{\frac{\sigma^2_{left,i}}{n_{left}}+\frac{\sigma^2_{right,i}}{n_{right}}}, \tag{4.15}$$

where $\mu_{left/right,i}$ and $\sigma^2_{left/right,i}$ are respectively the mean and variance of the signal on the left/right regime and dimension i, and $n_{left/right}$ the number of samples of left/right regimes. The higher $\hat{\Delta}_k$ is, the more different the two regimes are and the easier the change points are to detect. As some regimes are significantly longer than the average regime length, $\hat{\Delta}_k$ computations can show excessive values for some transitions. An additional $\hat{\delta}_k$ metric based on $\hat{\Delta}$ calculation (4.15) is also introduced to mitigate this excessive influence of regime length. If n_{right} or n_{left} exceed 100 frames (i.e. 10s), these numbers are fixed at 100.

$$\hat{\delta}^2 = \frac{1}{d}\sum_{i=1}^{d}\frac{(\mu_{left,i}-\mu_{right,i})^2}{\frac{\sigma^2_{left,i}}{n_{left}}+\frac{\sigma^2_{right,i}}{n_{right}}}, \text{with } n_{left} = min(n_{left},100) \text{ and } n_{right} = min(n_{right},100) \tag{4.16}$$

Table 4.1 displays the normalized mean-shift amplitude as well as additional information on all change points from Protocol 1 considered in the article (type of regimes and regimes' length). Table 4.2 displays the same characteristics for transitions from Protocol 2. It appears that transitions separating static phases from others tend to have high $\hat{\Delta}$ values (> 10). Change points separating slow medium phases from motion long phases also display high $\hat{\Delta}$ values. Transitions separating slow short phases from motion long phases present lower $\hat{\Delta}$ values (<9) as well as transitions dividing motion phases. On the other hand, lowest $\hat{\Delta}$ values are found for transitions separating two static phases (i.e. are the most difficult to detect) : transitions 6 and 7 from Table 4.1.

Transitions of type 1 display the biggest drop of value between Δ and δ which indicates they are to be also much more difficult to detect than other transitions. It correlates with the observation that mean shifts in spectrograms between two walking phases are less visible than for other transitions (between motion and static phases for instance).

4.3.4 Algorithms

In this study's context, the changes to detect are transitions between activities of the protocol and it is assumed that those transitions correspond to mean shifts in the concatenated spectrograms (see Section 4.3.2). Formally, let $y = \{y_1, y_2, \ldots, y_n\}$ denote a $\mathbb{R}^d$-valued

Transition number	Type of change points	change points characteristics			
		$\hat{\Delta}$	$\hat{\delta}$	Types of regimes	Regimes' length
1	W0→W1 W2→W3 W4→W5 W5→W1	6.63	5.45	Movement/Movement	Long/Long or Medium/Long
2	W1→A1	7.09	7.03	Movement/Slow	Long/Short
3	A1→W2	7.55	7.21	Slow/Movement	Short/Long
4	W3→A2	11.11	9.58	Movement/Slow	Long/Medium
5	A2 → A5	10.03	10.02	Slow/Static	Medium/Short
6	A5 → A4	4.88	4.91	Static/Static	Short/Short
7	A4 → A3	4.79	4.81	Static/Static	Short/Short
8	A3 → W4	13.11	12.76	Static/Movement	Short/Long or Short/Medium

Table 4.1: Details of change point's categories from Protocol 1. The length of the regimes segmented by each kind of change point is provided (short $<$ 5s,5s $<$ medium $<$ 10s and long $>$ 10s) as well as the type of these regimes and the average number of times they occur in a lap.

Transition number	Type of change points	change points characteristics			
		$\hat{\Delta}$	$\hat{\delta}$	Types of regimes	Regimes' length
1	W2→W3 W4→W5 W5→W1	6.69	5.33	Movement/Movement	Long/Long or Medium/Long
2	W1→A1	7.62	7.44	Movement/Slow	Long/Short
3	A1→W2	8.13	7.30	Slow/Movement	Short/Long
4	W3→A2	11.34	10.13	Movement/Slow	Long/Medium
5	A2 → A3	12.69	11.96	Slow/Static	Medium/Medium
6	A2 → A4	12.36	12.25	Slow/Static	Medium/Medium
7	A2 → A5	13.31	12.12	Slow/Static	Medium/Medium
8	A3 → W4	17.60	17.00	Static/Movement	Medium/Long or Medium/Medium
9	A4 → W4	17.82	16.36	Static/Movement	Medium/Long or Medium/Medium
10	A5 → W4	18.15	17.25	Static/Movement	Medium/Long or Medium/Medium

Table 4.2: Details of change point's categories from Protocol 2. The length of the regimes segmented by each kind of change point is provided (short $<$ 5s,5s $<$ medium $<$ 10s and long $>$ 10s) as well as the type of these regimes and the average number of times they occur in a lap.

signal with n samples associated to a spectrogram's signal. For K change point indexes

t_k ($\mathcal{T} = 1 < t_1 < t_2 < \cdots < t_K < T$), a common measure of approximation quality is the empirical quadratic risk that uses the chosen cost $c_{\mathcal{L}_2}$ for our segmentation method as detailed in Section 4.2. When using $c_{\mathcal{L}_2}$, the quantitative criterion $\mathcal{V}$ to minimize is defined as :

$$\mathcal{V}(\mathcal{T}, y) := \sum_{k=0}^{K+1} \left(\sum_{t=t_k}^{t_{k+1}-1} \|y_t - \bar{y}_{t_k..t_{k+1}}\|^2 \right) \tag{4.17}$$

where $\bar{y}_{t_k..t_{k+1}}$ is the empirical mean of $y_{t_k}, \ldots, y_{t_{k+1}-1}$ and $t_0 := 1$ and $t_{K+1} := n$ by convention. The risk (4.17) is simply the error when approximating y by a piecewise constant signal. The objective is to find the change points t_k that minimize this risk. When the number of breaks K is not known (which is the case here), the empirical quadratic risk is penalized with a linear penalty and the optimal change points are :

$$\hat{\mathcal{T}} := \arg\min_{\mathcal{T}} \left(\underbrace{\mathcal{V}(\mathcal{T}, y) + \beta K}_{R_\beta(\mathcal{T}, y)} \right) \tag{4.18}$$

This optimal partition is found by using PELT as detailed in Section 4.2.

Calibration of β

Instead of manually calibrating this β parameter by trial and error, a supervised approach described in [148] is used in this study to learn the optimal parameter to detect the changes we are interested in. In a nutshell, this procedure takes as input a collection of N labeled signals $y^{(1)}, ..., y^{(N)}$, meaning that for each $y^{(i)}$, an expert manually provided the set of true change point indexes $\mathcal{T}^{*(i)}$. The optimal smoothing parameter, denoted $\hat{\beta}_{\text{opt}}$, is such that the risk of true expert segmentation is closest to the one of the predicted segmentation, denoted $\hat{\mathcal{T}^{(i)}}$:

$$\hat{\beta_{\text{opt}}} := \arg\min_{\beta > 0} \frac{1}{N} \sum_{i=1}^{N} \varepsilon(y^{(i)}, \beta). \tag{4.19}$$

$$\varepsilon(y^{(i)}, \beta) = \left(R_\beta \left(\mathcal{T}^{*(i)}, y^{(i)} \right) - R_\beta \left(\hat{\mathcal{T}^{(i)}}, y^{(i)} \right) \right). \tag{4.20}$$

Intuitively, the algorithm searches for the penalty β that allows to reproduce the annotations, by forcing the $\beta-$optimal solution $\hat{\mathcal{T}^{(i)}}$ to be as close as possible to the ground truth partition $\mathcal{T}^{*(i)}$. This optimization is performed by minimizing the average excess penalized risk ε(y,β) (4.19) which is the average difference between risks from true expert segmentations and predicted segmentations as detailed in Equation (4.20). The excess penalized risk is by definition a convex function w.r.t. β. Indeed, since function $\beta \to R_\beta(y, \mathcal{T}^*)$ is affine and function $\beta \to R_\beta(y, \hat{\mathcal{T}})$ is a concave function (pointwise minimum of a set of affine functions as detailed in Figure 4.2), $\varepsilon(y, \beta)$ which is the difference of those functions is convex.

We therefore use Brent's method [26] as a convex optimization tool to minimize this component for each training signal. This approach uses two bound values of β (β_a and β_b). The minimum of a simplified parabola exactly containing three points : $\{\beta_a, \varepsilon(y, \beta_a)\}$,

$\{\beta_b, \varepsilon(y,\beta_b)\}$ and $\{\beta_c, \varepsilon(y,\beta_c)\}$ with $\beta_c = \frac{\beta_a+\beta_b}{2}$ is computed and replaces the previous β_b value. This "bound" method is then repeated : several minima of iterative parabolas between two specific provided beta bound values are found the same way to finally converge towards the optimal β value : $\hat{\beta}_{\text{opt}}$. A visualization of this method is detailed in Figure 4.3.

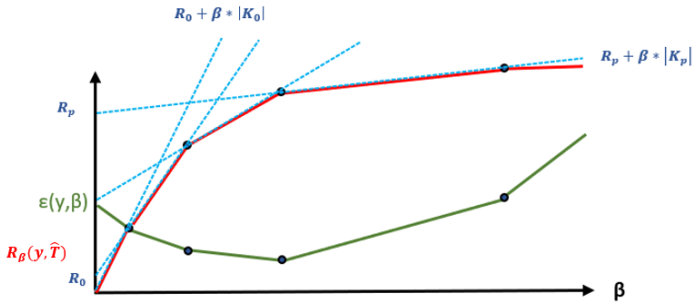


Figure 4.2: Excess penalized risk ε displayed as a convex function in green. Dashed lines correspond to penalized empirical risks obtained for some specific segmentations. These dashed lines are a set of affine functions whose components are fixed risks and the number of change points of each tested segmentation. The higher the number of these change points, the lesser is the value of fixed risks and inversely hence the displayed shapes of these dashed lines. Red line corresponds to $R_\beta(y,\hat{T})$ which is the pointwise minimum of this set of affine functions.

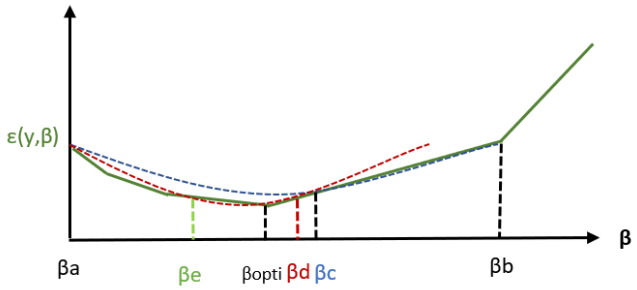


Figure 4.3: Brent's methods for convex minimization

4.4 Results

In the next section, we conduct an assessment of our supervised segmentation method with adapted metrics. All simulations are run with 3-fold cross-validation for Protocol 1 as well as for Protocol 2. To that aim we split the data set into 3 balanced sets (two training sets and one validation set) of 5 healthy participants in Protocol 1 and into 3 balanced sets (two training sets and one validation set) of 7 healthy subjects and 3 pathological subjects (2 participants having undergone or about to undergo orthopedical surgery and 1

neurological patient). This cross-validation allows to verify that the algorithm developed in this study can be used on new unseen data to apply the desired segmentation.

4.4.1 Evaluation metrics

Global metrics

IOU metrics ($F1_{IoU}$, $Recall_{IoU}$, $Precision_{IoU}$), usual metrics $F1_{usual}$, $Recall_{usual}$, $Precision_{usual}$ Haussdorf, Annotation Error, Rand Index can be used to assess segmentations on all identified change points : they are defined as global metrics as opposed to change points' metrics that assess segmentation only for a specific type of change point.

Concerning usual metrics, a predicted change point is a True Positive (TP) if it is within a margin of 3.5 seconds to a true change point. This margin corresponds to the maximum error for a transition observed in a human being from the recordings in this study.

Three IOU's threshold values (0.3, 0.4 and 0.5) which correspond to general thresholds used in other segmentation studies [97, 130] are tested. These thresholds allow for one predicted regime to be correctly associated to a true regime even if all samples from the predicted regime are not contained in the true regime and inversely.

Change points' metrics

In both Protocol 1 and Protocol 2, change points' categories can be evaluated separately by computing associated recall values from usual metrics. We have already noted in Section 4.3.3 that considered change points may not have the same degree of detection's complexity. We therefore perform an evaluation considering only one type of change points at a time. Only recall values can be computed to render metrics associated to specific categories of transitions since precision implies the use of the number of predicted change points which is bound to produce insignificant results. When a TP is detected, the difference in seconds between the calculated change point and the associated annotation can be computed: this corresponds to the Δ_{TE} metric which is computed and detailed in this section. Change points metrics' computations are displayed on Table 4.4 for Protocol 1 and on Table 4.6 for Protocol 2.

4.4.2 Protocol 1

Usual and IoU Metrics

Averaged usual metrics and IOU metrics for Protocol 1 are displayed in Table 4.3. Both usual and IOU metrics are satisfactory and indicate that this study's change point detection procedure is accurate for most change points, and also that there is no oversegmentation. Besides, IOU metrics' values obtained with the chosen overlap threshold (0.4) are close to usual metrics' values which indicates that regimes analysis (IOU metrics) provides the same interpretation outputs that a more specific analysis of the change points (usual metrics) offers.

It is noteworthy that as the overlap thresholds increase (i.e. the importance given to the intersection of the compared regimes in relation to their union increases), IOU metrics are likely to decrease. Moreover, for a tenth increase in threshold value (from 0.4 to 0.5), results that are similar to the usual metrics results at 0.4 (f1=0.75) differ quite significantly for the threshold set at 0.5 (f1=0.63). This indicates the difficulty in accurately estimating

change points. A threshold set below 0.5 allows predicted regimes to be associated with an annotated regime even if a turn transition, for example, has not been correctly predicted. Otherwise, with a higher threshold value, there would be an over-penalisation of slight undersegmentations for instance.

Usual Metrics	$F1_{usual}$	0.75 ± 0.02
	$Precision_{Usual}$	0.73 ± 0.09
	$Recall_{Usual}$	0.79 ± 0.04
IoU Metrics : Overlap = 0.3	$F1_{IoU}$	0.80 ± 0.03
	$Precision_{IoU}$	0.77 ± 0.09
	$Recall_{IoU}$	0.84 ± 0.04
IoU Metrics : Overlap = 0.4	$F1_{IoU}$	0.75 ± 0.05
	$Precision_{IoU}$	0.73 ± 0.10
	$Recall_{IoU}$	0.79 ± 0.03
IoU Metrics : Overlap = 0.5	$F1_{IoU}$	0.63 ± 0.07
	$Precision_{IoU}$	0.62 ± 0.11
	$Recall_{IoU}$	0.66 ± 0.03
Other Metrics	Haussdorf	1472.07 ± 213.00
	Rand Index	0.97 ± 0.001
	AE	7.60 ± 2.12
	$AE_{\%}$	22.61 %
Additional Characteristic	Number of true regimes	33.6 ± 1.7

Table 4.3: Change point detection results for Protocol 1.

Haussdorf, Randindex, AE, $AE_{\%}$

The value of the Haussdorf metric is high (14.72s) which is explained by the observation made in Section 4.2.4 : a single faulty detection of one change point on long regime portions (walking phases in the long corridors separated by turns in particular) could have induced this important deviation.

The percentage error annotation $AE_{\%}$ shows that the differences between the number of predicted and actual change points is around 20%. This difference is probably due to the difficulty of detecting one type of break by our segmentation method (especially turns as it is presented in Subsection 4.4.2). In one acquisition, an average of 12 turns are made by a participant: if the segmentation method fails to detect half of these turns for instance and assuming that there is no over-segmentation, the difference measured by $AE_{\%}$ is reached.

As for the Rand Index, the value obtained is extremely high (0.97) but this is due to the observations made in Section 4.2.4 which explain that this parameter is very sensitive to the length of our signals (here the number of samples is 36000).

Change points' Metrics

Transitions	Type of change points	change points characteristics				Accuracy results	
		$\hat{\Delta}$	$\hat{\delta}$	Types of regimes	Regimes' length	$Recall_{usual}$	ΔTE(s)
1	W2→W3 W4→W5 W5→W1	6.63	5.45	Movement/Movement	Long/Long or Medium/Long	0.71 ±0.07	1.92 ± 0.14
2	W1→A1	7.09	7.03	Movement/Slow	Long/Short	0.98 ±0.02	0.56 ±0.07
3	A1→W2	7.55	7.21	Slow/Movement	Short/Long	0.99 ±0.01	0.60 ±0.07
4	W3→A2	11.11	9.58	Movement/Slow	Long/Medium	0.98 ±0.02	1.70 ±0.32
5	A2 → A3	10.03	10.02	Slow/Static	Medium/Short	0.88 ±0.07	1.02 ±0.18
6	A3 → A4	4.88	4.91	Static/Static	Short/Short	0.30 ±0.21	0.96 ±0.65
7	A4 → A5	4.79	4.81	Static/Static	Short/Short	0.81 ±0.08	0.91 ±0.20
8	A5 → W4	13.11	12.76	Static/Movement	Short/Long or Short/Medium	0.93 ±0.01	1.48 ±0.28

Table 4.4: Details of change point's categories and recall results for PROTOCOL 1.

In protocol 1, change points associated to transitions between walking phases and activities are especially well detected (transition types 2, 3, 4 and 8) with recalls above 0.9. Turns between two walking phases (transition type 1) are less precisely detected (recall 0.71). The poorest results are obtained for transitions 6 and 7 dividing the static activities (leaning, sitting, standing). This result is probably due to the fact that these transitions occur between two short regimes: transitions 6 and 7 are within two seconds from each other on average. Therefore, when a change point is computed in their proximity, it is very challenging to define for which transition specifically it is computed and to precisely assess overlaps' influence. That is why their recall results can therefore not be considered separately but jointly. Besides, these transitions separate two sedentary regimes with similar spectral signatures hence the difficulty to perform an efficient segmentation for this category of change points. These results emphasize the correlation between the observations made on $\widehat{\Delta}$ values (made in Subsection 4.3.3) and the difficulty of detecting these particular change points.

One of the largest Δ_{TE} is obtained for Transition 4 (between the Walking phase (W3) and the Stairs activity (A2)). This is due to the fact that the deceleration phase of the participants before the stairs appears well upstream of the stairs whereas the manual annotations often indicate them very close to the stairs. Nevertheless, it appears that these errors are quite limited (<2s when the imposed margin to find a True Positive is fixed at

3.5s). The largest Δ_{TE} value is obtained for Transition 1 (turns). This type of transitions is among the hardest transitions to detect as $\widehat{\Delta}$ values suggested it hence this important Δ_{TE} value.

4.4.3 Protocol 2

Usual and IoU Metrics

Averaged usual metrics and IOU metrics for Protocol 2 are displayed in Table 4.5. Both usual and IOU metrics are satisfactory (respectively 0.76 and above 0.72 for all thresholds) and indicate that this study's change point detection procedure is accurate for most change points, and also that there is no oversegmentation.

The segmentation is currently performed using spectrograms whose hop size are 0.1 second, thus limiting the temporal resolution that can be achieved. In case we would like to lower it and to possibly increase segmentation accuracy, it is possible to do so, but at the expense of the computation time.

Usual Metrics	$F1_{usual}$	0.76 ± 0.01
	$Precision_{Usual}$	0.74 ± 0.02
	$Recall_{Usual}$	0.79 ± 0.02
IoU Metrics : Overlap = 0.3	$F1_{IoU}$	0.83 ± 0.02
	$Precision_{IoU}$	0.88 ± 0.04
	$Recall_{IoU}$	0.82 ± 0.01
IoU Metrics : Overlap = 0.4	$F1_{IoU}$	0.80 ± 0.02
	$Precision_{IoU}$	0.85 ± 0.03
	$Recall_{IoU}$	0.79 ± 0.01
IoU Metrics : Overlap = 0.5	$F1_{IoU}$	0.73 ± 0.02
	$Precision_{IoU}$	0.78 ± 0.04
	$Recall_{IoU}$	0.72 ± 0.01
Other Metrics	Haussdorf	1934.44 ± 64.25
	Rand Index	0.96 ± 0.002
	AE	5.72 ± 0.95
	$AE_{\%}$	20.78 %
Additional Characteristic	Number of true regimes	27.62 ± 0.25

Table 4.5: Change point detection results for Protocol 2.

Change points' Metrics

Transitions	Type of change points	change points characteristics				Accuracy results	
		$\hat{\Delta}$	$\hat{\delta}$	Types of regimes	Regimes' length	$Recall_{usual}$	Δ_{TE} (s)
1	W2→W3 W4→W5 W5→W1	6.69	5.33	Movement/Movement	Long/Long or Medium/Long	0.33 ±0.11	2.01 ±0.12
2	W1→A1	7.62	7.44	Movement/Slow	Long/Short	0.90 ±0.06	0.82 ±0.05
3	A1→W2	8.13	7.30	Slow/Movement	Short/Long	0.94 ±0.05	0.92 ±0.03
4	W3→A2	11.34	10.13	Movement/Slow	Long/Medium	0.95 ±0.05	1.08 ±0.17
5	A2 → A3	12.69	11.96	Slow/Static	Medium/Medium	0.99 ±0.05	1.12 ±0.07
6	A2 → A4	12.36	12.25	Slow/Static	Medium/Medium	0.99 ±0.01	1.45 ±0.46
7	A2 → A5	13.31	12.12	Slow/Static	Medium/Medium	0.99 ±0.01	1.44 ±0.03
8	A3 → W4	17.60	17.00	Static/Movement	Medium/Long or Medium/Medium	0.99 ±0.01	0.73± 0.12
9	A4 → W4	17.82	16.36	Static/Movement	Medium/Long or Medium/Medium	0.99 ±0.01	1.30 ±0.15
10	A5 → W4	18.15	17.25	Static/Movement	Medium/Long or Medium/Medium	0.99 ±0.01	1.10 ±0.14

Table 4.6: Details of change point's categories and recall results for Protocol 2.

In Protocol 2, transitions between end-of-lap activity phases (A3,A4,A5) and walking phases as well as transitions between (A2) activity (climbing up and down stairs) and end-of-lap activities are especially well detected with recall values close to 1. Those values are high since these transitions separate static and movement phases or slow phases that present highly different spectral signatures. As a matter of fact, these figures indicate that improved annotations enabled by the use of a wearable camera led to better learning data and thus to a better detection of those transitions that occur at the end of laps. The poorest results are obtained for transition 1 dividing two motion phases. It relates to previous observations on low δ values which correlates with the lack of clear mean-shifts in spectrograms at turns' true change points. Intermediary recall results are obtained for transitions between walking phases and (A1) (opening a fire door and performing a 90° turn). These transitions separate motion and slow phases with significantly different spectral signatures but (A1) (opening a fire door and performing a 90° turn) regimes are short which may induce observed errors. In a nutshell, these results are correlated with observations made on $\widehat{\Delta}$ values in Subsection 4.3.3) and the difficulty of detecting these particular change points. Transitions with $\widehat{\Delta}$>10 are perfectly detected, transitions with 7<$\widehat{\Delta}$<10 can be less well detected and transitions with $\widehat{\Delta}$<7 and notably $\widehat{\delta}$<6 display the poorest recall values.

Δ_{TE} values are limited as for Protocol 1 (<2s). Some of the largest Δ_{TE} values (>1s) are obtained for transitions 9 and 10 (separating sitting and leaning activities from walking phases). This is due to the fact that pathological subjects included in training and val-

idation sets tended to be slower to end their activities and resume walking with a more graduate way to perform this transition compared to healthy subjects. It induced disparity in annotations of true change points : transitions were slower to be performed by pathological subjects compared to annotations. The same observation is made for transitions 5,6 and 7 separating stairs activity (A2) from sitting, leaning standing activities with high Δ_{TE} values (>1s) : included pathological participants were slow to initiate these activities hence the observed time margins in detection. Transitions 8 separating standing phases do not display high Δ_{TE} values since annotations could be performed more efficiently for true change points : if the subject completely resumed his gait, a change point is annotated. Highest Δ_{TE} values are obtained for transitions 1 between two motion phases and are related to detection issues for this specific type of transition as mentioned above.

