



HAL
open science

Change point detection of physiological signals within in vivo settings via the FDpV method: case of the heart rate and the electrodermal activity of marathoners

Nadia Cheikh Khalfa

► **To cite this version:**

Nadia Cheikh Khalfa. Change point detection of physiological signals within in vivo settings via the FDpV method: case of the heart rate and the electrodermal activity of marathoners. Systems and Control [cs.SY]. Université Pierre et Marie Curie - Paris VI, 2015. English. NNT : 2015PA066653 . tel-01361322

HAL Id: tel-01361322

<https://theses.hal.science/tel-01361322>

Submitted on 7 Sep 2016

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

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

**Université Pierre et Marie Curie
EDITE**

Unité de Recherche Signaux et Systèmes (U2S)
Dpt. TIC-École Nationale d'Ingénieurs de Tunis (ENIT)

**Détection de ruptures de signaux physiologiques
en situation in vivo via la méthode FDpV :**

**Cas de la fréquence cardiaque
et de l'activité électrodermale de marathoniens**

Par

Nadia KHALFA

Thèse de doctorat en
Informatique, Télécom, Électronique (UPMC)

Dirigée par:

Pr. Mériem JAÏDANE & Dr. Raja GHOZI
Unité de Recherche Signaux et Systèmes (U2S-ENIT)

Pr. Pierre BERTRAND

Laboratoire de mathématiques, Université de Clermont Ferrand

Présentée et soutenue publiquement le 25 Sept. 2015

Devant un jury composé de:

BOUYEYRON, Charles	Professeur	Président
ABDELGHANI, Adnane	Professeur	Rapporteur
AZEVEDO, Christine	HDR	Rapporteur
CHAINAIS, Pierre	Maître de conférence	Examineur
GONCALVES, Paolo	HDR	Examineur
JAÏDANE, Mériem	Professeur	Encadrante
GHOZI, Raja	Maître assistante	Co-encadrante
BERTRAND, Pierre	Professeur	Encadrant

Remerciements

Ce travail de thèse a été effectué à l'Unité de recherche Signaux et Systèmes (U2S), Dpt. TIC¹ de l'École Nationale d'Ingénieurs de Tunis (ENIT), avec des stages réalisés à l'INRIA² Saclay, à l'UBIAE³ INSERM⁴ U902 et au laboratoire de mathématiques de l'université Blaise Pascal, Clermont Ferrand. Une partie du travail s'inscrit dans le cadre du projet ANR Do Well B ⁵.

Je tiens à présenter mes plus vifs et sincères remerciements à mes encadrants, Madame Mériem JAÏDANE, Professeur à l'ENIT, à Madame Raja GHOZI, Maître assistante à l'ENIT et à Monsieur Pierre BERTRAND, Professeur à l'Université de Clermont, pour leur encadrement, conseils et soutien tout au long de ce travail.

Je suis aussi reconnaissante à toutes les personnes et organismes avec qui j'ai collaboré notamment Monsieur Slaheddine DRISSI, enseignant à l'Issep⁶-Ksar Said, membre du laboratoire de recherche tunisien OPS⁷ du CNMSS⁸ et actuellement directeur technique à la QAF⁹, l'Association des parents et amis des autistes de Tunis "Pas à Pas" et le Centre Cordoue d'Education Spécialisée à Testour.

Je remercie les membres du jury de ma thèse:

- Mme Christine AZEVEDO du LIRMM¹⁰, Université Montpellier 2 et Pr. Adnane ABDELGHANI de l'INSAT¹¹, Université de Carthage pour avoir accepté de rapporter mon travail

¹Technologie de l'Information et de la Communication

²Institut National de Recherche en Informatique et en Automatique

³Unité de Biologie Intégrative des Adaptations à l'Exercice

⁴Institut national de la santé et de la recherche médicale

⁵<http://math.univ-bpclermont.fr/DoWellB/index-fr.html>

⁶Institut Supérieur de Sport et d'Education Physique

⁷Optimisation des Performances Sportives

⁸Centre National de la Médecine et Des Sciences Des Sports

⁹Qatar Athletic Federation

¹⁰Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier

¹¹l'Institut National des Sciences Appliquées et de Technologie

-
- Pr. Charles BOUVEYRON du Laboratoire MAP5¹², Université Paris Descartes, M. Paolo CONCALVES de l'ENS¹³ de Lyon et M. Pierre CHAINAIS, LAGIS¹⁴, Ecole Centrale Lille pour avoir participé à mon jury de thèse en tant que président et examinateurs.

Je remercie également le ministère de l'Enseignement supérieur Tunisien ainsi que l'Institut Français de Coopération (IFC) d'avoir financé une partie des travaux de ma thèse.

Je voudrais aussi saluer mes amis et collègues à l'U2S qui m'ont entourée et conseillée à savoir Faten BEN ALI, Olfa FRAJ, Slim ABDELBARI, Farouk MHAMDI, Houda BAHRI et Yosra MZAH. Je souhaite la bonne continuation à Bochra BOUCHHIMA, Amel AMMAMI, Nader MECHERGUI et Amira BEN JEMAA pour la suite de leurs thèses. Je remercie Guillaume PAUGAUM du laboratoire de Mathématiques de Clermont Ferrand 2 et Ahmed DHAOUADI de l'U2S pour leur soutien informatique apporté aux bons moments. Et je n'oublie pas de saluer l'adorable Sana MANSOURI pour son aide et sa présence préciseuse à l'U2S.

Par ailleurs, je voudrais remercier mes chers parents pour leur soutien et amour inconditionnels. Je salue mes soeurs, ma cousine Yosra et mon frère pour leur présence et amabilité. Je remercie ma belle famille et notamment ma belle mère et la charmante Molka pour leurs gentillesse et bienveillance.

J'en oublie certainement encore et je m'en excuse. Encore un grand merci à toutes les personnes qui m'ont inspirée et aidé à mener à bout ce travail.

Enfin, je remercie Faouzi, pour le mari qu'il est, pour son encouragement, sa compréhension, son attention et son soutien tout au long de ces années.

¹²Mathématiques Appliquées à Paris 5

¹³École Normale Supérieure

¹⁴Laboratoire d'Automatique, Génie Informatique et Signal

Contents

Introduction	1
1 Heart rate analysis through mean change point detection: Case study of marathoners	1
1.1 Introduction	1
1.2 HR analysis and change detection approach	2
1.2.1 Frequency domain analysis of HR series	3
1.2.2 Frequency and time domain analysis of HR series through change detection approach	6
1.2.2.1 Change detection on the mean	7
1.2.2.2 Change detection on the spectral density	7
1.3 Experiment and data	9
1.3.1 Subjects	10
1.3.2 Data acquisition and pre-processing	10
1.3.2.1 Case of marathon runners	10
1.3.2.2 Case of shift workers	11
1.3.3 Pre-processing HR data	11
1.4 Results	12
1.4.1 Results for marathon runners	13
1.4.2 Results for shift workers	14
1.4.3 The two cohorts study	16
1.4.3.1 Statistical study of the cohorts	18
1.4.3.2 Results	18
1.5 Discussion	19
1.6 Conclusion	20
2 Change point detection by Filtered Derivative on the mean with p-Value method (FDpV)	22
2.1 Introduction	22

2.2	FDpV for change detection: a two step procedure	23
2.3	Some FDpV's applications: examples of academic and real cases	25
2.3.1	Case of change detection on the mean	25
2.3.2	Case of change detection on the EDA signal	29
2.4	Discussion	31
2.5	Conclusion	32
3	Experimental protocol dedicated for marathoners state change assessment based on Electrodermal activity (EDA) measurement	33
3.1	In vivo pre-test experimentation with embedded sensors: Competition of Foulées du Lac	34
3.1.1	Population	34
3.1.2	Sensors for HR, EDA and respiration measurement	35
3.1.3	Faced issues	36
3.1.4	Towards a dedicated protocol for EDA measurement	38
3.2	EDA for state change detection	38
3.2.1	EDA versus HR	39
3.2.2	EDA indicators of state changes	39
3.2.3	EDA measurement	42
3.2.4	The Q sensor	44
3.3	Some possible EDA features and artefacts measurement	46
3.3.1	Motion artefact	47
3.3.2	Artefact due to mis-use of the sensor	48
3.3.2.1	Artefacts related to moved or detached electrodes	49
3.3.2.2	Missed data and zero values	50
3.3.3	Particular EDA feature: the storm	51
3.4	Comar Marathon with a well established protocol	53
3.4.1	Principle of the Comar experiment	53
3.4.2	Execution of the experience	55
3.4.3	Presentation of the protocol	56
3.4.4	Evaluation of the Comar protocol: data validation	58
3.5	Conclusion	59
4	Temporal signatures of electrodermal activity (EDA) for the evaluation of runners'performance: start and finish phases	60
4.1	Introduction	61
4.2	Protocol and data collection	61
4.2.1	Frame of the experiment	61

4.2.2	Population and classes of participants	62
4.2.3	Materials	63
4.3	Pre-processing	64
4.3.1	EDA artefacts identification	64
4.3.2	EDA special feature: The storm	65
4.4	Preliminary EDA analysis on a reference runner (P6)	65
4.5	Athletes EDA signature: the start and the finish phases	67
4.5.1	Temporal signature and EDA level of the start phase of the competition	67
4.5.1.1	Temporal signature of the start phase	67
4.5.1.2	EDA level at the start phase	68
4.5.2	Temporal signature and EDA level at the finish phase of a competition	69
4.5.2.1	Temporal signature of the finish phase	70
4.5.2.2	EDA level at the finish phase	70
4.6	Electrodermal reactions at the start and the finish phases	71
4.7	Conclusion	72
4.8	References	73
5	Tonic level and phasic activity extraction and motion artefact and special events detection	75
5.1	Introduction	76
5.2	EMD for EDA tonic level extraction	77
5.2.1	EDA tonic level extraction	77
5.2.2	EMD approach	79
5.2.3	IMF aggregation strategy for estimating EDA tonic component[73, 74]	80
5.3	EDA signal analysis: Case study of a marathon runner	81
5.3.1	EDA tonic level: extraction via EMD components aggregation	81
5.3.2	Pseudo periodic artefact detection via EMD components	84
5.4	Change point detection on IMF components of the EDA signal	87
5.5	Conclusion	90
	Appendices	97
	A Conception of an experimental protocol	98
	B Participant form	102
	C Sensor form	103
	Bibliography	104

List of Figures

- 1.1 An example of a normal ECG 2
- 1.2 Regulation of the heart rate through the two branch of the autonomous nervous system: Sympathetic ("Fight or flight" strategy) and Parasympathetic ("Rest and digest" strategy). LF band([0.04Hz, 0.15Hz]) reflects the activity of the two branches. HF band ([0.15Hz,0.4Hz]) is the response of the parasympathetic activity majorly. 5
- 1.3 *Steps of analysis* 9
- 1.4 *Polar RS 800, FI* 10
- 1.5 *A raw HR time series of a marathon runner B1, Paris Marathon 2006 (49,000 heartbeats)* 12
- 1.6 HR regulation of the marathon runner B1 in LF and HF bands. Green: the segmented LF energy computed with continuous wavelet transform. Red: segmented HF energy. For segmentation, we adopted the FDpV method . . . 13
- 1.7 HR regulation of the shift worker Y1 in LF and HF bands. Yellow: HR data of the shift worker Y1. Blue: The corresponding compressed HR data. Green: the segmented LF energy computed with continuous wavelet transform. Red: segmented HF energy. For segmentation, we adopted the FDpV method . . . 14
- 1.8 FDpV method allows details of the training of shift worker B1 namely the warming-up, the time for coach's recommendations, and the football game with two small breaks 15
- 1.9 Runners' HR index in HF and LF bands. Red points (LF,HF): effort condition (marathon race in red) are clearly separated from the green (LF,HF) points: resting condition before the race. $\log(LF/HF)$ does not provide any relevant information. 15
- 1.10 Shift workers' HR index in HF and LF bands 16
- 1.11 Shift workers and runners' HR index in HF and LF bands 18
- 1.12 HR regulation following the axis $\log LF - \log HF$ and $\log LF + \log HF$ 19

2.1	Red: A piecewise constant mean $\mu(t)$ following the change point vector τ . $\tau = [369, 3702, 3791, 3848, 4088]$ as defined in Table 2.1. Blue: The signal $S: S_t \in \mathcal{N}(\mu(t), 1)$	26
2.2	Red: the mean $\mu(t)$ as defined in Table 2.1. Black: The corresponding absolute value of the Filtered derivative $ FD_\mu(t) $. At each change point, $ FD_\mu(t) $ corresponds to a 'hat' of width $2A$; elsewhere, $FD_\mu = 0$	26
2.3	Blue: The signal $S: S_t \in \mathcal{N}(\mu(t), 1)$. Green: The corresponding absolute value of the Filtered derivative $ FD_S $. At each change point $\mu(t)$, $ FD_S $ corresponds to a 'hat' of width $2A$; elsewhere, $FD_S = 0$. Many 'hats' appear due to the white noise. Only 'the highest hats' are selected in <i>Step 1</i> of the FDpV method. In Step 2, selection is made by using a t test at each potential change point.	27
2.4	Change detection on HR series of a marathon runner. FDpV parameters: $A = 180, K_{max} = 50, pvalue = 0.01$	29
2.5	Electrodermal activity (EDA) of a semi-marathon runner during a competition	30
2.6	Change detection on the slope of an EDA signal of a semi-marathon runner. FDpV parameters: $A = 16min, Kmax = 10, p - value = 0.5$	31
3.1	Participants wearing the Q sensor	34
3.2	A participant wearing the Bioharness BT System with the help of two operators	34
3.3	The winner of the competition	35
3.4	Foulées du Lac: Used sensors and their specifications	35
3.5	A simultaneous SC record of two subjects S1(Blue) and S2(Red) of the same gender (women) in the crowded and "stressful" town center of Tunis. Difference in the EDA level of the two subjects yet being in the same conditions.	40
3.6	SC of a subject during sleep. Presence of "storms" between minutes 100 and 200. Apparent calm periods (sleep) can involve physiological arousal	41
3.7	SC of a student before (blue) and during (green) her defense of a master project	42
3.8	An example showing that changes in EDA can not always be linked to changes in temperature (skin temperature on the wrist was flat when EDA showed peaks, and there are no peaks when temperature climbs) [92].	44
3.9	(A) The used EDA sensor (Q sensor): composed of a wristband and a pod where are inserted electrodes. (B) The Q sensor electronic module. (Source: www.Affectiva.com)	45
3.10	The EDA sensor dedicated software: It offers to the user the ability to visualize different sensor's recorded signals and marking special events	46
3.11	EDA measurement of a semi-marathon runner: presence of null values and missing data	47

3.12	EDA and acceleration measurement of a semi-marathon runner: presence of oscillations due to the hand motion	48
3.13	Effect of a horizontal displacement on change detection processend on the mean of a simulated EDA measurement. $FDpVparameters : A = 1sec, K_{max} = 4, p - value = 0.1. obtainedresults = 4.10sec, 4.12sec, 12.42sec, 12.44sec$. . .	50
3.14	Missing data in about 1 sec recording of an EDA measurement of a semi-marathon runner	51
3.15	Presence of prominent rises in the EDA measurement	52
3.16	Change points obtained in the presence of storms	52
3.17	Experience form filled during the Comar protocol	57
3.18	Comar competition experimentation	58
4.1	The used electrodermal activity sensor. It is composed of a wristband and a pod where are inserted electrodes (credit: Affectiva)	63
4.2	Presence of storm and artefacts in electrodermal activity measures (EDA): case of P1 (42 km, 46 yrs old) and P3 (21 km, 26 yrs old); particularly high variability for P1 for about 10 minutes from the 50 th minute. Artefacts: “suspected” EDA measures for P3 at the 50 th minute and also from the 100 th minute till the end.	64
4.3	Measures of EDA (top), Temperature (middle) of the reference runner P6 before (Zone 1), during (Zone 2) and after (Zone 3) the competition. Key moments of the competition are indicated: $T_{start}, T_{stab}, T_{EDR}^1, T_{EDR}^2, T_{finish}$.	66
4.4	Temporal signature of the start phase: case of the reference runner P6. Three characteristic parameters $\Delta T_{stab}, EDA_{start}, EDA_{stab}$. Electrodermal reactions (EDR) are localized at T_{EDR} and correspond to the physiological activation occuring after a stimuli	68
4.5	Similarity of the temporal signature at the start phase with different $\Delta T_{stab}, EDA_{start}$ and EDA_{stab} for each runner and different T_{EDR}^1 of the restricted cohort of 21 km. Arrows show significant electrodermal reactions (EDR) between 15 min and 20 min from T_{start}	69
4.6	Electrodermal activity measure at the finish phase of the competition (semi-marathon). P3’s EDA is not presented due to the presence of many artefacts	70
4.7	An electrodermal reaction (EDR). Illustration of the different time parameters: $T_{lat}, T_{arousal}$ and $T_{\frac{1}{2}recovery}$. Source: [21]	72

5.1	10 minutes of a marathon runner EDA signal during a competition (Blue), the standard 15 sec-segments averaged EDA signal (Black), preliminary aggregation from EMD components of the EDA signal (Pink). More natural EMD result than averaged signal but need "intelligent choice"	77
5.2	Modeling an Electrodermal Reaction (EDR) by a first order system. An EDR is the results of a convolution between the stimulus $x(t)$ and the impulse response $h(t)$. The EDR figure is extracted from [21].	78
5.3	Simultaneous measurements of the runner P before, after and during the race (Sensor used on the wrist: The Q sensor). (Top) EDA recording. (Middle) Skin temperature. (Bottom) Wrist accelerations. The race period is well delimited through the accelerations. This study concerns the race period. . .	81
5.4	Results of the EMD decomposition of the EDA signal of participant P during the race (26 IMFs and a residue $r(t)$). (Top), The first four EMD components (4 IMFs). (Bottom), The last four EMD components (3 IMFs and final residue $r(t)$)	82
5.5	Correlation coefficient $\rho(i)$ between different aggregations $A(i)$ and the 30 sec-segments averaged EDA signal $y_m(t)$ for strategic aggregation for the extraction of EDA tonic level. The result is the aggregated tonic component (ATC). Maximum value of correlation with 30 sec-averaged EDA signal is obtained for $A(7)$ and $A(8)$, $i^* = 7, 8$	83
5.6	Superposition of the Aggregated Tonic Component (ATC) obtained with strategic aggregation of IMFs ($IMF_k, k = 1$ to 8) and the 20 sec-segments averaged of the EDA signal of the runner P	84
5.7	(Top), EDA signal of participant P: 13 minutes before the start and during the race. (Middle), Zoom on the 8 first seconds of the race and 8 seconds after the start. (Bottom), Power spectrum density Presence of frequency components due to the oscillations of hand motion	85
5.8	(Top), EDA signal during a simulation of a marathoner's hand motion. (Bottom), PSD of the corresponding EDA signal. Presence of frequency components due to motion oscillations	86
5.9	(Top), EDA signal of participant P without Aggregated Tonic Component (ATC) obtained by strategic aggregation of IMFs. (Middle), PSD of motion artefact (Aggregation of the first four IMF components, ($IMF_k, k = 1$ to 4)). (Bottom), PSD of the aggregated remaining IMF components ($IMF_k, k = 18$ to 5)	87

5.10	Change points detection with FDpV for different signals: original signal, filtered signal, ATC (A(8)) and A(5). FDpV parameters: $A = 2min, K_{max} = 20, p - value = 0.1$	88
5.11	Dates of change points obtained by FDpV for different signals: original signal, filtered signal, ATC (A(8)), A(5) and A(3). FDpV parameters: $A = 2min, K_{max} = 20, p - value = 0.1$	89
5.12	Similarity of change points on the original EDA signal and the filtered signal. FDpV parameters for the trend change detection: $A = 2min, K_{max} = 20, p - value = 0.1$	89
5.13	Superposition of the Original EDA signal (Blue) and the change points detected on aggregation A(5).	90
A.1	Resolution of a scientific problem [40]	98
A.2	Basics of an experimentation: does selected factors act on the studied phenomenon?	99

List of Tables

- 1.1 The shift worker’s (Y1) activities record. In this table, we denoted the diary of the day supplied by Y1 indicating the different activities he made while HR recording. 14
- 1.2 Log LF and Log HF of Marathon runners. 17
- 1.3 Log LF and Log HF of Shift workers 17

- 2.1 The values of the mean $\mu(t)$ by segment delimited by τ_i uniformly distributed ($U[1, N]$). $\tau = [369, 3702, 3791, 3848, 4088]$ 25
- 2.2 Details of p-values p_k calculation at each potential change point $\tilde{\tau}_k$ selected in Step 1 of signal $\mu(t)$ 28
- 2.3 Potential change point selected in Step 1 $\tilde{\tau}_k$ for signal S 28

- 3.1 List of participants of Foulées du Lac competition and their information (Age, Gender and Status), the sensors worn for each runner and his arrival time (Start: 8h04) 37
- 3.2 Some technical features of the Q sensor. (Source: User Guide of the Q sensor) 45
- 3.3 Possible artifacts due to the mis-use of the Q sensor: Example of moving or detaching electrodes 49
- 3.4 Factors to vary in the Comar experimental protocol 54
- 3.5 Expected results: EDA level following the factors of professionalism and competition phase for the Comar marathon 55
- 3.6 Other specifications related to the Comar experimental protocol 55

- 4.1 List of runners and specifications (gender, age, type of competition and rank). P2, P3, P5 and P6 are high level runners while P1, P4 and P7 are amateurs. NM: Not mentioned. P6 is our reference. Types of competitions: marathon (42 km), semi-marathon (21 km), 6 km 62

5.1 Selected IMF aggregations of the residue $r(t)$ and the i^* last IMFs for different averaging windows dt (sec). Selected aggregations correspond to highest correlation coefficients and are denoted by $A(i^*)$. Multiple value of i^* are possible. 83

Abbreviations

ATC	: Aggregated Tonic Component
SC	: Skin Conductance
EDA	: Electrodermal Activity
EDL	: Electrodermal level
EDR	: Electrodermal reaction
EMD	: Empirical Mode Decomposition
FDpV	: Filtered Derivative with p-value Method
FT	: Fourier Transform
HF	: High Frequency
HR	: Heart Rate
HRV	: Heart Rate Variability
LF	: Low Frequency
PSD	: Power Spectrum Density

Introduction

This thesis is in the frame of a multi-disciplinary work between protocol data acquisition and instrumentation, physiological signals' artefacts processing for the sport field and more specifically in the case of marathoners in in vivo conditions of competition. Furthermore, in order to detect instants of the physiological states changes, a dynamic approach for change-point detection was considered.

Recent technological development has offered the possibility of non invasive physiological assesment within in vivo conditions. It is actually an alternative to laboratory experiments carried with heavy protocols, huge instruments and inside laboratories or controlled environments [23]. Indeed, thanks to miniaturization of conventional physiological sensors, monitoring of physiological measurements has become possible at any life environment. The new generation of sensors is characterized by a long battery lifetime that supports weeks and months of continuous measurement by dint of low energy operation requirements. Some of them are even provided with realtime wireless transmission of the measurements. So, nowadays, we speak about wireless, small and wearable devices [19]. For instance, we can cite the polar chest band for heart rate and motion monitoring [27] and recently the Q sensor (a wrist band) for electrodermal activity measurement. Hence, applications of physiological monitoring have expanded to fields other than medicine such that personal [31], sports [18, 33] and military training [99, 3]. This permits to tackle "real" reactions within "real" situations.

On the other side, physiological signals may be an objective "inside look" into human reaction to external and internal stimuli. Indeed, it offers a look at the biological mechanisms of regulation and adaptation to the environment [29].

One of the most explored physiological measurement is the heart rate. In fact, heart rate variability (HRV) is the physiological indicator for heart rate regulation see [97] and is classically quantified through spectral energies on the two frequency bands LF ([0.04 Hz, 0.15Hz]) and HF ([0.15 Hz, 0.5Hz]) and more precisely their ratio LF/HF . It has been investigated for several case studies related to numerous diseases [37, 42, 52] and under certain conditions (sleep, rest, exercise, age, gender) [36].