4.4.4 Comparison between protocols

IOU and usual metrics

Both usual and IOU metrics are satisfactory and indicate that this study's change point detection procedure is accurate for most change points, and also that there is no oversegmentation. IOU metrics have been improved overall for Protocol 2. This is probably due to the changes in the end-of-lap activities to be performed. The annotated change points corresponding to transitions between sedentary activity phases are indeed close to each other in Protocol 1. If a segmentation error occurs for these change points (which happens for transitions between standing still (A3) and sitting still (A4) and between sitting still (A4) and leaning (A5) as displayed in Table 4.4), predicted regimes may be much longer than the annotated regimes and therefore induce lower IOU metric values. Changes made to the instructions to perform these end-of-lap activities therefore increase IOU metrics values in Protocol 2.

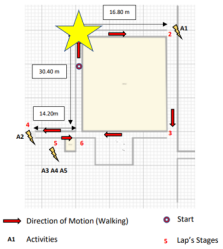
Haussdorf, Randindex, AE, $AE_{\%}$

In both protocols, the value of the Haussdorf metric is high (respectively 13.98s and 19.34) which is explained by the observation made in Section 4.2.4 : a single faulty detection of one change point on long regime portions (walking phases in the long corridors separated by turns in particular) could have induced this important deviation.

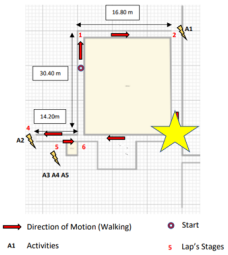
The percentage error annotation $AE_{\%}$ shows that the differences between the number of predicted and actual change points is around 20% for both protocols. This difference is probably due to the difficulty of detecting one type of break by our segmentation method (especially turns). In one acquisition, an average of 12 turns are made by a participant: if the segmentation method fails to detect half of these turns for instance and assuming that there is no over-segmentation, the difference measured by $AE_{\%}$ is reached.

As for the Rand Index, the value obtained is extremely high in both protocols (respectively 0.97 and 0.96) but this is due to the observations made in Section 4.2.4 which explain that this parameter is very sensitive to the length of our signals (here the number of samples is 36000).

((a)) Turns' type 1 : W0 to W1 turns



((c)) Turns' type 2 : W2 to W3 turns



((e)) Turns' type 3 : W4 to W5 turns

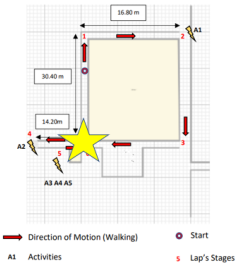


Figure 4.4: Screenshots of different types of performed turns in Protocol 2.Their locations on protocol's laps are also displayed.

Annotations and instructions

Annotations issues from Protocol 1 inducing high Δ_{TE} values for transitions between walking phases and A2 are corrected thanks to the use of the wearable camera and thanks to changes in protocols' structurations (Δ_{TE} values went from 1.60s in Protocol 1 to 1.13 s in Protocol 2). Annotations could be performed more accurately and lead to better segmentations' performances. Indeed, by providing precise annotations for transitions, algorithms are able to more efficiently perform change point detection and evaluation metrics (especially Δ_{TE} values) display better figures. Other Δ_{TE} values were not modified by protocols' changes which indicates that annotations were not highly influential on segmentations' algorithms for these specific transitions.

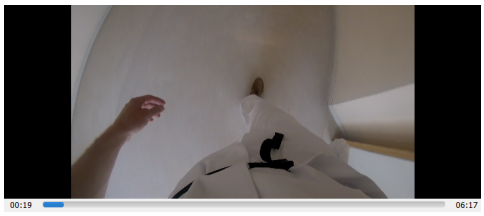
Besides, end-of-lap activities that were close one to another as end-of-lap activities (A3, A4 and A5) in Protocol 1 were better detected in Protocol 2. Indeed, changes in instructions (participants did not have to perform all these activities in a row in Protocol 2) induced better segmentation performances. Overlaps' issues in evaluation metrics between activities in Protocol 1were removed thanks to this new way of annotating. In the future, the minimum size enforced to the algorithm can be increased to reduce overlaps'

((a)) Start Turn Healthy

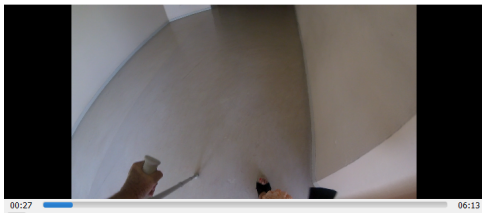
((b)) Start Turn Pathological

((c)) Mid Turn Healthy

((d)) Mid Turn Pathological



((e)) End Turn Healthy



((f)) End Turn Pathological

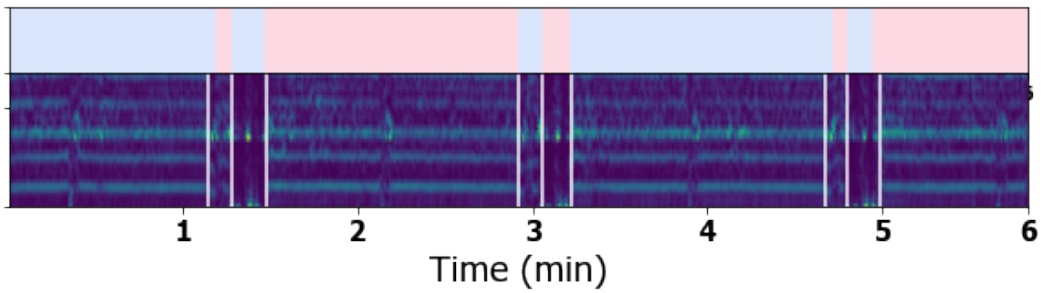
Figure 4.5: Screenshots of performed turns in Protocol 2 and of their timings. Healthy subject's turn is performed faster than pathological subject's (post-hip prothesis surgery) turn. Besides, the healthy subject performs a sharper turn by performing an important rotation mid-turn while the pathological subject performs a more "linear" turn.

influences.

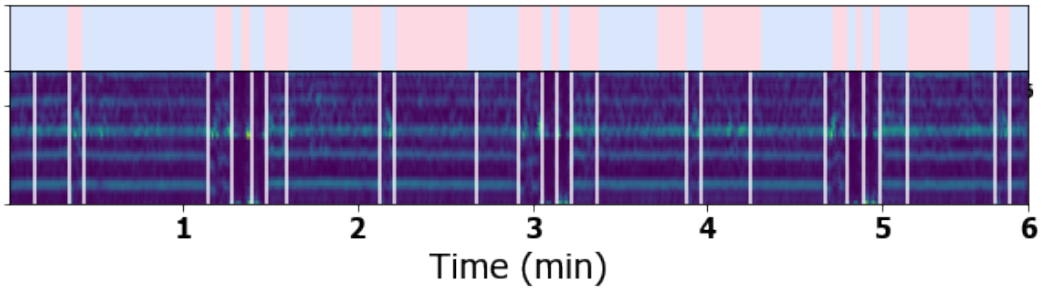
Turns detection

In both protocols, recall values for turn transitions (between two walking phases) are poor compared to other transitions. As detailed in Subsection 4.3.2, mean-shifts in spectrograms between those walking phases are not as evident as for other transitions. This detection difficulty was correctly predicted by the low δ (< 5.5) values displayed in Tables 4.1 and 4.2. Detections of turns occur at the end of annotated turns or at their beginnings : changes in turns impacts the means within signals of the walking regime predating or following the turn. Sometimes, when turns are long enough, change points can be found both at the beginning and at the end of turns.

This turn detection issue could be solved by using two segmentations relying on two annotations schemes that will be aggregated. Annotations for the first segmentation would be solely for starts and endings of turns with a reduced enforced minimum size while annotations for the second segmentation would be unchanged for all other transitions. As



((a)) Scenario 1 : F1=1.00



((b)) Scenario 2 : F1=0.85

Figure 4.6: Evolution of the change point's detection results depending on the annotations given as input for a control participant in Protocol 1. Displayed spectrograms are the concatenation of aAP and gCC spectrograms. White vertical lines correspond to the moments when the algorithm detects a change point. Annotations given as input are displayed above each spectrogram (alternating colors).

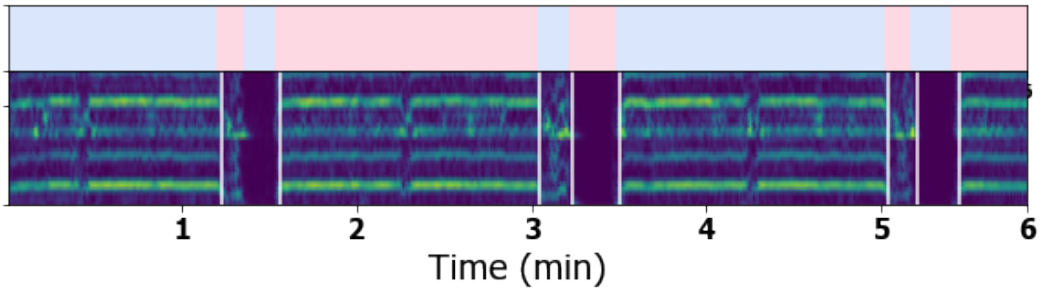
such, mean-shifts associated to turns could be more reliably detected.

Configurations of turns influence the way they are performed by the participants and therefore impact segmentations' results. As presented in Figure 4.4, turns' type 3 that separates W4 and W5 phases are structurally different from turns' types 1 and 2. Indeed, turns from type 3 occur around a 90° wall angle which leads participants to perform sharper turns while turns from type 1 and 2 occur around a 45° wall angle. Turns of type 3 were isolated and their detection was evaluated to assess the influence of these different configurations. In protocols 1 and 2, their averaged recall values are respectively 0.82 ± 0.03 and 0.54 ± 0.11 which are significantly higher than recall values computed for all types of turns (0.74 in Protocol 1and 0.32 in Protocol 2). It indicates that these various configurations indeed impacted change points' detection.

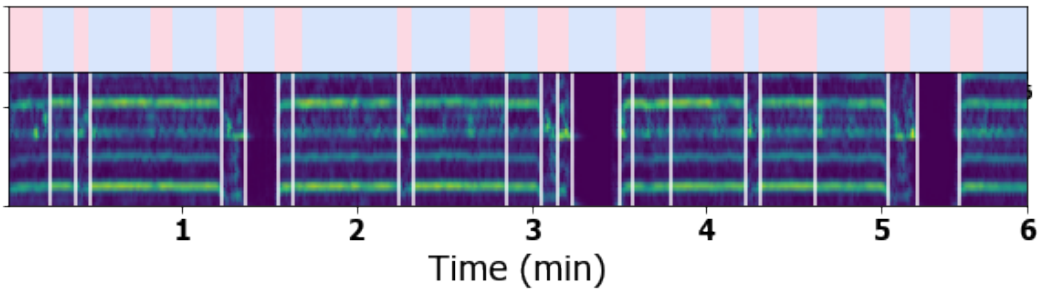
As displayed in Tables 4.4 and 4.6, adding signals from pathological subjects into cross-validation training folds induced higher confusion to detect turns in Protocol 2 than in Protocol 1which only includes healthy participants. In training folds for the cross validation from Protocol 2, pathological subjects were indeed included and mixed with healthy subjects with different ways of performing turns as displayed in Figure 4.5. Healthy subjects were indeed prone to perform sharper turns than pathological participants who turned more gradually due to their lower speed. Including those participants in training folds then induced poorer segmentation results for turn transitions. Indeed, signals at turns'

annotated change points differed and the calibration of β was thus altered. Besides, in protocol 1, participants that were followed by an examiner were more prone to perform sharp turns notably for turns' types 1 and 2 in order to accurately follow instructions. It led to more significant changes in signals retrieved at these turns' time samples and to an enhanced segmentation.

4.4.5 Influence of the Annotations



((a)) Scenario 1 : F1=0.94



((b)) Scenario 2 : F1=0.84

Figure 4.7: Evolution of the change point's detection results depending on the annotations given as input for a control participant in Protocol 2. Displayed spectrograms are the concatenation of aAP and gCC spectrograms. White vertical lines correspond to the moments when the algorithm detects a change point. Annotations given as input are displayed above each spectrogram (alternating colors).

To understand the influence of the annotations on the segmentation results, two scenarios are imagined where the method is fed with different labels in both protocols. These scenarios are as follows:

- Scenario 1: only label transitions with a $\widehat{\Delta}$ above 10;
- Scenario 2: label all transitions

In each protocol, our algorithm is fed with annotations that only displays the intended type of change points according to the scenario and the β value is calibrated on all included participants but one using the PELT approach as explained in Section 4.3.4. The remaining participant's signal is segmented using this β value. In Scenario 1, only the largest changes

(with respect to $\widehat{\Delta}$) are provided by the expert. In Scenario 2, all transitions are labelled by the expert. This corresponds to the setting of the previous section. Results of segmentations depending on these scenarios for one tested control participant are illustrated in Figure 4.6 for Protocol 1 and in Figure 4.7 for Protocol 2.

These figures clearly show that the algorithm adapts well to the different scenarios and to the experimenters' annotations . Indeed, each time the annotations are reinforced with a new type of label, changes are induced on the detection of change points and those changes fit to these input modifications (associated F1 scores in any tested protocol and scenarii is above 0.80). Moreover, it appears that this approach manages to adapt to the levels of granularity presented by the annotations structuring the scenarios. Thanks to this adaptability, the algorithm can reproduce the annotation strategy.

4.5 Conclusion

This method relying on an adapted learning of a penalty parameter through an optimization process enables a precise change points' detection in acquisitions performed in semi-FLEs. By optimizing β which defines a penalty feature, new acquisitions can be segmented to get partitions that are as close as possible to the strategy of annotation without knowing the number of change points to detect. This adaptive approach is based on a structured functioning that first enforces the detection of large change points before detecting smaller change points. This approach is a preamble to an innovative analysis of movements in FLEs and semi-FLEs. Homogeneous regimes are identified and can now be characterized as detailed in Chapter 5.

5 Classification Method

5.1 Context

The supervised change point detection method described in Chapter 4 allows to extract homogeneous regimes from raw data. Once these homogeneous regimes are delimited, they can be characterized and this characterization enables the evaluation of these regimes. Our first goal is therefore to define whether these regimes are walking regimes or regimes belonging to another activity. This task of defining walking regimes is known in the literature as Walking Bout (WB) detection. A first review of existing methods to detect Walking Bouts (WBs) with wearable sensors is performed here. Regimes belonging to other activities will also be characterized (sedentary or non-sedentary regimes). The results of this classification procedure are presented and discussed.

5.2 literature review of classification methods for walking bouts' detection

5.2.1 Rule-based, steps detection and wavelets decomposition methods

There are several approaches to perform WB detection. The first type of methods extracts some quantities of interest and rely on threshold-based detection methods : these are rule-based methods [124]. More specifically, recordings are analyzed frame-by-frame using sliding windows. Then, a quantity of interest is computed on each frame, such as the signal magnitude area performed on the normalized squared average of the 3 accelerations axes [161], the craniocaudal accelerations [157] or the magnitude of angular velocities. Finally, the WB detection is performed through thresholding : if a portion of the recordings exceeds a given threshold this portion is considered as belonging to a walking phase. Thresholds can be set arbitrarily or it can be adaptive relying on methods that update the value of the threshold depending on average values from linear accelerations and angular velocities [155]. These rule-based methods are widely used in the literature thanks to their intuitiveness and their simplicity.

The second type of methods detects successive steps and groups them to define WBs [127]. Steps can be detected by rule-based methods searching for thresholds in craniocaudal acceleration signals for instance. An other way to detect steps and associated WBs is to compare detected patterns with pre-existing templates via Dynamic Time Warping [18] or by performing cross-correlation [62]. WBs can also be detected by using the in-

verted pendulum model-based algorithm developed by [166] or by zero-crossing methods as detailed in [70].

Other methods use wavelet transform to perform automatic WB detection [64]. These are time-frequency domain methods that decompose signals by using basis functions which are wavelets (in contrary of sinusoids for Fourier Transform). Wavelets differ in both frequency and time domains while sinusoids are strictly frequency domain functions. Wavelet transforms rely therefore on windows whose sizes vary in time-frequency domain : time resolution decreases with low-frequency signals and inversely. Threshold-based detection is performed on discrete wavelet transforms for instance by relying on the ratio between the power of detail coefficients in specific chosen levels over the total power of detail coefficients [16].

5.2.2 Human Activity Recognition Methods

Another set of methods are HAR methods that rely on machine learning methods to classify signals' portions and to notably detect walking regimes.

Introduction to Human Activity Recognition methods

The HAR methods rely on features extracted from annotated signal portions in order to train classifiers that can associate new signal portions with specific output classes. Such classifications are performed for instance to assess the physical activity of specific patients with diseases in a detailed manner [112] (by evaluating the amount of time spent by a participant performing a specific activity associated to a high energy expenditure value for instance). It can also be used to compare efficiently laboratory acquisitions and FLEs' acquisitions [27] and to compare the impact of the placement or of the types of wearable sensors [154, 57]. It can validate the feasibility of a specific sensor to assess PA in free-living settings [104]. HAR methods are also valuable since they enable precise walking bouts' detection that can be characterized afterwards.

The aim of this subsection is to focus on HAR studies in free environments and to provide an overview of the current state-of-the-art performances. Table 5.1 displays the main features retained in specific subcategories (described in 5.2.2) while Table 5.2 summarizes characteristics of these studies. In the following subsections, features that are computed from raw signals will be detailed and insights on the classifiers used for activities' classification will be provided. Figure 5.1 summarises the different stages that HAR methods implement to classify portions of raw data.

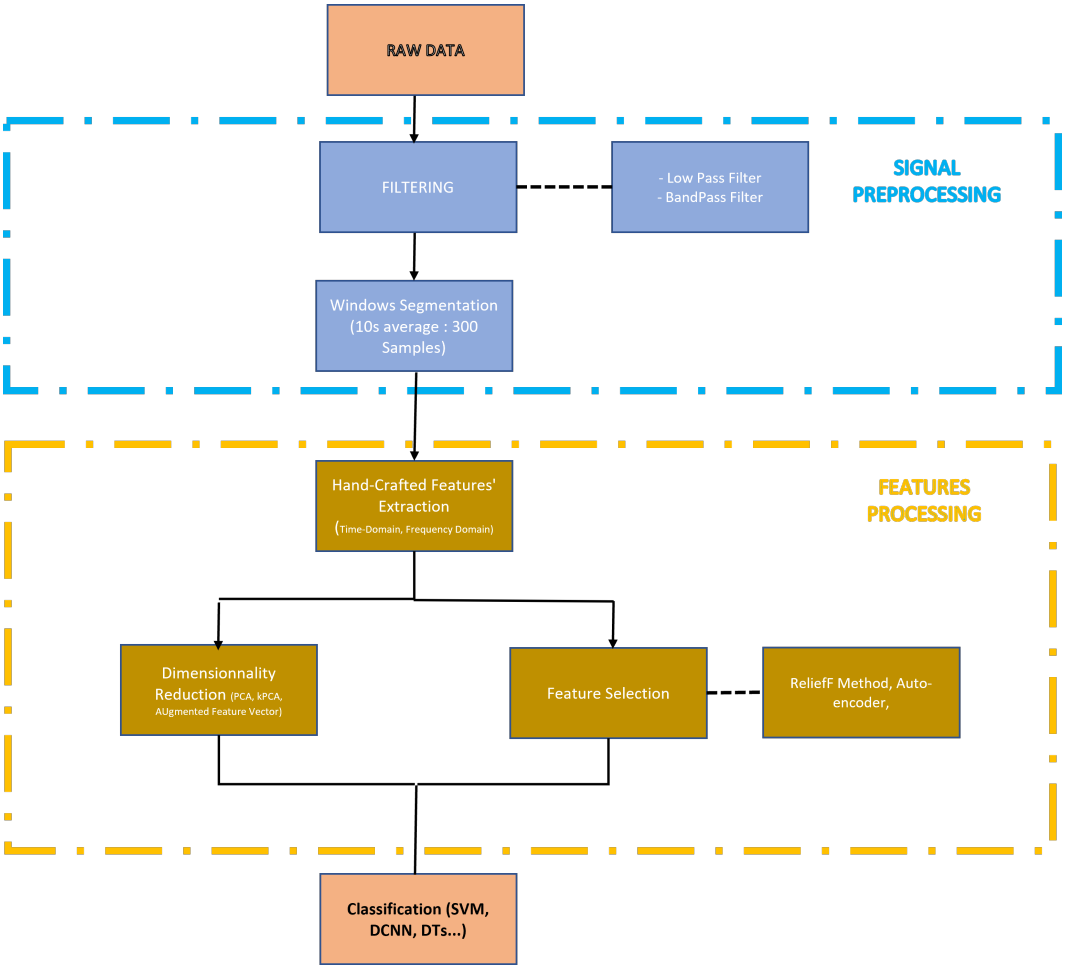


Figure 5.1: Description of axis on which raw signals are recorded from the lower back

Features

When portable sensors are attached to participants during their physical activity in free-living conditions, research teams recover raw signals whose nature differs according to the type of sensors implanted (linear accelerations for accelerometers, angular velocities for gyroscopes or both for IMUs for example). From these raw data, it is possible to compute numerous parameters that allow for a quantitative PA's analysis. These parameters can be basic statistics (mean or standard deviation of accelerations over a certain duration for example) or derived from advanced algorithms (step lengths, average swing time...). In the clinical field, it has been shown that those features can sometimes be used to define the quality of a walking category (rhythm, stability, springiness..) of a participant [157].

Those features are also extracted and used to train classifiers allowing to perform HAR. Most simple features are time domain features, that are only based on the timings of relevant gait events for instance. Time domain features are the parameters related to a notion of evolution in time of certain particularities of the obtained signal (standard deviation, means...). More sophisticated features can be computed in the frequency domain, by focusing for example on some relevant frequency bands: those are the frequency domain features. Frequency domain analysis dwells upon the number of times some events occur

in the recorded signals and to the notion of periodicity. From Table 5.2, it appears that the majority of the features retained for the classification of activities are time domain features (31 studies use time domain features [162, 44, 122, 105] while only 19 studies use frequency domain features [57, 27, 4, 120]). Table 5.1 details the most used parameters for each of these feature categories (time domain features, frequency domain features, derived parameters) retained to set up the activity classifiers.

In addition, some specific parameters are used by several studies (Vector Magnitude (VM) [38, 105], Signal Magnitude Area (SMA)[10, 85] ...). It is also to be noted that the sizes of the sample windows on which signals are analyzed are various (from 1s [104] to 5 mins [162]) but are often non-overlapping (only one study uses overlapping windows [51]). Using sliding windows to perform features' extraction can be restricting for computational costs when working in free environments at high frequencies (around 100Hz) since it can induce long retrieved signals.

time domain Features	frequency domain Features	Derived Parameters
Mean Coefficient of Variation Standard Deviation Variance Min, Max Median Autocorrelation Coefficients 25th and 75th percentile	Dominant Frequency Dominant Frequency Magnitude Spectral Power Spectral Energy	Speed Step/Stride Time Step/Stride Velocity

Table 5.1: Details of the features mostly selected to feed activity classifiers according to their associated categories (time domain Features, frequency domain Features, derived parameters)

Algorithms

When these parameters are calculated, they can be used as variables for each observed activity in order to train classifiers. Redundant features or irrelevant features can be discarded to perform the classification through a feature selection step. Almost half of the studies classifying activities (18 of them) used feature selection or dimensionality reduction methods such as Principal Component Analysis (PCA) [44, 4] or others [146, 67, 109]) before setting up their classification system. The vast majority of these studies obtained classification accuracies (accuracy or sensitivity) above 85%. All algorithms referenced in Table 5.2 are supervised learning systems. This corresponds to a learning where the inputs, i.e. features, and the outputs, i.e the associated types of activity, are known when data is processed. Unsupervised learning is much less predominant than supervised learning [3] for HAR methods. The fact that the labelling of activities can be carried out through annotations whose implementation is increasingly becoming more practical over the years (better quality cameras, use of trajectory reconstruction, etc.) may explain this difference.