Furthermore, one of the physiological measurement of human affective arousal is the EDA. It can be measured by monitoring skin conductance caused by internal or external stimuli which induces a physiological activation. This activation is an indicator of the sympathetic nervous system state [32, 81]. It can also be interpreted as the stress/motivation state of the subject[?, 47, 49].

To sum up, physiological monitoring allows to understand regulation processes, and helps to enhance the performances or/and the well being of subjects. This is what we pretend to explore in the case of runners through heart rate and EDA. And to our best knowledge, it is rarely explored in the case of marathoners ¹⁵. So, for that purpose, we try to detect state changes through these two physiological indicators in the case of marathon runners during competition. For that we adopt a method of segmentation the Filtered derivative with p-value (FDpV) method that offers a dynamic view of the data and has the advantage of low computation time and memory allocation [15].

So, it is worth to note again that the studied problematic is multidisciplinary involving applied mathematics through change detection technique, signal by dealing with artefact aspects, instrumentation through physiological sensors and also protocols and applications specific to the cohort of marathon runners. In this work, we are concerned by handling the different aspects without going in detail onto each one.

We then divide this report into five chapters.

In chapter 1, we propose to focus on heart rate data series of marathon runners recorded during the Paris marathon (2006). We aim to explore the regulation of the heart rate according to the intensity of physical activity. For this, we consider spectral energy computation in the LF and HF frequency bands as they are considered indicators of the HRV. We then propose to detect different levels of spectral energy. We then propose a locally stationary gaussian process model for mean and spectral density of the heart rate (HR) data series and adopt the Filtered Derivative with p-value method for detecting changes on the LF and HF frequency bands; to validate our approach we make a comparison with the population of shift workers carrying similar conditions of rest and relatively intense physical activity. At the end, we propose The electrodermal activity (EDA) an alternative to heart rate measurement.

In chapter 2, we put the light into the segmentation method FDpV that we adopt. We give academic and real examples for illustration of the different steps of the method for mean change detection. We then propose to process change detection on the electrodermal activity and figure out possible changes.

¹⁵www.scienceofrunning.com: a website belonging to Mr. Steve Magness, author of the book, *The Science of Running* and writer for *Running Times* magazine.

In chapter 3, we propose to deal with the EDA as a physiological indicator of motivation and stress management for runners during a competition. We then propose to lead a pre-test experimentation for deciding of the best way to carry such an experiment. We then give an insight on the EDA as a physiological signal and its components. Moreover, we describe the wireless wearable sensor adopted for EDA measurements. We will detail the faced issues and namely measurement artefacts and the challenges of change detection in this case.

In chapter 4, we propose to concentrate on the start and finish phases of a semi-marathon competition since they present key moments in the race of motivation and stress management. We try to extract some features for elite runners through temporal analysis of EDA measurement.

In chapter 5, we call the attention to the tonic component of the EDA signal of a marathoner during a competition since it reveals different states of the overall physiological arousal. We propose for that the EMD decomposition for EDA tonic component extraction with a strategy based on correlation coefficient. And for additional temporal focus, change detection with FDpV approach is then processed on the tonic component and the filtered EDA signal for a comparison purpose.

A conclusion sums up the main results and issues faced and presents important perspectives with first results. Indeed, even if we were concerned in this work with one application : the marathon runners, the measurement protocol and change detection states are re-thought and then applied to the autistic population. This is done in the frame of the ANR project “Design of Well Being Monitoring Systems (Do Well B.)” in which we deal with the physiological reactions of autistic subjects to different in-vivo environmental stimuli.

Chapter 1

Heart rate analysis through mean change point detection: Case study of marathoners

1.1 Introduction

In this work we are concerned with detection of state changes in physiological signals. Indeed, state change reflects modification in the way we react to stimulations. For instance, HR regulation expresses these modifications if we are assessing heart rate; the level of arousal, stress and motivation can also be quantified through electrodermal activity (EDA) measurement.

Hence, as an application, we propose to assess within in vivo conditions and via mobile and wireless devices the HR and EDA of marathon runners during a competition. For that, we use a method of change detection called Filtered derivative with p Value method (FDpV) introduced in [15]. So, in this chapter, we give an illustration of HR assessment of marathon runners during a race. And we propose through a new HRV index obtained by change detection on the energies of LF and HF frequency bands a classification according to physical effort intensity. This is validated in comparison with another population dealing with other levels of physical activity intensity i.e. shift workers.

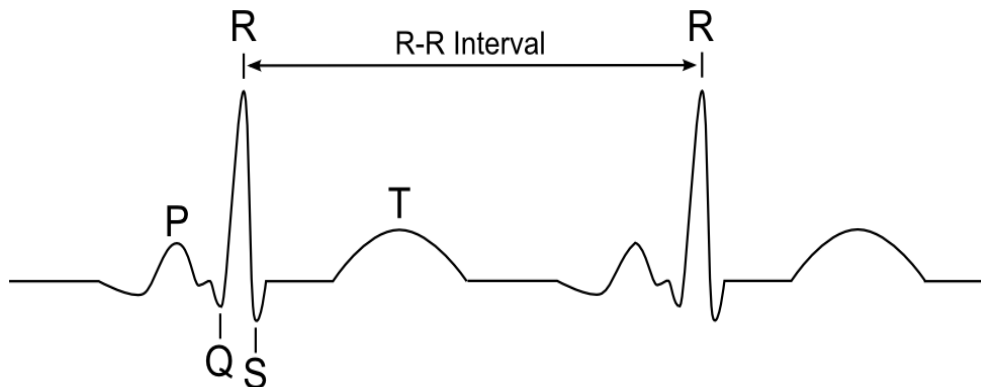
It is worth to note that in this chapter, we present a part of the study published as an article "*Heart Rate Regulation Processed Through Wavelet Analysis and Change Detection: Some Case Studies*", Acta Biotheoretica, 2012, Volume 60, Issue 1-2, pp 109-129, Springer Netherlands.

Hence, in the sequel, after an introduction to HR and change detection in section 1, we make an overview of the state of the art for modelling and processing heartbeat series in section 2. In the same section, we introduce probabilistic models for that type of time series, both in the stationary case and in the locally stationary case. In section 3, we describe the

experimental protocol used in this study. In Section 4, we summarize the results obtained by the application of our technologies on the data, completed by a short statistical study. Eventually, in Section 6 we make a discussion on the physiological meaning of these first results, and we propose future study perspectives in the conclusion.

1.2 HR analysis and change detection approach

Physical activities induce variation of the heart rate (HR), which is defined as the number of heartbeats per minute (bpm). However, devices do not measure the heart rate, but instead the time interval between two successive R-waves. Indeed, the RR interval corresponds to the duration (in seconds) of each single cardiac cycle, see Fig.1.1. The two quantities are linked by the equation $HR = 60/RR$, and in this work, we will use indifferently both notations.



Source: Prof. Rob MacLeod works:

<http://www.sci.utah.edu/~macleod/bioen/be6000/labnotes/bp/descrip.html>

Figure 1.1: An example of a normal ECG

It is worth to note that the succession of RR intervals is called a tachogram. Except in the case of some very severe diseases, a heartbeat series fluctuates. This phenomenon is known as heart rate variability (HRV). HRV has retained the attention of cardiologists, see [97] or [52] for a recent survey. Indeed, HRV reflects the regulation of HR and can contain invisible but relevant information. Therefore, heartbeat series are stochastic time dependent signals and almost as fluctuating and complicated as financial processes. One novelty of this work is the use of a theoretical probabilistic model combined with statistics for time series.

During the last two decades, many studies of HRV have been done. Some forget the dynamic structure of heartbeat series. Others analyze the dynamic structure of heartbeat series: A first family of methods was introduced during the 1990's and is based on spectral analysis. It consists in the computation of the spectral density in different bands of frequency

for instance High Frequency (HF) and Low Frequency (LF) ones by using the Fourier transform or the wavelet transform. The regulation of the HR is linked to the variation of the two quantities HF and LF , that is the variation in normalized units i.e. n.u (see [97]). An apparently alternative set of methods is based on fractal analysis, see [59] or [52] and the references therein.

All methods assume, at least implicitly, that a heartbeat series is a stochastic process X depending on a d -dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$, where Θ is a subset of the whole space. The parameter θ , which varies depending on the experimental conditions, is assumed to provide relevant informations on HR or HR regulation. Saying that the process X , also denoted by X_θ , depends on parameter θ is equivalent to say that X_θ belongs to a certain class of models $\mathcal{M} = \{X_\theta, \theta \in \Theta\}$. Next, the logical question becomes: Which stochastic model shall we use for heartbeat series? As pointed out by G. Box: *"All models are wrong, but some are useful"*, and particularly simple models are easier and therefore more useful. In care centers, hospitals or laboratories, experimental conditions are under control and can be assumed to stay fixed. As a corollary, the environment remains independent of time and a mathematical translation is provided by stationary process, namely a process whose structural parameters stay fixed during time. In order to remain simple, we will also assume that the process is Gaussian.

The simplest stochastic model is provided by locally stationary Gaussian process, introduced by [38]. *"Locally"* means the existence of a segmentation $\tau = (\tau_1, \dots, \tau_K)$ such that the process X is a Gaussian stationary process on each sub-interval (τ_k, τ_{k+1}) for $k = 0, \dots, K$, where by convention τ_0 and τ_{K+1} are respectively the initial time and the final time. Stress that K is the number of changes and can be equal to zero.

From a statistical point of view, this way of modelling addresses two sets of statistical questions: (1) Estimating the structural parameters θ of stationary Gaussian process, and (2) Fast change point analysis of these parameters knowing that recent technological progress allows recording and fast processing of large time series (40,000 or 100,000 or more).

Most of the methods suppose the stationarity of the signal, especially in time and frequency domains. The first following subsection summarizes the static methods used to quantify HRV. The necessity of the dynamic approach in this framework is highlighted in the second subsection. Eventually, in the third subsection, we propose a new method of analysis for quantifying HRV.

1.2.1 Frequency domain analysis of HR series

The most widely used methods to assess HRV are frequency domain methods. Indeed, heartbeat signal is a time series, thus we can note another structural parameter which is time dependence or correlation structure. This correlation structure varies following the

different scales of time. This could be analyzed easily through Fourier Transform. Following recommendations of the *Task Force (1996)* (see [97]), we consider Fourier Transform of tachogram and the corresponding energy in Very Low Frequency, Low Frequency and High Frequency bands, denoted in the sequel by VLF, LF, HF. The VLF corresponds to frequencies under 0.03Hz, the LF corresponds to the frequency band [0.04Hz, 0.15Hz] and the HF deals with frequencies of the frequency band [0.15Hz,0.4Hz]. Actually, the state of the art in cardiology is to consider the energy of Fourier Transform into the three frequency bands VLF, LF, HF, and the ratio HF/LF. These different frequency bands are interpreted as referring to different regulation systems: the VLF have been attributed to the renin-angiotensin system (a hormone system that regulates blood pressure and fluid balance), other humoral factors and thermoregulation, the LF is supposed to reflect the activity of the two components of the autonomous nervous system the orthosympathetic and parasympathetic while the HF band is the response of the parasympathetic activity majorly. In Fig.1.2, we can find an illustration of these two systems.

Many studies have shown the inability of these methods known as traditional measures of HRV to detect subtle changes in RR series, see *e.g.* [84].

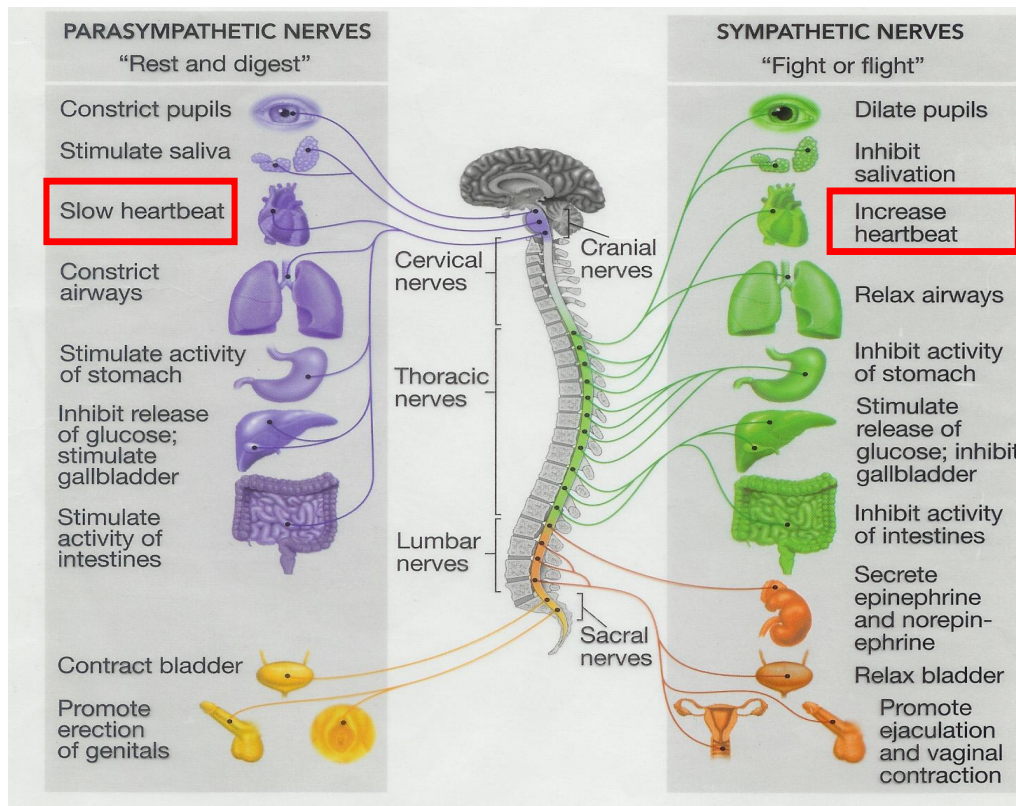
We prefer to base our statistical analysis on probabilistic modelling. The mathematical translation of a random signal with constant structural parameters is furnished by the notion of stationary process. This kind of process is well known in probability and widely used in engineering since the 1970's. Its main feature is the existence of a spectral representation ([34]). However, in this study, we restrict ourselves to Gaussian process and recall that every zero mean stationary Gaussian process admits a harmonizable representation ([101])

$$X(t) = \int_{\mathbb{R}} e^{it\xi} f^{1/2}(\xi)W(d\xi), \quad \text{for all } t \in \mathbb{R}, \quad (1.1)$$

where $W(d\xi)$ is a complex Brownian measure chosen such that X is real valued, and f is an even function called the power spectral density. By adding the mean value μ , we shift to the case of stationary Gaussian process and expression (1.1) becomes:

$$X(t) = \mu + \int_{\mathbb{R}} e^{it\xi} f^{1/2}(\xi)W(d\xi), \quad \text{for all } t \in \mathbb{R}, \quad (1.2)$$

Roughly speaking, the power spectral density corresponds to Fourier Transform of the correlation structure. In this framework, the energy in a given frequency band $[\omega_1, \omega_2]$ corresponds to the quantity $\int_{\omega_1}^{\omega_2} f(\xi)d\xi$. More precisely, $LF = \int_{0.04}^{0.15} f(\xi)d\xi$ and $HF = \int_{0.15}^{0.5} f(\xi)d\xi$. Moreover, these quantities can be estimated as the modulus of Fourier transform of the process X multiplied by the inverse Fourier transform of the rectangular spectral window localized on frequency band $(0.04 \text{ Hz}, 0.15 \text{ Hz})$, resp. $(0.15 \text{ Hz}, 0.4 \text{ Hz})$. However, using a rectangular spectral window induces a bad time localization. This latter will be a major drawback for statistical study of locally stationary process (see subsection 2.3 below).



Source: Prof. W.S. Stark works: <http://starklab.slu.edu/Physio/Autonomic.jpg> (with modification)

Figure 1.2: Regulation of the heart rate through the two branch of the autonomic nervous system: Sympathetic ("Fight or flight" strategy) and Parasympathetic ("Rest and digest" strategy). LF band([0.04Hz, 0.15Hz]) reflects the activity of the two branches. HF band ([0.15Hz,0.4Hz]) is the response of the parasympathetic activity majorly.

For this reason, in the following, we prefer to use wavelet transform rather than Fourier transform.

Recall on wavelet coefficient

Let ψ a mother wavelet with a good localization in time and frequency. Stress that there exists no function (except 0) with both finite time support and finite frequency support, but we can found mother wavelet with 0.999 of its $L^2(\mathbb{R})$ norm inside finite support, both in time and frequency. Roughly speaking, we can assume that the mother wavelet ψ has a time support $(0, L)$ and a frequency support (Λ_1, Λ_2) . Then by modulation/scaling, we can deduce two wavelets ψ_1 and ψ_2 with their frequency support include in the orthosympathetic, resp. parasympathetic frequency band ([17], [4]).

Next, for every shift $b \in \mathbb{R}$, we define the associated continuous wavelet coefficient by

$$W_{\psi_j}(b) = \int_{\mathbb{R}} \psi_j(t - b)X(t) dt \quad . \quad (1.3)$$

for $j = 1$ or 2 . Then, from Formula (1.6), we can deduce the following harmonizable representation of wavelet coefficients ([2], [6], [4]) for all $(b, b + L_j) \subset (\tau_k, \tau_{k+1})$

$$W_j(b) = \int_{\mathbb{R}} e^{ib\xi} \widehat{\psi_j}(\xi) f_k^{1/2}(\xi) dW(\xi) \quad (1.4)$$

which turns to be a zero mean complex valued Gaussian process with variance $\mathbb{E} (|W_j(b)|^2) = \int_{\mathbb{R}} |\widehat{\psi_j}(\xi)|^2 f_k(\xi) d\xi$ where $|z|$ denotes the modulus of the complex number z . Thus we have

$$\log (|W_j(b)|^2) = \log \left(\int_{\mathbb{R}} |\widehat{\psi_j}(\xi)|^2 f_k(\xi) d\xi \right) + \zeta(b) \quad (1.5)$$

where $\zeta(b)$ is a random variable with law $\ln \chi^2$, for all b . Stress that in Formula (1.5), the modulus of the Fourier transform of the wavelet $|\widehat{\psi_j}(\xi)|^2$ appears as a spectral window. Let us refer to [2] for a didactic introduction in discrete wavelet transform.

In this framework, we are concerned by a problem of change point analysis. More precisely, we are concerned by two change points problems: detecting change on the mean and detecting change on power spectral density.

1.2.2 Frequency and time domain analysis of HR series through change detection approach

The aim of this work is to detect change of the structural parameters of a heartbeat series that would mention a change in behaviour and so in biological functions. More precisely, we want to detect changes on the regulation of the heartbeats. The regulation is measured

through power spectral energy in the HF and LF bands. This real life analysis may be interesting in other cases such as athletes or in case of some pathologies needing continuous monitoring like neuropathologies. A first step in this direction is reached in Subsection 1.4.3.1, where we build an index of physiological stress derived from change detection on the spectral power density in HF and LF frequency bands.

To this end, the theoretical framework used is that provided by the notion of locally stationary process ([38]). To avoid technicality, a signal X is locally stationary if there exists time segmentation $\tau = \tau_1, \dots, \tau_{K+1}$, such that X is stationary on each interval (τ_k, τ_{k+1}) . K is the number of change points, and by convention the case $K = 0$ corresponds to stationary process with zero change points. For such locally Gaussian stationary process, we have the following spectral representation theorem

$$X(t) = \mu(t) + \int_{\mathbb{R}} e^{it\xi} f^{1/2}(t, \xi) dW(\xi), \quad \text{for all } t \in \mathbb{R}, \quad (1.6)$$

where

- the spectral density $f(t, \xi)$ is a piecewise constant function of time, i.e., there exists a partition τ_1, \dots, τ_K such that $f(t, \xi) = f_k(\xi)$ for $t \in [\tau_i, \tau_{i+1}[$
- the mean RR interval $\mu(t)$ is also piecewise constant for another partition $\tilde{\tau}_1, \dots, \tilde{\tau}_L$ with $\mu(t) = \mu_\ell$ if $t \in [\tilde{\tau}_\ell, \tilde{\tau}_{\ell+1})$.

In this framework, we are concerned by a problem of change point analysis. More precisely, we are concerned by two change points problems: detecting change on the mean and detecting change on power spectral density.

1.2.2.1 Change detection on the mean

For heartbeat series, changes on the mean are easy to detect, when they are large enough. Moreover, change point analysis on the mean is a well documented problem, see the monographs [11], or [24]. However, fast change point analysis with big dataset was a challenging problem in the beginning of 21st century, let us refer to the introduction of [15] and the references therein.

1.2.2.2 Change detection on the spectral density

On the other hand, change on the spectral density corresponds to change on hidden structural parameter, so they cannot be detected by eyes. Moreover there exist few references on the detection of changes on the spectral density, to our best knowledge only [7] or [70]. The first natural idea is to localize the static estimation procedure and then to test the existence of change points: On one hand, we need two pass band filters localized on the HF and

LF frequency bands respectively. On the other hand, we need a good time localization in order to detect time change on the spectral density. To sum up, the specification is as follows: Find a filter with both good time and frequency localization. The family of so called wavelets provide well-suited solution. In this paper, we used Gabor wavelet which corresponds to a gaussian window in both time and frequency. Due to its exponential decrease. The support of a Gabor wavelet is concentrated in the compact interval $(-3,3)$, that is “numerically”: $\|G\|_{L^2(-3,3)} = \|G\|_{L^2(\mathbb{R})}$. To sum up, wavelets are well localized both in time and in frequency and they provide well-suited tool for processing signal with a time varying frequency behaviour ([48], [70], [17], [4]).

Change point problem

The transposition from static to dynamic study, leads us to consider the change point problem of the mean of the multivariate time series $Z_1(b)$, $Z_2(b)$ where

$$Z_1(b) = \log (|W_1(b)|^2), \quad \text{and} \quad Z_2(b) = \log (|W_2(b)|^2).$$

A huge literature exists on change point problem, see the monograph [11], [24], [75], [35] or the model selection approach ([20]). But in this study, we prefer to use a simple method well suited for big datasets, that is the Filtered Derivative with p-value (FDpV) method ([15]). To put it into a nutshell, FDpV is a two step procedure, in which a so-called filtered derivative method is applied first (Step 1) and then a test is carried out to remove the false detection from the list of change points found in Step 1. This method detects changes with a precision similar to other methods, its main advantage is its complexity: Both time and memory complexity of FDpV method are of order $\mathcal{O}(n)$, where n denotes the size of the dataset. Recall that Least Squares method with or without penalization are of order $\mathcal{O}(n^2)$, and that for a marathon runner $n \simeq 40,000$ and for a day in the life of a shift worker $n \simeq 110,000$.

Calibration of FDpV method

However, we have to check the good adequation of data to the assumption that the heartbeat series is a locally stationary Gaussian process, that is to test the stationarity on each interval (τ_k, τ_{k+1}) resulting on the segmentation on wavelet coefficient. Such an adequation test can be done by using KPSS test (Kwiatkowski *et al.* 1992). Moreover, FDpV method depends on extra-parameters as the window size, the maximum number of change and the critical p-values. So, the choice of extra parameter is assumed to be correct if the KPSS test accepts more than 80% of segmentation segments.

Summary

To sum up, for modelling the heartbeat series, we choose a locally stationary gaussian process. We then use the continuous wavelet transform for calculating the spectral density energy in the HF and LF bands. Finally, for detecting changes on heart regulation, we adopt the FDpV segmentation. The choice of extra-parameters of FDpV is done in a way such that KPSS test validates more than 80% of segments. The following diagram of Fig.1.3 shows the different described steps of our analysis:

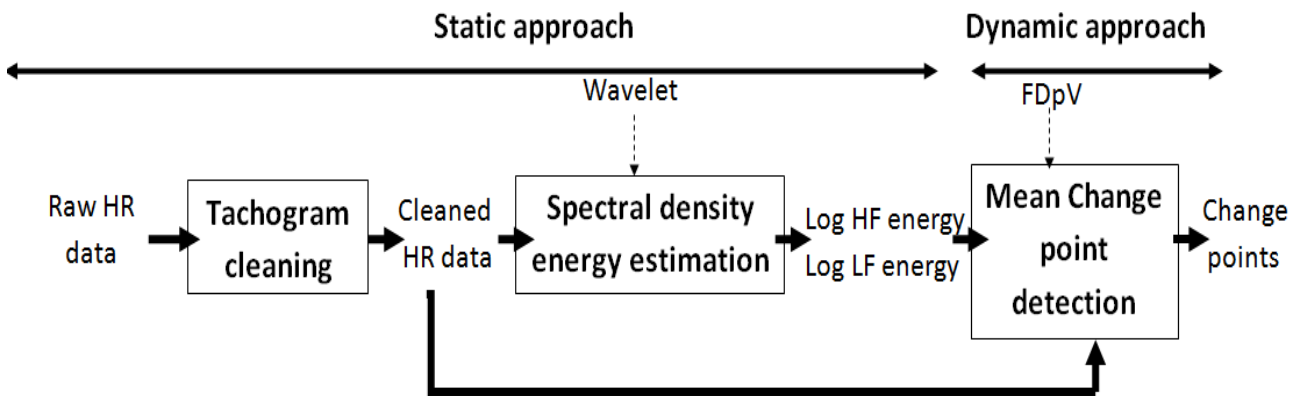


Figure 1.3: *Steps of analysis*

Flow chart summarizing steps of processing HR data for change detection points

Raw data of HR are first pre-processed through "Tachogram cleaning" application. Then, spectral energy in both bands HF and LF are computed. Finally, we search for change on the mean of the spectral energy via FDpV method. Change detection through FDpV on the mean of "cleaned" data is also processed but not presented in this diagram.

1.3 Experiment and data

HR measurements of ten runners who participated on Paris marathon 2006 were conducted by the laboratory Inserm Team U.902 and University Évrý Génopole. HR recordings of the second cohort of eleven shift workers of a compagny are supplied by our collaborators from the Institute of Industrial Hygiene, Occupational Medicine and Ergonomics, Faculty of Medicine of Clermont-Ferrand.

1.3.1 Subjects

The marathon of Paris is 42.195 km long. It lasts from 3 to 4 h and 30 min and plus. Runners need to get a medical certificate affirming their physical fitness before participating in the race. Note that the marathon race is no longer reserved to elite runners. For the greatest marathon (Boston, New York, Paris, London), the average performance is 3 h 50 min and the average marathon age is 45 years (both for the females and males) with a standard deviation of 4 years. The marathon race is run at a high percentage (80-100%) of maximal heart rate reserve HR_{max} and oxygen uptake VO_{2max} (Billat et al. 2009; Boudet et al. 2004; Douglas et al. 1987; Maron et al. 1976). On the other side, shift workers are in charge of several activities. It is mainly divided into two parts: loading and unloading of bulk. The activity is very repetitive: lifting, carrying, transporting and depositing loads.

1.3.2 Data acquisition and pre-processing

1.3.2.1 Case of marathon runners

The studied cohort of marathoners is composed of ten non elite marathon runners. Before participation, all subjects were informed of the risks and the stress associated with the protocol, and gave their written voluntary informed consent. The protocol conformed to the standards set by the Declaration of Helsinki and its procedures were approved by the local ethics committee of the Saint Louis Hospital of Paris. Heart rate signal was recorded during Paris marathon in April 2006 using a cardio-frequency meter (Polar RS 800, FI). The race began at 8:45 am and a temperature of 17°C without wind (< 2 m/s, anemometer, Windwatch, ALBA, Silva, Sweden) was observed. Two weeks prior to the race, the subjects performed a test to determine individual physical parameters (VO_{2max} and HR_{max}).



Figure 1.4: *Polar RS 800, FI*

1.3.2.2 Case of shift workers

For eleven shift workers, measures are made using Spiderview of ELA Medical (Sorin Group Company), a very compact numeric Holter recorder. It runs on batteries and can make recordings of 24 hours continuously at the workplace. It is worn on a belt or a shoulder, and is not cumbersome. Its 5 adhesive electrodes receive the signal by two derivations and insure reliable recording. Controlling the quality of the recording is possible from the installation of the electrodes and the viewing patterns on screen graphics [?]. The data are stored on a memory card. High-resolution ECGs (1000Hz, 2.5 μ V) are obtained.

1.3.3 Pre-processing HR data

Heart rate recording during exercise is not an easy work because it is realized in “free-running” conditions and not within controlled clinical conditions. Indeed, bad contact between the worker’s skin and the frequency meter added to the possible bad manipulation of the device may induce the presence of aberrant data.

It is worth noting that the size of our data series ranges from 30,000 (for marathon runners) to 150,000 (for shift workers), so it would take much time to carry on all the changes and corrections in order to get an exploitable series of data for further study. While focusing on the previous works, we found that most of them make a part of the corrections of what they call “*ectopic beats*” ([36]), “*premature beats*” or noise ([85]) manually with the help of an “*experienced observer*” ([93]). So, the use of these techniques is different from a work to another and is not practical for huge data sets. That being said, it exists programs used to analyze such recordings in hospitals which are implemented for sick people; consequently interpretations can be inappropriate or wrong in case of healthy subjects. For example, for a high level athlete, the interval between two heartbeats can easily reach 450 ms (133 bpm) which is abnormal in the case of people lying in hospital.

For that purpose, and by referring to physiological arguments, we developed a technique named “*Tachogram cleaning*”¹ capable to automatically identify artifacts and correct them in healthy subjects’ heartbeat series. “TACHOGRAM CLEANING” corrects aberrant data by referring to physiological arguments rather than statistical ones.

Let us take a look on a tachogram (recorded following specifications of Paragraph 1.3.2.1) of a runner in Fig.5.1 before processing artefacts.

We can denote in Fig.5.1 some “unexpected” data values that we designate by “artefacts” and for which it is difficult to determine the cause. The result of pre-processing data series of B1 is shown in Fig.5.7 in yellow line.

¹ “Tachogram cleaning” has been developed by Nadia Khalfa and Pierre R. Bertrand. The software was deposited in 2009.

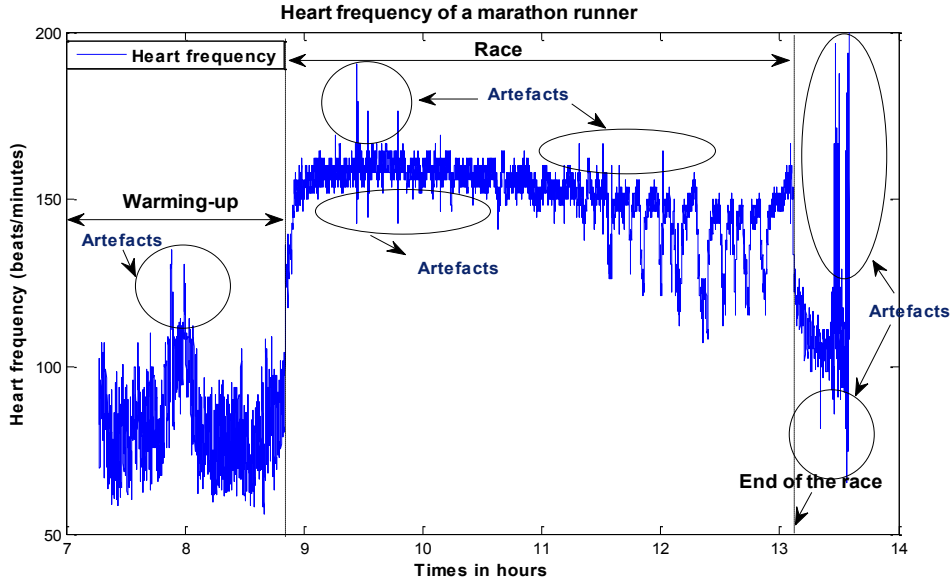


Figure 1.5: A raw HR time series of a marathon runner B1, Paris Marathon 2006 (49,000 heartbeats)

As a by-product, we can automatically measure the rate of aberrant data and then correct them. The average rate of aberrant data was around 7% for heartbeat series of marathon runner and around 1% for shift workers. Following the standard ([97] or [36]), we reject datasets with a rate of aberrant data greater than 10%. By doing so, 7 datasets of marathon runners and 11 datasets of shift workers were retained. However, only 5 shift workers have intense physical activity outside working hours, which corresponds to heart rate greater than 130 beats/minute during at least 10 minutes.

1.4 Results

Results of the analysis of HR series following the scheme of Fig.1.3 for the two cohorts (marathon runners and shift workers) are explained in the subsequent sections.

In this section, we give result of FDpV segmentation on heart rate and on log-wavelet energy into HF and LF bands for marathon runners and shift workers. Then, we present the difference between usual conditions and effort conditions, on the two cohorts, namely marathon runners and shift workers. In subsection 1.4.1 we present a paradigmatic example of a marathon runner, whereas in subsection 1.4.2 we present an example of a shift worker. Eventually, subsection 1.4.3 provides analysis of log energy effort versus resting for the two cohorts.

It is worth noting that in this study we consider only absolute values of spectral power

for LF and HF (in ms^2) and not the normalized units. Hence, the two quantities can vary in the same sense (increase and decrease together) and do not necessarily change in opposite ways as it is the case for the normalized values.

1.4.1 Results for marathon runners

As mentioned in section 1.3.3, seven marathon runners data sets out of ten were retained after the pre-processing step. Then, heart rate is segmented by FDpV method. Actually, we segment the RR interval time series by FDpV method and then we translate the results into HR's corresponding ones, by using the equation $HR = 60/RR$. We also calculate the wavelet coefficient in HF and LF bands, then the corresponding log-wavelet energies, and finally, we use FDpV method to segment them. Segmentation results of HR and LF and HF energies of the runner B1 are shown in Fig. 5.7.

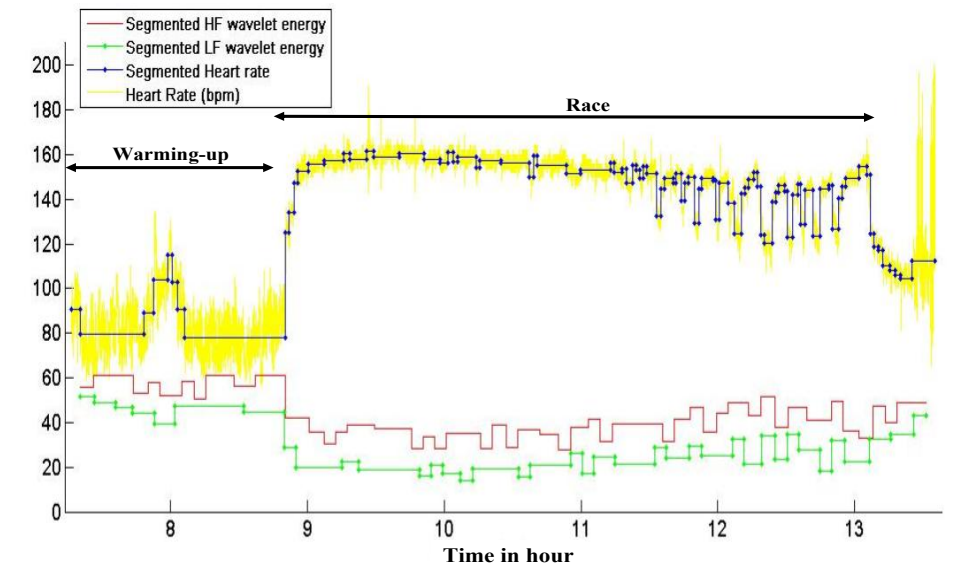


Figure 1.6: HR regulation of the marathon runner B1 in LF and HF bands. Green: the segmented LF energy computed with continuous wavelet transform. Red: segmented HF energy. For segmentation, we adopted the FDpV method

In Fig. 5.7, we can note the blue line that represents the segmented cleaned HR time series. The original signal (yellow) is compressed from around 40,000 heartbeat to 100 points. The compressed signal detects well the changes namely those occurring at the beginning of the race (just before 9 a.m) by an abrupt increase and at the end of the race by an abrupt decrease (after 1 p.m). The beginning of the race is equally very well detected in HF (green line) and LF (red line) energies by an abrupt decrease. Other changes are detected throughout the race. A notable increase in the HF energy is detected at the end of the race.

1.4.2 Results for shift workers

For the cohort of the eleven shift workers, only five ones have intense physical activity outside working hours and were retained for HRV analysis. Let us take the example of Y1 a shift worker who manually reported changes of daily activity on a diary, as shown in Table 1.1. This dataset is first preprocessed using ‘Tachogram cleaning’ to avoid the maximum of

Time	7:24-8:23	8:23-11:05	11:05-12:53	12:53-20:00	20:00-21:38	22:30-4:30
Activity	Task 1	Task 2	Picking	Free time	Football	Sleeping

Table 1.1: The shift worker’s (Y1) activities record. In this table, we denoted the diary of the day supplied by Y1 indicating the different activities he made while HR recording.

artifacts and noise. Then, wavelet coefficient in LF and HF bands were computed and FDpV segmentation was applied. For instance, Y1 has reported “*Football training*” from 20h to 21h38, which is supported by our analysis. Indeed, with FDpV method, we can see more details of the training namely the warming-up, the time for coach’s recommendations, and the football game with two small breaks, see Fig.5.8 and Fig.1.8.

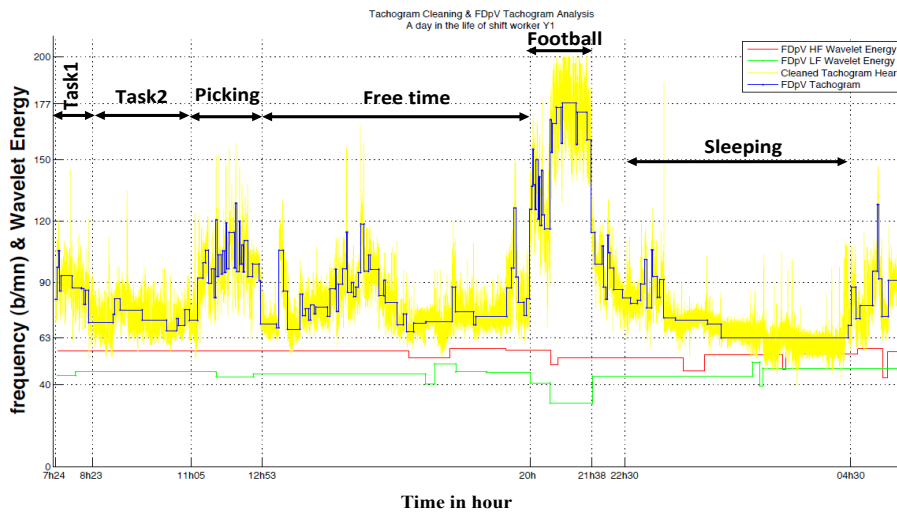


Figure 1.7: HR regulation of the shift worker Y1 in LF and HF bands. Yellow: HR data of the shift worker Y1. Blue: The corresponding compressed HR data. Green: the segmented LF energy computed with continuous wavelet transform. Red: segmented HF energy. For segmentation, we adopted the FDpV method

1.4. Results

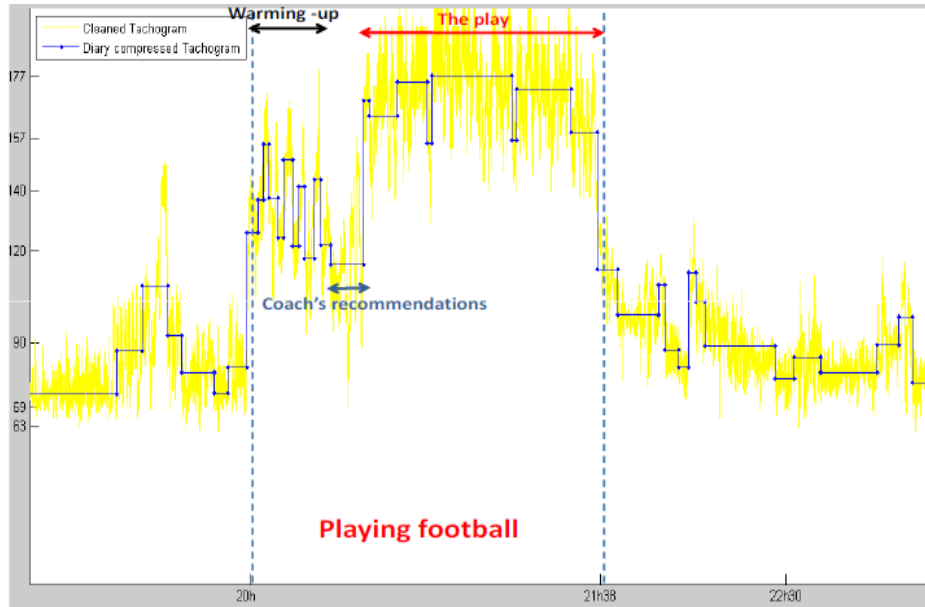


Figure 1.8: FDpV method allows details of the training of shift worker B1 namely the warming-up, the time for coach's recommendations, and the football game with two small breaks

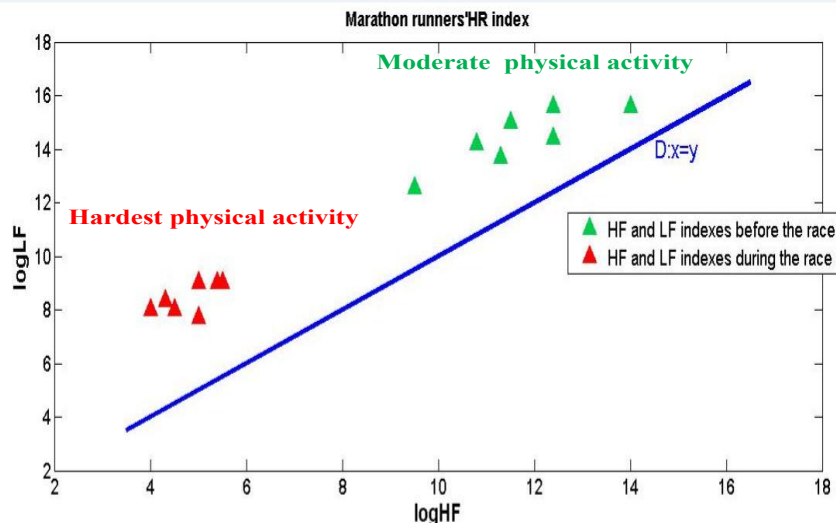


Figure 1.9: Runners' HR index in HF and LF bands. Red points (LF, HF): effort condition (marathon race in red) are clearly separated from the green (LF, HF) points: resting condition before the race. $\log(LF/HF)$ does not provide any relevant information.

Wavelet energies in *LF* and *HF* bands relative to Y1 are represented in Fig. 5.4. Note the decrease of HF wavelet energy during the football game.

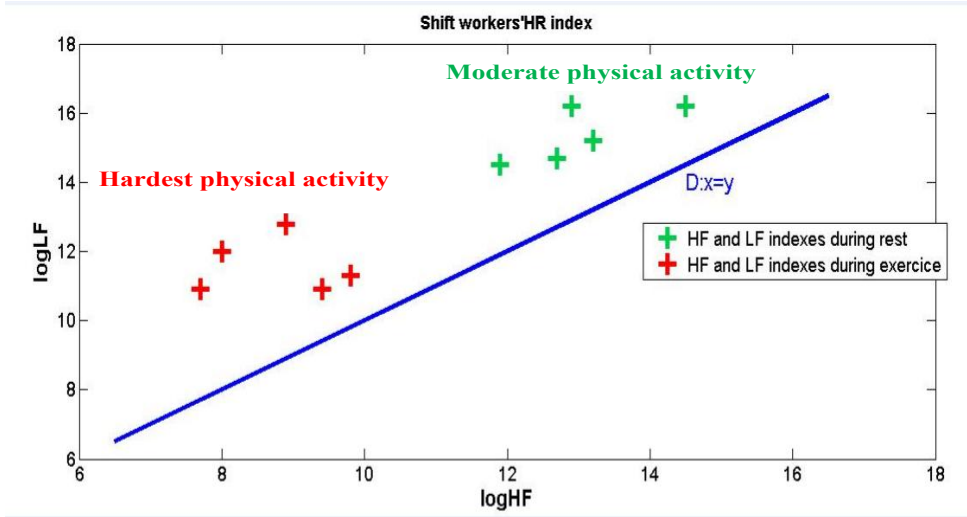


Figure 1.10: Shift workers' HR index in HF and LF bands

1.4.3 The two cohorts study

In Fig. 5.7 and Fig. 5.8, we can denote a variation of HF and LF energy following the type of activity for the marathon runner B1 and the shift worker Y1 respectively. This led us to represent the activity by a point in the plane of HF and LF energy. More precisely, we use log-log coordinates where the x-axis corresponds to HF wavelet energy and the y-axis to LF wavelet energy.

The characteristic values related to the cohort of marathon runners are gathered in Table 1.2 and presented in Fig.5.4 where each green point represents the maximum of wavelet energy reached by a runner before the beginning of the race, while each red point represents the minimum of wavelet energy reached during the race.

The distance between any point and the straight line \mathcal{D} of equation ($y = x$) or equivalently $\log LF = \log HF$ is the quantity $(y - x)$ or equivalently $\log(LF/HF)$.

As we can notice in Fig. 5.4, the (LF,HF) points corresponding to effort condition (marathon race in red) are clearly separated from the (LF,HF) points corresponding to so-called resting condition before the race (green points). The points are differentiated following an axis almost parallel to the first bissectrice \mathcal{D} with equation ($y = x$). Therefore the quantity $(y - x)$ or equivalently $\log(LF/HF)$ remains almost constant for these two sets of points. In conclusion, $\log(LF/HF)$ does not provide any relevant information, as previously noted for instance by [36].

Marathon runner	Max Log HF	Max Log LF	Min Log HF	Min Log LF
1	12.4	14.4	5	9
2	12.4	15.6	5.4	9
3	14	15.6	4	8
4	9,5	12.56	4.32	8,32
5	11.3	13.7	5,5	9
6	11.5	15	4.5	8
7	10.8	14.2	5	7.7

Table 1.2: Log LF and Log HF of Marathon runners.

Shift worker	Max Log HF	Max Log LF	Min Log HF	Min Log LF
1	12.7	14.7	7.7	10,9
2	11.9	14.5	8	12
3	14.5	16.2	8.9	12,8
4	12.9	16.2	9.8	11.3
5	13.2	15.2	9.4	10.9

Table 1.3: Log LF and Log HF of Shift workers

The same method is used for shift workers. Their characteristic values are gathered in Table 1.3 and plotted in Fig.1.10. Let us point a slight difficulty: For marathon runners, we denote a clear difference between physical activity and normal condition. For shift workers, sport activity was not included in the experimental protocol. Fortunately, we have found five shift workers with physical activity outside of working time. The best example is the shift worker Y1, who reported football training from 8 p.m to 9 p.m 38. We can notice the same difference between effort conditions (sport) and normal condition. Eventually, we can observe that there exist no difference between marathon runners and shift workers during normal or resting condition, but there exists a difference between sport condition as football

game and marathon race.