Depending on studies, classifications are carried out on a different number of activities carried out under conditions that sometimes differ drastically from one study to another (instructions, sensors ...). All these factors, added to others (positioning of sensors, cohorts, etc.) do not allow entirely valid comparisons to be made between studies with regard to their classifiers and their associated algorithms.

It appears that the majority of the studies have an interesting average rate of classification precision (31 analyses obtain a precision or sensitivity higher than 85 percent [100, 68, 27, 120]). It should also be noted that certain types of classifiers are recurrent, notably Random Forest (RF) (12 studies [146, 76, 83]) and Decision Trees (DT) (8 studies [104, 32]), SVM (13 studies [164, 68, 84, 63]), Bayesian approaches (Naive Bayes (NB)) (3 studies [42, 164, 68]), K-Nearest Neighbor (KNN) (5 studies [146, 67, 68, 84]) and Neural Network (NN) (6 studies [85, 63, 164, 38] with notably Multi-Layer Perceptron (MLP)). Some teams decided to compare or test several types of already known classifiers such as those cited above [51], while other teams combined some of these classifiers with other algorithms specific to their work in order to set up custom classification systems [100, 104].

Table 5.2: Summary of all the specificities of the studies selected for this study which classify performed activities. The features retained as well as the nature of the classifiers and their precision are detailed (acc : accuracy, se : sensitivity).

				Features								Algos	
Date	Ref	Activities Nb	Time Res	Features Selection / Reduction	T-D Features	F-D Features	VM	SMA	Angle ROM	Corr	MAD	Classifier(s) Type	Accuracy
2010	[85]	15	3.2s		√	√		√	√			NN	Acc=97.9%
2012	[44]	2	4s	PCA	√	√						Other	Acc=90% → 95%
	[146]	NC	NC	Performed								K-M, KNN, SVM, RDF SV	Acc=71% → 96%
	[164]	15	NC			√						DT, SVM, NB, LR, NN	Semax = 97.20% avg
2013	[101]	NC	NC									Wavelet	Se=96.78%
	[100]	13	NC		√	√						KNN/SVM	Acc=89.6% avg
	[122]	3	5s		√				√			Other	Se =77% → 94%
2014	[68]	8	1s	Performed	√	√		√	√	√		SVM,NB,kNN	89.5% → Acc=96.8%
	[51]	8	1.25s									DT	Se =69.2% → Acc=82%
2015	[38]	13	NC	FC	√	√	√					MLP,SVM, RBF, NN	Acc=97.20% avg
	[105]	8	NC									Other	Acc=90.40%
	[119]	NC	NC	NC								Other + ADC	

Table 5.2 –*Continued from previous page*

Date	Ref	Activities Nb	Time Res	Features Selection / Reduction	T-D Features	F-D Features	VM	SMA	Angle ROM	Corr	MAD	Classifier(s) Type	Accuracy
2016	[12]	8	160ms	FC	√				√	√		ADC	Se = 97% avg
	[76]	4 Groups	30s	RF	√							RF (100 DT)	Acc=99.71% avg, Se= 84.62% → 99.9%
	[27]	1	1.2s	Performed	√	√						DT	Acc>97%
	[57]	4	1min	RF	√	√	√		√	√		RF,HMM	Acc=84.6% → 89.4%
	[83]	5	1min	PCA	√		√					RF	Acc=51% → 77%
	[4]	13	NC		√	√						RF	Acc=91.80%
	[162]	3	5 mins		√	√						LR	Acc=90.2% → 94.3%
2017	[120]	4 Groups	10s	Performed	√	√						RF	Acc=92.70%
	[112]	7	NC		√				√			Other	Acc=100% avg Se= 97.6%
	[67]	8	3s	Performed	√	√	√	√				DT,SVM,kNN	Acc>95%
	[42]	5	5s	Correlation Based						√		NB	Acc=65% → 79%
2018	[113]	7	NC		√				√			Other	Acc=90% avg, Se=90.8%
	[154]	(FLE) 4	6s		√				√		√	ROC Curve	Other

Table 5.2 –*Continued from previous page*

Date	Ref	Activities Nb	Time Res	Features Selection / Reduction	T-D Features	F-D Features	VM	SMA	Angle ROM	Corr	MAD	Classifier(s) Type	Accuracy
	[154] (Semi-FLE)	4	6s		√				√		√	ROC Curve	Other
	[52]	51	1min		√							LR	Se=80%
	[110] (FLE)	12	NC			√			√			SVM	Acc=99%
	[84]	6	5s		√		√					SVM, NB, DT, RF, HMM, kNN	Acc>80% (RF)
	[63]	8	3.5s		√						√	DT, SVM, NN	Acc>92%
	[32]	8	NC		NC	NC						DT	Acc=93.50%
	[2]	4 Groups + Others	10s	RF	√	√						RF, SVM, BDT	Acc=76.1% → 89%
2019	[10]	1	1s to 10s	Three Methods	√	√	√	√				SVM	Acc>80%
	[109]	12	5s	Performed	√	√	√		√	√		RF	Acc=53% → 99%
	[46]	8	2s		√							Custom	Se=63% → 100%
2020	[104]	1	1s		√		√					RF,DT	Se=60.4% → 93.5%
	[69]	4	1s	Performed	√							SVM	Acc=67.22% avg

Table 5.2 – *Continued from previous page*

Date	Ref	Activities Nb	Time Res	Features Selection / Reduction	T-D Features	F-D Features	VM	SMA	Angle ROM	Corr	MAD	Classifier(s) Type	Accuracy
	[165]	10	5s		√	√						RF, MLP, SVM, KNN,ET	Acc=78.6% max
	[47]	5	20s	CNN Extractor, PCA	√	√						CNN	Acc=67.51%
2021	[11]	4	5s	CBFS, FCBF, ReliefF	√	√	√	√				LSTM	Acc=97%

5.3 Proposed Method

We will now present the classification method we will be using : a method derived from HAR studies. However, contrary to state-of-the-art methods used for HAR, instead of using (possibly overlapping) frames, it is proposed to perform this classification at the regimes level. The advantages are twofold : first, because of the segmentation procedure : each regime is stationary, which is a valuable theoretical property for computing robust features. Second, the average length of the regimes is often longer than typical frame durations, which provides more data for computing features.

Before applying classifiers to our gathered data, the latter are filtered and features are extracted from annotated regimes as presented in Subsection 5.3.1. Then, these features are standardized to apply PCA as presented in Subsection 5.3.2.

5.3.1 Features extraction

First, for each annotated regime an extensive list of both temporal and frequency features (displayed in Table 5.3) such as variances, means, dominant frequencies, power at dominant frequencies is extracted : features are computed on pre-filtered signals as presented in Section 3.6. Features used to train classifiers have been selected in accordance with the performed state-of-the-art dedicated to activity classification from IMU signals. These 135 features are retained because their computation is convenient for long FLEs signals as they do not require any detection of events (heel strikes, toe strikes...). The features list is presented in 5.1. Formulas to compute all features are also explained. In the classification process, regimes are defined as observations delimited by expert's annotations and features correspond to variables.

5.3.2 Standard scaling and Features Dimensionality Reduction

Beforehand, standardization on data features' vectors is performed to apply PCA. Indeed since PCA is sensitive to variables' variances, standardization enables to efficiently compare features' vectors which can be originally expressed in various units and in various orders of magnitude.

PCA is then applied to transform the high-dimensional data into lower dimensions while keeping as much information as possible. Principal components are linear combinations of the initial variables values in training data and are computed so that they are uncorrelated and that most information within initial variables values are compressed within the first components. In this study with 135 variables, 135 components are rendered. Each component causes a specific amount of variance : the explained variance. Thus, each component is associated to a specific value of explained variance ratio : the ratio of its induced explained variance to the total value of explained variance induced by all components.

In this study, where the classification is intended to be as accurate as possible, the principal components causing 99% of the cumulative explained variance (an explained variance ratio of 0.99) are retained for the PCA. Figures 5.2 and 5.3 illustrate the value of this cumulative explained variance for the set of data used to train our classifiers in both Protocol 1 and Protocol 2. Data sets are composed of all segmented regimes from the associated protocol as observations which are labeled as walking or non walking regimes. According

Table 5.3: Details of features used for classification. For each formula, $X = [x_1, x_2..., x_n]$ is assumed to be one of the 6 dimensions (3 linear accelerations and 3 angular velocities) from the IMU signal. In total, 135 features are used. Notations: $\overline{x}$ is the empirical mean of X, $\hat{\sigma}$ is the empirical unbiased standard deviation of X, FFT(X) is the Fourier transform of X, ConjFFT(X) denotes the complex conjugate of FFT(X).

Features	**Description**	**Domain**	**Formulas**
mean_signal	Mean	Time	$\overline{x}$
std_signal	Standard Deviation	Time	$\hat{\sigma}$
var_signal	Variance	Time	$\hat{\sigma}^2$
min_signal	Minimum	Time	$\min(x)$
max_signal	Maximum	Time	$\max(x)$
PD0_signal	Power at the first dominant frequency	Frequency	$max(\frac{FFT(X)ConjFFT(X)}{N})$
F0_signal	First Dominant frequency	Frequency	$argmax(\frac{FFT(X)ConjFFT(X)}{N})$
PD2_signal	Power at the second dominant frequency	Frequency	$max_{2nd}(\frac{FFT(X)ConjFFT(X)}{N})$
F2_signal	Second dominant frequency	Frequency	$argmax_{2nd}(\frac{FFT(X)ConjFFT(X)}{N})$
Ent	Spectral Entropy	Frequency	Let $P(\omega_i)$ be coefficients of PSD : $Ent = -\sum_{i=1}^{n} p_i ln(p_i)$ with $p_i = \frac{P(\omega_i)}{\sum_i P(\omega_i)}$
CV_signal	Coefficient of variation	Time	$\hat{\sigma}/\overline{x}$
p75_signal	75th percentile (on absolute values)	Time	Let R be the 75th percentile rank : $R = \frac{75*n}{100}$, p75 corresponds to the Rth value on the sorted X array
p25_signal	25th percentile (on absolute values)	Time	Let R be the 25th percentile rank : $R = \frac{25*n}{100}$, p25 corresponds to the Rth value on the sorted X array
p85_signal	85th percentile (on absolute values)	Time	Let R be the 85th percentile rank : $R = \frac{85*n}{100}$, p85 corresponds to the Rth value on the sorted X array
p15_signal	15th percentile (on absolute values)	Time	Let R be the 15th percentile rank : $R = \frac{15*n}{100}$, p15 corresponds to the Rth value on the sorted X array
p95_signal	95th percentile (on absolute values)	Time	Let R be the 95th percentile rank : $R = \frac{95*n}{100}$, p95 corresponds to the Rth value on the sorted X array
p5_signal	5th percentile (on absolute values)	Time	Let R be the 5th percentile rank : $R = \frac{5*n}{100}$, p5 corresponds to the Rth value on the sorted X array
p75m_signal	75th percentile at the middle of the signal (2/3 of the signal) (on absolute values)	Time	Let R be the 75th percentile rank : $R = \frac{75*n}{100}$, p75 corresponds to the Rth value on the sorted X array
p25m_signal	25th percentile at the middle of the signal (2/3 of the signal) (on absolute values)	Time	Let R be the 25th percentile rank : $R = \frac{25*n}{100}$, p25 corresponds to the Rth value on the sorted X array
p85m_signal	85th percentile at the middle of the signal (2/3 of the signal) (on absolute values)	Time	Let R be the 85th percentile rank : $R = \frac{85*n}{100}$, p85 corresponds to the Rth value on the sorted X array
p15m_signal	15th percentile at the middle of the signal (2/3 of the signal) (on absolute values)	Time	Let R be the 15th percentile rank : $R = \frac{15*n}{100}$, p15 corresponds to the Rth value on the sorted X array
RMSR_signal	Root Mean Square Ratios	Time	$RMS_{aAP} = \sqrt{\frac{1}{n}(\sum_{i=1}^{n} xAP_i^2)}$ and $RMSR_{aAP} = \frac{RMS_{aAP}}{\sqrt{RMS_{aAP}^2 + RMS_{aCC}^2 + RMS_{aML}^2}}$
$P1_{CC}$	First peak of autocorrelation coefficients for CranioCaudal Acceleration	Time	$ACF = iFFT[FFT(X)ConjFFT(X)]$, P1 is the first peak of ACF
$P2_{CC}$	Second peak of autocorrelation coefficients for CranioCaudal Acceleration	Time	$ACF = iFFT[FFT(X)ConjFFT(X)]$, P2 is the second peak of ACF
VM	Vector magnitude of all accelerations (CranioCaudal aCC, MedioLateral aML and Anteroposterior aAP)	Time	$VM = \sqrt{aCC^2 + aAP^2 + aML^2}$

to the fixed threshold and to this figure, it is possible to define the number of components to be used in the study : 60.

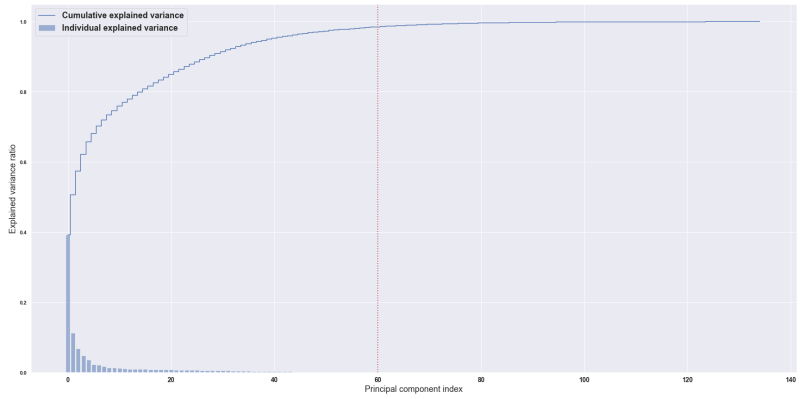


Figure 5.2: Evolution of the explained variance ratio in Protocol 1 according to the individual principal components' associated explained variance on the training data set

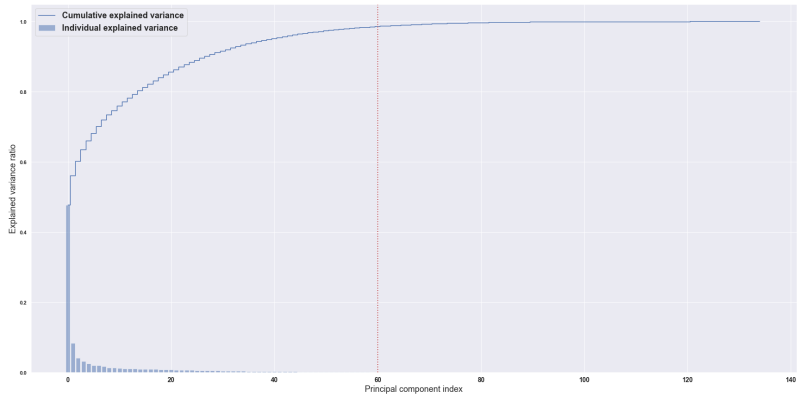


Figure 5.3: Evolution of the explained variance ratio in Protocol 2 according to the individual principal components' associated explained variance on the training data set

5.3.3 Proposed pipeline method

In our pipeline two tasks are performed and used in cascade. Details of this pipeline are presented in Figure 5.4. In task 1, we perform a walking/non-walking binary classification, while only non-walking phases detected in the first task are considered in Task 2. In this task, we classify non walking regimes as sedentary or non sedentary regimes. Performing these two tasks successively was motivated by the imbalance in observations (one observation is one segmented regime) classes used for the training sets in both protocols. Indeed 77% of the total regimes correspond to walking regimes, 12% are associated with non-sedentary regimes and 11% to sedentary regimes in Protocol 1. In Protocol 2, 79% of the total regimes correspond to walking regimes, 11% are associated with non-sedentary regimes and 10% to sedentary regimes Thus, by performing Task 1 to classify walking and non walking regimes, regimes' classes are automatically rebalanced. Regimes' classes

	Number of annotated regimes (observations)	
Data sets used	**Task 1**	**Task 2**
Protocol 1	489	242
Protocol 2	742	254

Table 5.4: Details for observations number (annotated regimes' number) in data sets used to train classifiers in both protocols

used to perform Task 2 are already properly balanced (as many sedentary regimes as non sedentary regimes). In both protocols, Task 1 and Task 2 are performed by using the same kind of classifiers (two linear SVM, two Gaussian kernel SVM...) : this is a cascade classifier. Details for observations number (annotated regimes' number) in data sets used to train classifiers in both protocols are displayed in Table 5.4.

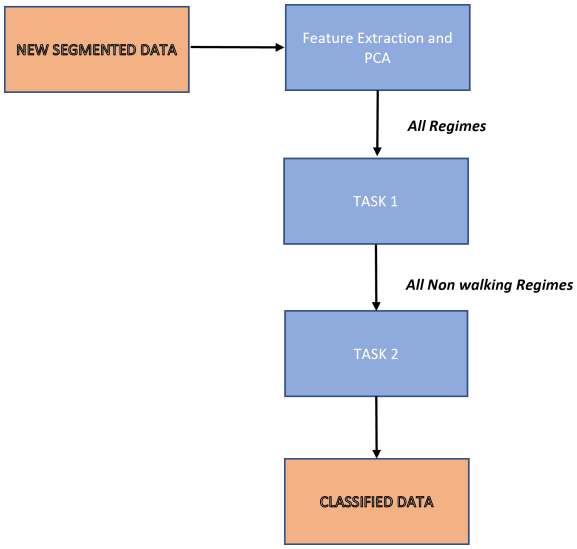


Figure 5.4: Pipeline method used to classify new segmented data

5.4 Results

5.4.1 Introduction

Performances of several classifiers among the most used in the state of the art as displayed in Table 5.2 are compared when applied to our signals. Retained classifiers for evaluation are :

- Linear kernel Support Vector Machine SVM : a generalized linear classifier that relies on the hypermarginal hyperplane to classify data.
- 3rd degree polynomial kernel SVM : this kernel function can separate non linearly separable data by mapping the into higher-dimensional space using a polynomial function

- Gaussian kernel SVM : this kernel function can separate non linearly separable data by mapping the input vector to Hilbert space.

- k Nearest Neighbours kNN : find the nearest neighbour of one input data point to label it. The number of nearest neighbours to use is fixed at 3.

- Decision Tree DT : such a classifier relies on an algorithmic flow chart to perform decisions according to specific criteria values and to label data points

- Random Forest RF : an ensemble learning method for classification using 100 decision trees.

5.4.2 Evaluation metrics

In both protocols, we evaluate classifiers' performances by performing a 3-fold cross-validation with the same participants' distribution as for the cross-validation used in Chapter 4. Data sets are split into three sub data sets which correspond to training and testing folds (all sets are built with annotated regimes as observations).

Receiver operating characteristic Curve

Classifiers are trained on two data folds and applied to the third data fold with annotated regimes whose targeted labels are known. The Receiver Operation characteristic ROC curve can therefore be plotted for each classifier and for each tested fold. The higher the average Area Under Curve AUC of these curves is for a classifier on the 3 testing data sets, the more successful this classifier is in classifying the regimes correctly according to the labels used. ROC curve plots the True Positive Rate TPR against the False Positive Rate FPR defined in (5.1). This evaluation method is presented in Figure 5.5.

$$\begin{aligned} TPR(Recall) &= \frac{TP}{TP+FN} \\ Specificity &= \frac{TN}{TN+FP} \\ FPR &= 1 - Specificity \end{aligned} \tag{5.1}$$

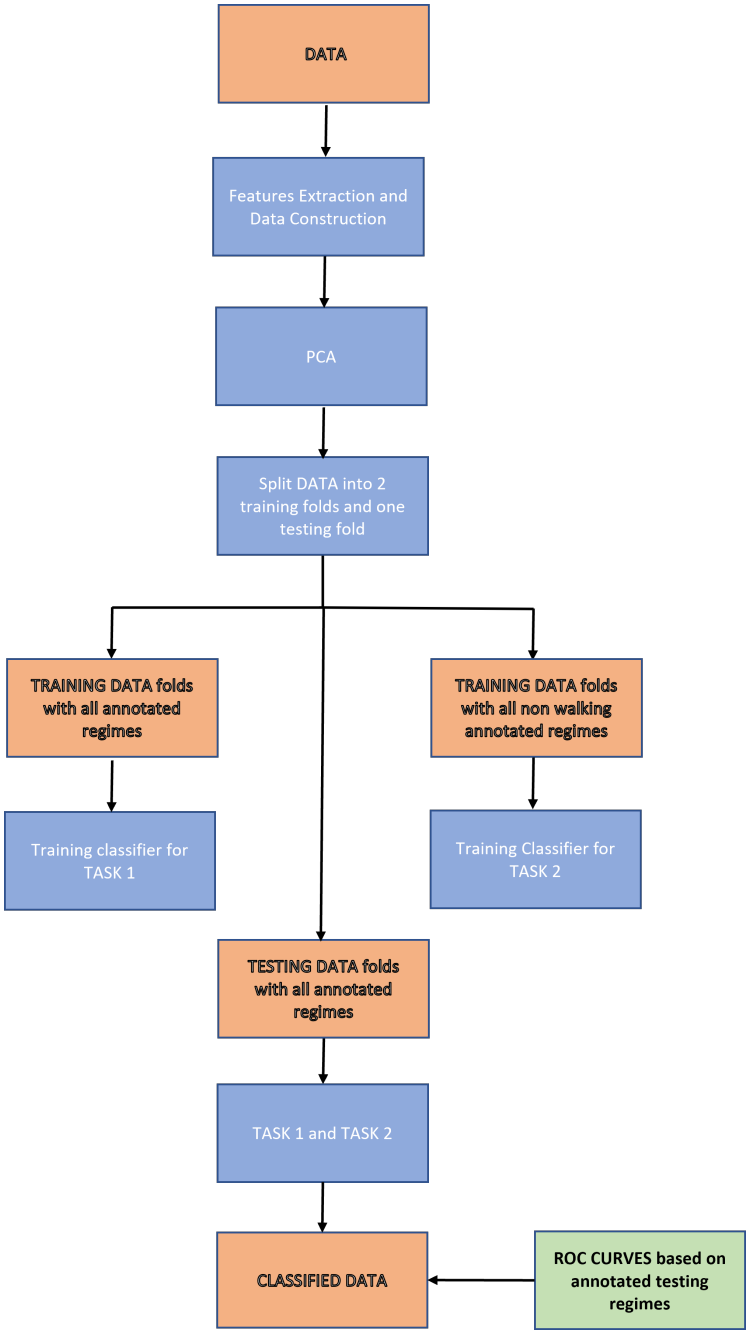


Figure 5.5: Evaluation method to render ROC Curves

Joint Evaluation of Segmentation and Classification Steps

Intuitively, the regime-scale classification performances depend both on the segmentation step and on the regime classification step hence the need to jointly evaluate these two methods. The performances of the cascade classifiers from both tasks are tested when applied to regimes predicted by our segmentation method. The optimal β values learned in Chapter 4 for each testing fold are used again here to segment participants' signals. Once the regimes are segmented, the classifiers trained as explained above on the training

data sets with annotated regimes are applied to these segmented regimes.

After the segmentation and classification tasks, each data sample is labelled as Walking, Non-walking/Non-sedentary or Sedentary for all participants : all signals from the testing fold are segmented and all associated regimes are classified. More precisely, all individual samples from these regimes are associated to one of the three types of samples so that they can be compared to "real" individual annotated samples.

To jointly evaluate classification and segmentation, we compute the confusion matrix between all 3 labels. Each coefficient of the matrix represents the percentages of samples annotated as belonging to the row activity, that have been classified as the column activity. Perfect performances would correspond to a diagonal matrix. Classifiers' accuracy are also computed. The higher accuracy's figures are, the more efficient are the associated classifiers. This evaluation method is presented in Figure 5.6.