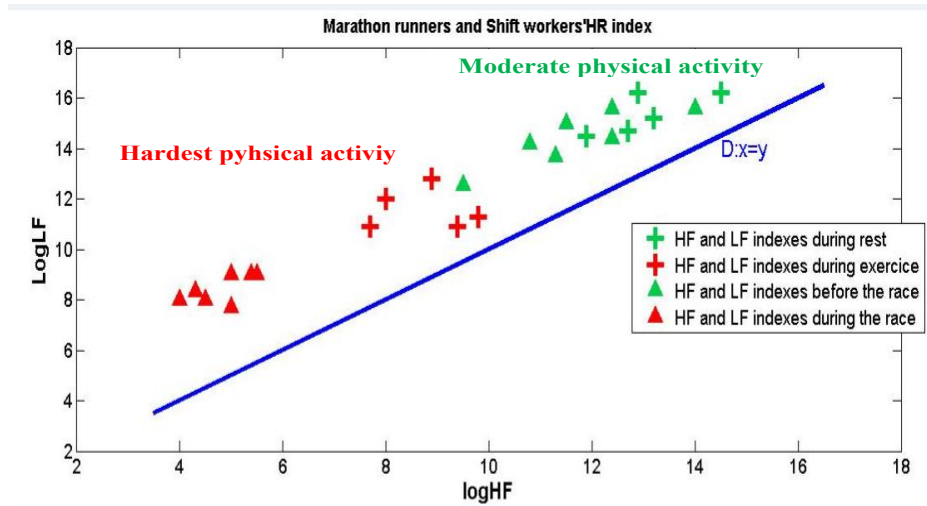


Figure 1.11: Shift workers and runners' HR index in HF and LF bands

1.4.3.1 Statistical study of the cohorts

The change in the characteristic values of the ratio $\log(LF/HF)$ before the beginning of the race and during the race i.e. $(Max \log LF - Max \log HF)$ and $(Min \log LF - Min \log HF)$ were analyzed using a T-test of student for paired data ($\alpha = 0.05$) with R software. Similarly, variations in the characteristic values of the quantity $(\log LF + \log HF)$ i.e. $(Max \log LF + Max \log HF)$ before the race and $(Min \log LF + Min \log HF)$ during the race were analyzed using the same T-test. The characteristic values of shift workers during rest and 'extreme' effort were analyzed following the same procedure.

1.4.3.2 Results

The difference between $\log(LF/HF)$ before the beginning of the race and $\log(LF/HF)$ during the race is not significant (p-value = 0.08, (Mean = 2.74, SD = 0.74) and (Mean = 3.61, SD = 0.47) respectively). Equivalently, for shift workers, this ratio is not different in rest and during high effort (p-value = 0.53, (Mean = 2.32, SD = 0.64) and (Mean = 2.82, SD = 1.23) respectively). While the quantity $(\log LF + \log HF)$ is significantly different between the beginning of the race (Mean = 26.13, SD = 2.41) and the 'extreme moment' of the race (Mean = 13.25, SD = 1.02) and even for those of shift workers with (Mean = 28.4, SD = 1.64) and (Mean = 20.34, SD = 1.18) with p-values that are well under α and respectively equal to 1.310^{-5} and 6.110^{-5} .

These results come to support visual conclusions already obtained from Fig.5.4 and Fig.1.10. Indeed, in Fig.1.12 we can see that the quantity $(\log HF + \log LF)$ increases following the first bisectrice \mathcal{D} with equation $(x = y)$. Following this direction, the points are well differentiated while the projection following the second bisectrice corresponding to the increase of quantity $(\log LF - \log HF)$ or equivalently $\log(LF/HF)$ is extremely narrow.

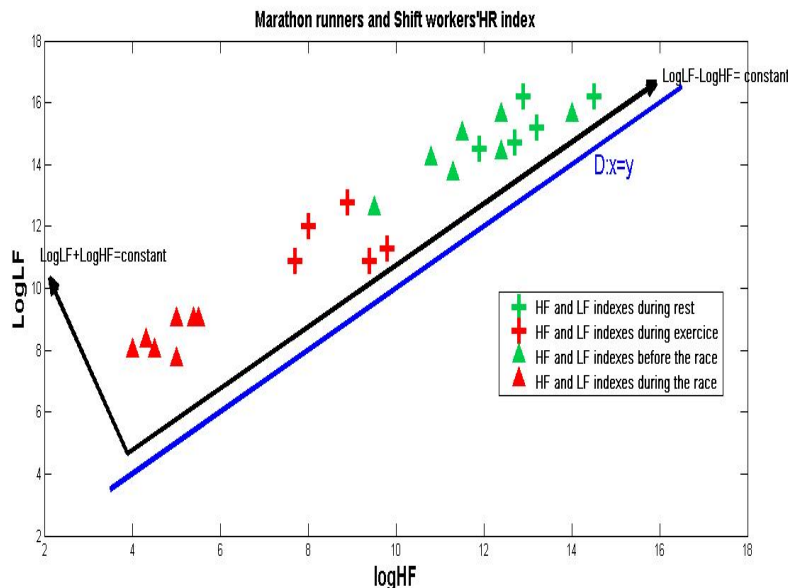


Figure 1.12: HR regulation following the axis $\log LF - \log HF$ and $\log LF + \log HF$

1.5 Discussion

The juxtaposition of the data measured at rest of two cohorts of healthy subjects (sportsmen and working people) shows the same behaviour while activities of moderate intensity (professional context) and vigorous activities (sport context) can be ordered in a logical continuum. This variation is measured through a new index $(\log LF + \log HF)$. Let us point out that the lines of equation $(\log LF + \log HF = \text{constant})$ are orthogonal to the lines of equation $(\log LF - \log HF = \text{constant})$.

The interpretation of data in this form has already opened new perspectives, complementing the classical scheme of the activation of the cardio-vascular system. Indeed, the classical scheme assumes that the regulation of the cardio-vascular system is gradually resulted from the reduction of the parasympathetic influence, then the stimulation of the sympathetic system. Spectral analysis of HRV revealed that the sympathetic system can be principally assessed through the components of the LF band while the parasympathetic activity is either reflected in the LF or both LF and HF frequency bands (see [97] and the

reference therein). However, representing the HR power spectral density of heartbeat series by just two parameters appeared as a simplistic modeling. For instance, the ratio LF/HF has been considered for some to reflect sympatho-parasympathetic balance and for others to express only sympathetic modulations. Moreover, the ratio LF/HF appeared as non a significant index for a number of studies ([84], [86], [85], [90], [36], [53]). Indeed, it is rarely significant while changes in experimental conditions occur. For example, in [84], atropine progressive injections made on 6 healthy male subjects did not show significance for the ratio LF/HF while HF energy (2nd, 4th and 8th wavelet level computation) and LF energy (32nd level) were significant. For this reason, alternatives to that 'classical' parameter were recommended such as non linear methods ([85]). However, in this work, we use statistical tools for studying the variations of the HF and LF energy for two cohorts of subjects outside hospital. It appears on the example of a marathon runner that the two components in the HF and LF bands calculated in absolute values can be strongly correlated with concomitant, parallel and fast changes at the beginning of the race. The so-called neuro vegetative regulation might well have been described in other ways and according to a chronology much finer representing an instant rather than a state.

Other measures, in different circumstances (sleep, wakefulness, physical overtraining, burnout, drug treatment with agonists and antagonists of the autonomic nervous system) and on subjects with clearly defined characteristics (fitness, age, gender, cardiac, neurological or psychiatric disease) are needed to advance the understanding and interpretation of this new informations.

1.6 Conclusion

In conclusion, in the present study we calculate wavelet energies in LF and HF bands combined with the FDpV method for segmentation of the heartbeat series. On one hand, it compresses the heartbeat signal by a factor of 1,000 while keeping important informations and detecting what may not be obvious visually (see the shift worker's example above). On the other hand, it permits to follow up with the wavelet energy of the heartbeat signal in low and high frequency bands. In light of these statistical studies of the heartbeat time series, we were able to reconsider the classical parameter LF/HF and we opt for a new one in order to assess HR regulation during physical effort.

In chapter 1, we were concerned with one signal analysis (HR); meanwhile, in the rest of this thesis we propose to deal with another physiological signal i.e. electrodermal activity (EDA) that we introduce in chapter 3. Data acquisition and protocol establishment is also discussed in the same chapter

In chapter 4, we will test the reliability of EDA for reflecting changes in arousal and

stress/motivation levels in the cohort of marathoners during a race. This being said, we propose in the following chapter to put the light onto the change detection method FDpV. We will also focus on which modifications are necessary for adapting the change detection method FDpV to the data series of EDA.

Chapter 2

Change point detection by Filtered Derivative on the mean with p-Value method (FDpV)

2.1 Introduction

In the first chapter, we propose to detect change points on the mean of datasets of HR and on the corresponding spectral density. This allows to detect state changes following the intensity of physical activity and to propose an index for cardiac regulation.

Furthermore, change point analysis has been proven relevant and widely used in financial and economic time series analysis, see e.g.[95] but also in audio and image processing [41], computer networks [96], in wind energy analysis [56] biology and medicine [82].

For statisticians, the change point problem is treated with several approaches: online and offline, for single or multiple change points, with parametric and non parametric sampling model.

A Change detection method that is based on Filtered derivative approach was introduced in [14] and [11] with memory and time complexity of order $O(n)$. In [16], a second step was added in order to eliminate false detection so-called Filtered derivative with p-Value (FDpV).

Hence, in this chapter, we detail the specificities and give academic and real examples of the FDpV. In section 2, we describe the change detection method as a two step procedure. In section 3, we give an academic example of mean change detection on a piece-wise mean constant toy model and we re-illustrate change detection on the HR signal of a marathon runner. We show that change detection on electrodermal activity (EDA) signal is rather processed in the slope. In section 4, we make a discussion on the results obtained, the parametrization made; then, we summarize in the conclusion.

2.2 FDpV for change detection: a two step procedure

In real life situation, we do not observe a continuous time process but a time series, which can be viewed as the observation of a continuous time process at discrete times, namely the observation of $X(t_i)$ for $i = 0, \dots, N$ where by notational convention $t_0 = 0$ and $t_N = T$

Hence, we deal with series X indexed by the time $t = 1, \dots, N$ with a segmentation $\tau = (\tau_1, \dots, \tau_K)$ such that X_t is a family of independent identically distributed (iid) random variables for $t \in (\tau_k, \tau_{k+1}]$ and $k = 0, \dots, K$, where by convention $\tau_0 = 0$ and $\tau_{K+1} = N$.

In [15], a method well suited for detecting changes for big datasets is proposed, that is the Filtered Derivative with p-value (FDpV).

It is a two steps procedure, in which a so-called filtered derivative is first applied in Step 1 and then, in Step 2, a test is carried out to remove the false detection from the list of change points found in Step 1.

Step 1: Selection of the potential change points Let θ be a variable extracted from the data series $(X_t)_{t=1, \dots, N}$ of length N (that can be the mean, the variance, the trend, etc.). Given the parameter θ and a sliding window of size A , we can compute the filtered derivative FD for the series X_t such that:

$$FD(k, A) = \begin{cases} \hat{\theta}(k+1, k+A) - \hat{\theta}(k-A+1, k) & \text{if } k \in [A, N-A] \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

Potential change points $\tilde{\tau}_k$ are selected as the global maximum of the absolute value of the function filtered derivative $|FD_k(t, A)|$. Then, we define the function FD_{k+1} by putting $|FD_k(t, A)|$ points lying within a distance of $\pm A$ of $\tilde{\tau}_k$ to 0. Hence, we iterate this algorithm either by reaching a fixed number of change points K_{max} (what we adopt in this work) or while $|FD_k(\tilde{\tau}_k, A)|$ is superior to a certain threshold (see[16]). Hats of width $2A$ and height $\theta_{k+1} - \theta_k$ at each change point $\tilde{\tau}_k$ are obtained. Consequently, from *Step 1*, a vector of potential change points is obtained:

$$(\tilde{\tau}_1, \tilde{\tau}_2, \dots, \tilde{\tau}_{K_{max}})$$

Step2: Elimination of false alarms After the selection of potential change points $\tilde{\tau}_k$ such that $k \in (1, \dots, K_{max})$, we test whether the parameter θ is the same for $t \in (\tau_{k-1}, \tilde{\tau}_k)$ and $t \in (\tilde{\tau}_k, \tau_{k+1})$ or not. More formally, we test for all $k \in (1, \dots, K_{max})$ the following hypothesis:

$$(H_0) : \theta_1 = \theta_2 = \dots = \theta_{N-1} = \theta_N$$

against

$$(H_1) : \exists K \geq 1 \text{ and } 0 = \tau_0 < \tau_1 < \dots < \tau_K < \tau_{K+1} = N$$

such that,

$$\theta_1 = \dots = \theta_{\tau_1} \neq \theta_{\tau_1+1} = \dots = \theta_{\tau_2} \neq \theta_{\tau_2+1} = \dots = \theta_{\tau_K} \neq \theta_{\tau_K+1} = \dots = \theta_{\tau_{K+1}}$$

and calculate the corresponding p-values: $p_1, \dots, p_{K_{max}}$ for $\tilde{\tau}_1, \dots, \tilde{\tau}_{K_{max}}$ expressing the probability of the relevance of H_0

P-value calculation in the case of change detection on the mean For P-value calculation and since we are comparing two quantitative parameters of a supposed iid gaussian distribution with unknown (equal) variances, we consider the static Student T. So, for all $k \in (1, \dots, K_{max})$

$$t_k = \frac{\hat{\theta}_k - \hat{\theta}_{k-1}}{\sqrt{\frac{S_k^2}{N_k} + \frac{S_{k-1}^2}{N_{k-1}}}} \quad (2.2)$$

S_k^2 is the variance of parameter θ

$$N_k = \tau_k - \tau_{k+1} + 1 \quad (2.3)$$

Then,

$$p_k = 2 \times (1 - St_d(|t_k|)) \quad (2.4)$$

where St_d denotes the cumulative distribution function of a Student law of degree d such that $d = N_k + N_{k-1} - 2$. For $A > 30$, the distribution of t_k is approximately Gaussian and since we are dealing with a two tailed test, we can set:

$$p_k = 2 \times (1 - \Phi(|t_k|)) \quad (2.5)$$

such that Φ is the cumulative distribution function of the zero mean standard Gaussian law.

Remark A localisation error estimation of change points can be considered, see [15] for more details

2.3 Some FDpV's applications: examples of academic and real cases

We have already seen the example of a piecewise mean change in the first chapter with heartbeat series of marathon runners and shift workers. In this section, we take the case of change detection on the mean of a piecewise constant gaussian model. In the following, we illustrate mean change detection on a toy model and on a real HR runner's data series. In a second part, we take the case of an other physiological signal : electrodermal activity (EDA) and we check change detection on it.

2.3.1 Case of change detection on the mean

Toy model The most simple example is S_t a sequence of independent standard gaussian variables of length $N = 5000$, with $S_t \in \mathcal{N}(\mu(t), 1)$ where $\mathcal{N}(\mu, \sigma)$ denotes the Gaussian law with mean μ and standard deviation $\sigma = 1$ and $t \rightarrow \mu(t)$ is a piecewise constant map such that $\mu(t) = \mu_k$ for $t \in (\tau_k, \tau_{k+1}]$ such that the mean μ follows the subsequent scheme:

$$\mu(t) = \sum_{k=0}^K \chi_{(\tau_k, \tau_{k+1}]}(t) \times \mu_k \quad (2.6)$$

We choose that τ_k are uniformly distributed in $(1, N)$ i.e $\tau_k \in \mathcal{U}(1, N)$ where $\mathcal{U}(1, N)$ denotes the Uniform law in $(1, N)$

A realization of of the vector τ_0 is:

$$\tau_0 = [369, 3702, 3791, 3848, 4088] \quad (2.7)$$

Segment	$[1, \tau_1]$	$[\tau_1, \tau_2]$	$[\tau_2, \tau_3]$	$[\tau_3, \tau_4]$	$[\tau_5, \tau_6]$	$[\tau_6, N]$
value of $\mu(t)$	3	2.5	1.5	2	3.5	3

Table 2.1: The values of the mean $\mu(t)$ by segment delimited by τ_i uniformly distributed ($\mathcal{U}[1, N]$). $\tau = [369, 3702, 3791, 3848, 4088]$

Change detection is then separately processed in μ_t and S_t by FDpV method.

FDpV's Step 1: Selection of potential change points *Step 1* consists on the calculation of the filtered derivative FD following scheme 2.1 and the selection of potential change points.

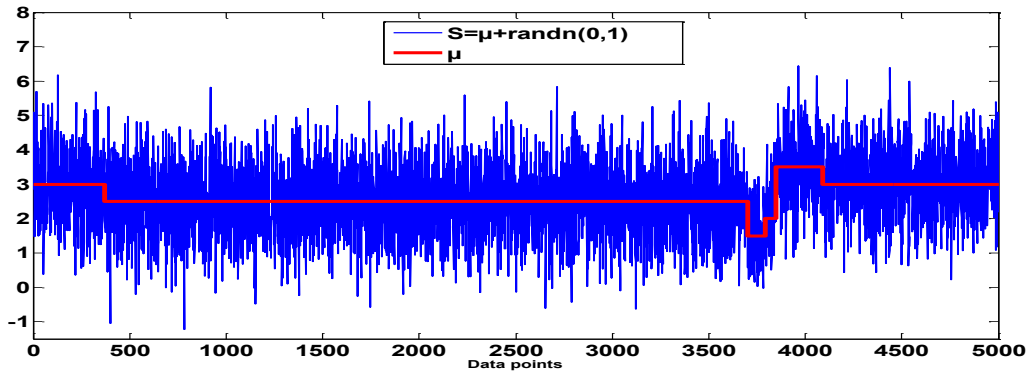


Figure 2.1: Red: A piecewise constant mean $\mu(t)$ following the change point vector τ . $\tau = [369, 3702, 3791, 3848, 4088]$ as defined in Table 2.1. Blue: The signal S : $S_t \in \mathcal{N}(\mu(t), 1)$

So, the filtered derivative $FD_\mu(t)$ for the mean μ is computed and presented in Fig. 2.2. Second, five local maximum of the absolute value of $FD_\mu(t)$ are selected. A vector of potential change points: $\tilde{\tau}$ is obtained: $\tau_0 = [369, 3702, 3791, 3848, 4088]$

The vector of change points exactly corresponds to the original vector τ_0 .

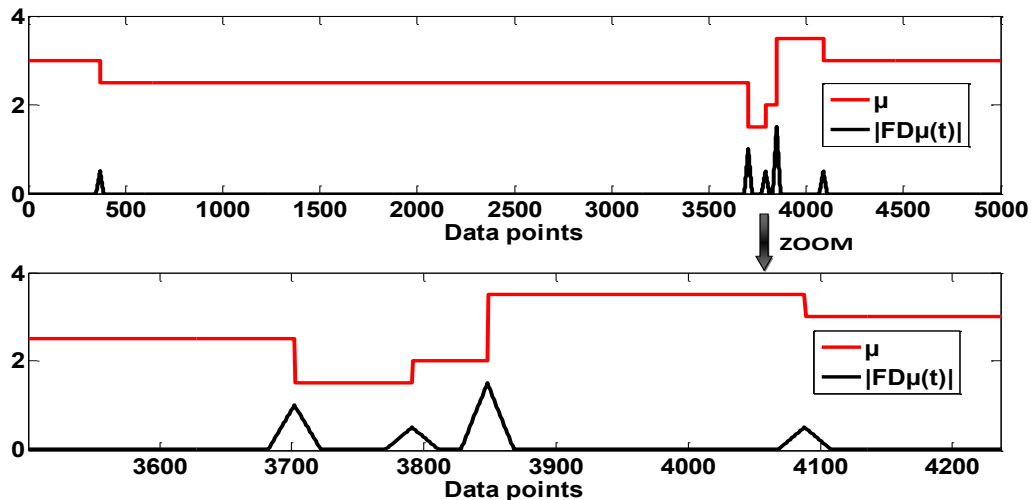


Figure 2.2: Red: the mean $\mu(t)$ as defined in Table 2.1. Black: The corresponding absolute value of the Filtered derivative $|FD_\mu(t)|$. At each change point, $|FD_\mu(t)|$ corresponds to a 'hat' of width $2A$; elsewhere, $FD_\mu = 0$

For the signal S_t , the absolute value of the filtered derivative FD_S is shown in Fig. 2.3. It has many hats compared to that of the mean μ_t due to the white noise.

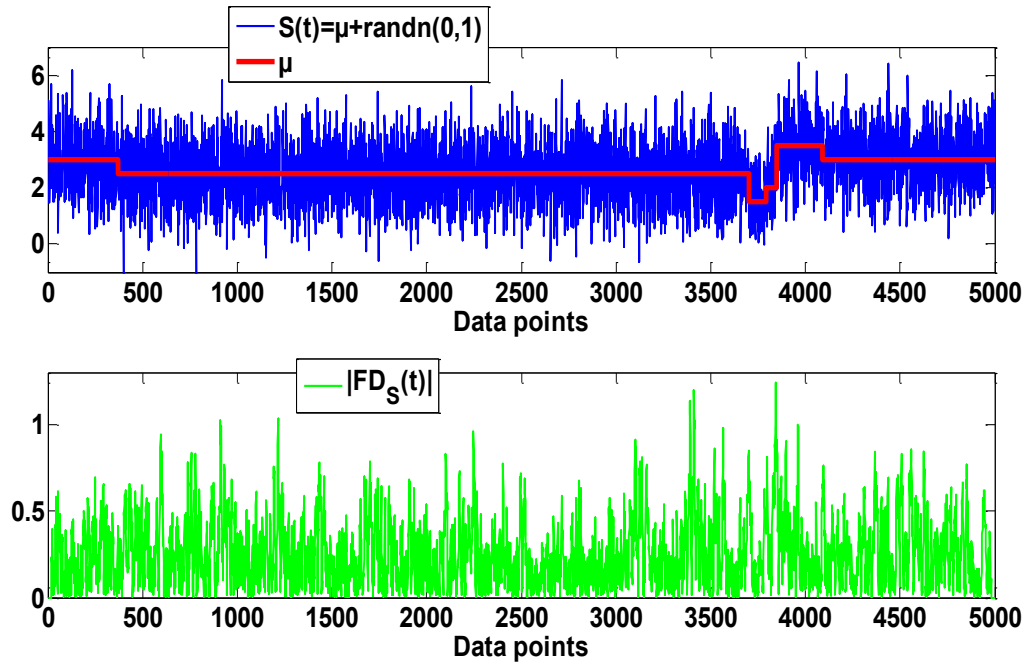


Figure 2.3: Blue: The signal S : $S_t \in \mathcal{N}(\mu(t), 1)$. Green: The corresponding absolute value of the Filtered derivative $|FD_S|$. At each change point $\mu(t)$, $|FD_S|$ corresponds to a 'hat' of width $2A$; elsewhere, $FD_S = 0$. Many 'hats' appear due to the white noise. Only 'the highest hats' are selected in *Step 1* of the FDpV method. In *Step 2*, selection is made by using a t test at each potential change point.

FDpV's *Step 2*: Elimination of false alarms After selection of the potential change points $\tilde{\tau}$ from *Step 1* for both signals μ and S , elimination of potential change points can be processed through p-value calculation as shown in Table 2.2 and Table 2.3 following scheme 2.2 .

$(\tilde{\tau}_\mu)_k$	$m_x = \theta_k - \theta_{k-1} $	$s_x = \frac{S_k^2}{N_k} + \frac{S_{k-1}^2}{N_{k-1}}$	$p_k = 2 * (1 - \phi(\frac{m_x}{\sqrt{s_x}}))$
369	0.500	2.03×10^{-7}	0
3702	0.989	41.310^{-7}	0
3791	0.480	2×10^{-4}	0
3848	1.502	2×10^{-4}	0
4088	0.006	0.0093	0.948

Table 2.2: Details of p-values p_k calculation at each potential change point $\tilde{\tau}_k$ selected in Step 1 of signal $\mu(t)$

$\tilde{\tau}_{S_k}$	$\tilde{\tau}_{S1}$	$\tilde{\tau}_{S2}$	$\tilde{\tau}_{S3}$	$\tilde{\tau}_{S4}$	$\tilde{\tau}_{S5}$
	911	1215	3390	3411	3843

Table 2.3: Potential change point selected in Step 1 $\tilde{\tau}_k$ for signal S

Most change points obtained $\tilde{\tau}_S$ in the case of the signal S are different from right ones. This is due to the presence of the white noise. Hence, the choice of extra-parameters have to be re considered.

Case of a real HR data series Here, we go back to a real example of HR data series of a marathon runner, see Fig.2.4. As for the toy model, we compute the filtered derivative (FD) of the original signal. In Fig.2.4 (bottom), we can see the temporal evolution of $|FD(t)|$. The best K_{max} are then selected. In a second step, p-values for these potential change points are computed. Then, change points with p-values under a threshold are retained. The result is shown in Fig.2.4 (Top).

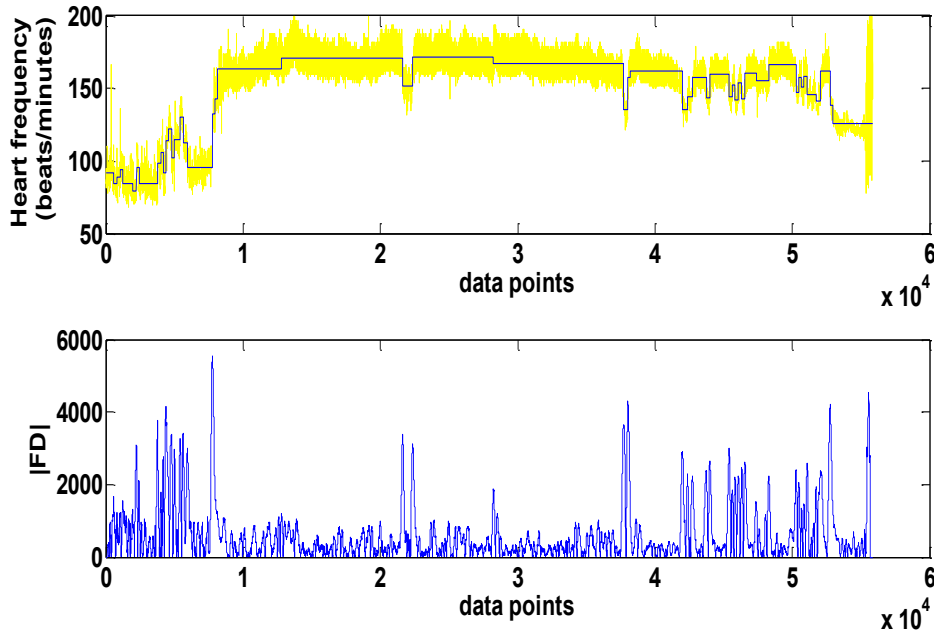


Figure 2.4: Change detection on HR series of a marathon runner. FDpV parameters: $A = 180$, $K_{max} = 50$, $pvalue = 0.01$

2.3.2 Case of change detection on the EDA signal

Now, let us take the example of another physiological signal: the electrodermal activity (EDA). The measurement we show in Fig.2.5 is that of a marathon runner during a race. We want to detect changes on this signal. We superimpose the original signal and the pre-processed one. It is important to note that a low pass FIR filter of 32 order is used for that as recommended in [81].

By looking at the signal, we can note that it is not about changes on the mean but it rather deals with changes on the slope.

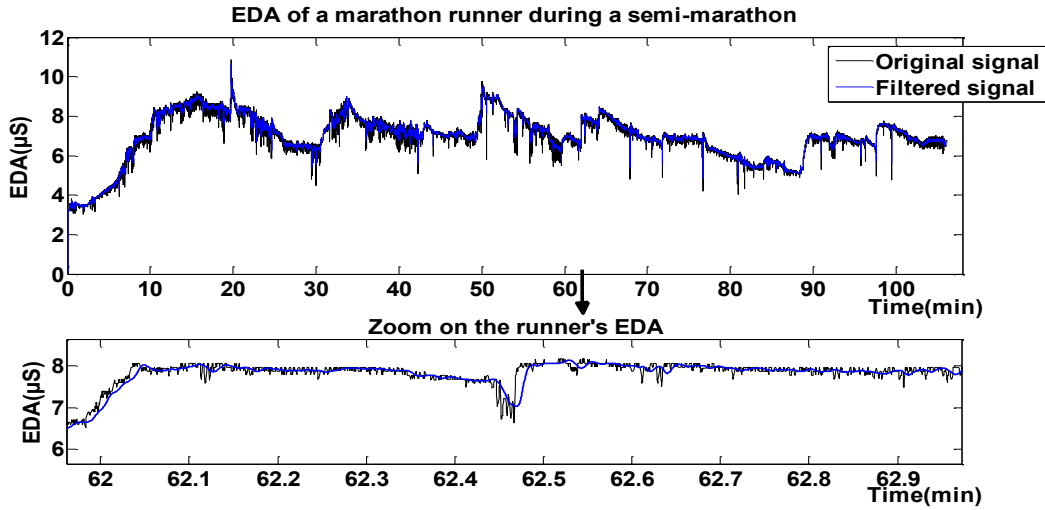


Figure 2.5: Electrodermal activity (EDA) of a semi-marathon runner during a competition

FDpV for change detection on the slope of a piecewise linear model Similarly, the idea is to consider a piece-wiselinear model: Considering the linear model:

$$Y_k = at_k + b + e_k, 1 \leq t \leq N \quad (2.8)$$

where the intercept b remains constant.

Let us recall that the filtered derivative method applied to the slope is based on the differences between estimated values of the slope computed on two sliding windows at the right and left of the index k , both of size A . For $k \in (A, N - A)$, these differences are given by:

$$|FD_k(t, A)| = \hat{a}(k, A) - \hat{a}(k - A, A) \quad (2.9)$$

such that,

$$\hat{a}(k, A) = \left[A \sum_{j=k+1}^{k+A} t_j Y_j - \sum_{j=k+1}^{k+A} t_j \sum_{j=k+1}^{k+A} Y_j \right] \left[A \sum_{j=k+1}^{k+A} t_j^2 - \left(\sum_{j=k+1}^{k+A} t_j \right)^2 \right]^{-1} \quad (2.10)$$

P-value calculation according to 2.2 remains reliable in the linear model.

For the example of the EDA measurement of the marathon runner, we process change detection on the filtered signal. Results are shown in Fig.2.6. We can note that results are not optimized. FDpV parameters have to be reviewed.

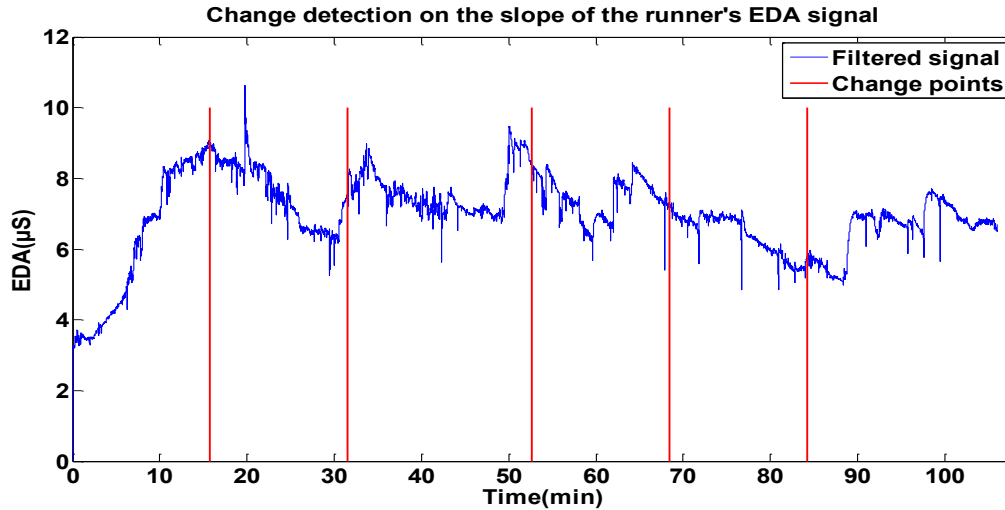


Figure 2.6: Change detection on the slope of an EDA signal of a semi-marathon runner. FDpV parameters: $A = 16min$, $K_{max} = 10$, $p - value = 0.5$

2.4 Discussion

The main advantage of FDpV is its complexity: both time and memory complexity of FDpV method are of order $\mathcal{O}(N)$, where N denotes the size of the dataset. We can cite that Least Squares method with or without penalization [] are of order $\mathcal{O}(N^2)$, and that for our case $N \simeq 50,000$ in the case of marathon runners for HR data and about 200,000 for EDA data of a semi-marathon.

On the other side, for *Step 1* of the method we have to fix a number of change points to detect K_{max} , however we don't necessarily know their number in advance. But, it can be estimated with the aid of the physiologist who can know approximately the number of changes and their "localization". The situation become similar to an academic example where change points are known.

The last point to be discussed is the iid normal model necessary for carrying such a t test at *Step 2* of the FDpV method. Data may not be normal but the CLT theorem insure convergence to the gaussian distribution. We also suppose that data are weakly dependent and that results remain valid in this case.

As for EDA change detection, we can note that an other model than that applied to HR series was needed, namely the piece-wise linear model. Moreover, extra-parameters need to be reviewed for an optimized result.

2.5 Conclusion

In this chapter, we explain the principle of the change detection method: Filtered Derivative with p Value method (FDpV) for mean change detection. We present an academic example for that. We show how it is suited for HR series and we give an application of mean change point on an example of HR as it is the case in the application of Chapter 1. We show that this is not adequate in the case of another type of physiological signal namely the EDA. Consequently, we show that FDpV method with a change on the slope is better fitted to this signal and that extra-parameters optimization is needed.

In the sequel, in chapter 3, we shed light on the features of EDA signal and the experimental protocol dedicated for marathoners EDA measurement. In chapter 4, we expose temporal analysis of these measurements. Finally, in chapter 5, we go further in preprocessing EDA signals for a more efficient change detection in relation with our application.

Chapter 3

Experimental protocol dedicated for marathoners state change assessment based on Electrodermal activity (EDA) measurement

In Chapter 1, we detected changes on the spectral density of the series of heart rate of marathoners and then we introduced a HR regulation index. HR data were recorded in France during Paris marathon in 2006 by the laboratory Inserm Team U.902 and University Évry Génopole. We then highlighted the importance of the experimental protocol and so the importance of data acquisition for a better data analysis. We also propose an other physiological measurement whose mobile applications have just emerged with the last technological development i.e. electrodermal activity (EDA).

So, at present, we want to make the measurement of the EDA on a cohort of tunisian runners. We want to stress that this application is also rarely explored. So, for leading our measurements, we choose a wireless sensor manufactured by the spin-Off of the Massachusetts Institute of Technology (MIT) interested in affective computing. It is important to note that we do not find usage of this sensor in other french laboratories. So, we find that it will be interesting to explore this new application with a very comfortable sensor.

Hence, in this chapter, we focus on the establishment of an experimental protocol in order to assess emotional engagement of runners during the start and finish phases of a race through physiological indicators and for instance through EDA.

In the sequel, we detail in section 1 a first experimentation (pre-test) led on runners during the competition of Foulées du Lac. This pre-test resulted in two important out-comings: a reflexion on the experimental protocol and an attention given to the EDA as a physiological indicator of stress/motivation. Thus, in section 2, we present the EDA and the

3.1. In vivo pre-test experimentation with embedded sensors: Competition of Foulées du Lac

reasons behind its use. In section 3, possible EDA measurement artefacts specific and not to marathon runners are illustrated and solutions of pre-processing artefacts are proposed, others are detailed in Chapter 5. Section 4 gives a description of the experience held during the Comar marathon following a proper experimental protocol. Finally, in the conclusion we summarize important outputs.

3.1 In vivo pre-test experimentation with embedded sensors: Competition of Foulées du Lac

The idea behind this experience is to focus on the physiological variation of runners during a competition through physiological measurements (electrodermal activity, respiration, heart rate, acceleration) and try to find the effect of audio visual environment assessed through video recording on physiological parameters.

3.1.1 Population

We decided to participate in the event of Foulées du Lac of 2012. It consists on a competition of semi-marathon (21 km to run) organized every year and that took place in the 12th of April in the Lac region of Tunis. We can see in Fig 5.8, 5.4 and 3.3 some photos of this event. We did not have great prior preparation due to the lack of time, though, we succeeded to make relation with interested organisms as Foulées du Lac 2012 organizers and Fédération Tunisienne d'athlétisme (FTA). This made the recruitment of subjects easier. Indeed, with their help, runners recruitment was made in the same day of the competition, about 30 minutes before the beginning. The cohort was composed of 12 runners (9 men and 3 women). We tried to get both professionals and amateurs for both men and women, juniors and seniors. It is also worth to note that the female winner of the competition is one of the runners who accepted to wear the Q sensor, see Fig.3.3. And so, she is considered as the reference runner.



Figure 3.1: Participants wearing the Q sensor



Figure 3.2: A participant wearing the Bioharness BT System with the help of two operators

3.1. In vivo pre-test experimentation with embedded sensors: Competition of Foulées du Lac



Figure 3.3: The winner of the competition

3.1.2 Sensors for HR, EDA and respiration measurement

To lead this experimentation and collect measurements of physiological signals specifically EDA, HR and respiration and also those of audio-visual environment, we had already chosen a number of sensors that have the advantage to be wireless and not cumbersome and so permit in vivo experiments. In Table 3.4, we summarize specifications of the different used sensors.







Sensor	Manufacturer	Measurements	Sampling rate	Autonomy
 <p>Qsensor</p>	 Affectiva / MIT- Media Lab, USA	<ul style="list-style-type: none"> - Skin conductance - Actigraphy (x,y,z) - Dermal temperature 	2,4,8,16,32 Hz	2 months of continuous recording at 8 Hz
 <p>BioHarness BT System</p>	 http://www.zephyranywhere.com/	<ul style="list-style-type: none"> - Heart Rate - Breathing Rate - Activity level - Subject orientation (posture) - Skin temperature - Blood Pressure (optional external sensor) - SpO₂ (optional external sensor) 	250 Hz	24 h max
 <p>Frequency meter</p>	 http://www.polar.com/fr	<ul style="list-style-type: none"> - Heart rate 	Sampling instant FC every 1, 5, 15 and 60 seconds	Recording capacity of up to 99 files

Figure 3.4: Foulées du Lac: Used sensors and their specifications

Before leading the experience, sensors have to be prepared by synchronization and by adjusting the sampling rate for some of them and recharging.

- Qsensor: Synchronization of the Q sensor happens through the configuration tool of the software of the sensor (Q software) which is connected to the computer clock. Sampling frequency must be set similarly for all subjects. It is equal to 32 Hz in this experience (we opt for the maximum of frequency to get a better temporal resolution), see [89] for more details.
- Frequency meter: It involves its own clock, see [9] for more details.
- BioHarness BT System: Synchronization happens using the configuration tool of the software of the sensor (OmniSense Analysis Module), see [19] for more details
- Spy glasses: As for this sensor, there is no indication on the time. Synchronization can occur by quick slamming hands for example.

3.1.3 Faced issues

So, we can see details of the cohort and their arrival time of each participant is noted in Table 3.1.

At the end of the competition and the collection of the different sensors, we noted a number of observations related to the experimental protocol that we summarize in the following points:

1. Not all subjects tolerate the wearing of the different sensors: Professionals who are not acquainted with the frequencymeter, the BioHarness BT System and spy glasses do not accept them in order to be as comfortable as possible to win the competition. Indeed, for the first two cited sensors the chest belt may interfere with the runner, see Fig5.4. This induces the fact that almost all professionals did wear only Q sensors that are light and easy to use. Only one professional, participant 10, accept spy glasses, see Table 5 for more details. Add to this, the logistic reason consisting on the good number of the disposable Q sensors and the shortage of the rest of the sensors. Hence, the recorded physiological signal that is more affordable is the EDA measurements.
2. Due to the misuse of some sensors (Q_4 for the participant 7 and also L_2 for participant 10) there was some failing recording.

3.1. In vivo pre-test experimentation with embedded sensors: Competition of Foulées du Lac

Participant	Age	Gender	Status*	Sensors			Arrival time
				BioHarness BT System	Q sensor	Frequency meter	
1	48	M	A	$x(Z_1)$	$x(Q_5)$		10h56
2	58	M	A	$x(Z_2)$			11h02
3		M	A	$x(Z_5)$			NM
4	61	M	A	$x(Z_3)$			10h57
5	57	M	A			$x(P_2)$	NM
6	62	M	A	$x(Z_4)$			11h45
7	24	F	P		$x(Q_4)$		Registration Failure
8	27	F	P		$x(Q_3)$		10h26
9	34	F	P		$x(Q_2)$		10h50
10	23	M	P		$x(Q_6)$		10h11
11	38	M	P	$x(Z_2)$	$x(Q_1)$		Cancellation
12	26	M	A			$x(P_1)$	NM

Table 3.1: List of participants of Foulées du Lac competition and their information (Age, Gender and Status), the sensors worn for each runner and his arrival time (Start: 8h04)

(*):A:Amateur and P:Professional. Q_i : Q sensor number i. Z_i : Bioharness Zephyr belt number i. P_i : Frequency meter number i.

3.1.4 Towards a dedicated protocol for EDA measurement

The pre-test held during the competition of Foulées du Lac lead us to draw a number of conclusions:

- The more “affordable” measurement is the EDA: we have an important number of Q sensors and runners accept easily to wear it. Whereas, the Bioharness BT System or the frequency meter are not as easy to wear (there is the chest belt to wear). Similarly, for spy glasses: their movement on one’s face can annoy the runner if there is no solution to fix it (which we tried to find before the competition). So, we choose to consider the EDA measurement for runners physiological assessment.
- The importance of a well established experimental protocol: Many aspects have to be taken into account in order to decide of the quality of obtained measurements namely the different factors influencing the result and the good questions to ask

In the light of these previous issues, we made a research on the bibliographic sources on how to conceive an experimental protocol. A summary of the important recommendations and steps for the establishment of an experimental protocol is given in Appendix 1.

Hence, ideas about the conception and leading of an experimental protocol become clearer and then we conducted a second experience: The Comar Marathon. Details of this experiment is given in section 4.

But, first, let focus on the EDA for physiological assessment and state change detection.

3.2 EDA for state change detection

In chapter 1, physiological changes on the cohort of marthoners were assessed through HRV. An indicator varying with the intensity of physical activity is then proposed. In the present chapter, we propose to deal with another physiological signal: electrodermal activity (EDA) which is an indicator of the stress/motivation state of a subject[87, 91]. We think that the two indicators can be considered complementary for the (Autonomous Nervous System) ANS functioning understanding. So, in this chapter, we choose to present the embedded sensor used for EDA measurement. This type of measurements become possible with the recent technological development. It is worth to note that this sensor is used and proved relevant in different applications such autism or driving [22, 87].

Hence, in paragraph 3.2.1, we give the physiological background for our suggestion. In paragraph 3.2.2 we explain how EDA reflects state change through its two indicators (tonic level and phasic activity) and illustrate it by examples. In the following paragraph 3.2.3, we explain the measurement technique of EDA. And in paragraph 3.2.4, we show the used sensor in our experiments.

3.2.1 EDA versus HR

Let us recall that the ANS is responsible of the unconscious functioning of the body (breathing, digestion, blood circulation, etc.). It is composed of two branches (the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS)). While SNS acts in the case of danger or emergency through the called "fight-or-flight" response, the PNS is responsible of the "rest and digest" response.

In Chapter 1, we propose a HRV indicator ($\log LF + \log HF$) which expresses HR regulation induced by the action of these two antagonistic branches as an alternative to the classical indicator LF/HF .

On the other side, EDA was reported to reflect only the activity of the SNS [32, 91]. So EDA is frequently reported to express arousal, stress and motivation[]. Thus, it is one of the proposed solution to enable people to communicate (and learn about) their ANS arousal changes as declared R.Picard at the 15th annual Future of Health Technology Summit.

To sum up, EDA is specific to one aspect of our brain (nervous system) functioning governed by the SNS. Hence, in this work we justify this choice of physiological measurement and its reliability. Let us note that EDA and skin conductance (SC) are sometimes used interchangeably [47]. This will be clarified in paragraph 3.2.3.

3.2.2 EDA indicators of state changes

EDA or SC is generally assessed through its two components: skin conductance level (SCL) or tonic level and skin conductance responses (SCR's) or phasic activity [21]. Tonic level generally refers to the signal's slow variation or global tendency, while phasic activity reflects the rapid changes on the EDA/SC signal. In the following paragraphs, we give a bibliographic overview of these two components illustrated by some examples.

Skin conductance level (SCL) or tonic level SCL describes the overall conductivity of the skin over longer time intervals, typically ranging from tens of seconds to tens of minutes. Indeed, tonic changes in the SCL typically occur in that time intervals [47].

SCL is typically obtained by taking an average of several discrete measurement points distributed across the time window of interest [22]. These measurement points should not be taken during a SCR (as this would lead to an overestimation of SCL). This is due to the fact that SCL or tonic level is generally considered to be the level of SC in the absence of any particular discrete environmental event or external stimuli.

In contrast to the SCRs, it is assumed that the SCL is not directly related to stimuli. But it is indicative of a more general level of arousal or evaluates the overall physiological arousal. In other references, it is designated as a base signal, see [31].

3.2. EDA for state change detection

Tonic level can slowly vary over time in an individual depending upon his or her psychological state (skin blood flow, skin temperature...), hydration, skin dryness, autonomic activation and fractionally certain demographic characteristics (age, gender, ethnic or heritability) [21].

For SCL's physiological interpretation, we find the following in Affectiva documentations:

Stress or interest or excitement or any activity requiring increased level of emotion, cognition, or brain activity is reflected physiologically by an arousal/activation or responsiveness/attentiveness. This can be assessed on the SC by a higher than usual baseline or level.

However, a low level can mean relaxation, and no activation, whether good or bad (calm and at peace, or disengaged/bored).

As an example, let us see in Fig.3.5 the SC of two subjects S1 (Blue) and S2 (Red). Measurements were recorded while the two subjects walked in the crowded and stressful town center of Tunis. Even if variability of the two signals is similar to each other but SCL is constantly different: that of S1 is higher than that of S2 and practically with the same difference except for the period between 20 and 40 min of the journey. This refers to the concept of individual arousal baseline and level of physiological activation.

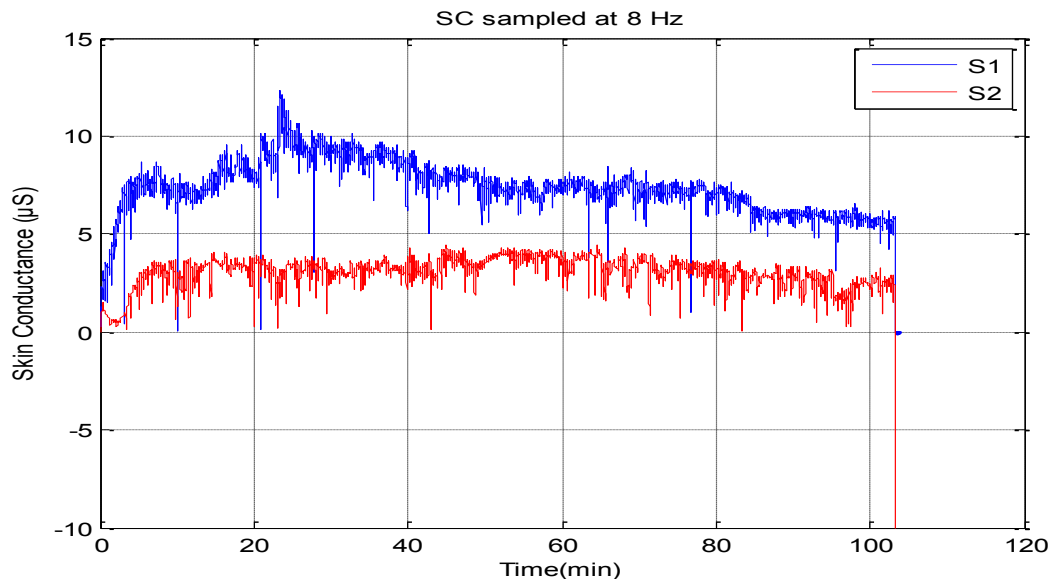


Figure 3.5: A simultaneous SC record of two subjects S1(Blue) and S2(Red) of the same gender (women) in the crowded and "stressful" town center of Tunis. Difference in the EDA level of the two subjects yet being in the same conditions.