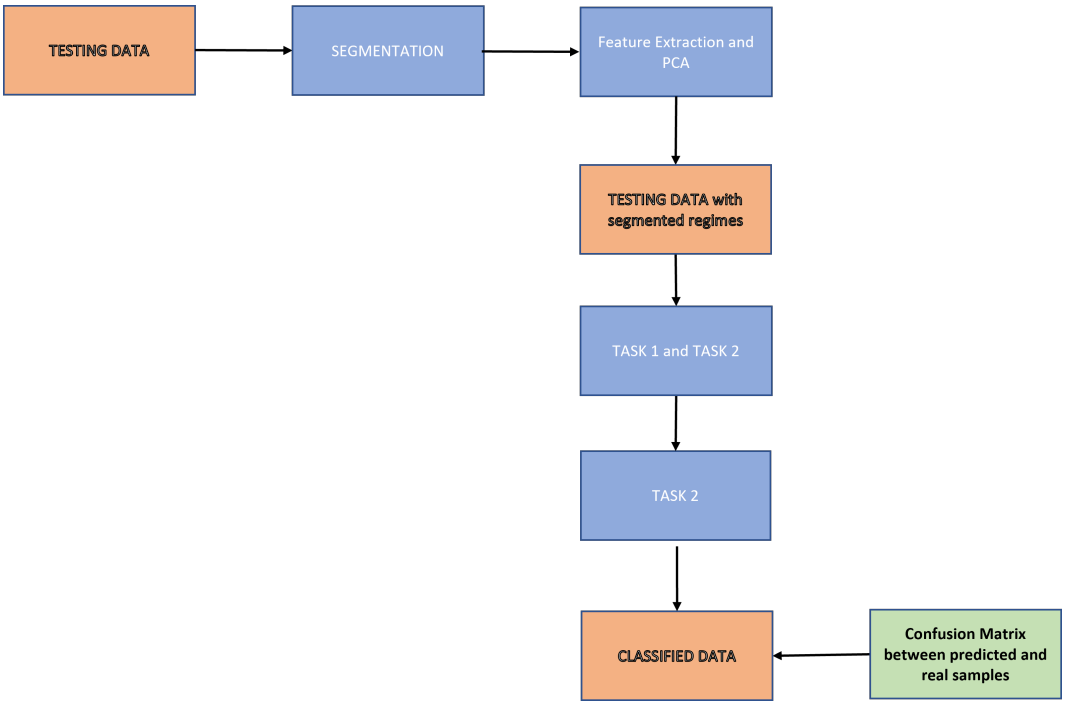
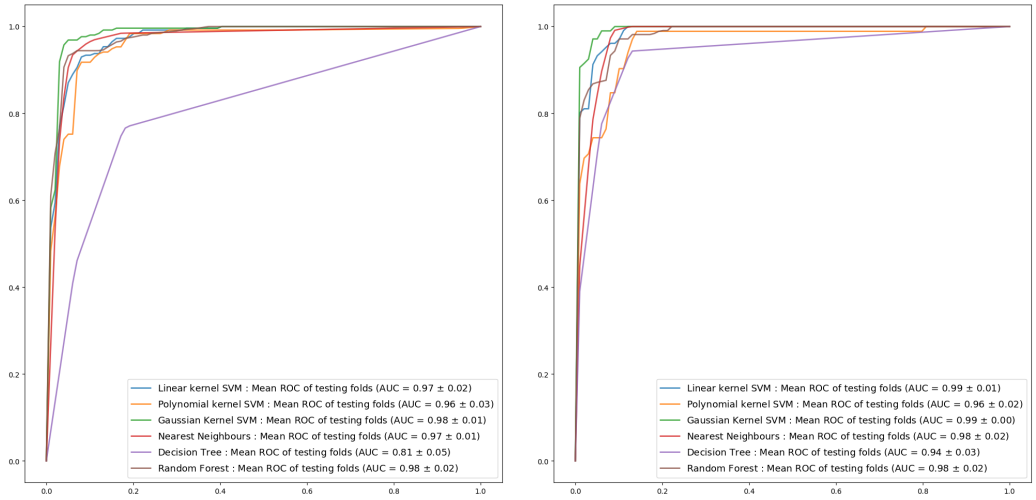


Figure 5.6: Evaluation method to assess both segmentation and classification

5.4.3 Protocol 1

ROC Curve and AUC

ROC Curves for each performed task (Task 1 and Task 2) and for each tested fold are plotted in Figure 5.7. AUC values are satisfactory for all types of classifier : the lowest values are found when using the Decision Tree classifers (notably for the first task : AUC=0.81).



((a)) Task 1 ((b)) Task 2

Figure 5.7: ROC curves plotted for each classifier applied to all testing folds in Protocol 1

Joint Evaluation of Segmentation and Classification Steps

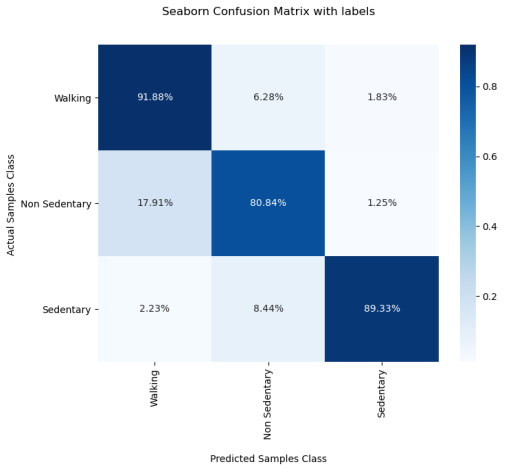
Table 5.5 displays accuracy's values for each tested cascade classifier. Accuracy results are satisfactory for all types of classifiers (>0.80). Linear SVM and kNN classifiers are particularly accurate (accuracy >0.90). Accuracy results are difficult to compare considering the imbalance between actual samples in different classes: there are many more samples of walking than samples of sedentary or non-sedentary activity for example.

Type of Classifiers	Linear Kernel SVM	Polynomial Kernel SVM	Gaussian Kernel SVM	kNN	Decision Tree	Random Forest
Accuracy	0.90 ± 0.06	0.85 ± 0.17	0.86 ± 0.09	0.91 ± 0.06	0.78 ± 0.11	0.84 ± 0.14

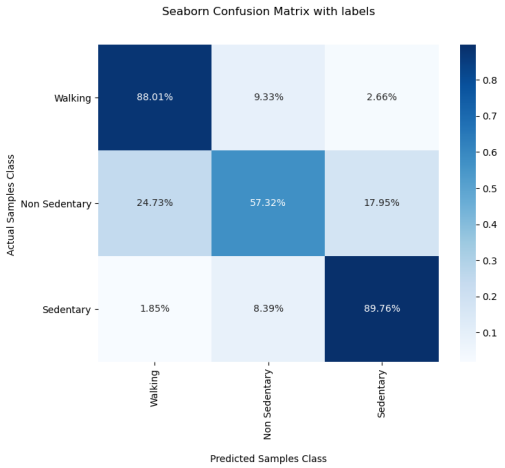
Table 5.5: Accuracy values for each tested cascade classifier in Protocol 1

Confusion matrices for each type of classifier are shown in Figure 5.8. In all confusion matrices, walking phases are well predicted (>79%). Sedentary phases are also well discriminated for all classifiers (>89%). Non-sedentary phases are well discriminated for almost all classifiers (>75%). Only Decision Tree (60.03%) and Polynomial SVM classifiers (57.32%) display poorer results. Imprecision margins are consequences of errors in the segmentation. In addition to segmentation error, inaccuracy can be introduced by the cascade classifiers. Walking regimes are well detected because this activity is structured and made up of repetitive and precise patterns that therefore manifest themselves with intense spectral signatures. Since a significant proportion of the features used for classification are spectral features, this probably facilitates the classification process. Sedentary and non-sedentary activities are inherently more difficult to differentiate. For example, activities where subjects open fire doors (A1) include movements similar to those observed during walking (stomping and some slow steps) that have spectral signatures closer to those of walking activities. As a result, non-sedentary activities are often mistaken for walking

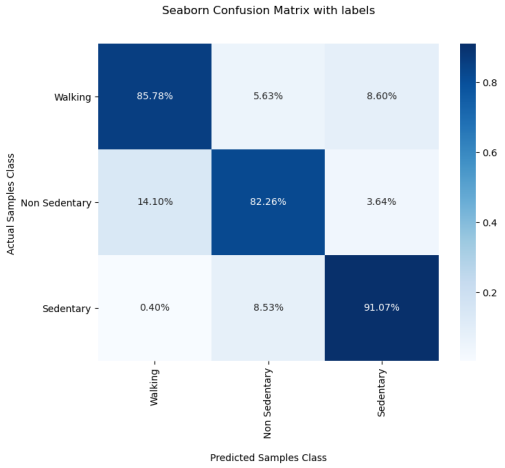
regimes as displayed in all presented confusion matrices and especially in the one associated to Decision Tree classifiers (27.65 % of non sedentary real samples are misclassified as walking samples). On the contrary, the intensity of sedentary activities tends to be very low, which may lead to confusions with some low-energy non-sedentary activities as it is shown in the confusion matrix associated to Random Forest classifiers (around 10% of real sedentary regimes are misclassified as non sedentary regimes). All of these confusions are often encountered in other studies aimed at classifying activities and especially walking activities [5].



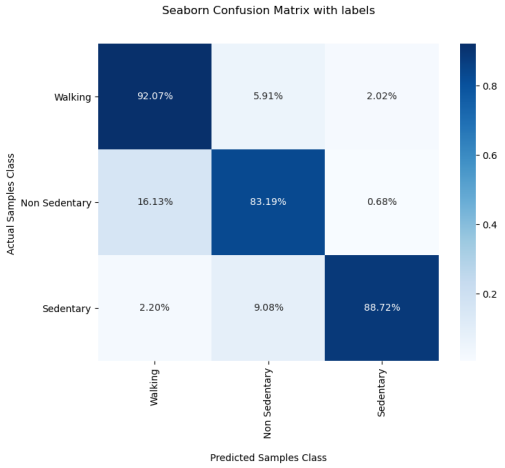
((a)) Linear SVM



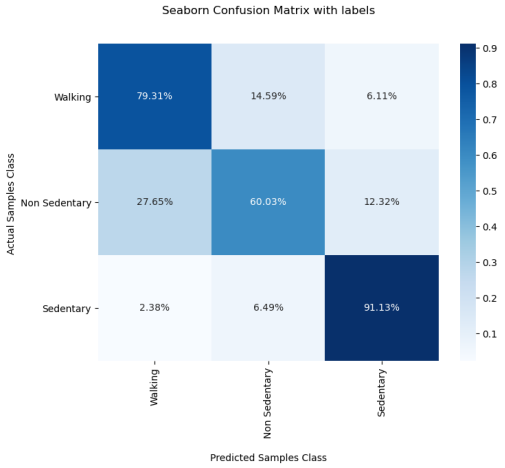
((b)) Polynomial SVM



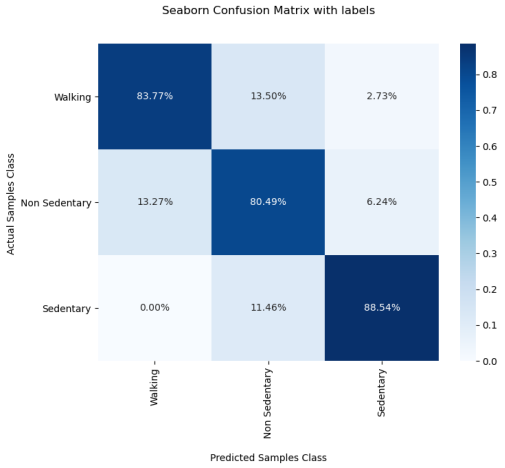
((c)) Gaussian SVM



((d)) kNN



((e)) Decision Tree



((f)) Random Forest

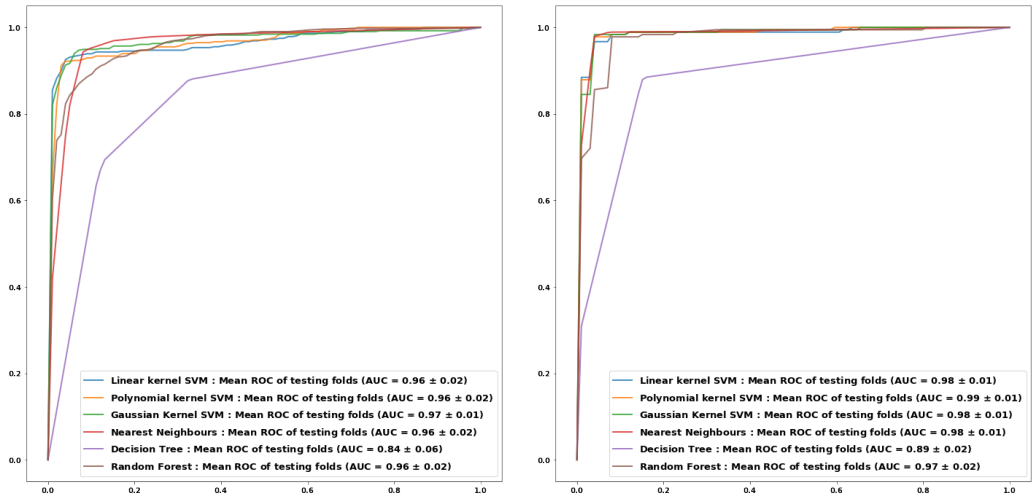
Figure 5.8: Confusion matrices for each tested cascade classifier in Protocol 1

5.4.4 Protocol 2

As detailed in Chapter 3, end-of-lap activities from Protocol 1 were changed to build Protocol 2 and participants' transitions between activity and walking phases were annotated with a wearable camera. Pathological participants were also included.

ROC Curve and AUC

ROC Curves for each type of classifier and for each tested fold are plotted in Figure 5.9. AUC values are satisfactory for all types of classifier but are significantly lower when testing Decision Tree classifier on Task 1 and Task 2) : AUC<0.9 on both tasks.



((a)) Task 1 ((b)) Task 2

Figure 5.9: ROC curves plotted for each classifier applied to all testing folds in Protocol 2

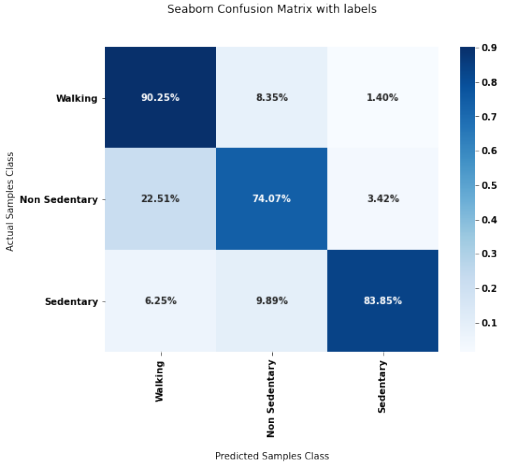
Joint Evaluation of Segmentation and Classification Steps

Table 5.6 displays accuracy's values for each tested cascade classifier. Accuracy results are satisfactory for all types of classifiers (>0.70). Linear SVM, Gaussian kernel SVM and kNN classifiers are particularly accurate (around 0.90).

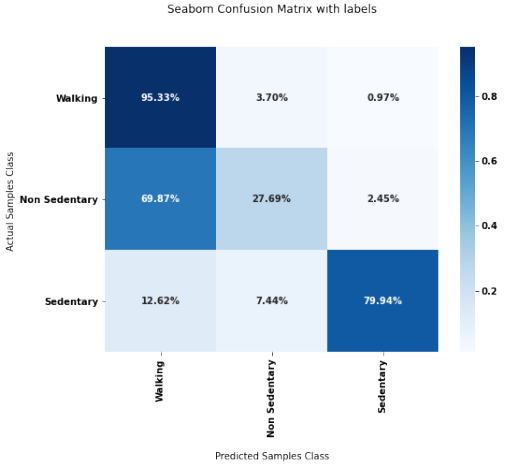
Type of Classifiers	Linear Kernel SVM	Polynomial Kernel SVM	Gaussian Kernel SVM	kNN	Decision Tree	Random Forest
Accuracy	0.88 ± 0.14	0.86 ± 0.05	0.88 ± 0.15	0.89 ± 0.06	0.77 ± 0.17	0.85 ± 0.16

Table 5.6: Accuracy values for each tested cascade classifier in Protocol 2

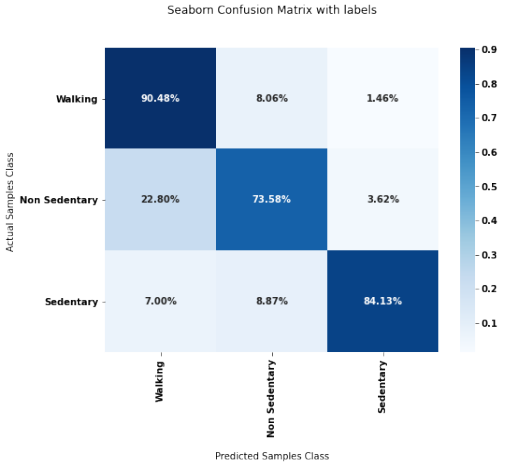
Confusion matrices for each type of classifier are shown in Figure 5.10. Non-sedentary phases are less well discriminated for all classifiers than for Protocol 1 (only three classifiers discriminate correctly around 75% of sedentary samples : Linear SVM, Gaussian SVM and Random Forest). Misclassifications are more numerous than for Protocol 1 and are prohibitive in several cases : respectively 69.87% and 38.05% of non sedentary samples are



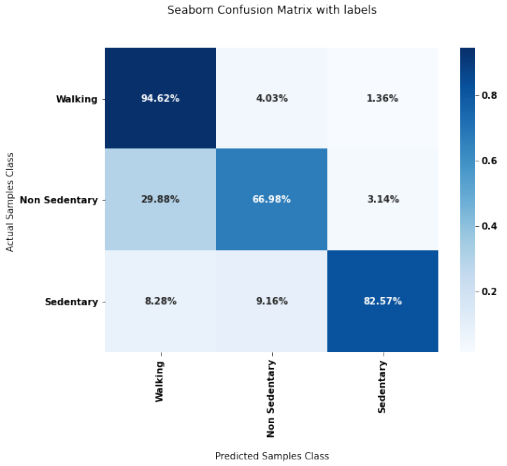
((a)) Linear SVM



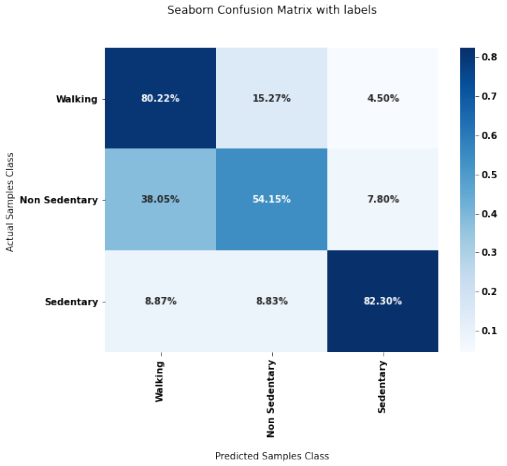
((b)) Polynomial SVM



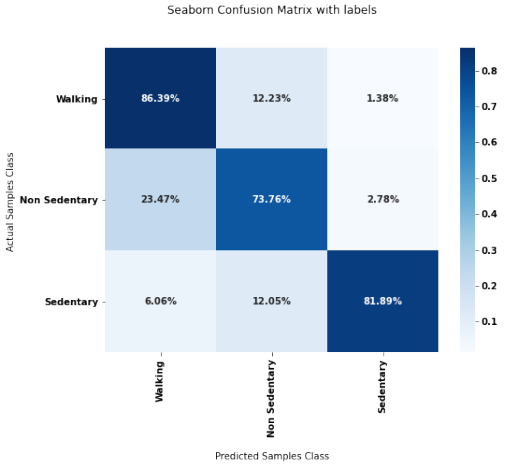
((c)) Gaussian SVM



((d)) kNN



((e)) Decision Tree



((f)) Random Forest

Figure 5.10: Confusion matrices for each tested cascade classifier in Protocol 2

misclassified as walking samples when testing polynomial SVM and Decision Tree classifiers. These increased imprecision margins compared to Protocol 1 can be explained by segmentation errors in Protocol 2 detailed in Chapter 4. Besides, pathological participants from Protocol 2 showed smaller differences in spectral signatures, for example between walking and non-sedentary activities or between non-sedentary and sedentary activities which led to classification errors.

5.5 Conclusion

By comparing the results of the comparison of the classifiers obtained for both protocols, two types of classifiers stand out to be selected in order to characterize the regimes segmented by the method described in Chapter 4 : linear SVM and gaussian kernel SVM. Indeed, these two classifiers present similar AUC values, accuracy values close to 0.90 on both protocols and their matrix confusion presents satisfactory values on both protocols on all classes. Nevertheless, linear SVM induces lower computational costs and is therefore selected to characterize segmented regimes.

Based on the segmentation procedure that Chapter 4's method uses, we have developed a HAR approach that makes it possible to characterise segmented regimes. This method works directly at the regime scale by relying on a feature extraction built from a state of the art, a dimensionality reduction method (PCA) and on a linear SVM cascade classifier to perform accurate classifications of segmented regimes. These two tasks allow for the classification of regimes into three different types of labels : walking regimes, non-sedentary regimes and sedentary regimes. This characterisation of regimes then allows them to be compared with each other and to be evaluated according to healthy subject standards as it is presented in Chapter 6.

6

Visual Feedback Rendering

6.1 Introduction

Once the regimes within recorded signals have been segmented and characterised as described in Chapters 4 and 5, it is then possible to provide a quantitative and qualitative assessment of the detected walking regimes. In this chapter we describe the method leading to an intuitive, didactic and innovative visual feedback of participants' physical activity based on these segmented and characterised regimes. This graphical output is a detailed, adaptive and structured visualization which helps better understand the salient events in a complex gait protocol. Besides, it constitutes a novel tool for clinicians to assess the physical activity of their patients in a different context from their usual medical consultations.

A quick review of existing approaches that have built visual graphical tools for the assessment of physical activity of subjects in free environments using inertial sensors will first be conducted. Then, we will detail scores generation for several gait categories. We need these scores to be able to evaluate the walking regimes according to healthy standards. Then, examples of our visual feedback will be detailed and discussed (with detailed inputs from the clinicians who take care of the pathological participants included in the study).

6.2 Review of visual feedback methods for physical activity's assessment

Several approaches have been used in the literature to display the output of physical activity assessment algorithms.

A first group of studies (which involves the vast majority of the methods listed in this section) lists their aggregated final computed parameters. These features can be of various nature: number of walking regimes in a day [123], time spent in each activity, the proportion of activities associated with ranges of METs that were achieved by the participants [53], accuracy of the classifiers tested for the methods of HAR [65]. This listing is done in order to compare results on different cohorts or to evaluate data processing approaches. A second group of studies plots graphs to analyse the evolution of a specific parameter over time and notably the previously detailed aggregated features. For instance, expenditure rates over a defined period of time can be plotted by assessing the time spent in activities associated with specific Metabolic Equivalent Task MET values [118]. Other methods can display the evolution of more precise metrics such as feet angles [13] over time. In HAR studies, plots can allow users to quickly identify the types of activity carried

out by participants, notably on timelines [111]. A final group of studies set up advanced diagrams to provide a general interpretation of their subjects' physical activity : these can be histograms assessing the quality of walking regimes according to specific parameters such as the variability of signals within a step [123]. These methods can be online and allow immediate monitoring of the participants' physical activities [58].

6.3 Proposed Method

6.3.1 Introduction

It appears that few studies based on the use of IMUs in FLEs have endeavored to provide a macro-analysis displayed in the form of an easy-to-understand visual legend that fully assesses the entire timeline of FLEs signals. The studies reviewed in the previous section provide visual summaries that either focus on metrics that are too specific, which prevents clear and didactive visual feedback, or on metrics that are too general, which prevent a complete assessment of a subject's physical activity in free environments. In this kind of studies, the influence of time is often erased by computing features that are often agglomerated over the whole of the measurements as explained above. Rather than knowing the percentage of time spent in each activity, we can for example be interested in the impact of transitions between activities on the quality of walking regimes, on the evolution of this quality over several consecutive regimes. In this section, we develop a graphical tool by circumventing these pitfalls of physical activity assessment. Our visual summary is built as follows:

- Features characterizing four categories of walking (Stability, Steadiness, Sturdiness, Symmetry) are extracted from the segmented walking regimes (see Section 6.3.2)
- A normalised score is set up in order to be able to compare the evolution of a gait regime on the four categories with a model built on data from healthy subjects (see Section 6.3.4)
- A final graphical tool linking these scores to pre-defined colour codes is produced as output (see Section 6.3.5)

6.3.2 Features extraction

In our visual summary, a walking regime is assessed according to four standard criteria : stability, steadiness, sturdiness and symmetry. The four features to be extracted for regimes' evaluation, detailed in Table 6.1, are chosen because of their ability to accurately characterize the gait, their recurrent use in the literature and ease of computation. All four criteria have been defined by the literature :

Stability [75, 137]: criterion evaluating postural balance used to prevent falls for instance [14]. Stable walking can be defined as *gait that does not lead to falls despite perturbations* [29]. This aspect is evaluated by using the Root Mean Square Ratio computed on the mediolateral acceleration $RMSR_{ML}$. It corresponds to the ratio of the Root Mean Square of the mediolateral accelerations RMS_{ML} to the Root Mean Square vector magnitude computed on all axes RMS_A as displayed in (6.1). RMS evaluates the magnitude

of the acceleration on one specific axis. The higher $RMSR_{ML}$ is, the higher the values of mediolateral accelerations tend to be compared to other accelerations. This indicates an instability on the mediolateral axis and therefore a postural instability. Thus $RMSR_{ML}$ is selected for our study since it has been proven to be uncorrelated with walking speed [137].

$$\begin{aligned} RMS_A &= \sqrt{RMS_{ML}^2 + RMS_{CC}^2 + RMS_{AP}^2}, \\ RMSR_{ML} &= \frac{RMS_{ML}}{RMS_A} \end{aligned} \tag{6.1}$$

Sturdiness [19, 157]: criterion *evaluating gait amplitude* [157]. For instance, sturdiness can be assessed to quantify observed defects in patients with Parkinson's disease with low amplitude movements [92]. This category is evaluated by using the Root Mean Square Ratio computed on the anteroposterior acceleration $RMSR_{AP}$ whose computation is done the same way it is performed for $RMSR_{ML}$. The higher it is, the higher the anteroposterior accelerations' values compared to other accelerations and the higher the sturdiness is. Indeed, high anteroposterior acceleration values mean that steps impulsions are vigorously performed by the participant. $RMSR_{AP}$ is used instead of RMS_{AP} in order to limit the influence of the walking speed.

$$RMSR_{AP} = \frac{RMS_{AP}}{RMS_A} \tag{6.2}$$

Steadiness [90, 106] : criterion evaluating step regularity : *regularity is related to similarity of consecutive strides* [153]. This category can be analyzed in order to quantify locomotion flaws in targeted cohorts with lower limbs defects (such as transfemoral amputees). This category is evaluated by using the second peak of the autocorrelation coefficients calculated on craniocaudal accelerations via the Wiener-Khinchin theorem $P2_{CC}$. This unbiased autocorrelation function uses both Fast Fourier Transform FFT and inverse Fast Fourier Transform iFFT as detailed in Table 6.1. This feature compares similarity between strides within a walking regime since it occurs with a time lag of two steps. The higher $P2_{CC}$ is, the more similar the performed strides are. Let aCC be the associated craniocaudal acceleration signal, ConjFFT(aCC) the complex conjugate of FFT(aCC) and ACF the autocorrelation coefficients : $P2_{CC}$ is defined as detailed in Table 6.1. Figure 6.1 shows a craniocaudal acceleration signal associated to its autocorrelation : $P1_{CC}$ and $P2_{CC}$ locations are presented.

$$ACF = iFFT[FFT(aCC)ConjFFT(aCC)] \tag{6.3}$$

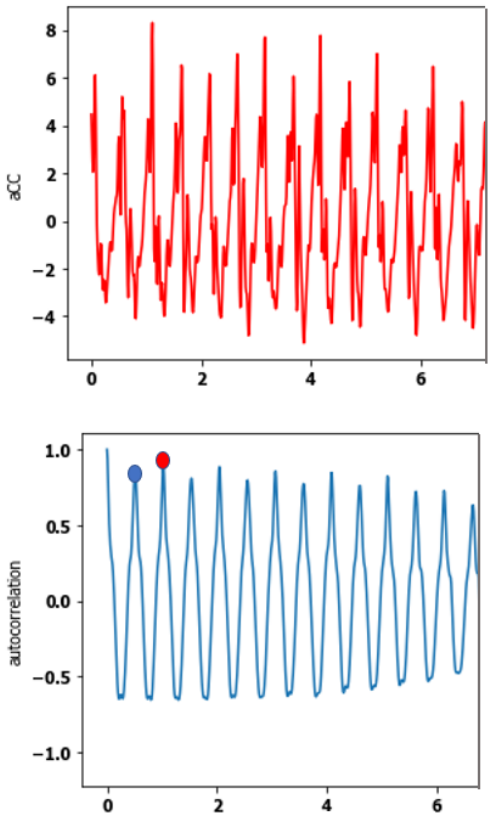


Figure 6.1: One aCC signal and its associated unbiased autocorrelation. Definition of $P1_{CC}$ (blue dot) and $P2_{CC}$ features (red dot)

Symmetry [90, 106] : criterion evaluating step symmetry : *A symmetric gait pattern for humans is characterized by the almost identical behavior of bilateral limbs during a gait cycle* [31]. This category is evaluated by using the first peak of the autocorrelation coefficients calculated on craniocaudal accelerations via the Wiener-Khinchin theorem $P1_{CC}$. It compares similarity between steps within a walking regime since it occurs with a time lag of one step. $P1_{CC}$ evaluates the ability to maintain vertical correspondence between right and left hemi-bodies during walking regimes. The higher it is, the more similar steps from both sides are.

6.3.3 Features robustness

In the final graphical feedback, four features are used to characterize the gait activity. The robustness of the feedback depends mainly on the robustness of those features, especially when confronted with segmentation errors. To investigate this issue, we conducted an additional experiment where we intentionally degrade the segmentation process (e.g., by voluntary lowering the number of samples for feature computation), in order to assess the robustness of the features. In total, 10 degraded configurations are tested, as described in Table 6.2. Figure 6.2 shows the distribution of features in all categories over the 10 configurations in walking sections of one healthy subject (*HSU*) and two pathological subjects:

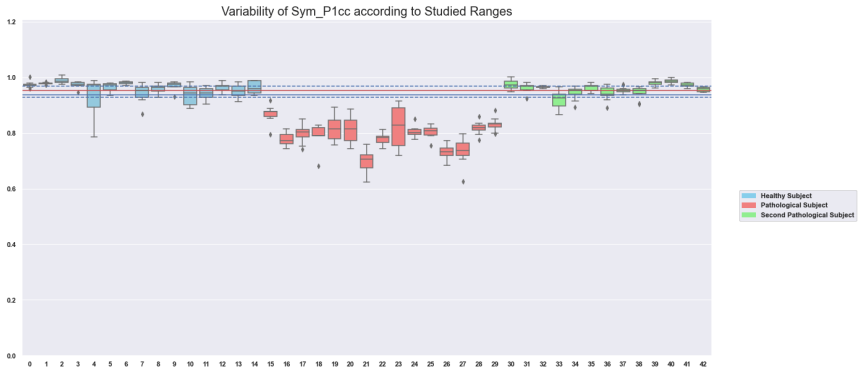
Table 6.1: Features used to establish scores for the graphical feedback. In total, 4 features are used. ConjFFT(X) denotes the complex conjugate of FFT(X).

Categories	Features	Description	Mathematical Computation
Steadiness	$P2_{CC}$	The second peak of the autocorrelation coefficients calculated on Craniocaudal accelerations via the Wiener-Khinchin theorem: the higher it is, the more similar the steps are	$ACF = iFFT[FFT(X)ConjFFT(X)]$, P1 is the first peak of ACF while P2 is the second peak
Symmetry	$P1_{CC}$	The first peak of the autocorrelation coefficients calculated on Craniocaudal accelerations via the Wiener-Khinchin theorem: the higher it is, the more similar the strides are	P1 is the first peak of ACF while P2 is the second peak
Sturdiness	$RMSR_{AP}$	Root Mean Square Ratio on anteroposterior acceleration. The higher it is, the higher the sturdiness is.	$RMS_A = \sqrt{RMS_{ML}^2 + RMS_{CC}^2 + RMS_{AP}^2}$, $RMSR_{AP} = \frac{RMS_{AP}}{RMS_A}$
Stability	$RMSR_{ML}$	Root Mean Square Ratio on mediolateral acceleration. The lower it is, the higher the stability is.	$RMS_A = \sqrt{RMS_{ML}^2 + RMS_{CC}^2 + RMS_{AP}^2}$, $RMSR_{ML} = \frac{RMS_{ML}}{RMS_A}$

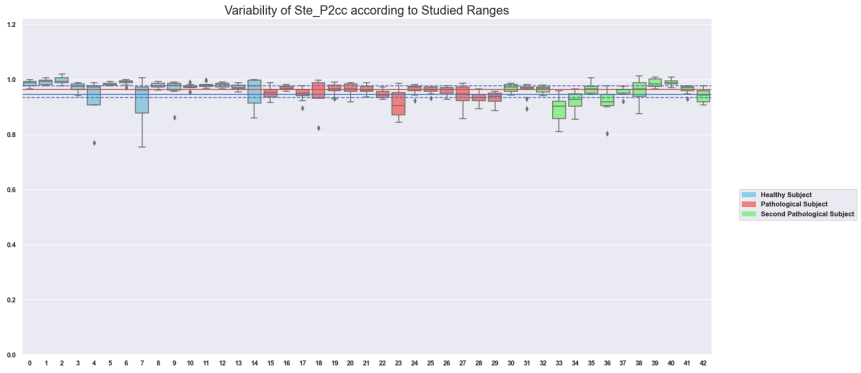
Configurations
All the regime is used (normal configuration)
Only the first 3 seconds of the regime are used
Only the first 3.5 seconds of the regime are used
Only the first 4 seconds of the regime are used
Only the first 5 seconds of the regime are used
Only the first 40% of the regime are used
Only 40% of the regime is used (with start at 20% of the total duration)
Only 40% of the regime is used (with start at 30% of the total duration)
Only 40% of the regime is used (with start at 40% of the total duration)
Only the last 40% of the regime are used

Table 6.2: Degraded configurations for the computation of the features

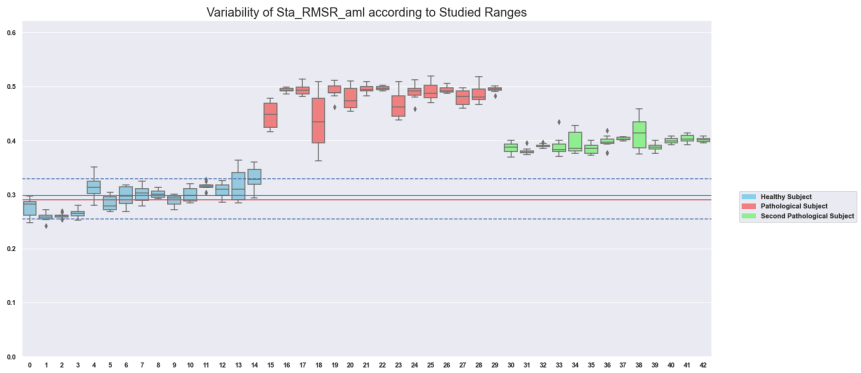
PSU1 (gluteus medius deficiency) and *PSU2* (Post-radiation left brachial plexitis). *PSU1* has shown the highest instability and lack of symmetry in his deambulation, *PSU2* has shown degraded sturdiness. For each subject, we have extracted all walking regimes, and computed the features according to the different configurations. Each box contains the distribution of the different values of this feature on all 10 tested configuration in a given walking regime. The walking regimes for *HSU* are displayed in blue, and the ones of the first, second and third pathological subject respectively in red and green.



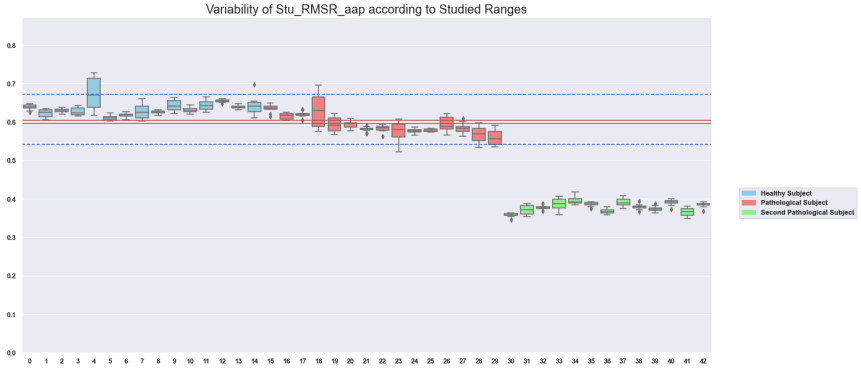
((a)) Symmetry $P1_{CC}$



((b)) Steadiness $P2_{CC}$



((c)) Stability $RMSR_{ML}$



((d)) Sturdiness $RMSR_{ML}$

Figure 6.2: Evaluation of the robustness of selected features. Features with low dispersion and high discrimination between classes. The blue horizontal line shows the average value of the feature for all healthy subjects, the red horizontal line shows the median value of the feature for all healthy subjects and the dotted lines correspond to the $75^{th}/25^{th}$ percentiles. Each boxplot corresponds to 10 computations of the feature on a walking regime on 10 degraded ranges. Boxplots are displayed with specific colors depending on their associated subject: blue for an healthy subject, red for *PSU1* and green for *PSU2*.

One first observation is that all boxes display little spread over all recorded subjects, which suggests that the computation process is robust. The blue horizontal line shows the average value of the feature for all walking regimes from healthy subjects and the dotted lines correspond to the $75^{\text{th}}/25^{\text{th}}$ percentiles. It is interesting to note that the differences between the three subjects are clearly visible for all walking regimes in all categories but steadiness (no recorded participants displayed an affected regularity). Moreover, the patient with the most impact on his stability (*PSU1*) displays boxes associated to the $RMSR_{ML}$ feature that are even more detached than *PSU2* from the figures of the healthy subjects, which confirms the different visual impacts observed on the gait of each pathological subject. This feature thus presents satisfactory robustness' results in terms of dispersion on the degradation ranges as well as in terms of discrimination between subjects. This confirms the relevance of using this feature to evaluate the stability of walking regimes. The calculation of this feature remains indeed constant on all the ranges presented in Table 6.2, which allows our method to be correctly applied despite eventual segmentation errors that may occur. No patient presented a continuous affection in steadiness and it was thus difficult to estimate the discrimination power from $P2_{CC}$: it must be evaluated in further works. Other figures and additional experiments show that all other features listed in 6.1 display the same consistency and robustness, which is an important asset of our proposed approach.

6.3.4 Scores' generation with healthy models

Using a database of healthy walking phases taken from the healthy subjects from Protocol 1 and Protocol 2, statistics for the different features are computed (means, percentiles). These models are then used to assess each novel walking phase with the scoring procedure described as follows.

Considering a feature with mean μ and standard deviation σ on all walking regimes from healthy subjects, we compute the z-score normalized feature

$$z = \frac{x - \mu}{\sigma}. \tag{6.4}$$

The z-score normalized features are then displayed with a color bar of boundaries $[x_{min}, x_{max}]$ where x_{min} and x_{max} are respectively the 10% and 90% percentiles of the normalized features on the healthy subjects. This score is used to build evaluation of walking regimes which are displayed in our visual feedbacks.

6.3.5 Visual feedback

Graphical outputs based on scores generation (examples are shown in the following section) display the evaluation of the whole protocol segmented in regimes in a clockwise manner. The first outer circle specifies the nature of the segmented regimes: dark blue for non-sedentary activities, standard blue for sedentary activities and light blue for walking regimes. The four next inner concentric circles are each associated with a gait criterion: stability, steadiness, sturdiness and symmetry. Each portion of these circles delimited by black lines corresponds to a segmented regime, whose length is proportional to the duration of the regime. For a given evaluation criterion, each walking regime is then assigned a color from dark red to dark green. This color depends on the comparison of this regime to the average on healthy walking regimes. Non-walking regimes are not evaluated and are

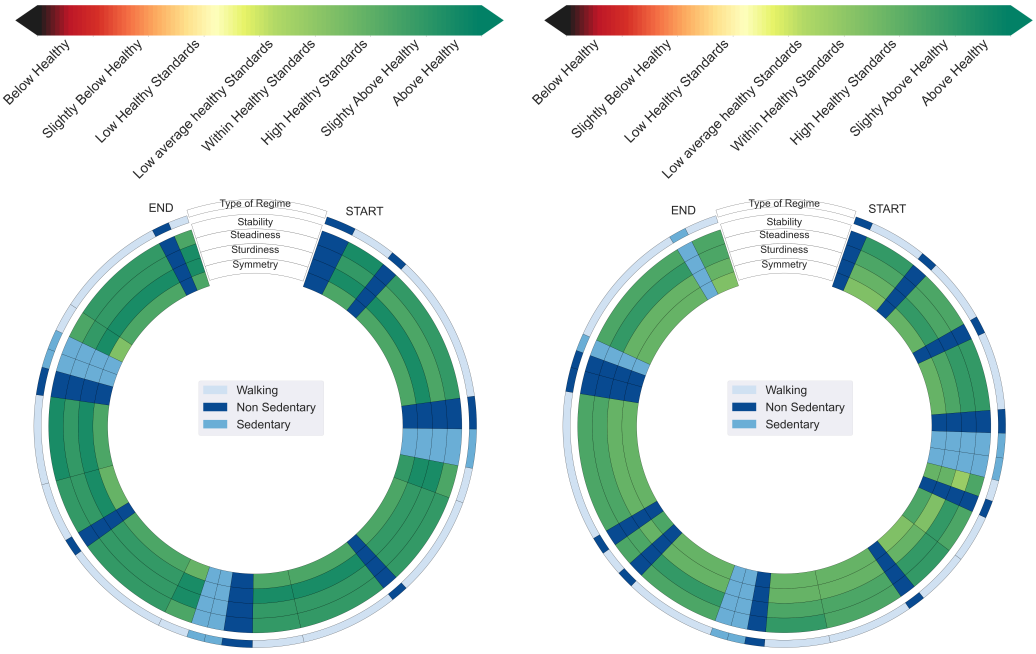
displayed in dark blue for non-sedentary activities and light blue for sedentary activities (as in the first outer circle).

6.4 Results

Graphical outputs are plotted for 2 healthy subjects in Protocol 1, 2 healthy subjects in Protocol 2 and for all included pathological subjects in Protocol 2 as detailed in Chapter 3. Visual summary for one specific participant was built by training methods on all other included participants (penalty learning, classifers training...). For neurological patients, figures are provided with comments from a neurologist (Hopital d'Instruction des Armées HIA de Percy, Service de Santé des Armées SSA). For orthopedical patients, figures are provided with comments from an orthopedical surgeon (Hopital d'Instruction des Armées HIA de Percy, Service de Santé des Armées SSA). Both clinicians were asked to assess the reliability between rendered visual feedbacks and the actual condition of the patients they follow on a daily basis.

6.4.1 Examples of visual feedbacks from Protocol 1

Figure 6.3 shows graphical outputs rendered for healthy participants from Protocol 1. Walking regimes from all participants are deemed to be of satisfactory quality according to all evaluated criteria. These graphical tools make it possible to see the course of the entire protocol, with alternating periods of walking, sedentary activities, and non-sedentary activities. It can also show how many times the subject has completed an entire lap of the protocol (three times for the healthy subjects presented here), and differences in the chronology between each performed activity.



((a)) Healthy subject 1

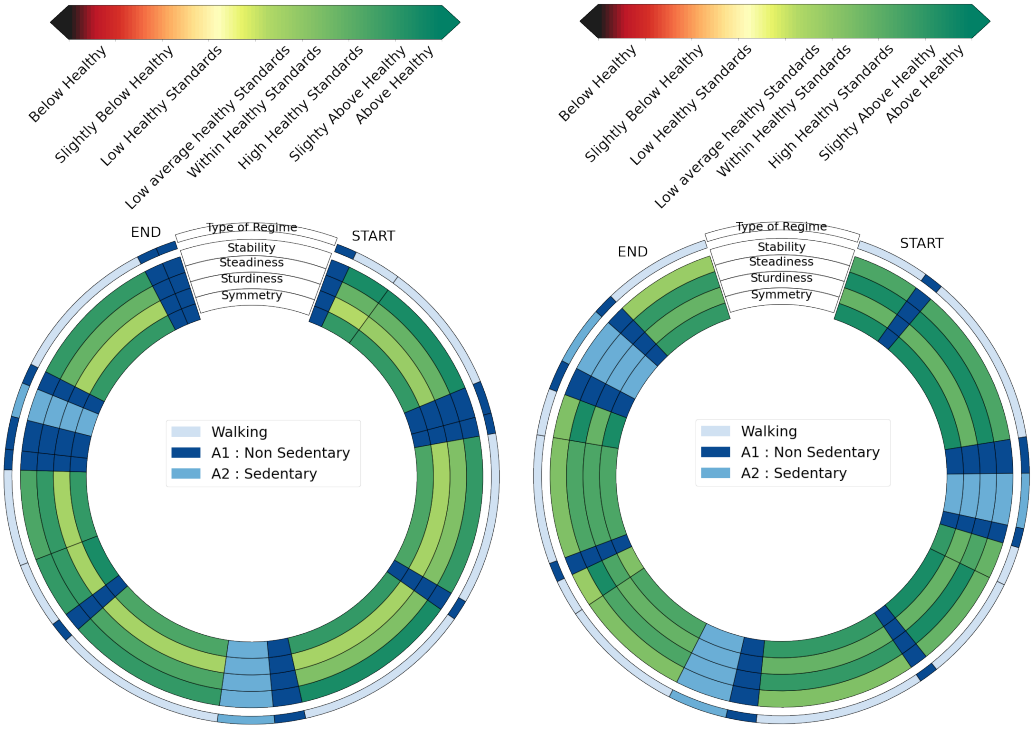
((b)) Healthy subject 2

Figure 6.3: Visual Feedbacks from Healthy Subjects in Protocol 1

6.4.2 Examples of visual feedbacks from Protocol 2

Healthy subjects

Figure 6.4 shows graphical outputs rendered for healthy participants from Protocol 2. Walking regimes from all participants are deemed to be of satisfactory quality according to all evaluated criteria.



((a)) Healthy subject 1 ((b)) Healthy subject 2

Figure 6.4: Visual Feedbacks from Healthy Subjects in Protocol 2

Neurological patients

Figures 6.5, 6.6, 6.7 and 6.8 show graphical output rendered for neurological patients from Protocol 2. All visual feedbacks from participants included to train segmentation and classification methods are displayed. One additional visual feedback taken from a newly recorded neurological participant (Post Radiation Leukopathy after treatment of a cerebral recurrence of lymphoma) is also displayed.