Skin conductance reaction (SCR) or phasic activity The skin conductance responses (SCRs), designated also as "peaks", are known as the phasic activity correspond to short

3.2. EDA for state change detection

fluctuations that last several seconds (from 2.5 sec to 18 sec after and due to stimuli). SCRs reflect the higher-frequency variability of the signal that is modulated on top of the slower changes in SCL [47]. A SCR has a characteristic shape [22, 61] happening generally following external environmental stimuli (sight, sound, smell, etc.). It is then called "specific" SCR. It can also occur with no apparent stimuli. Hence, it is designated by "spontaneous" or "nonspecific" (NS.SCR).

A SCR is a short fluctuation in skin conductance that lasts several seconds and usually follows a characteristic pattern of an initial, relatively steep rise, a short peak, and then a relatively slower return to baseline. High frequency of SCRs is frequently seen and is known as "storm". This particular pattern in the EDA is rather defined and studied during sleep as we verified it by our own in some recordings. We give here an example of a "storm" illustrated in Fig.3.6. Prominent rises can be seen in this recording between minutes 100 and 200.

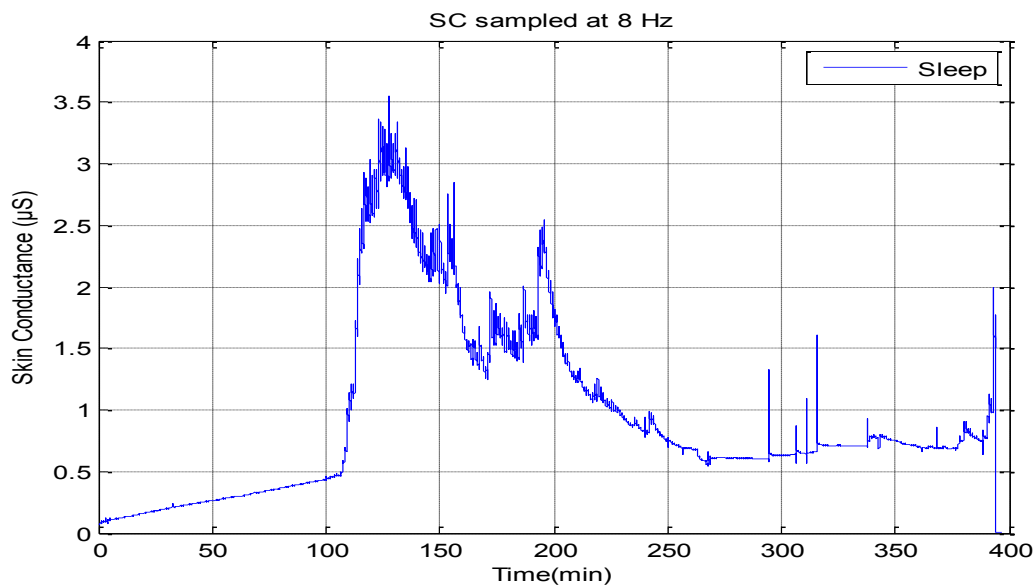


Figure 3.6: SC of a subject during sleep. Presence of "storms" between minutes 100 and 200. Apparent calm periods (sleep) can involve physiological arousal

But, we have also seen it in our recordings during waking periods. This is actually what we observe practically during the whole recording of a master student who accepted to wear the sensor just before her defense when she was repeating and during the defense, see Fig.3.7.

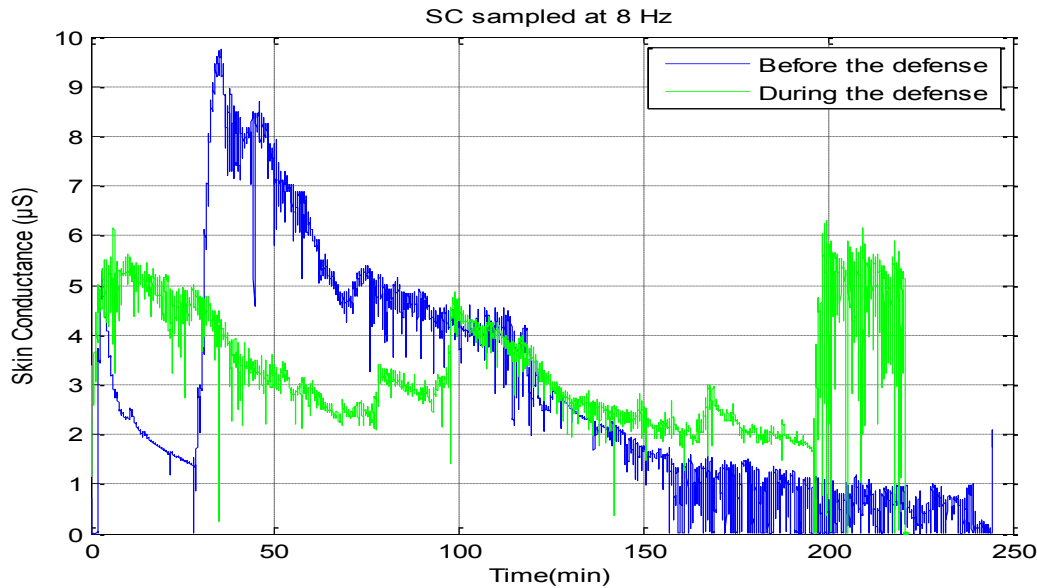


Figure 3.7: SC of a student before (blue) and during (green) her defense of a master project

To sum up, EDA can be evaluated through its two components: tonic level and phasic activity. Tonic level refers to slow or long term change. The most common measure of this aspect is skin conductance level (SCL) or electrodermal level (EDL). On the other hand, phasic changes refer to short changes and result on higher frequency variability of the signal. They occur after a stimulation or not.

Towards a case study approach Furthermore, apparently calm situations may hide arousal in physiological measurement like the case of sleeping shown in Fig.3.6). Moreover, for the same situation and the same stimuli, subjects can react differently as illustrated in the case of the two subjects in the town center of Tunis, see Fig.3.5. Consequently, EDA can be analyzed in a individualized approach. This is what we do in this thesis.

3.2.3 EDA measurement

When the subject is exposed to internal or external stimuli inducing physiological activation, his/her skin immediately becomes a slightly better conductor of electricity. This electrical activity (property) of the skin is called Electrodermal activity (EDA). For instance, we can assess skin conductance or SC (how well the skin conducts a current). And that is why EDA and SC are used interchangeably.

According to [22], the basic measurement method is to apply a small voltage E (e.g., 0.5 V) to two electrodes placed on the intact palmar surface of the skin and include a small resistor (e.g., 200 to 1000 Ω) in series with the skin.

Skin resistance is of order $100\text{ k}\Omega$ or more, the small series resistor is then considered negligible and so can be ignored when measuring current flow I . Hence, by applying Ohm's law, the current (I) measured between the electrodes and necessarily through the resistor is $I = E/R_p$ such that R_p is the resistance of the participant's skin. Since E is constant, I varies following $1/R_p$, the reciprocal of resistance called conductance.

The unit of conductance is siemens (S) such that $1/1\Omega = 1\text{S}$. Because skin resistances are very large (of order of the $\text{k}\Omega$ or $\text{M}\Omega$), skin conductance are small and so measured in units of μ siemens.

Many studies recommend the palmar surface of the fingers or hand rather than the ventral wrist for EDA measurement. This is argued by the fact that palmar surface of the fingers or hand are more sensitive to emotional effects on SC than ventral wrist. This last is rather affected by thermoregulation. In the same time, it was reported that palmar site is influenced by external and internal changes in temperature [22]. Moreover, physical activity can also induce higher sweating and then affect SC.

But, an other recent study [92] compared for the same external conditions (namely temperature) wrist SC to palmar SC. Greater activity in terms of amplitude and number of peaks in favour of the first measure (i.e wrist SC). Add to that, wrist SC recording is more convenient for long term measurement than palmar SC recording. It is worth to note that an adaptation time for the sensor measurement is recommended.

Furthermore, research studies have also recommended to assess ambient temperature and physical activity simultaneously with SC. This would allow to give attention to periods when temperature is significantly high and when hard physical activity is conducted. In this context, we want to note here that high temperature does not always cause elevation in SC. Indeed, an example of SC and temperature measurement shows that changes in EDA are not always caused by changes in temperature, [92]. In fact, in Fig.3.8 skin temperature on the wrist was flat when SC was elevated showing many peaks, and there are no important SC activity when temperature climbs.

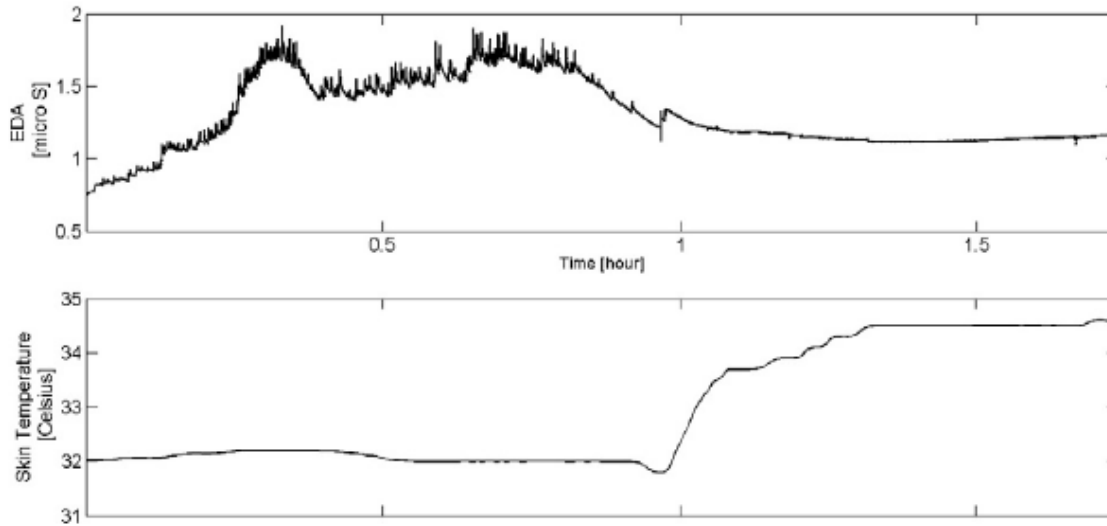


Figure 3.8: An example showing that changes in EDA can not always be linked to changes in temperature (skin temperature on the wrist was flat when EDA showed peaks, and there are no peaks when temperature climbs) [92].

Recent technological development has offered the possibility to assess SC within *in vivo* conditions. As far as we are concerned, we opted for the practical measurement solution via a sensor within which we can track SC, temperature and 3-axis accelerometer data. Details about the used sensor used is given in the following paragraph.

3.2.4 The Q sensor

The chosen sensor consists on a wrist wearable wireless biosensor like a watch manufactured by the compagny Affectiva ¹, a spin-Off of the Massachusetts Institute of Technology (MIT). The sensor (labelled Q sensor) is composed of a wristband equipped with a pod where two electrodes can be inserted and are in contact with the skin on the outer or the inner wrist, see Fig.3.9.

The Q sensor allows the assessment of three different types of measures: SC (in μ siemens), skin temperature in *Celsius* and accelerations of the hand movements in the three directions (x,y,z) .

¹www.affectiva.com

3.2. EDA for state change detection



Figure 3.9: (A) The used EDA sensor (Q sensor): composed of a wristband and a pod where are inserted electrodes. (B) The Q sensor electronic module. (Source: www.Affectiva.com)

Type of measurements	Sampling frequency options	Autonomy	Dedicated software
Skin Conductance, skin temperature and 3-D accelerations	2, 4, 8, 16 and 32 Hz	2 months of continuous recording at 8 Hz	Q software

Table 3.2: Some technical features of the Q sensor. (Source: User Guide of the Q sensor)

As pointed out in Table 3.2, the sensor is provided with a dedicated software allowing to visualize different measurements by connecting the sensor to the computer via a USB cable. In Fig.3.10 a print screen of visualized SC signal is shown. This software allows also to export data in .csv format so that it can be used for further analysis using Matlab or an other software.

One can also see in Fig.3.10 event marks (litte lines) due to marking of instants that are jugded special by the user. This marking could be very useful in understand physiological meaning especially in absence of video/audio recordings or precise activity reports.

The portability, the easy use of the sensor make it convenient for long measurement, add to that the simple transfer of data from the sensor to the computer and also its important autonomy of measurement were the reasons for choosing it for conducting our experiments. Let us point out that the aesthetic side of the sensor and its resemblance to the watch make it attractive for most users.

3.3. Some possible EDA features and artefacts measurement



Figure 3.10: The EDA sensor dedicated software: It offers to the user the ability to visualize different sensor's recorded signals and marking special events

3.3 Some possible EDA features and artefacts measurement

Artefact can be defined as undesired alteration in data introduced by a technique and/or technology [21]. Artefacts are also defined as changes in the recorded signals that do not stem from the signal source in question. It can be due to recording procedure or physiological response of the organism other than the EDA.

A visual inspection of the recordings is always recommended as well as taking notes of possible mis-use during the experience. This is doable in the case of controlled experiments but not easy in the frame of an "invivo" measurements.

We also find that a particular feature of the EDA signal namely "the storm" is, in certain application of change detection, "annoying".

Furthermore, we searched for techniques used in processing EDA artefacts. We majorly find that low pass filtering is used for this purpose. For example in [87], a low pass-filter with a cut-off frequency of 3 Hz is used for reducing motion artifacts and electrical noise. In [91] and [92], a low pass FIR filter of order 32 with a cut-off frequency of 0.4 Hz is used in the pre-processing step. Linear interpolation is also mentioned for dealing with aberrant data.

3.3. Some possible EDA features and artefacts measurement

In this section, we focus our attention into possible EDA measurement artefacts for runners and choose solution in relation with the analysis tool to be applied on them. So, we observed the pre-test (Fouléss du Lac) recordings and reported mainly 2 types of artefacts: those related to runners' motion and those due to the mis-use of the sensor that are possible for any type of users. The storm feature that we noted in several measurements of runners and others is also highlighted.

3.3.1 Motion artefact

Let us take a look at the sensor measurements of a marathon runner in Fig.3.11. There are three principle time zones that can be distinguished:

- Adaptation to the sensor
- Warm-up time
- The race



Figure 3.11: EDA measurement of a semi-marathon runner: presence of null values and missing data

3.3. Some possible EDA features and artefacts measurement

In the case of marathon runners, we have noted oscillations of the hand during the race. This can be observed in the corresponding hand accelerations recordings (oscillations within the three axes). In Fig.3.12, we show these oscillations at the beginning of the marathon race in acceleration in Z axis.

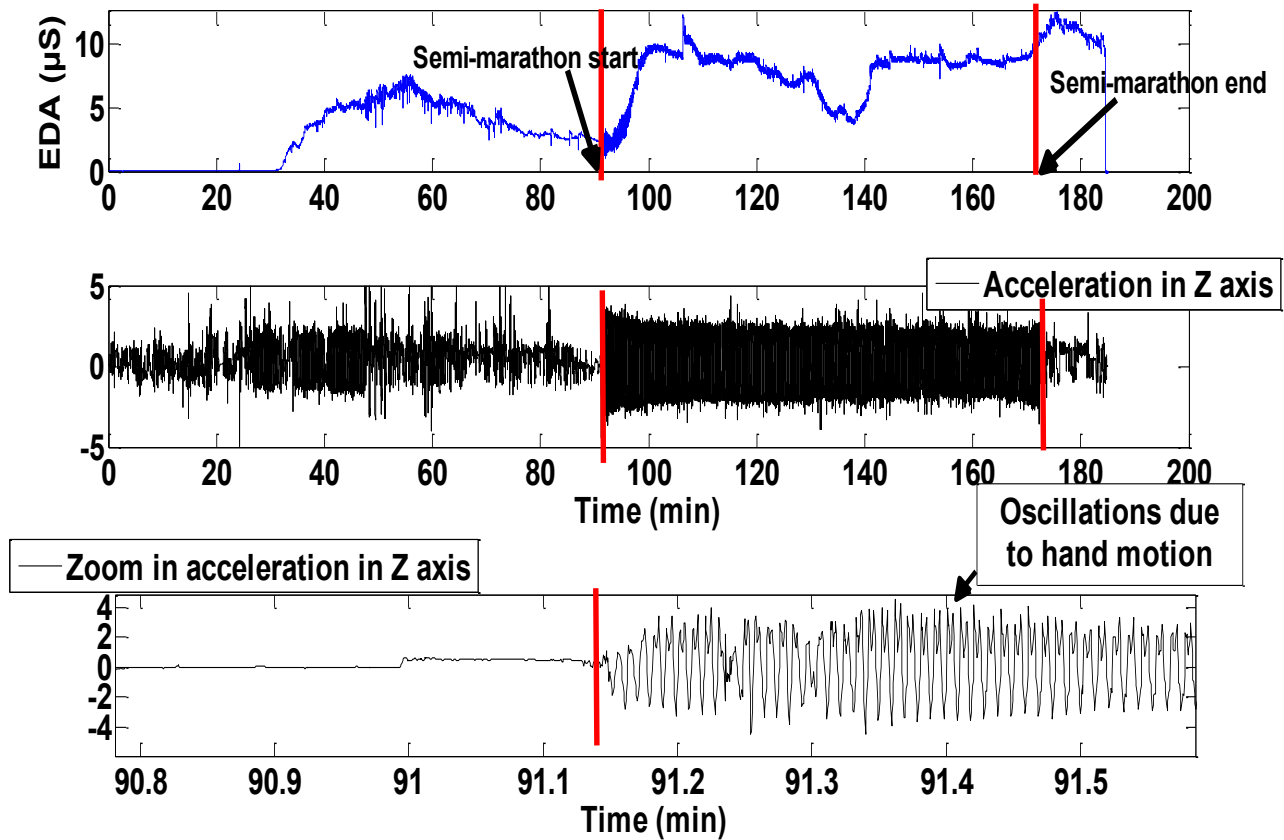


Figure 3.12: EDA and acceleration measurement of a semi-marathon runner: presence of oscillations due to the hand motion

The spectrum of the corresponding EDA recording show harmonic components that are not present when the marathoner is not running. Hence, EDA data are altered in the case of motion oscillations. This will be detailed later on chapter 5.

3.3.2 Artefact due to mis-use of the sensor

We searched for possible artefacts that can be due to the mis-use of the sensor and that can occur during a race.

3.3. Some possible EDA features and artefacts measurement

3.3.2.1 Artefacts related to moved or detached electrodes

We tried to simulate these artefacts in order to be easily identified. Results of these experiments are gathered in Table 3.3.

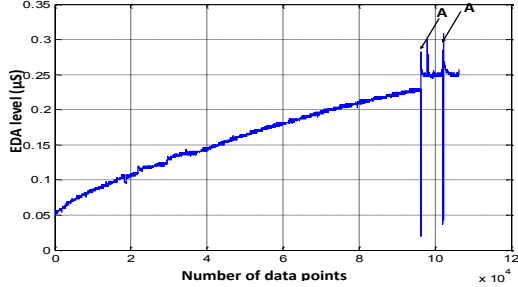
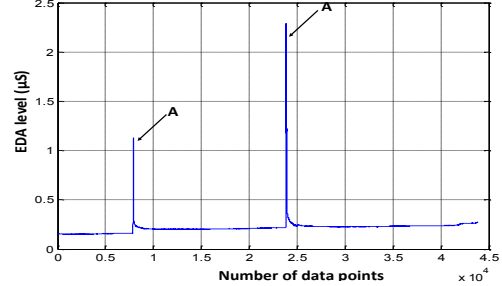
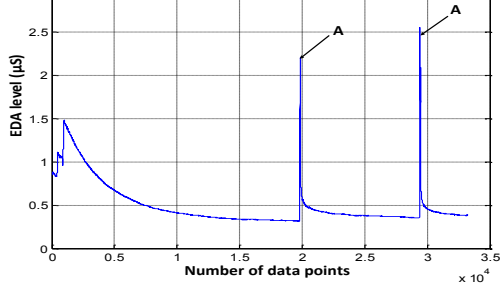
Artefact	Illustration
Moving up	
Horizontal translation to the front	
Horizontal translation to the rear	

Table 3.3: Possible artifacts due to the mis-use of the Q sensor: Example of moving or detaching electrodes

We check the result of change detection on the mean of an EDA signal and we illustrate in Fig.3.13 the effect of the presence of such artefact on the result of change detection.

For such artefacts that are not very frequent, we make a visual inspection of the signal and treat the similar cases by linear interpolation.

3.3. Some possible EDA features and artefacts measurement

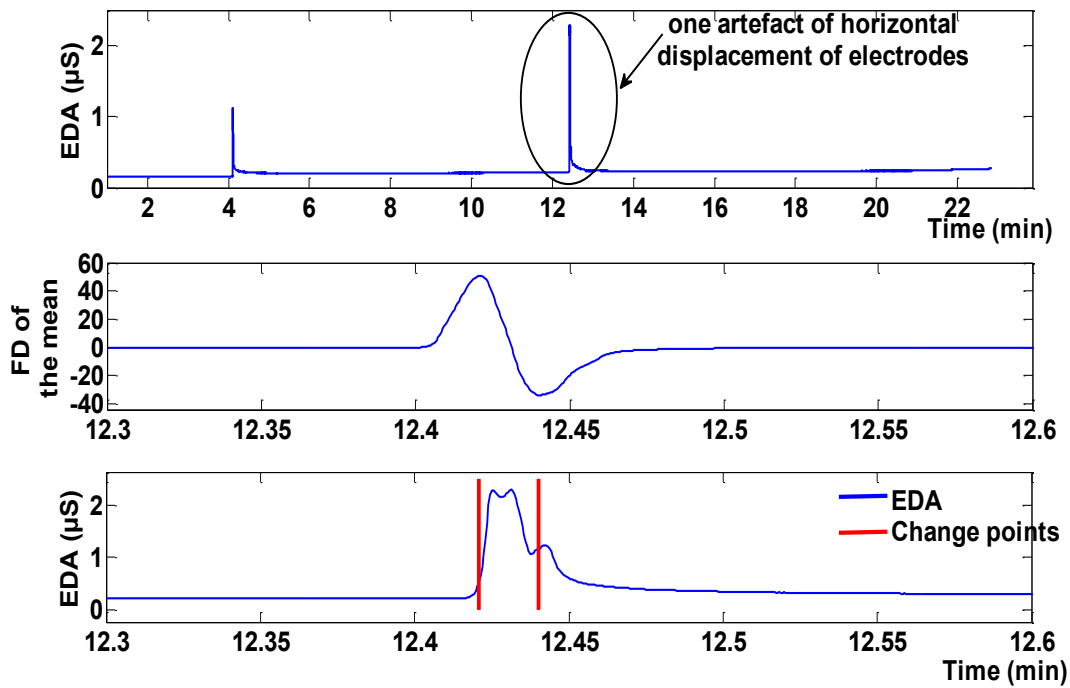


Figure 3.13: Effect of a horizontal displacement on change detection processend on the mean of a simulated EDA measurement. $FDpVparameters : A = 1sec, K_{max} = 4, p-value = 0.1.$ $obtainedresults = 4.10sec, 4.12sec, 12.42sec, 12.44sec$

3.3.2.2 Missed data and zero values

We noted also that we can be confronted to missing and null EDA values data in our recordings, see Fig.3.11. After the end of the semi-marathon, the runner removed the sensor (as noted in the top of the figure). After that, we can see null EDA values and missed ones.

Missed data are find as "NaN" values in EDA recordings and found at any time of the recording (beginning, middle or end), see for instance Fig.3.11. "NaN" values induce "NaN" computed statistics (for instance slope values in the case we search for change points in the EDA slope).