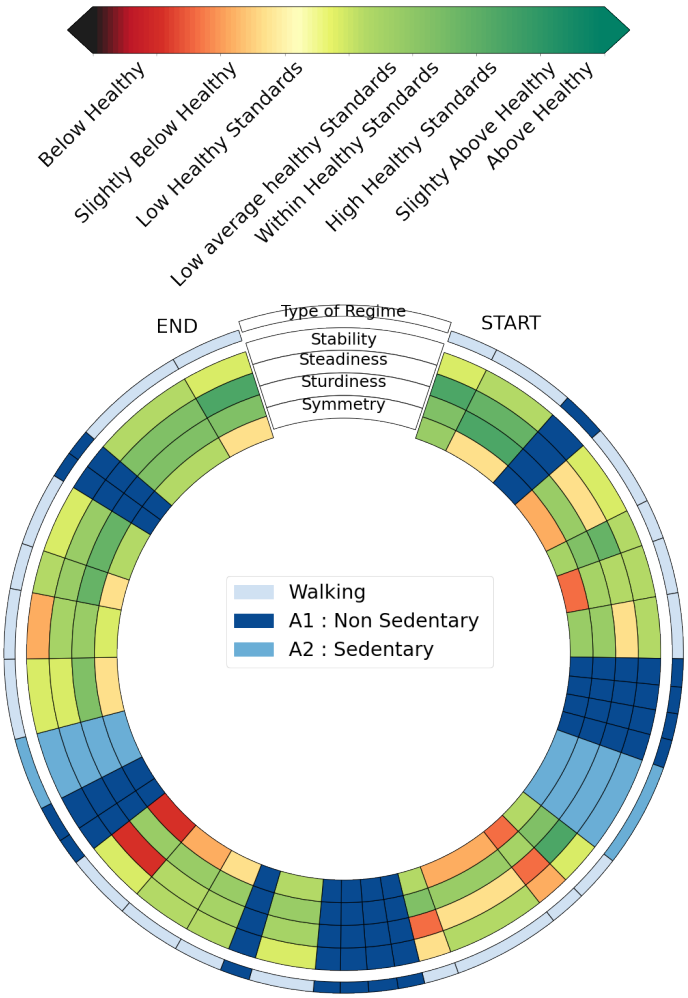


Figure 6.5: Visual feedback from Neurological Subject 1 : *patient with post-radiation leukopathy after treatment of a left temporal glioma*

Comment from the neurologist : "This is a patient with post-radiation leukopathy after treatment of a left temporal glioma. The patient presents a dysexecutive cognitive impairment with difficulties in controlling her behavior : loss of planning, attention, and of high-level control of her behavior. What is observed by the study of walking in this visual feedback reflects the disorder of planning and control of behavior with distractibility."

Additional comment: The observation from the neurologist correlates with the rendered graphical output : there are numerous detected change points within the signal which shows erratic walking. Besides, walking evaluations are slightly degraded.

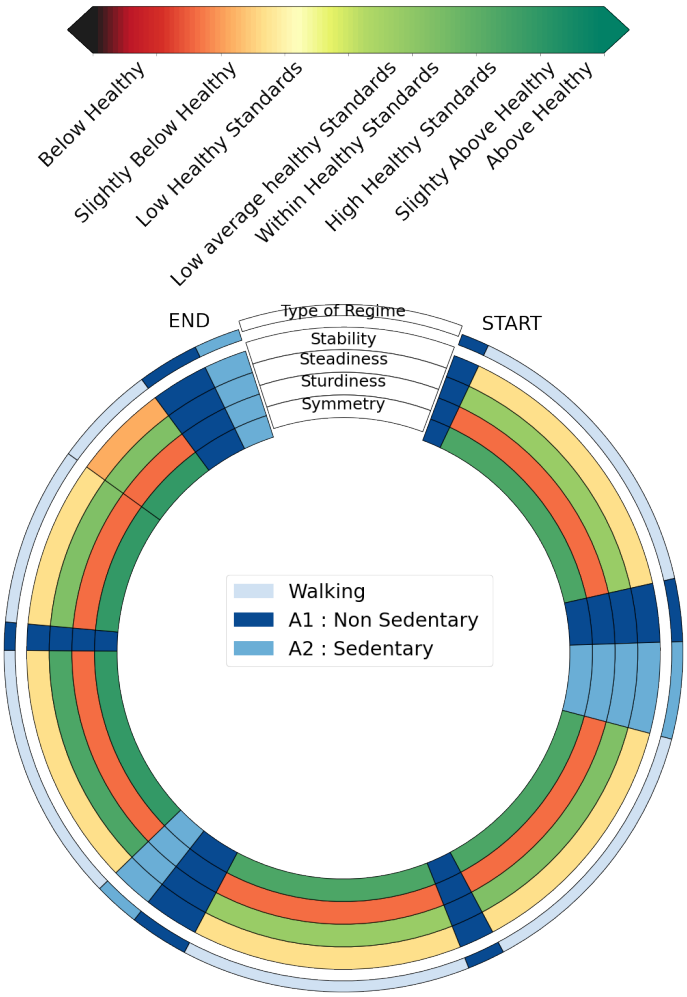


Figure 6.6: Visual feedback from Neurological Patient 2 : *patient with post-radiation left brachial plexitis 20 years after radiation treatment for breast cancer*

Comment from the neurologist : "The patient suffered from post-radiation left brachial plexitis 20 years after radiation treatment for breast cancer. Complete paralysis of the entire left upper limb was observed. The patient's stability and sturdiness are affected. The impairment of balance may be due to the imbalance related to the dead weight of her left arm, which hangs from the shoulder and weighs at least 10 kg."

Additional comment: The observation from the neurologist correlates with the rendered graphical output which displays a degraded sturdiness (patients performed steps with little amplitude) and a slightly degraded stability.

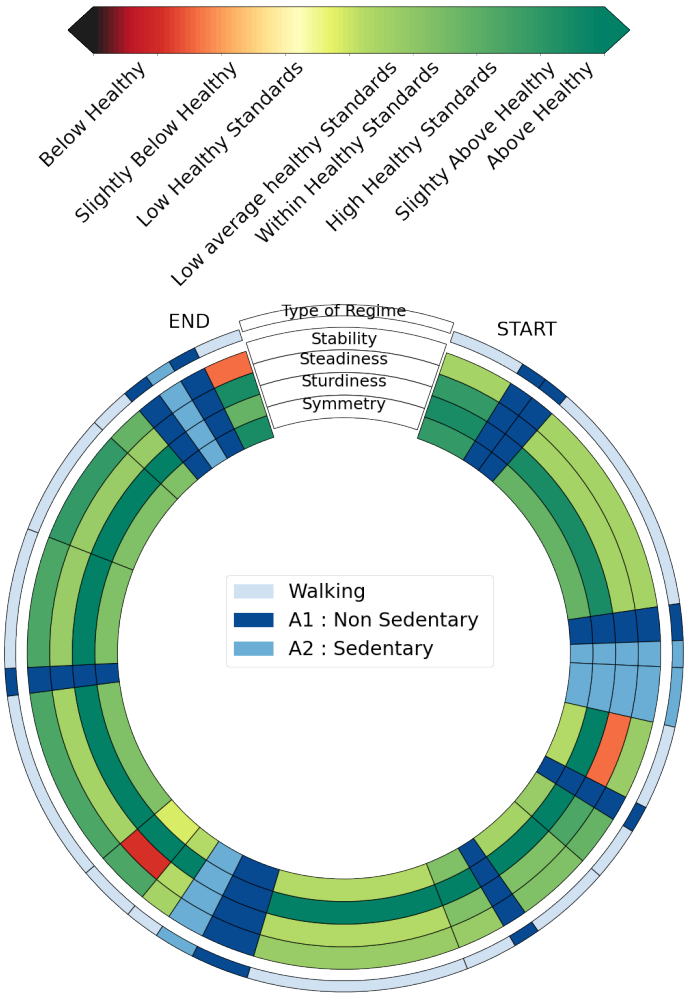


Figure 6.7: Visual feedback from Neurological Patient 3 : *patient with damage to the cranial pairs and neck muscles after radiotherapy of a Chordoma*

Comment from the neurologist : "The patient has damage to the cranial pairs and neck muscles after radiotherapy of a Chordoma (tumor of the sphenoid bone which is the bone of the skull behind the visage) without locomotor affection: this is verified on this graph with notably a good walking sturdiness."

Additional comment: The observation from the neurologist correlates with the rendered graphical output except for a lack of stability and steadiness in few regimes.

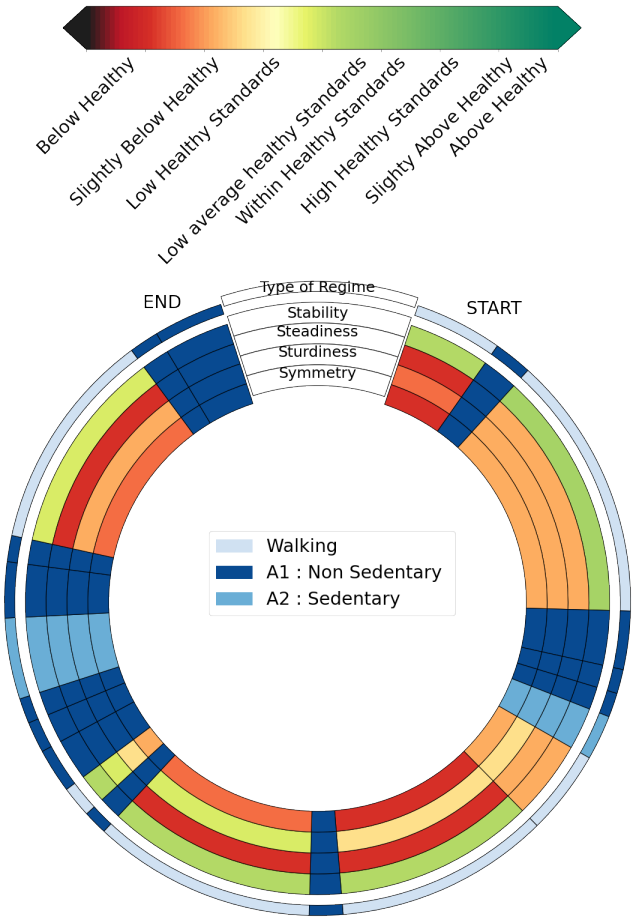


Figure 6.8: Visual feedback from Neurological Patient 4 : *patient with Post Radic Leukopathy after treatment of a cerebral recurrence of a lymphoma*

Comment from the neurologist : "Post Radic Leukopathy after treatment of a cerebral recurrence of a lymphoma. In addition to the loss of behavioral control identical to that of patient 1, he presents a severe balance control disorder: affection of all the categories of walking."

Additional comment: The observation from the neurologist correlates with the rendered graphical output : all categories are affected on at least one regime.

Orthopedical patients

Figures 6.9, 6.10, 6.11, 6.12, 6.13 and 6.14 show graphical output rendered for orthopedical patients from Protocol 2. All visual feedbacks from participants included to train segmentation and classification methods are displayed.

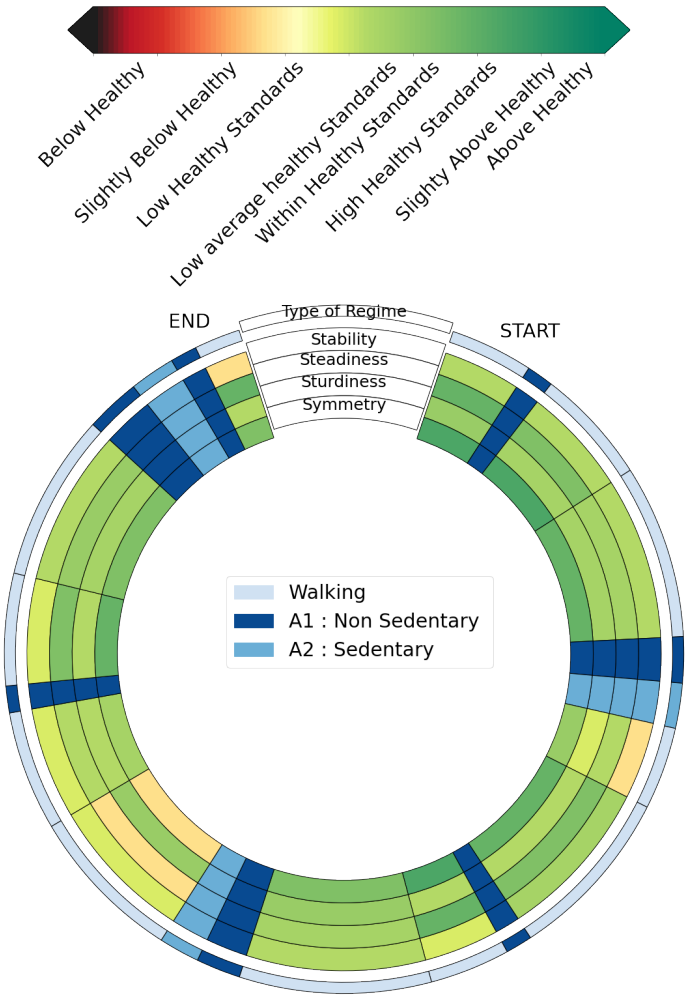


Figure 6.9: Visual feedback from Orthopedical Subject 1 : *patient in immediate preoperative phase of a knee ligamentoplasty*

Comment from the orthopedical surgeon : "This patient was examined in the immediate preoperative phase of a knee ligamentoplasty. Clinically, she presented with a flexible, mobile, painless knee and an overall normal gait. However, she had knee instability, which justified the operation and could explain the poorer results on certain parameters."

Additional comment: The observation from the orthopedical surgeon correlates with the rendered graphical output : categories' evaluations are satisfactory even if they display low healthy standards

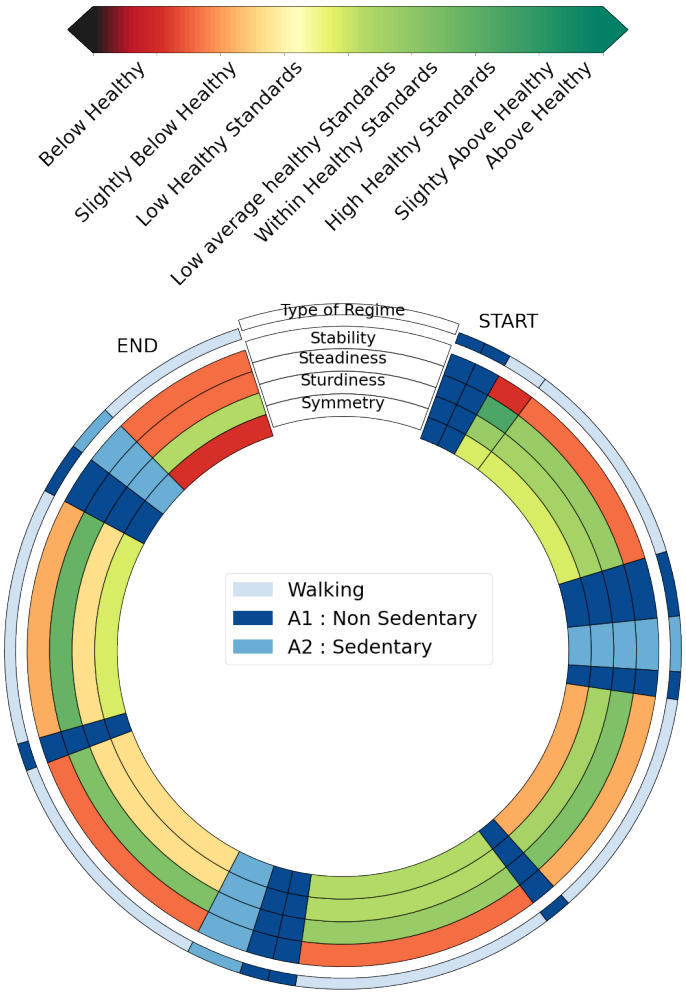


Figure 6.10: Visual feedback from Orthopedical Subject 2 : *patient is in the immediate postoperative period of a knee ligamentoplasty*

Comment from the orthopedical surgeon : "This is the same patient recorded as the one in the previous Figure 6.9. Here, the patient is in the immediate postoperative period. At this time, there was significant quadricipital sideration which fully explains the alteration in stability and symmetry."

Additional comment: The observation from the orthopedical surgeon correlates with the rendered graphical output : stability and symmetry are affected especially in the last regime : patient displayed highly irregular foot placements in this last regime. Besides, sturdiness results are changing depending on the regimes. This succession of figure shows how efficient this graphical tool is to follow-up longitudinally patients.

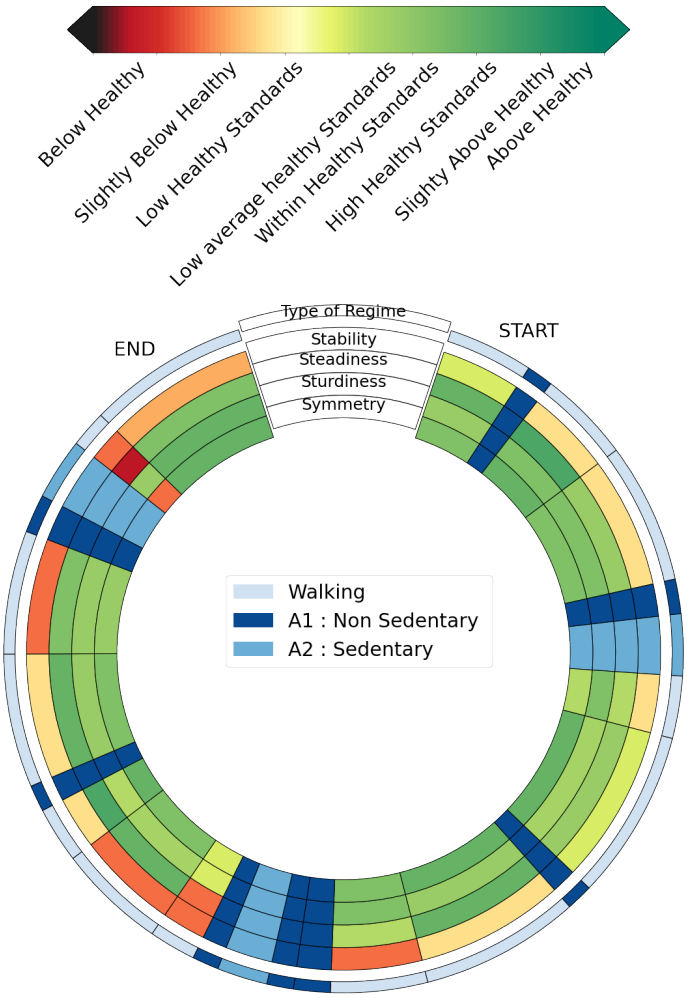


Figure 6.11: Visual feedback from Orthopedical Subject 3 : *patient in the immediate preoperative phase of an Antero Cruciate Ligaments ACL ligamentoplasty*

Comment from the orthopedical surgeon : "This patient was also examined in the immediate preoperative phase of an Antero Cruciate Ligaments ACL ligamentoplasty. However, he had two associated meniscal tears that caused pain and minimal joint limitation, which could explain this slightly degraded gait despite the preoperative rehabilitation."

Additional comment: The observation from the neurologist correlates with the rendered graphical output : stability was degraded. It could be added that the patient was using his smartphone during the whole acquisition which could have affected his stability.

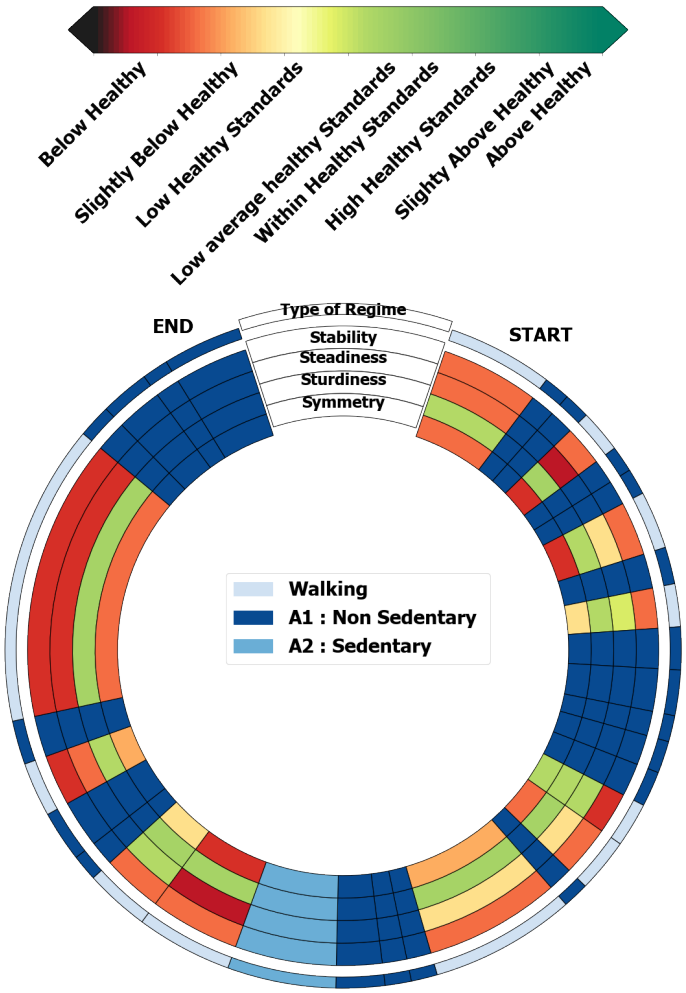


Figure 6.12: Visual feedback from Orthopedical Subject 4 : *patient examined the day after a total hip replacement*

Comment from the orthopedical surgeon : "This is a patient examined the day after a total hip replacement. There is a lameness linked to a moderate pain and a transient muscular deficit."

Additional comment: The observation from the neurologist correlates with the rendered graphical output : symmetry is degraded as well as stability and steadiness.

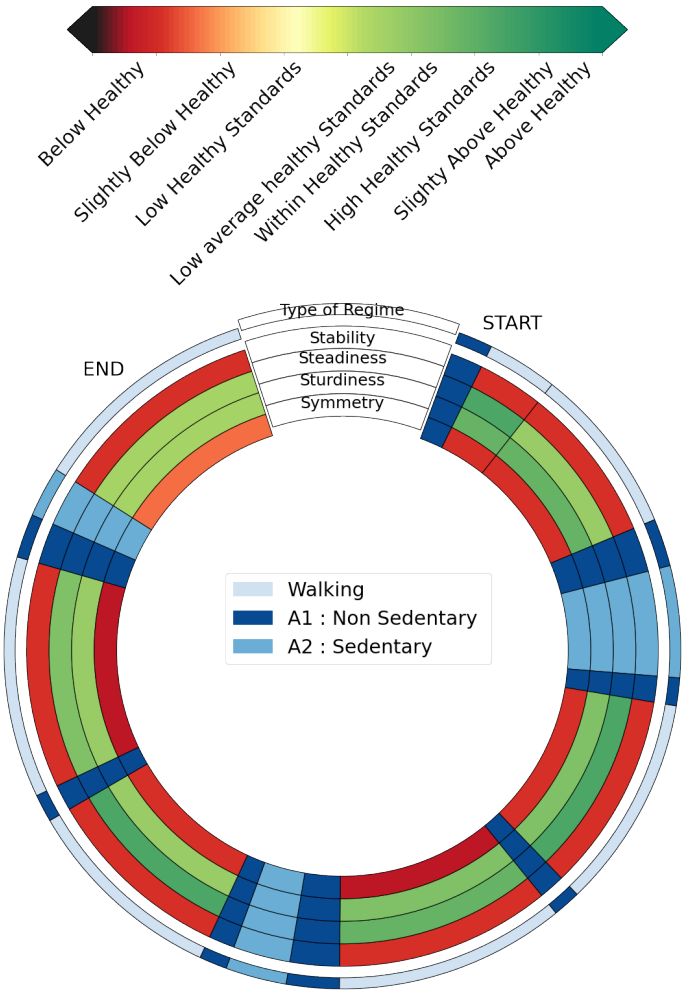


Figure 6.13: Visual feedback from Orthopedical Subject 5 : *the patient has a history of multiple right hip surgeries resulting in chronic gluteus medius insufficiency*

Comment from the orthopedical surgeon : "This patient has a history of multiple right hip surgeries resulting in chronic gluteus medius insufficiency causing Trendelenburg type lameness. This explains the low scores for symmetry and stability despite a preserved speed."

Additional comment: The observation from the neurologist correlates with the rendered graphical output : highly affected symmetry and stability.

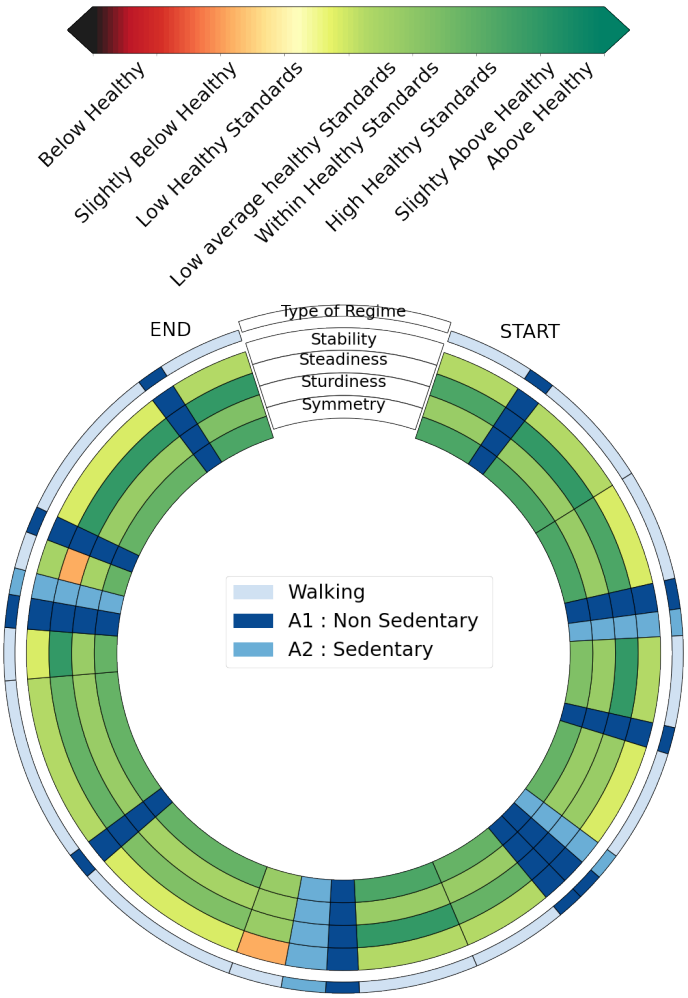


Figure 6.14: Visual feedback from Orthopedical Subject 6 : *the patient was undergoing rehabilitation after a knee sprain with ACL rupture*

Comment from the orthopedical surgeon : "The patient was undergoing rehabilitation after a knee sprain with ACL rupture. She had no clinical instability and did not require surgical treatment. The slightly low results are explained by her post-traumatic condition, which caused moderate pain and minimal muscle weakness, which was in the process of recovery."

Additional comment: The observation from the neurologist correlates with the rendered graphical output : all categories display average to good evaluations.

6.5 Conclusion

The method presented in this chapter leads to a reliable, innovative and didactic graphical tool for assessing the physical activity of subjects. We believe that such a visualization will help clinicians to perform more accurate and comprehensive longitudinal tracking of locomotion in the natural environment of their patients. This could for instance allow to evaluate rehabilitation procedures or treatment choices for specific diseases and to assess the impact of treatments on pathologies such as musculoskeletal tumors of the lower limbs

or neurological disorders (Parkinson's disease...). The current monitoring of the effects of these FLEs treatments is only carried out via calculations of general (steps/day, ambulatory bouts/day...) or specific metrics (variations withing steps, angles...) when our graphical tool will enable a refined follow up displaying an enhanced macro analysis of gait phases.

Conclusion and Perspectives

In this thesis, several contributions were proposed in order to provide practitioners with a didactic graphical tool for an innovative assessment of physical activity and in particular walking in free environments. The primary motivation of this study was to address the difficulties of constructing interpretation tools in such environments. Indeed, the existing evaluation methods aggregate the evaluation features or compute excessively precise metrics, which prevents the macro-analysis that we wish to implement with our new visual feedback.

In Chapter 3, we conceive and implement three protocols in collaboration with clinical experts (two in semi-FLEs including 15 healthy participants and one in FLE including 21 healthy subjects, 9 patients with an orthopedic pathology and 3 patients with cerebral lesions) based on the findings detailed in the literature review detailed in Chapter 2. In every measurement, IMUs are attached to participants while they perform several activity and walking phases. These protocols allow to perform longitudinal follow-ups and interindividual comparisons on included cohorts. New data need to be recorded in order to, for example, support our methods of longitudinal follow-up before and after surgery.

In Chapter 4, we develop an adaptive change point detection algorithm to process signals measured by the IMUs. This method looks for significant changes in the time-frequency space at a given scale, i.e., the moments when the subject has modified his behavior/activity. This method relying on an adapted learning of a penalty parameter through an optimization process enables a precise change point detection in acquisitions performed in semi-FLEs. By optimizing β which defines a penalty feature, new acquisitions can be segmented to get partitions that are as close as possible to the strategy of annotation without knowing the number of change points to detect. This adaptive approach is based on a structured functioning that first enforces the detection of large change points before detecting smaller change points. Signals are thus segmented into several homogeneous regimes that allow us to extract relevant information from the experiments. Obtained results are competitive when the method is applied to both semi-FLEs protocols. Inaccuracies observed in the results of our segmentation method, especially for turns, could be corrected by changing the annotation strategy. Indeed, we could for example label these turns as an additional type of activity by annotating the beginning and end of these regimes. Furthermore, in our method, signals must be fully annotated in order to obtain satisfactory results. This process can be time consuming if it has to be performed on even longer signals often found in FLEs and semi-FLEs. An interesting research direction would be to extend the experts' annotation mechanisms on portions of signals (containing the targeted change points) in order to automatically reproduce this annotation strategy to segment longer signals. In addition, we could reinforce representation learning: we

could learn, for example, which lines of the spectrograms are the most relevant in our segmentation methods.

In Chapter 5, we conceive a HAR approach that enables to characterise segmented regimes. This method works directly at the regime scale by relying on a feature extraction process built from a selective review, a dimensionality reduction method (PCA) and on a linear SVM cascade classifier to perform accurate classifications of segmented regimes. These two tasks allow for the classification of regimes into three different types of labels : walking regimes, non-sedentary regimes and sedentary regimes. One possibility for improving the study would be to compare additional classifiers and other tuning parameters in order to provide even more reliable interpretations as for the final choice of the cascade classifier.

In Chapter 6, we build a reliable, innovative and didactic graphical tool for assessing the physical activity of subjects and especially their walking phases. This visual output relies on four walking regimes' evaluation critera : Stability, Steadiness, Sturdiness, Symmetry. Features associated to each criterion are extracted within walking regimes and models are built from healthy standards computed on those features. New walking regimes' assessments can then be compared to these pre-defined scores. Such a visual summary will help practitioners to perform more accurate and comprehensive longitudinal tracking of locomotion in free environments. This could for instance allow to evaluate rehabilitation procedures or treatment choices for specific diseases and to assess the impact of treatments on pathologies such as musculoskeletal tumors of the lower limbs or neurological disorders (Parkinson's disease...). The current monitoring of the effects of these FLEs treatments is only carried out via calculations of general (steps/day, ambulatory bouts/day...) or specific metrics (variations withing steps, angles...) when our graphical tool will enable a refined follow up displaying an enhanced macro analysis of gait phases. New walking criteria could be added to our graphical feedback as well as assessments on regimes other than walking regimes (non-sedentary activities for instance). Another possible research direction would be to measure more healthy subjects in order to refine the comparison models. Furthermore, the colour codes of the colormap used for the final visual summary could be modified by including more pathological subjects. Indeed, adding extreme values in the range of possible values would help to refine the comparison results. In addition, including more subjects with conditions would allow them to be clustered together to build pathology-specific comparison models. Besides, a distance parameter between two graphical feedbacks could be introduced. It could enable a reliable quantified inter-individual comparison in addition to the visual comparison allowed by the graphical tools. Finally, a new training of our processing pipeline must be launched on new data acquired from long signals. Protocol 3 has enabled us to acquire some data, but the data sets need to be larger in order to refine our methods on annotation strategies that can be different because of the length of the signals.

Bibliography

[1] Manuel Abbas and Régine Le Bouquin Jeannès. Acceleration-based gait analysis for frailty assessment in older adults. *Pattern Recognition Letters*, 161:45–51, 2022.

[2] Matthew Ahmadi, Margaret O'Neil, Maria Fragala-Pinkham, Nancy Lennon, and Stewart Trost. Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. *Journal of NeuroEngineering and Rehabilitation*, 15(1):105, dec 2018.

[3] Matthew N Ahmadi, Margaret E O'neil, Emmah Baque, Roslyn N Boyd, and Stewart G Trost. Machine learning to quantify physical activity in children with cerebral palsy: Comparison of group, group-personalized, and fully-personalized activity classification models. *Sensors*, 20(14):3976, 2020.

[4] Mohammad Arif Ul Alam, Nirmalya Roy, Sarah Holmes, Aryya Gangopadhyay, and Elizabeth Galik. Automated Functional and Behavioral Health Assessment of Older Adults with Dementia. In *Proceedings of the 1st International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE, Washington DC, USA, 2016*, pages 140–149, Washington DC,USA, aug 2016. IEEE.

[5] Javier Andreu-Perez, Luis Garcia-Gancedo, Jonathan McKinnell, Anniek Van der Drift, Adam Powell, Valentin Hamy, Thomas Keller, and Guang-Zhong Yang. Developing fine-grained actigraphies for rheumatoid arthritis patients from a single accelerometer using machine learning. *Sensors*, 17(9):2113, 2017.

[6] D. Angelosante and G. B. Giannakis. Group lassoing change-points piece-constant ar processes. *EURASIP Journal on Advances in Signal Processing*, 70, 2012.

[7] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. Physical human activity recognition using wearable sensors. *Sensors*, 15(12):31314–31338, 2015.

[8] Darren Au, Andrew G Matthew, Paty Lopez, William J Hilton, Rashami Awasthi, Guillaume Bousquet-Dion, Karim Ladha, Franco Carli, and Daniel Santa Mina. Prehabilitation and acute postoperative physical activity in patients undergoing radical prostatectomy: a secondary analysis from an rct. *Sports medicine-open*, 5(1):18, 2019.

[9] Muhammad Awais, Luca Palmerini, Alan K Bourke, Espen AF Ihlen, Jorunn L Helbostad, and Lorenzo Chiari. Performance evaluation of state of the art systems for physical activity classification of older subjects using inertial sensors in a real life scenario: A benchmark study. *Sensors*, 16(12):2105, 2016.

[10] Muhammad Awais, Lorenzo Chiari, Espen Alexander F. Ihlen, Jorunn L. Helbostad, and Luca Palmerini. Physical Activity Classification for Elderly People in Free-Living Conditions. *IEEE Journal of Biomedical and Health Informatics*, 23(1):197–207, jan 2019.

[11] Muhammad Awais, Lorenzo Chiari, Espen AF Ihlen, Jorunn L Helbostad, and Luca Palmerini. Classical machine learning versus deep learning for the older adults free-living activity classification. *Sensors*, 21(14):4669, 2021.

[12] Fouaz S. Ayachi, Hung P. Nguyen, Catherine Lavigne-Pelletier, Etienne Goubault, Patrick Boissy, and Christian Duval. Wavelet-based algorithm for auto-detection of daily living activities of older adults captured by multiple inertial measurement units (IMUs). *Physiological Measurement*, 37(3):442–461, feb 2016.

[13] Joonbum Bae and Masayoshi Tomizuka. A tele-monitoring system for gait rehabilitation with an inertial measurement unit and a shoe-type ground reaction force sensor. *Mechatronics*, 23(6):646–651, 2013.

[14] Hosein Bahari, Juan Forero, Jeremy C Hall, Jacqueline S Hebert, Albert H Vette, and Hossein Rouhani. Use of the extended feasible stability region for assessing stability of perturbed walking. *Scientific Reports*, 11(1):1026, 2021.

[15] Jushan Bai. Estimating multiple breaks one at a time. *Econometric theory*, 13(3): 315–352, 1997.

[16] Pierre Barralon, Nicolas Vuillerme, and Norbert Noury. Walk detection with a kinematic sensor: Frequency and wavelet comparison. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1711–1714. IEEE, 2006.

[17] Jens Barth, Cacilia Oberndorfer, Patrick Kugler, Dominik Schuldhaus, Jurgen Winkler, Jochen Klucken, and Bjorn Eskofier. Subsequence dynamic time warping as a method for robust step segmentation using gyroscope signals of daily life activities. In *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 2013*, pages 6744–6747, Osaka,Japan, jul 2013. IEEE.

[18] Jens Barth, Cäcilia Oberndorfer, Cristian Pasluosta, Samuel Schülein, Heiko Gassner, Samuel Reinfelder, Patrick Kugler, Dominik Schuldhaus, Jürgen Winkler, Jochen Klucken, et al. Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data. *Sensors*, 15(3):6419–6440, 2015.

[19] Khaireddine Ben Mansour, Philippe Gorce, and Nasser Rezzoug. The multifeature gait score: An accurate way to assess gait quality. *PLoS one*, 12(10):e0185741, 2017.

[20] Mourad Benoussaad, Benoît Sijobert, Katja Mombaur, and Christine Azevedo Coste. Robust foot clearance estimation based on the integration of foot-mounted imu acceleration data. *Sensors*, 16(1):12, 2015.

[21] Lauren C Benson, Anu M Räisänen, Christian A Clermont, and Reed Ferber. Is this the real life, or is this just laboratory? a scoping review of imu-based running gait analysis. *Sensors*, 22(5):1722, 2022.

[22] Federico Bianchi, Stephen J Redmond, Michael R Narayanan, Sergio Cerutti, and Nigel H Lovell. Barometric pressure and triaxial accelerometry-based falls event detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18 (6):619–627, 2010.

[23] Thomas Bikias, Dimitrios Iakovakis, Stelios Hadjidimitriou, Vasileios Charisis, and Leontios J Hadjileontiadis. Deepfog: an imu-based detection of freezing of gait episodes in parkinson's disease patients via deep learning. *Frontiers in Robotics and AI*, 8:537384, 2021.

[24] Carlijn VC Bouten, Karel TM Koekkoek, Maarten Verduin, Rens Kodde, and Jan D Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE transactions on biomedical engineering*, 44(3): 136–147, 1997.

[25] Leif Boysen, Angela Kempe, Volkmar Liebscher, Axel Munk, and Olaf Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics*, 37(1):157–183, 2009.

[26] Richard P Brent. *Algorithms for minimization without derivatives*. Courier Corporation, 2013.

[27] Matthew A.D. Brodie, Milou J.M. Coppens, Stephen R. Lord, Nigel H. Lovell, Yves J. Gschwind, Stephen J. Redmond, Michael Benjamin Del Rosario, Kejia Wang, Daina L. Sturnieks, Michela Persiani, and Kim Delbaere. Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. *Medical and Biological Engineering and Computing*, 54(4):663–674, apr 2016.

[28] Boris E Brodsky, Boris S Darkhovsky, Alexander Ya Kaplan, and Sergei L Shishkin. A nonparametric method for the segmentation of the eeg. *Computer methods and programs in biomedicine*, 60(2):93–106, 1999.

[29] Sjoerd M Bruijn, OG Meijer, PJ Beek, and Jaap H van Dieen. Assessing the stability of human locomotion: a review of current measures. *Journal of the Royal Society Interface*, 10(83):20120999, 2013.

[30] Bill Byrom and David A Rowe. Measuring free-living physical activity in copd patients: deriving methodology standards for clinical trials through a review of research studies. *Contemporary clinical trials*, 47:172–184, 2016.

[31] Silvia Cabral. *Gait Symmetry Measures and Their Relevance to Gait Retraining*, pages 429–447. Springer International Publishing, Cham, 2018. ISBN 978-3-319-14418-4. doi: 10.1007/978-3-319-14418-4_201. URL https://doi.org/10.1007/978-3-319-14418-4_201.

[32] Gabriela Cajamarca, Iyubanit Rodríguez, Valeria Herskovic, Mauricio Campos, and Juan Carlos Riofrío. Straightenup+: monitoring of posture during daily activities for older persons using wearable sensors. *Sensors*, 18(10):3409, 2018.

[33] Lena Carcreff, Corinna N. Gerber, Anisoara Paraschiv-Ionescu, Geraldo De Coulon, Christopher J. Newman, Kamiar Aminian, and Stéphane Armand. Comparison of gait characteristics between clinical and daily life settings in children with cerebral palsy. *Scientific Reports*, 10(1):2091, dec 2020.

[34] Alain Celisse, Guillemette Marot, Morgane Pierre-Jean, and GJ Rigaill. New efficient algorithms for multiple change-point detection with reproducing kernels. *Computational Statistics & Data Analysis*, 128:200–220, 2018.

[35] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications*, 76 (3):4405–4425, feb 2017.

[36] Jie Chen and Arjun K Gupta. Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical association*, 92(438): 739–747, 1997.

[37] Jie Chen and Arjun K Gupta. Parametric statistical change point analysis: with applications to genetics, medicine, and finance. 2012.

[38] Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu. Genetic algorithm-based classifiers fusion for multisensor activity recognition of elderly people. *IEEE Journal of Biomedical and Health Informatics*, 19(1):282–289, jan 2015.

[39] Herman Chernoff and Shelemyahu Zacks. Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35(3):999–1018, 1964.

[40] Siddhartha Chib. Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241, 1998.

[41] Nethra Ganesh Chigateri, Ngaire Kerse, Laurian Wheeler, Bruce MacDonald, and Jochen Klenk. Validation of an accelerometer for measurement of activity in frail older people. *Gait & posture*, 66:114–117, 2018.

[42] Alok Kumar Chowdhury, Dian Tjondronegoro, Vinod Chandran, and Stewart G. Trost. Ensemble Methods for Classification of Physical Activities from Wrist Accelerometry. *Medicine and Science in Sports and Exercise*, 49(9):1965–1973, sep 2017.

[43] Ian Cleland, Basel Kikhia, Chris Nugent, Andrey Boytsov, Josef Hallberg, Kåre Synnes, Sally McClean, and Dewar Finlay. Optimal placement of accelerometers for the detection of everyday activities. *Sensors (Basel, Switzerland)*, 13(7):9183–9200, jul 2013.

[44] Cynthia M. Clements, Mark J. Buller, Alexander P. Welles, and William J. Tharion. Real time gait pattern classification from chest worn accelerometry during a loaded road march. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Hong Kong, China, 2012*, volume 2012, pages 364–367, Hong Kong, China, 2012.

[45] Kirsten Corder, Esther MF van Sluijs, Antony Wright, Peter Whincup, Nicholas J Wareham, and Ulf Ekelund. Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? *The American journal of clinical nutrition*, 89(3):862–870, 2009.

[46] Patrick Crowley, Jørgen Skotte, Emmanuel Stamatakis, Mark Hamer, Mette Aadahl, Matthew L Stevens, Vegar Rangul, Paul J Mork, and Andreas Holtermann. Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: a proof-of-concept for the prospective physical activity, sitting, and sleep consortium (propass). *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):65, 2019.

[47] Federico Cruciani, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. Feature learning for human activity recognition using convolutional neural networks. *CCF Transactions on Pervasive Computing and Interaction*, 2(1):18–32, 2020.

[48] Eling D De Bruin, Antonia Hartmann, Daniel Uebelhart, Kurt Murer, and Wiebren Zijlstra. Wearable systems for monitoring mobility-related activities in older people: a systematic review. *Clinical rehabilitation*, 22(10-11):878–895, 2008.

[49] Ihana Thaís Guerra de Oliveira Gondim, Caroline de Cássia Batista de Souza, Marco Aurélio Benedetti Rodrigues, Izaura Muniz Azevedo, Maria das Graças Wanderley de Sales, Otávio Gomes Lins, et al. Portable accelerometers for the evaluation of spatio-temporal gait parameters in people with parkinson's disease: an integrative review. *Archives of Gerontology and Geriatrics*, page 104097, 2020.

[50] Silvia Del Din, Alan Godfrey, Brook Galna, Sue Lord, and Lynn Rochester. Free-living gait characteristics in ageing and Parkinson's disease: Impact of environment and ambulatory bout length. *Journal of NeuroEngineering and Rehabilitation*, 13(1): 46, may 2016.

[51] Michael B. Del Rosario, Kejia Wang, Jingjing Wang, Ying Liu, Matthew Brodie, Kim Delbaere, Nigel H. Lovell, Stephen R. Lord, and Stephen J. Redmond. A comparison of activity classification in younger and older cohorts using a smartphone. *Physiological Measurement*, 35(11):2269–2286, nov 2014.

[52] Adrian Derungs, Corina Schuster-Amft, and Oliver Amft. Longitudinal walking analysis in hemiparetic patients using wearable motion sensors: Is there convergence between body sides? *Frontiers in Bioengineering and Biotechnology*, 6(MAY): 57, may 2018.

[53] Adrian Derungs, Corina Schuster-Amft, and Oliver Amft. Physical activity comparison between body sides in hemiparetic patients using wearable motion sensors in free-living and therapy: A case series. *Frontiers in bioengineering and biotechnology*, 6:136, 2018.

[54] Cour des Comptes. Rapport sur l'application des lois de financement de la sécurité sociale, 2017.

[55] Aiden R Doherty, Paul Kelly, Jacqueline Kerr, Simon Marshall, Melody Oliver, Hannah Badland, Alexander Hamilton, and Charlie Foster. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, 10(1):22, 2013.

[56] Tristan Dot, Flavien Quijoux, Laurent Oudre, Aliénor Vienne-Jumeau, Albane Moreau, Pierre-Paul Vidal, and Damien Ricard. Non-linear template-based approach for the study of locomotion. *Sensors*, 20(7):1939, 2020.

[57] Katherine Ellis, Jacqueline Kerr, Suneeta Godbole, John Staudenmayer, and Gert Lanckriet. Hip and wrist accelerometer algorithms for free-living behavior classification. *Medicine and Science in Sports and Exercise*, 48(5):933–940, may 2016.

[58] Jill Emmerzaal, Arne De Brabandere, Yves Vanrompay, Julie Vranken, Valerie Storms, Liesbet De Baets, Kristoff Corten, Jesse Davis, Ilse Jonkers, Benedicte Vanwanseele, et al. Towards the monitoring of functional status in a free-living environment for people with hip or knee osteoarthritis: Design and evaluation of the jolo blended care app. *Sensors*, 20(23):6967, 2020.

[59] Rosana Esteller, George Vachtsevanos, Javier Echauz, and Brian Litt. A comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 48(2):177–183, 2001.

[60] Nighat Farooqi, Frode Slinde, Lena HAaglin, and Thomas Sandström. Validation of sensewear armband and actiheart monitors for assessments of daily energy expenditure in free-living women with chronic obstructive pulmonary disease. *Physiological reports*, 1(6):1–12, 2013.

[61] Yuri Feito, David R. Bassett, and Dixie L. Thompson. Evaluation of activity monitors in controlled and free-living environments. *Medicine and Science in Sports and Exercise*, 44(4):733–741, apr 2012.

[62] Frank Feldhege, Anett Mau-Moeller, Tobias Lindner, Albert Hein, Andreas Markschies, Uwe Klaus Zettl, and Rainer Bader. Accuracy of a custom physical activity and knee angle measurement sensor system for patients with neuromuscular disorders and gait abnormalities. *Sensors*, 15(5):10734–10752, 2015.

[63] Laura Fiorini, Manuele Bonaccorsi, Stefano Betti, Dario Esposito, and Filippo Cavallo. Combining wearable physiological and inertial sensors with indoor user localization network to enhance activity recognition. *Journal of Ambient Intelligence and Smart Environments*, 10(4):345–357, 2018.

[64] Anthony Fleury, Norbert Noury, and Michel Vacher. A wavelet-based pattern recognition algorithm to classify postural transitions in humans. In *2009 17th European Signal Processing Conference*, pages 2047–2051. IEEE, 2009.

[65] Anthony Fleury, Michel Vacher, and Norbert Noury. Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. *IEEE transactions on information technology in biomedicine*, 14(2):274–283, 2009.