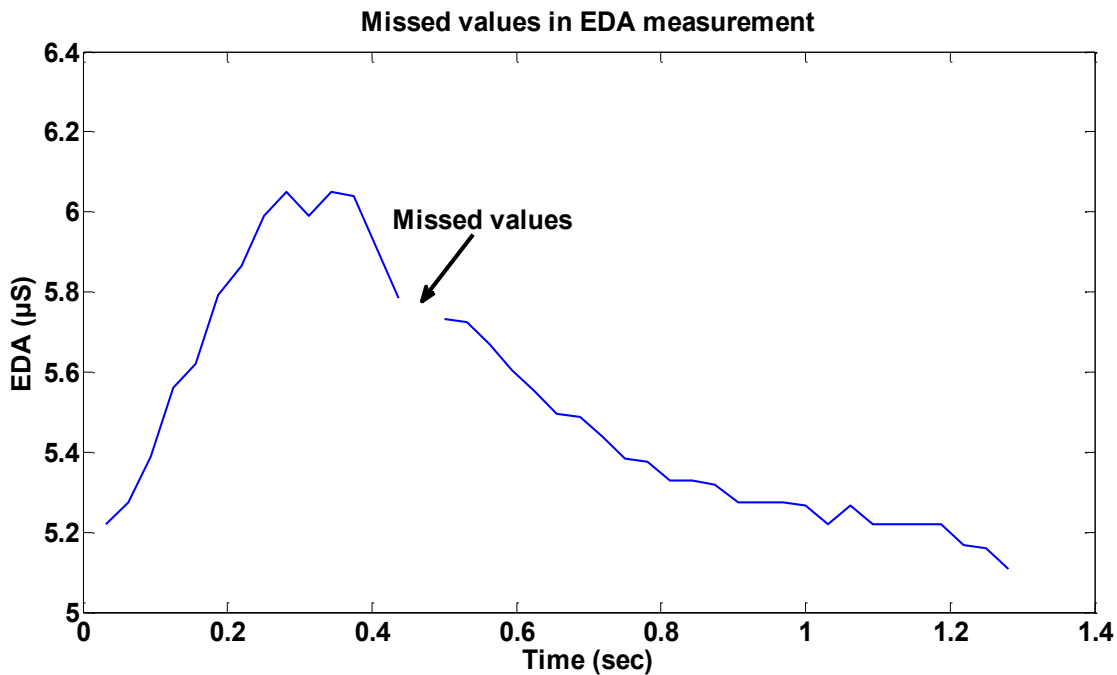


Figure 3.14: Missing data in about 1 sec recording of an EDA measurement of a semi-marathon runner

Null EDA values are noted in the case of removing the sensor without switching it off. So, they are not the measurement of the physiological signal and we have to not consider them in our analysis.

In the case of missed data, linear interpolation can be a good solution to overcome this case.

3.3.3 Particular EDA feature: the storm

Prominent rises, SCRs or peaks in SC are known as "storms". In the studies we examined and which tackled this phenomena, the attention was given to storms during sleep periods rather than waking ones. In [92] storms are defined as regions of continuous peaks or EDA peak epochs such that an EDA storm has a minimum duration of 1 minute and has at least two peaks during that minute. The former and first definition of Burch in [25] presented the storm as a sequence of more than 5 peaks per minute for more than 10 consecutive minutes of sleep. In the following we can see that this feature can also be seen in the case of marathon runners. Indeed, in Fig.3.15, we illustrate an EDA signal of a runner at the beginning of the race. We can note at this signal the presence of prominent rises. Their presence induces peaks in the filtered derivative function (FD) at the time they occur, see Fig.3.16 (Top) and the most potential change points are then localized in the region of storms even if the FDpV

3.3. Some possible EDA features and artefacts measurement

eliminates iteratively potential change points situated within A window before or after the first selected change points.

Resulted change points, see Fig.3.16 (bottom) are then selected with a p-value under then 0.1.

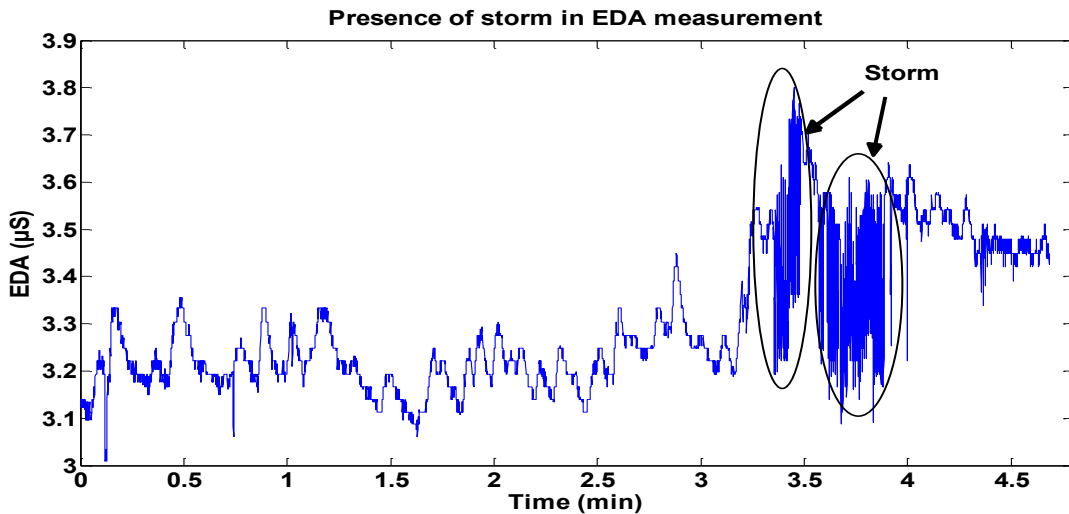
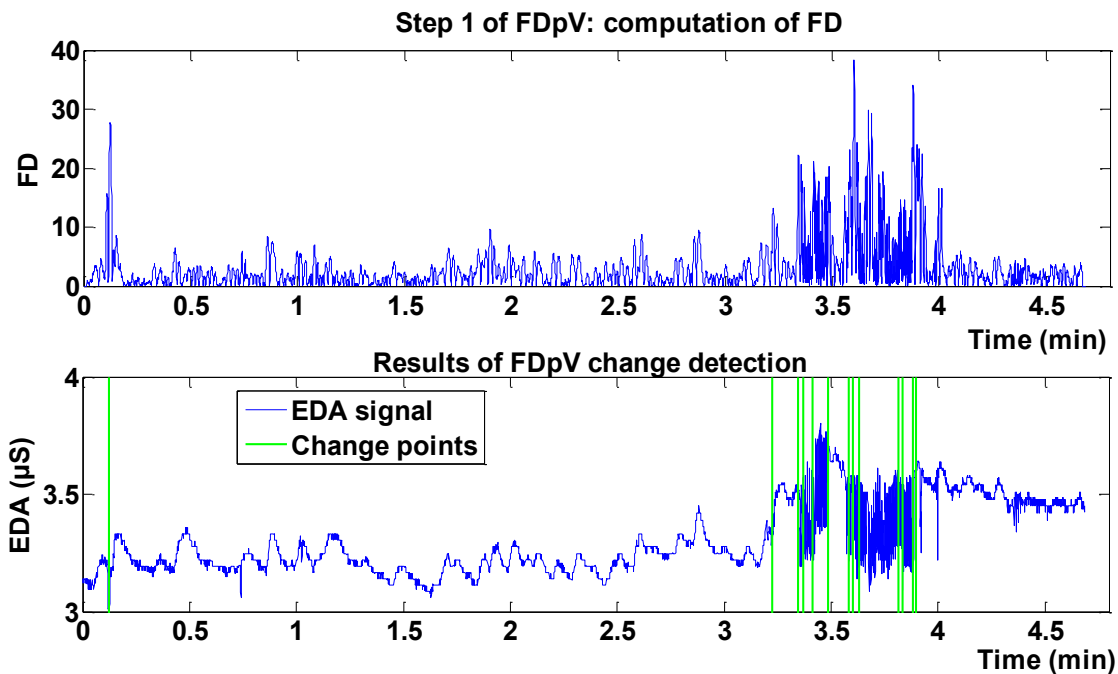


Figure 3.15: Presence of prominent rises in the EDA measurement



FDpV parameters:

$A = 1\text{sec}$, $K_{max} = 20$, $p\text{-value} = 0.1$, Nbr of change points obtained= 13

Figure 3.16: Change points obtained in the presence of storms

3.4 Comar Marathon with a well established protocol

We have had a second chance to lead our EDA measurements on marthoners by taking into account the outputs of the pre-test, see Section 3.1. Indeed, we restricted ourselves to the only measurement of EDA and following a well established experimental protocol.

This experiment was led in the occasion of the 27th COMAR² marathon that took place on October, 21st, 2012 in Tunis. Hence, in this section, we first give the principle of the experimentation and the procedure as recommended in [40]. Second, we detail how the experience was held. Then, in the third paragraph, we present the experimental protocol. Finally, we evaluate its execution and the measurements obtained.

3.4.1 Principle of the Comar experiment

In the following, we precise the objective of the experiment and so the questions that we want to answer (purpose of the experiment), the mean that we choose to search for that answer (studied phenomenon) and factors that can influence the studied phenomenon.

- Purpose of the experimentation: The study of the emotional engagement of runners during a marathon competition specifically start and finish phases.
- Studied phenomenon: The studied phenomenon is EDA. We suppose that EDA can characterize motivation and emotional engagement for a runner during key phases (start and finish phases).
- Causal relation to be tested:
 - Professionalism affects positively emotional engagement i.e professional runners will show higher arousal than amateurs
 - The competition phase affects the emotional engagement and so EDA

Principle of the procedure: In the case of our study, we propose to focus on differences in EDA variations between amateurs and professional marthoners during two different phases of the competition (At the starting phase/At the final phase).

- Factors

Let us precise some notions:

- Professionals: subjects who are trained frequently

²The Mediterranean company for insurance and reinsurance, www.comar.com.tn

3.4. Comar Marathon with a well established protocol

- Amateurs: subjects who practise sports from time to time

For a competition of a semi-marathon, the masculine records is of 58 min 23 s and the female one is 1 h 05 min 12 s. Then, we defined the starting and final phases as follows:

- The starting phase: about 30 minutes after the beginning of the competition
- The final phase: about 30 minutes before the end of the competition

Table 3.4 summarizes the varying factors.

Factors	Levels and specifications
Professionalism	Amateurs and Professional
Competition phase	<p>At the starting phase: It begins at the starting of the competition and lasts about 30 minutes.</p> <p>At the finish phase: It ends when the participant crossed the finish line and begins about 30 minutes before. The organizing team marks every arrival of the participants at the finish line in each corresponding form</p>

Table 3.4: Factors to vary in the Comar experimental protocol

- Expected results:

EDA usually increases when a subject is “more aroused” in case of stress, engagement or excitement situations. However, it tends to stay low or decrease when the subject is “less aroused” if he is in a calm and no engaging situation [21], [78].

Referring to this reflection, a number of expected results are summarized in Table 3.5.

3.4. Comar Marathon with a well established protocol

	At the beginning phase	At the final phase
Amateurs	EDA level would be important (effect of the event)	Depends on the engagement and ambition
Professionals	Important EDA level	Important EDA level

Table 3.5: Expected results: EDA level following the factors of professionalism and competition phase for the Comar marathon

3.4.2 Execution of the experience

In this part, we specify tools and materials used to lead the experience:

- The Q sensor is the sensor used to assess EDA
- The factor of professionalism is controlled through the recruitment of the participants
- Competition phase is determined according to the start time and the arrival time of each subject.

In addition, a number of specifications have to be detailed in relation with factors, duration of the experience, security measures, see Table 3.6.

Other specifications	Description
Duration of the experience	Duration of the competition (marathon) for each participant
Measurement frequency of the observed parameter	One time(during the competition)
Security measures	Mentioned in the user guide

Table 3.6: Other specifications related to the Comar experimental protocol

Moreover, steps of the experimentation that were followed are defined as follows:

1. At the beginning, do wear the Q sensor for runners (already synchronized) one by one
 - (a) Fill the form corresponding to each subject

3.4. Comar Marathon with a well established protocol

- (b) Fill the form corresponding to each Q sensor
- 2. At the end, retrieve sensors and take feedback notes from runners on the form of each participant

3.4.3 Presentation of the protocol

The protocol is presented in the file: Experience form (see Fig.3.17)

3.4. Comar Marathon with a well established protocol

Date: 21/10/2012
Experience form
Filled by: Nadia Khalfa

Date of the experience	May, 13, 2012
Frame/Event	Comar marathon
Place	Tunis center
Administrators/Operators	Three operators: Nadia Khalfa Wisseem Ouislati Slaheddine Drissi
Purpose of the experience	Study of the emotional engagement during a marathon competition
Population	1-Type <input checked="" type="checkbox"/> Runners 2- Size : 5 3- Age : from 17 to 55 4- Gender : <input checked="" type="checkbox"/> Man (4) <input checked="" type="checkbox"/> Woman (1)
Measurements	EDA Temperature Actigraphy Presence of stimulus : <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If 'Yes', precise the frequency: 2 (at least) At the beginning- At the end
Sensors	Q sensor
Description of the experience	1) At the beginning , do wear the Q sensor for runners (already synchronized) one by one 1.1) Fill the form corresponding to each subject 1.2) Fill the form corresponding to each Q sensor 2) At the end , retrieve sensors and take feedback notes from runners on the form of each participant
Duration of the experience	Depends on each runner

Figure 3.17: Experience form filled during the Comar protocol

3.4. Comar Marathon with a well established protocol

Two other files: Participant form.pdf (Annex 2) and Sensor form.pdf (Annex 3) are involved in this protocol. We respectively present the form related to each participant and that of each sensor worn in order to guarantee the best tracking of subjects and sensors during the competition.

In Fig3.18 we illustrated main points related to Comar competition experimentation.

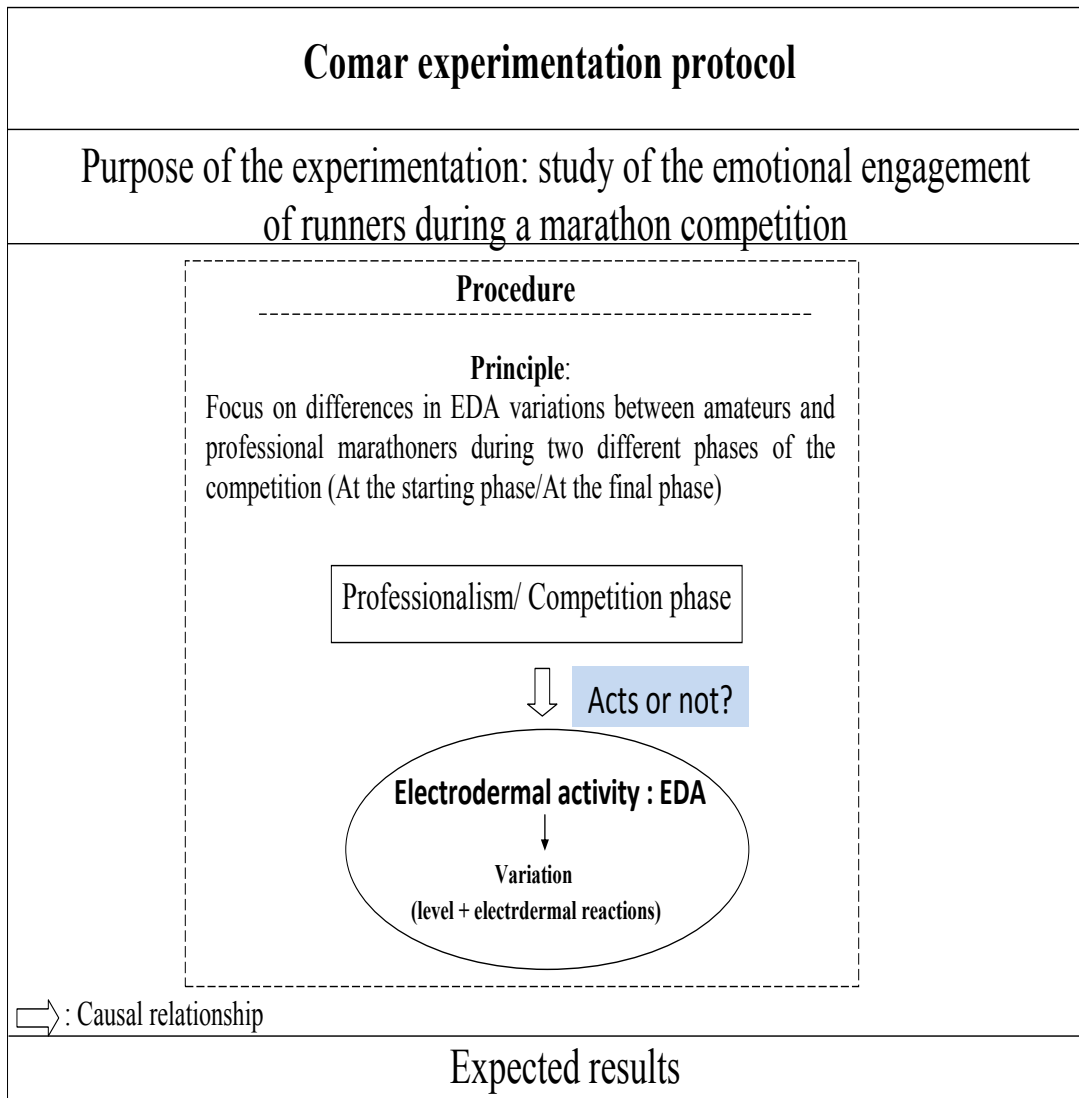


Figure 3.18: Comar competition experimentation

3.4.4 Evaluation of the Comar protocol: data validation

1. Relevance

- External relevance can be tested through the question: Do EDA evaluate emo-

tional engagement during a marathon competition? Response: Signal analysis following the two phases of the competition decides of the relevance of this choice.

- Internal relevance can be evaluated through the choice of the sensors to collect data. Response: The Q sensor permits the collection of EDA efficiently and easily
- Is the protocol adapted to the population of marathoners? Response: Yes, runners accepted wearing the Q sensors and did not report any disturbance in the end of the competition.

2. Quality

- Accuracy: (needs comparison with an other device)
- Fidelity/Repeatability: Measurements were taken just one time

3. Executability:(needs external revision)

4. Communicability:(needs external revision)

After leading the Comar marathon measurements, we get EDA data series with our own material and that we can process and analyze more easily than an other signal. We store obtained data in a base for runners' physiological signals.

3.5 Conclusion

In this chapter, we tackled the experimentation aspects related to data collection. A reflexion on the protocol establishment was then made to allow, after a first pre-test experiment (Foulées du Lac) a second experience (Comar Marathon) more adapted to marathoners' population and more targeted by focusing into start phase and finish phase.

The EDA was proposed for detecting state changes through its two indicators (tonic level and phasic activity). We give arguments and examples showing the reliability of EDA measure to reflect motivation and arousal.

For assessing EDA, an ambulatory measurement method was choosen. It affords wireless EDA recordings. Possible recording artefacts were also discussed in accordance with the processing method. In chapter 4, we analyze temporal evolution of the EDA of this cohort of runners at the beginning and finish phases and highlight temporal signatures at those key phases and give an EDA profile of an elite marathon runner. In Chapter 5, for these EDA signals, we propose a method for the tonic level extraction and the motion artefact detection.

Chapter 4

Temporal signatures of electrodermal activity (EDA) for the evaluation of runners' performance: start and finish phases

Preamble

Motivation is a key factor in most human activities; especially when these are related to performance [32, 99]. In the world of sports, athletes are under the pressure to achieve best results. For that, the motivation is a major affective dimension that needs to be developed, nourished and maintained [60]. In this work we have considered the case of marathoners when motivation is needed to boost their performance and get better rankings. For that, the approach we propose is to depict some physiological important changes that a marathoner faces during a competition which may not necessarily be visible or easily perceived. Then, he/she can with the help of specialists (trainer, sport physiologists or psychologists) interpret and deal with.

So, let us recall that a reflection on a protocol measurement establishment for the cohort of marathoners in two competition situations (Foulées du Lac 2012 and Comar marathon 2012) was detailed in Chapter 3. As a result, EDA was chosen as the physiological signal that can reflect motivation state changes. Chapter 4 constitutes a study that deals with the recorded EDA measurements of marathoners during these two competitions. Hence, the aim of the study is to assess the ability of EDA to characterize the performance of few amateur and professional Tunisian runners. We focus on the start and finish phases of the competition since they are key elements in the determination of the athletes' performance[54, 55]. So, we examine the EDA temporal signature during those key phases. We note that the overall EDA performance tends to be similar for all subjects during the starting phase of the competition. However, the end phase seems to differentiate among them: specifically, we note with the winner of the competition (the reference

runner) that there is a "better management of stress level" in the sense of EDA variability.

This study is a conference paper titled "*Temporal signatures of electrodermal activity (EDA) for the evaluation of runners' performance: start and finish phases*", co-worked by N. Khalfa, S. Drissi¹, R.Ghozi, M. Jaïdane. It was presented and published in the conference of The 9th International Workshop on Systems, Signal Processing and their applications (WOSSPA, Algeria, 2013)².

4.1 Introduction

The electrodermal activity (EDA) is examined by monitoring skin conductance caused by internal or external stimuli which induces a physiological activation. This activation is an indicator of the sympathetic nervous system state [81]. It can also be interpreted as the stress/motivation state of the subject. The aim of the paper is to assess the ability of the EDA to characterize the performance of runners during a competition. Indeed, we examine the EDA measurement at specific time periods that are important for the evaluation of the motivation of a runner: that of the start phase and also the finish phase of the competition. We then propose to study the EDA of seven runners. The performance criteria (ranking) will be analyzed in conjunction with the EDA measurement. In section 2, we provide a description of the work frame and the experimental protocol of the study. In section 3, we will explain the necessary pre-processing related to EDA measurement artefacts. Then, we will detail the temporal analysis of the EDA signal for semi- marathon runners in section 4 in particular for the reference runner P6. An analysis of the start and the finish phases of the competition for that cohort will be held in the following section defining temporal signatures in both phases. And finally, we summarize first results and propose perspectives to this study.

4.2 Protocol and data collection

4.2.1 Frame of the experiment

Measurements analyzed in this work and relative to our cohort (see Table 4.1) were gathered during two major Tunisian athletic events:

- The 27th COMAR³ marathon on October, 21, 2012.

¹S. Drissi is a teacher in Issep-Ksar Said, University of Manouba and member of the Tunisian Research Laboratory OPS.CNMSS

²In the sequel, we keep the expression "in this paper" instead of "in this chapter"

³The mediterranean company for insurance and reinsurance, www.comar.com.tn

- The semi-marathon of Foulées du Lac⁴ during its 8th edition on May, 13, 2012.

These competitions took place in Tunis, Tunisia.

4.2.2 Population and classes of participants

Seven subjects were selected in this study. Details about each participant are indicated in Table 4.1, but some informations for the rank are not indicated because the runner in question is not classed within the five ones. In this case it is not mentioned (NM). In term of time, tunisian semi-marathoners spend about 1h02 min/1h06 min (men) and 1h15min/1h25min (women) in this type of competitions.

Professionals and amateurs

Professionals: It is worth to note that P2, P3, P5 and P6 can be classed as high level runners and for instance both P2 and P6 participated with the determination to win. Indeed, quite important sums of money are at stake. We may stress that a majority of runners belong to disadvantaged areas and that such competitions represent a real challenge to gain money.

Amateurs: On the other hand, P1, P4 and P7 are rather amateur athletes who are participating to enhance their last results.

Participant	Gender	Age	Competition	Rank	Competition event
P1	M	46	42 km	NM	Comar
P2	F	21	21 km	1	Comar
P3	M	26	21 km	3	Comar
P4	M	55	6 km	NM	Comar
P5	F	34	21km	NM	Foulées du Lac
P6 (reference)	F	27	21 km	1	Foulées du Lac
P7	M	48	21 km	NM	Foulées du Lac

Table 4.1: List of runners and specifications (gender, age, type of competition and rank). P2, P3, P5 and P6 are high level runners while P1, P4 and P7 are amateurs. NM: Not mentioned. P6 is our reference. Types of competitions: marathon (42 km), semi-marathon (21 km), 6 km

⁴www.lesfouleesdulac.org

Cohort of interest and the reference runner

Cohort of interest: P2, P3, P5, P6 and P7 is the concerned cohort of our analysis. All of them participated in the semi-marathon. Indeed, in order to insure coherence, we decided to focus only on participants of the 21 km competition i.e semi-marathoners (P1 and P4 participated in 42 km and 6 km).