[66] Mikaela L Frechette, Brett M Meyer, Lindsey J Tulipani, Reed D Gurchiek, Ryan S McGinnis, and Jacob J Sosnoff. Next steps in wearable technology and community ambulation in multiple sclerosis. *Current neurology and neuroscience reports*, 19(10): 80, 2019.

[67] Elliott Fullerton, Ben Heller, and Mario Munoz-Organero. Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers. *IEEE Sensors Journal*, 17(16):5290–5297, aug 2017.

[68] Lei Gao, A. K. Bourke, and John Nelson. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering and Physics*, 36(6):779–785, 2014. ISSN 18734030. doi: 10.1016/j.medengphy.2014.02.012.

[69] Daniel Garcia-Gonzalez, Daniel Rivero, Enrique Fernandez-Blanco, and Miguel R Luaces. A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors*, 20(8):2200, 2020.

[70] Rafael C González, Antonio M López, Javier Rodriguez-Uría, Diego Alvarez, and Juan C Alvarez. Real-time gait event detection for normal subjects from lower trunk accelerations. *Gait & posture*, 31(3):322–325, 2010.

[71] E Gorman, HM Hanson, PH Yang, KM Khan, T Liu-Ambrose, and MC Ashe. Accelerometry analysis of physical activity and sedentary behavior in older adults: a systematic review and data analysis. *European Review of Aging and Physical Activity*, 11(1):35–49, 2014.

[72] André Henriksen, Jonas Johansson, Gunnar Hartvigsen, Sameline Grimsgaard, and LAILA HOPSTOCK. Measuring physical activity using triaxial wrist worn polar activity trackers: A systematic review. *International Journal of Exercise Science*, 13 (4):438, 2020.

[73] Toby Hocking, Guillem Rigaill, Jean-Philippe Vert, and Francis Bach. Learning sparse penalties for change-point detection using max margin interval regression. In *International conference on machine learning*, pages 172–180. PMLR, 2013.

[74] Chengli Hou. A study on imu-based human activity recognition using deep learning and traditional machine learning. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, pages 225–234. IEEE, 2020.

[75] Chia-Yu Hsu, Yuh-Show Tsai, Cheng-Shiang Yau, Hung-Hai Shie, and Chu-Ming Wu. Differences in gait and trunk movement between patients after ankle fracture and healthy subjects. *Biomedical engineering online*, 18:1–13, 2019.

[76] Maogui Hu, Wei Li, Lianfa Li, Douglas Houston, and Jun Wu. Refining time-activity classification of human subjects using the global positioning system. *PLoS ONE*, 11 (2):e0148875, feb 2016.

[77] Milla Jauhiainen, Juha Puustinen, Saeed Mehrang, Jari Ruokolainen, Anu Holm, Antti Vehkaoja, and Hannu Nieminen. Identification of motor symptoms related to parkinson disease using motion-tracking sensors at home (käveli): Protocol for an observational case-control study. *JMIR research protocols*, 8(3):e12808, 2019.

[78] Sylvain Jung, Mona Michaud, Laurent Oudre, Eric Dorveaux, Louis Gorintin, Nicolas Vayatis, and Damien Ricard. The use of inertial measurement units for the study of free living environment activity assessment: A literature review. *Sensors*, 20(19): 5625, 2020.

[79] Sylvain Jung, Laurent Oudre, Charles Truong, Eric Dorveaux, Louis Gorintin, Nicolas Vayatis, and Damien Ricard. Adaptive change-point detection for studying human locomotion. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2020–2024. IEEE, 2021.

[80] Dean M Karantonis, Michael R Narayanan, Merryn Mathie, Nigel H Lovell, and Branko G Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE transactions on information technology in biomedicine*, 10(1):156–167, 2006.

[81] D Kelly, J Condell, J Gillespie, K Munoz Esquivel, J Barton, S Tedesco, A Nordstrom, M Åkerlund Larsson, and A Alamäki. Improved screening of fall risk using free-living based accelerometer data. *Journal of Biomedical Informatics*, page 104116, 2022.

[82] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. An online algorithm for segmenting time series. In *Proceedings 2001 IEEE international conference on data mining*, pages 289–296. IEEE, 2001.

[83] Jacqueline Kerr, Ruth E. Patterson, Katherine Ellis, Suneeta Godbole, Eileen Johnson, Gert Lanckriet, and John Staudenmayer. Objective assessment of physical activity: Classifiers for public health. *Medicine and Science in Sports and Exercise*, 48(5):951–957, may 2016.

[84] Jacqueline Kerr, Jordan Carlson, Suneeta Godbole, Lisa Cadmus-Bertram, John Bellettiere, and Sheri Hartman. Improving Hip-Worn Accelerometer Estimates of Sitting Using Machine Learning Methods. *Medicine and Science in Sports and Exercise*, 50(7):1518–1524, jul 2018.

[85] Adil Mehmood Khan, Young Koo Lee, Sungyoung Y. Lee, and Tae Seong Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1166–1172, sep 2010.

[86] Rinat Khusainov, Djamel Azzi, Ifeyinwa E Achumba, and Sebastian D Bersch. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors*, 13(10):12852–12902, 2013.

[87] Rebecca Killick, Paul Fearnhead, and Idris A Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.

[88] Raphael Knaier, Christoph Höchsmann, Denis Infanger, Timo Hinrichs, and Arno Schmidt-Trucksäss. Validation of automatic wear-time detection algorithms in a free-living setting of wrist-worn and hip-worn ActiGraph GT3X+. *BMC Public Health*, 19(1):244, feb 2019.

[89] Stanley IM Ko, Terence TL Chong, and Pulak Ghosh. Dirichlet process hidden markov multiple change-point model. *Bayesian Analysis*, 10(2):275–296, 2015.

[90] Hiromitsu Kobayashi, Wataru Kakihana, and Tasuku Kimura. Combined effects of age and gender on gait symmetry and regularity assessed by autocorrelation of trunk acceleration. *Journal of neuroengineering and rehabilitation*, 11(1):1–6, 2014.

[91] Kai Kunze and Paul Lukowicz. Sensor placement variations in wearable activity recognition. *IEEE Pervasive Computing*, 13(4):32–41, 2014.

[92] Ombeline Labaune, Thomas Deroche, Caroline Teulier, and Bastien Berret. Vigor of reaching, walking, and gazing movements: on the consistency of interindividual differences. *Journal of neurophysiology*, 123(1):234–242, 2020.

[93] Rémi Lajugie, Francis Bach, and Sylvain Arlot. Large-margin metric learning for constrained partitioning problems. In *International Conference on Machine Learning*, pages 297–305. PMLR, 2014.

[94] Marc Lavielle. Optimal segmentation of random processes. *IEEE Transactions on Signal Processing*, 46(5):1365–1373, 1998.

[95] Marc Lavielle. Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their applications*, 83(1):79–102, 1999.

[96] Marc Lavielle. Using penalized contrasts for the change-point problem. *Signal processing*, 85(8):1501–1510, 2005.

[97] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–165, 2017.

[98] Breiffni Leavy, Niklas Löfgren, Maria Nilsson, and Erika Franzén. Patient-reported and performance-based measures of walking in mild–moderate parkinson's disease. *Brain and behavior*, 8(9):e01081, 2018.

[99] Ka Yiu Lee, Duncan J. Macfarlane, and Ester Cerin. Comparison of three models of actigraph accelerometers during free living and controlled laboratory conditions. *European Journal of Sport Science*, 13(3):332–339, may 2013.

[100] Heike Leutheuser, Dominik Schuldhaus, and Bjoern M. Eskofier. Hierarchical, Multi-Sensor Based Classification of Daily Life Activities: Comparison with State-of-the-Art Algorithms Using a Benchmark Dataset. *PLoS ONE*, 8(10):e75196, oct 2013.

[101] Thurmon E. Lockhart, Rahul Soangra, Jian Zhang, and Xuefang Wu. Wavelet based automated postural event detection and activity classification with single IMU. *Biomedical Sciences Instrumentation*, 49:224–233, 2013.

[102] Gary Lorden. Procedures for reacting to a change in distribution. *The annals of mathematical statistics*, pages 1897–1908, 1971.

[103] Robert Maidstone, Toby Hocking, Guillem Rigaill, and Paul Fearnhead. On optimal multiple changepoint algorithms for large data. *Statistics and computing*, 27(2):519–533, 2017.

[104] Robert T Marcotte, Greg J Petrucci Jr, Melanna F Cox, Patty S Freedson, John W Staudenmayer, and John R Sirard. Estimating sedentary time from a hip-and wrist-worn accelerometer. *Medicine and Science in Sports and Exercise*, 52(1):225–232, 2020.

[105] Fabien Massé, Roman R. Gonzenbach, Arash Arami, Anisoara Paraschiv-Ionescu, Andreas R. Luft, and Kamiar Aminian. Improving activity recognition using a wearable barometric pressure sensor in mobility-impaired stroke patients. *Journal of NeuroEngineering and Rehabilitation*, 12(1):72, aug 2015.

[106] Rolf Moe-Nilssen and Jorunn L Helbostad. Estimation of gait cycle characteristics by trunk accelerometry. *Journal of biomechanics*, 37(1):121–126, 2004.

[107] Mareike Münch, Robert Weibel, Alexandros Sofios, Haosheng Huang, Denis Infanger, Erja Portegijs, Eleftheria Giannouli, Jonas Mundwiler, Lindsey Conrow, Taina Rantanen, Arno Schmidt-Trucksäss, Andreas Zeller, and Timo Hinrichs. MOBIlity assessment with modern TEChnology in older patients' real-life by the General Practitioner: The MOBITEC-GP study protocol. *BMC Public Health*, 19(1):1703, dec 2019.

[108] Susan L Murphy. Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct. *Preventive medicine*, 48(2):108–114, 2009.

[109] Anantha Narayanan, Lisa Mackay, and Tom Stewart. Application of machine learning in the measurement of free-living physical activity behaviours Human Potential Centre Supervisors. Technical report, Auckland University of Technology, 2019.

[110] Milad Nazarahari and Hossein Rouhani. Detection of daily postures and walking modalities using a single chest-mounted tri-axial accelerometer. *Medical Engineering and Physics*, 57:75–81, jul 2018.

[111] Hung Nguyen, Karina Lebel, Sarah Bogard, Etienne Goubault, Patrick Boissy, and Christian Duval. Using inertial sensors to automatically detect and segment activities of daily living in people with parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(1):197–204, 2017.

[112] Hung Nguyen, Karina Lebel, Patrick Boissy, Sarah Bogard, Etienne Goubault, and Christian Duval. Auto detection and segmentation of daily living activities during a Timed Up and Go task in people with Parkinson's disease using multiple inertial sensors. *Journal of NeuroEngineering and Rehabilitation*, 14(1), apr 2017.

[113] Hung Nguyen, Karina Lebel, Sarah Bogard, Etienne Goubault, Patrick Boissy, and Christian Duval. Using Inertial Sensors to Automatically Detect and Segment Activities of Daily Living in People with Parkinson's Disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(1):197–204, jan 2018.

[114] Norbert Noury, Anthony Fleury, Pierre Rumeau, Alan K Bourke, GO Laighin, Vincent Rialle, and Jean-Eric Lundy. Fall detection-principles and methods. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1663–1666. IEEE, 2007.

[115] Lauro V Ojeda, John R Rebula, Arthur D Kuo, and Peter G Adamczyk. Influence of contextual task constraints on preferred stride parameters and their variabilities during human walking. *Medical engineering & physics*, 37(10):929–936, 2015.

[116] Laurent Oudre, Rémi Barrois-Müller, Thomas Moreau, Charles Truong, Aliénor Vienne-Jumeau, Damien Ricard, Nicolas Vayatis, and Pierre-Paul Vidal. Template-based step detection with inertial measurement units. *Sensors*, 18(11):4033, 2018.

[117] ES Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527, 1955.

[118] Nikhil Panda, Ian Solsky, Emily J Huang, Stuart Lipsitz, Jason C Pradarelli, Megan Delisle, James C Cusack, Michele A Gadd, Carrie C Lubitz, John T Mullen, et al. Using smartphones to capture novel recovery metrics after cancer surgery. *JAMA surgery*, 155(2):123–129, 2020.

[119] Amy Papadopoulos, Nicolas Vivaldi, Cindy Crump, and Christine Silvers. Differentiating Walking from other Activities of Daily Living in Older Adults Using Wrist-based Accelerometers. *Current Aging Science*, 8(3):266–275, nov 2015.

[120] Toby G. Pavey, Nicholas D. Gilson, Sjaan R. Gomersall, Bronwyn Clark, and Stewart G. Trost. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *Journal of Science and Medicine in Sport*, 20(1):75–80, jan 2017.

[121] D Pérennou, A El Fatimi, M Masmoudi, C Benaim, M Loigerot, JP Didier, and J Pélissier. Incidence, circonstances et conséquences des chutes chez les patients en rééducation après un premier accident vasculaire cérébral. In *Annales de réadaptation et de médecine physique*, volume 48, pages 138–145. Elsevier, 2005.

[122] Bruno Perriot, Jerome Argod, Jean-Louis Pepin, and Norbert Noury. Characterization of physical activity in copd patients: Validation of a robust algorithm for actigraphic measurements in living situations. *IEEE journal of biomedical and health informatics*, 18(4):1225–1231, 2013.

[123] Paola Pierleoni, Sara Raggiunto, Alberto Belli, Michele Paniccia, Omid Bazgir, and Lorenzo Palma. A single wearable sensor for gait analysis in parkinson's disease: A preliminary study. *Applied Sciences*, 12(11):5486, 2022.

[124] Hari Prasanth, Miroslav Caban, Urs Keller, Grégoire Courtine, Auke Ijspeert, eike Vallery, and Joachim Von Zitzewitz. Wearable sensor-based real-time gait detection: A systematic review. *Sensors*, 21(8):2727, 2021.

[125] Zhongjun Qu and Pierre Perron. Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502, 2007.

[126] Roberto A Rabinovich, Zafeiris Louvaris, Yogini Raste, Daniel Langer, Hans Van Remoortel, Santiago Giavedoni, Chris Burtin, Eloisa MG Regueiro, Ioannis Vogiatzis, Nicholas S Hopkinson, et al. Validity of physical activity monitors during daily life in patients with copd. *European Respiratory Journal*, 42(5):1205–1215, 2013.

[127] Fabian Marcel Rast and Rob Labruyère. Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments. *Journal of NeuroEngineering and Rehabilitation*, 17(1):1–19, 2020.

[128] Rana Zia Ur Rehman, Silvia Del Din, Jian Qing Shi, Brook Galna, Sue Lord, Alison J Yarnall, Yu Guan, and Lynn Rochester. Comparison of walking protocols and gait assessment systems for machine learning-based classification of parkinson's disease. *Sensors*, 19(24):5363, 2019.

[129] Rana Zia Ur Rehman, Philipp Klocke, Sofia Hryniv, Brook Galna, Lynn Rochester, Silvia Del Din, and Lisa Alcock. Turning detection during gait: Algorithm validation and influence of sensor location and turning characteristics in the classification of parkinson's disease. *Sensors*, 20(18):5377, 2020.

[130] Alexander Richard and Juergen Gall. Temporal action detection using a statistical language model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3131–3140, 2016.

[131] Guillem Rigaill. A pruned dynamic programming algorithm to recover the best segmentations with 1 to k_{max} change-points. *Journal de la Société Française de Statistique*, 156(4):180–205, 2015.

[132] Verónica Robles-García, Yoanna Corral-Bergantiños, Nelson Espinosa, María Amalia Jácome, Carlos García-Sancho, Javier Cudeiro, and Pablo Arias. Spatiotemporal gait patterns during overt and covert evaluation in patients with parkinson's disease and healthy subjects: is there a hawthorne effect? *Journal of applied biomechanics*, 31(3):189–194, 2015.

[133] Majd Saleh and Régine Le Bouquin Jeannès. Elderly fall detection using wearable sensors: A low cost highly accurate algorithm. *IEEE Sensors Journal*, 19(8):3156–3164, 2019.

[134] Jeffer Eidi Sasaki, Amanda M. Hickey, John W. Staudenmayer, Dinesh John, Jane A. Kent, and Patty S. Freedson. Performance of activity classification algorithms in free-living older adults. *Medicine and Science in Sports and Exercise*, 48(5):941–949, may 2016.

[135] Lars Schwickert, C Becker, U Lindemann, C Maréchal, A Bourke, L Chiari, JL Helbostad, W Zijlstra, K Aminian, C Todd, et al. Fall detection with body-worn sensors. *Zeitschrift für Gerontologie und Geriatrie*, 46(8):706–719, 2013.

[136] Joseph J. Scott, Alex V. Rowlands, Dylan P. Cliff, Philip J. Morgan, Ronald C. Plotnikoff, and David R. Lubans. Comparability and feasibility of wrist- and hip-worn accelerometers in free-living adolescents. *Journal of Science and Medicine in Sport*, 20(12):1101–1106, dec 2017.

[137] Masaki Sekine, Toshiyo Tamura, Masaki Yoshida, Yuki Suda, Yuichi Kimura, Hiroaki Miyoshi, Yoshifumi Kijima, Yuji Higashi, and Toshiro Fujimoto. A gait abnormality measure based on root mean square of trunk acceleration. *Journal of neuroengineering and rehabilitation*, 10(1):1–7, 2013.

[138] Sung Yul Shin, Robert K. Lee, Patrick Spicer, and James Sulzer. Quantifying dosage of physical therapy using lower body kinematics: A longitudinal pilot study on early post-stroke individuals. *Journal of NeuroEngineering and Rehabilitation*, 17(1):15, feb 2020.

[139] Marco Sica, Salvatore Tedesco, Colum Crowe, Lorna Kenny, Kevin Moore, Suzanne Timmons, John Barton, Brendan O'Flynn, and Dimitrios-Sokratis Komaris. Continuous home monitoring of parkinson's disease using inertial sensors: A systematic review. *PloS one*, 16(2):e0246528, 2021.

[140] Marcin Straczkiewicz, Nancy W. Glynn, and Jaroslaw Harezlak. On Placement, Location and Orientation of Wrist-Worn Tri-Axial Accelerometers during Free-Living Measurements. *Sensors (Basel, Switzerland)*, 19(9), may 2019.

[141] Samuel Stuart, Lucy Parrington, Douglas N. Martini, Nicholas Kreter, James C. Chesnutt, Peter C. Fino, and Laurie A. King. Analysis of free-living mobility in people with mild traumatic brain injury and healthy controls: Quality over quantity. *Journal of Neurotrauma*, 37(1):139–145, jan 2020.

[142] Akihito Sugino, Yoshiaki Minakata, Masae Kanda, Keiichiro Akamatsu, Akira Koarai, Tsunahiko Hirano, Hisatoshi Sugiura, Kazuto Matsunaga, and Masakazu Ichinose. Validation of a compact motion sensor for the measurement of physical activity in patients with chronic obstructive pulmonary disease. *Respiration*, 83 (4):300–307, 2012.

[143] Ashley Sustakoski, Subashan Perera, Jessie M VanSwearingen, Stephanie A Studenski, and Jennifer S Brach. The impact of testing protocol on recorded gait speed. *Gait & posture*, 41(1):329–331, 2015.

[144] Yuta Tawaki, Takuichi Nishimura, and Toshiyuki Murakami. Monitoring of gait features during outdoor walking by simple foot mounted imu system. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 3413–3418. IEEE, 2020.

[145] Salvatore Tedesco, John Barton, and Brendan O'Flynn. A Review of Activity Trackers for Senior Citizens: Research Perspectives, Commercial Landscape and the Role of the Insurance Industry. *Sensors*, 17(6):1277, jun 2017.

[146] Kinh Tran, Tu Le, and Tien Dinh. A high-accuracy step counting algorithm for iPhones using Accelerometer. In *Proceedings of the International Symposium on Signal Processing and Information Technology, ISSPIT, Saigon, Vietnam, 2012*, pages 213–217, Saigon,Vietnam, 2012. IEEE.

[147] Charles Truong. *Détection de ruptures multiples–application aux signaux physiologiques.* PhD thesis, Université Paris-Saclay, 2018.

[148] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Penalty learning for changepoint detection. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1569–1573. IEEE, 2017.

[149] Charles Truong, Rémi Barrois-Müller, Thomas Moreau, Clément Provost, Aliénor Vienne-Jumeau, Albane Moreau, Pierre-Paul Vidal, Nicolas Vayatis, Stéphane Buffat, Alain Yelnik, et al. A data set for the study of human locomotion with inertial measurements units. *Image Processing On Line*, 9:381–390, 2019.

[150] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Greedy kernel change-point detection. *IEEE Transactions on Signal Processing*, 67(24):6204–6214, 2019.

[151] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Supervised kernel change point detection with partial annotations. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3147–3151. IEEE, 2019.

[152] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[153] Andrea Tura, Michele Raggi, Laura Rocchi, Andrea G Cutti, and Lorenzo Chiari. Gait symmetry and regularity in transfemoral amputees assessed by trunk accelerations. *Journal of neuroengineering and rehabilitation*, 7(1):1–10, 2010.

[154] H. Vähä-Ypyä, P. Husu, J. Suni, T. Vasankari, and H. Sievänen. Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer. *Scandinavian Journal of Medicine and Science in Sports*, 28(3):1092–1102, mar 2018.

[155] Luan Van Nguyen and Hung Manh La. Real-time human foot motion localization algorithm with dynamic speed. *IEEE Transactions on Human-Machine Systems*, 46 (6):822–833, 2016.

[156] Paolo Verdecchia, Giuseppe Schillaci, Claudia Borgioni, Antonella Ciucci, Ivano Zampi, Roberto Gattobigio, Nicola Sacchi, and Carlo Porcellati. White coat hypertension and white coat effect similarities and differences. *American journal of hypertension*, 8(8):790–798, 1995.

[157] Aliénor Vienne, Rémi P Barrois, Stéphane Buffat, Damien Ricard, and Pierre-Paul Vidal. Inertial sensors to assess gait quality in patients with neurological disorders: a systematic review of technical and analytical challenges. *Frontiers in psychology*, 8:817, 2017.

[158] Martijn Vooijs, Laurence Alpay, Jiska Snoeck-Stroband, Thijs Beerthuizen, Petra Siemonsma, Jannie Abbink, Jaap Sont, and Ton Rövekamp. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with copd, 2015.

[159] Weixin Wang and Peter Gabriel Adamczyk. Analyzing gait in the real world using wearable movement sensors and frequently repeated movement paths. *Sensors*, 19 (8):1925, 2019.

[160] Zhihua Wang, Zhaochu Yang, and Tao Dong. A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors*, 17(2):341, 2017.

[161] Aner Weiss, Sarvi Sharifi, Meir Plotnik, Jeroen PP van Vugt, Nir Giladi, and Jeffrey M Hausdorff. Toward automated, at-home assessment of mobility among patients with parkinson disease, using a body-worn accelerometer. *Neurorehabilitation and neural repair*, 25(9):810–818, 2011.

[162] Aner Weiss, Marina Brozgol, Nir Giladi, and Jeffrey M Hausdorff. Can a single lower trunk body-fixed sensor differentiate between level-walking and stair descent and ascent in older adults? preliminary findings. *Medical engineering & physics*, 38(10): 1146–1151, 2016.

[163] Che-Chang Yang and Yeh-Liang Hsu. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8):7772–7788, 2010.

[164] Shaoyan Zhang, Alex V. Rowlands, Peter Murray, and Tina L. Hurst. Physical activity classification using the GENEA wrist-worn accelerometer. *Medicine and Science in Sports and Exercise*, 44(4):742–748, apr 2012.

[165] Sijie Zhuo, Lucas Sherlock, Gillian Dobbie, Yun Sing Koh, Giovanni Russello, and Danielle Lottridge. Real-time smartphone activity classification using inertial sensors—recognition of scrolling, typing, and watching videos while sitting or walking. *Sensors*, 20(3):655, 2020.

[166] Wiebren Zijlstra and At L Hof. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait & posture*, 18(2):1–10, 2003.