The reference runner: P6 can be considered as *a reference runner*. She is of international scale, experienced and well prepared for the event. She presents a typical profile that can be verified through the analysis.

4.2.3 Materials

Runners were asked to wear on their wrist a wireless biosensor⁵ that is a wristband equipped with a pod where two electrodes can be inserted, see Fig.4.1.



Figure 4.1: The used electrodermal activity sensor. It is composed of a wristband and a pod where are inserted electrodes (credit: Affectiva)

It allows a tracking of three different types of measurements: EDA, skin temperature and accelerations of the wrist movements in the three directions (x,y,z) . All subjects wore the sensor for at least 30 minutes before the start of the marathon including the warm-up period. They took it off few minutes after they reach the arrival line. Recorded data are sampled at 32 Hz. It is worth to note that for any experiment with this sensor a period of 15 minutes of record before the experiment is recommended to obtain a baseline measurement of the EDA. Indeed, EDA levels may have significant differences among individuals and a period of time adaptation for calibration is needed as recommended in different studies [31, 32].

⁵The qsensor is designed by Affectiva, a spin-Off of MIT, www.affectiva.com

4.3 Pre-processing

4.3.1 EDA artefacts identification

Eventhough artefacts are inherently present in any data collection, they tend to be specific for the situation and sensor use. In this study, we wish to take a brief look at the kind of artefacts that can appear with the Q sensor on athlete during a competition as it is the one considered here. We will see later the coupling between motion and EDA as a source of artefacts. Fig.4.2 shows a sample EDA plot of P3. We can note the presence of “suspected” values for example at the 50th minute and also from the 100th minute till the finish phase, that we propose to discard the whole EDA series from the analysis. These artefacts can be due to:

- Possible miss-use of the sensor by moving it around the wrist or hand movements due to the rubbing of the hands with the body
- Refreshment with water since it is possible in each number of kilometres to cool with water

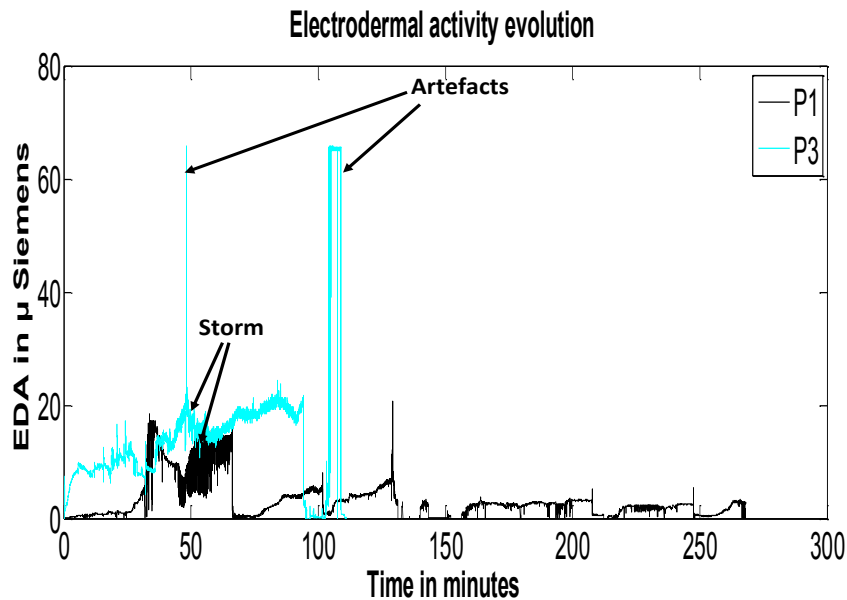


Figure 4.2: Presence of storm and artefacts in electrodermal activity measures (EDA): case of P1 (42 km, 46 yrs old) and P3 (21 km, 26 yrs old); particularly high variability for P1 for about 10 minutes from the 50th minute. Artefacts: “suspected” EDA measures for P3 at the 50th minute and also from the 100th minute till the end.

4.3.2 EDA special feature: The storm

A special feature of the EDA measurement was seen for practically all the subjects but especially for P1 who is a 46 years old participant of the 42 km competition for the tenth time. It is about the high frequency and variability of EDA peaks during the competition and also the warm-up. This phenomenon is known as 'storms' introduced by Burch who defines it as five electrodermal peaks/min for at least ten consecutive minutes of sleep, see also [21, 81, 91].

We can also notice for the EDA measure of P1 a high variability from approximately the 45th minute (when the competition begun) to the 66th minute of the record. P1 was determined to enhance his last result although he was suffering from injuries in the leg especially at the end of the 42 km, when EDA values were under 3μ Siemens. This can explain the absence of this feature at the finish phase of the marathon see Fig.4.2.

To summarize, we can circumvent the artefacts by detection and interpolation while for the 'storm' feature and high variability, we can use a simple averaging or low-pass filter [81, 91].

However, in this work, our analysis will not be preceded by any pre-processing step. In fact, we choose not pre-process to see all variations and artefacts and then specify our own strategy of treating artefacts.

Next, we will concentrate on temporal analysis and temporal signature of the start and the finish phases of a competition in the EDA raw signal.

4.4 Preliminary EDA analysis on a reference runner (P6)

In Fig.4.3, we illustrate the EDA, temperature and actigraphy of one runner (P6) during the semi-marathon of Foulées du Lac⁶. These three types of measurements are taken simultaneously with the wearable sensor and visualized using the appropriate software.

- On the bottom of the same figure, we can see the acceleration measurements with three different colors: red (for the x axis), green (for the y axis) and blue (for z axis).

In order to track the EDA changes, we examine the athletes motion information namely these accelerations where we can identify 3 periods of activities.

- Period 1: Before the competition. We can see period of general and specific warm-up of the runner.

⁶On May, 13, 2012 at Tunis

4.4. Preliminary EDA analysis on a reference runner (P6)

- Period 2: The competition time duration, delimited by T_{start} and T_{finish} . During which, accelerations are very regular and relatively high.
 - Period 3: After the end of the competition.
- In the middle of Fig.4.3, we can see the measured temperature that progressively increase just before the start of the competition and abruptly decrease just after T_{start} . The increase of temperature can be explained by the warm up process, due to the stimulation and operating of the cardio-respiratory, endocrine, and neuromuscular systems, and the decrease seems to be due to the convection process, factor of the air blowing over the body, less covered than prior to the competition [29].
 - On the top, we find EDA evolution in time marked by an increased level during the warm up followed by a decrease during a “rest” period (during the specific warm-up) when the temperature increases. One last observation shows that temperature and electrodermal activity do not necessarily progress in the same manner.

For all participants, the trend of EDA evolution within time was very similar but we will closely examine the EDA signals especially at the start phase (known) and the finish phase (unknown for an athlete) of the competition. Graphically, it concerns *Period 2* in Fig.4.3.

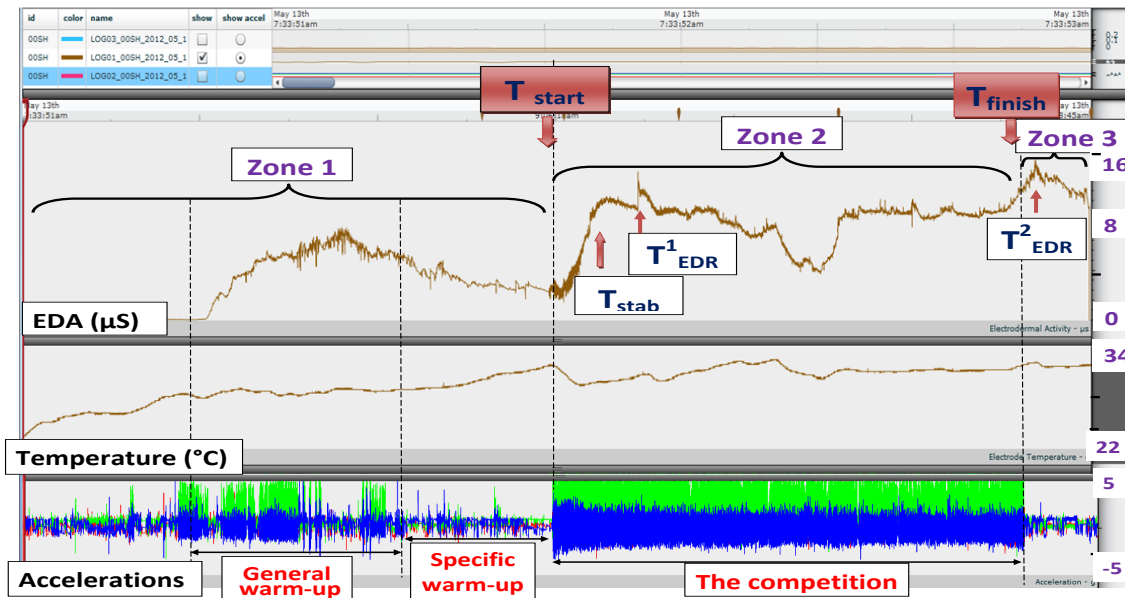


Figure 4.3: Measures of EDA (top), Temperature (middle) of the reference runner P6 before (Zone 1), during (Zone 2) and after (Zone 3) the competition. Key moments of the competition are indicated: T_{start} , T_{stab} , T^1_{EDR} , T^2_{EDR} , T_{finish}

4.5 Athletes EDA signature: the start and the finish phases

We recall that we are examining the EDA records of the semi-marathon competition. It concerns runners from 21 to 34 years old (P2, P3, P5, P6) and a middle aged competitor of 48 years old who can be considered as an amateur, P7, see Table4.1.

4.5.1 Temporal signature and EDA level of the start phase of the competition

We begin our analysis by examining the EDA of P6 10 minutes before the start of the competition and about 30 minutes after, see Fig.4.4. We can note that the EDA level takes a certain time (from 3 to 7 minutes in the case of P6) at the start to reach high values and to maintain a relatively constant level. So, we can distinguish three key moments as annotated in Fig.4.4:

- T_{start} : The starting instant that marks the beginning of the marathon
- $T_{start} + \Delta T_{stab}$: It is the moment when EDA level become relatively stable after a progressive increase

This increase can also be seen in heart rate frequency measures as we have noted in other studies [15, 63]. And if the heart rate is a mirror of the functioning of one organ: the heart, skin conductance may be the result of the whole organism functioning. In fact, from a physiological point of view, ΔT_{stab} reflects the adaptations of both metabolic and physiologic systems of runners, until it reaches the tempo of aerobic steady state [46, 69].

Following this temporal marking of EDA periods in the reference athlete, we can introduce the notion of temporal signature of the start phase in an athletic competition.

4.5.1.1 Temporal signature of the start phase

The start phase of the competition relative to the reference runner P6 as illustrated in Fig.4.4 can be characterized by three temporal parameters namely:

- EDA_{start} is the EDA level measured at T_{start}
- ΔT_{stab} which is the time duration needed for EDA to stabilize from the start of the race, see paragraph 4.5.1
- EDA_{stab} is the EDA level measured at T_{stab}

4.5. Athletes EDA signature: the start and the finish phases

These temporal parameters marking the start phase of a race in EDA plot can also be identified for the other athletes, see Fig4.5.

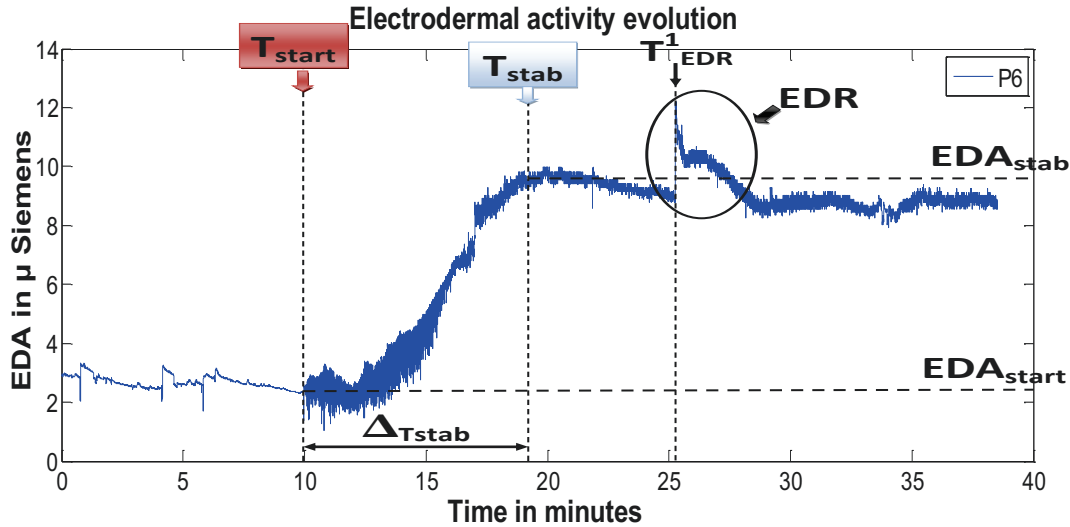


Figure 4.4: Temporal signature of the start phase: case of the reference runner P6. Three characteristic parameters ΔT_{stab} , EDA_{start} , EDA_{stab} . Electrodermal reactions (EDR) are localized at T_{EDR} and correspond to the physiological activation occurring after a stimuli

4.5.1.2 EDA level at the start phase

Moreover, we can distinguish different groups of close EDA level (reached just after T_{stab}), and can be arranged in descendant EDA values:

- P3 present EDA measure between 14 and 15 μ Siemens.
- P2, P5 and P6 : their EDA level vary between 6 to 11 μ Siemens
- P7 has the lowest values of EDA (5 to 8 μ Siemens)

Based on these observations, some hypotheses can be emitted. In deed, motivation can explain the lowest values of P7 (a non athlete) who participated in the competition without high challenge as he reported. Also, gender can be a parameter of classification. In fact, the group level of P2, P5 and P6 is composed only of women athletes whose the stress level at the start phase is lower than that of P3 who is a man athlete.

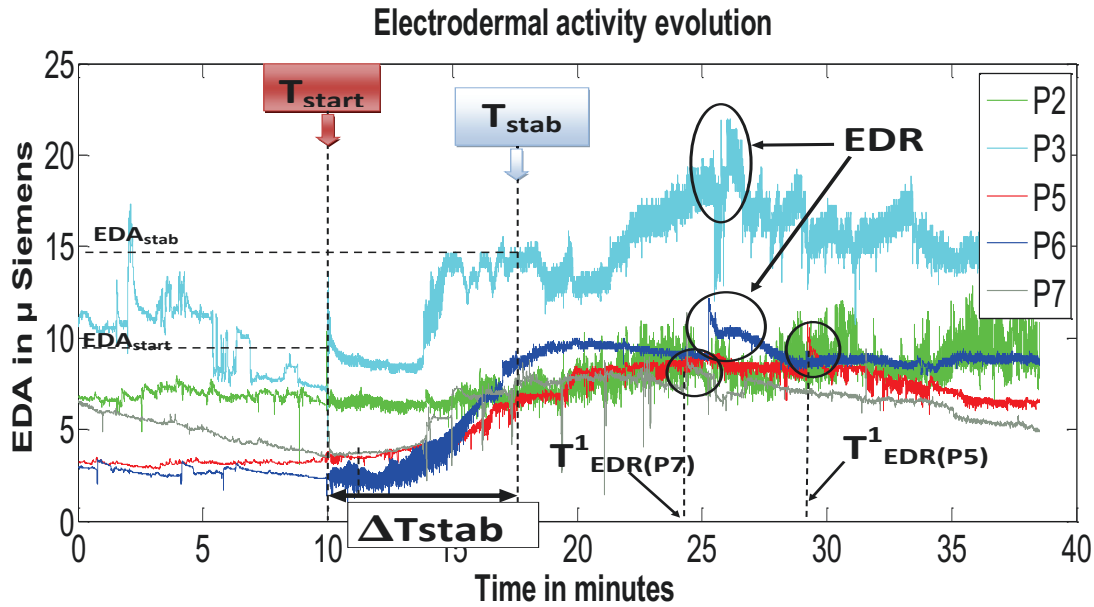


Figure 4.5: Similarity of the temporal signature at the start phase with different ΔT_{stab} , EDA_{start} and EDA_{stab} for each runner and different T_{EDR}^1 of the restricted cohort of 21 km. Arrows show significant electrodermal reactions (EDR) between 15 min and 20 min from T_{start}

4.5.2 Temporal signature and EDA level at the finish phase of a competition

The EDA measures of the finish phase of the competition are different from the start phase and may also reveal some aspects. In Fig.4.6, we illustrate about 28 minutes before the end and 2 minutes after.

There are two key moments in the EDA measurement at the finish phase of the semi-marathon:

- T_{finish} : is the moment of the arrival
- $T_{finish} - \Delta T_{finish}$: is the time that precedes the moment of reaching the arrival line and from which the EDA level seem to increase.

Indeed, at about 5 minutes before the end of the marathon, we can note an increase of the EDA for P3, P6 and P7. This is an expectable result especially for P6 who won the competition and reflects her increased excitement and joy. P2 also won the competition but we do not see this augmentation of the electrodermal activation at the finish phase. This may be explained by the level of preparation that precedes the semi-marathon. In fact, P2 unlike P6 was not highly trained for the event. Add to that, P6 is a marathoner of international

scale and has enough experience to “manage” a competition from the start to the end. So, we can consider that her EDA level as a typical measure of electrodermal activation for a professional runner.

4.5.2.1 Temporal signature of the finish phase

The temporal signature of the finish phase of the competition can be characterized by two parameters, annotated in Fig.4.6, namely ΔT_{finish} (explained in paragraph 4.5.2) and EDA_{finish} is the EDA level measured at T_{finish}

4.5.2.2 EDA level at the finish phase

As we can see in Fig.4.6 the relative position of EDA signals for the studied cohort in comparison to the start phase is almost the same. Although the EDA level for all subjects remain practically the same, P2 who is the youngest of this cohort is marked by a higher level of electrodermal activation. Normally, toward the finish phase of a marathon, when the speed and effort of running increase, the body becomes less efficient at using energy, which produces more excess heat, that in turn drives the body temperature and arousal (EDA level) to become higher.

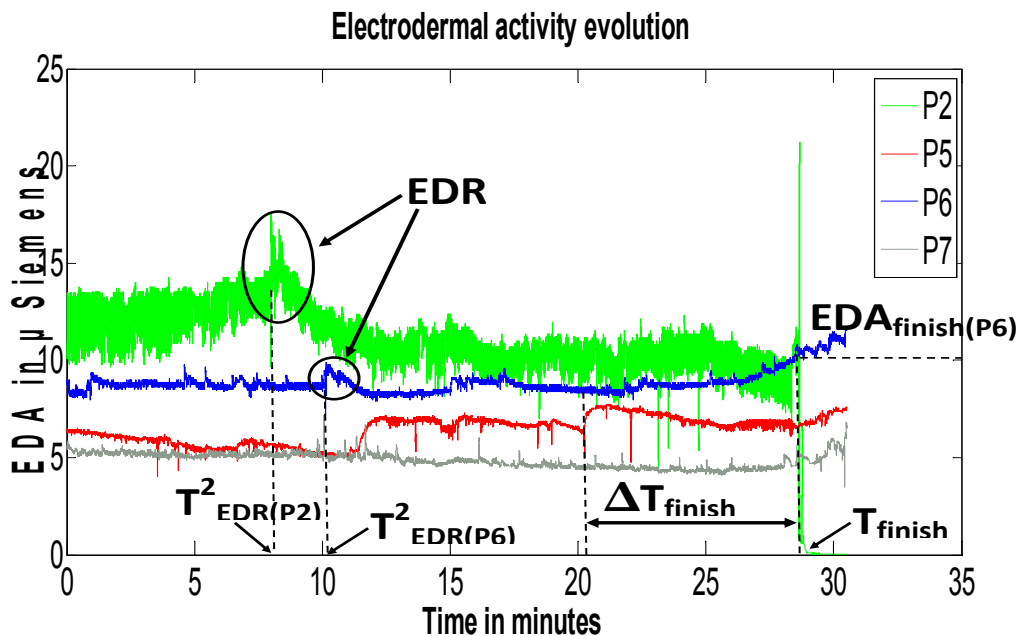


Figure 4.6: Electrodermal activity measure at the finish phase of the competition (semi-marathon). P3’s EDA is not presented due to the presence of many artefacts

4.6 Electrodermal reactions at the start and the finish phases

The time evolution of the EDA either at the start or at the finish phases of the competition shows special events at specific moments.

We noted “special” electrodermal activity between the 15th and the 20th minute from T_{start} is registered for all subjects. The arrows in Fig.4.5 indicate these specific activation. Also in Fig.4.6, at the finish phase of the competition, this phenomenon is observed for some subjects. This is known in the literature as ‘Electrodermal reactions’ (or EDRs).

The time instants when these special events occur are annotated by T_{EDR} . These unknown instants are associated to EDA specific variations. In the literature, we could recognize this description through what is called tonic and phasic variations. Indeed, slow variations are called tonic variations. However, phasic changes or electrodermal reactions (EDR) are due to observable or non observable stimulations and then respectively designated as specific and non specific EDRs [31].

Interestingly, an electrodermal reaction, see Fig.4.7, can be approximated mathematically via several formulas and we select that of [31]:

$$EDR(t > \tau_{lat}) = K \times \exp\left[\frac{-(t - \tau_{lat})}{t_{arousal}}\right] \times \frac{1}{\left[1 + \left(\frac{t - \tau_{lat}}{t_{recovery}}\right)^{-2}\right]^2}$$

where:

- K : gain. It determines the elevation of an EDR.
- τ_{lat} : the time interval is the time that separates the mental activation from its skin manifestation. It is usually of the order of 1 to 2 s [21, 31]
- $T_{arousal}$: It is the time between the beginning of the EDR and the maximum (amplitude). It is between 0.5 to 5 s [21, 31].
- $T_{recovery}$: It is the time needed to return to its initial value before the stimulation which is not easy to find since the return to the initial value does not always occur. This is can for example be due to the succession of EDRs. In that case, we usually consider the time to reach the half of the amplitude or also its quarter. $T_{\frac{1}{2}recovery}$ is in general between 3 to 5 sec.

It was revealed that $T_{arousal}$ and $T_{\frac{1}{2}recovery}$ are correlated (between .54 and 0.8). This can be useful in case of EDR s succession.

The time duration of an EDR is ideally defined by the time between its beginning and the return to the initial value. If the duration is important that means that the stimulation was intense and/or lasted in time.

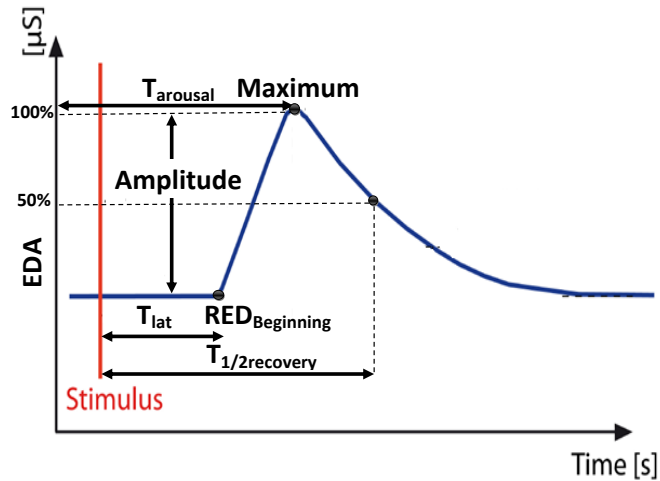


Figure 4.7: An electrodermal reaction (EDR). Illustration of the different time parameters: T_{lat} , $T_{arousal}$ and $T_{\frac{1}{2}recovery}$. Source: [21]

We notice that the shape of the EDR can be found in the EDA signal in three scales namely the overall EDR, the unit of a storm and finally the slow variation of a signal and for instance the signatures of the start and the finish phases already described in a previous paragraph.

4.7 Conclusion

In this study, we aimed to characterize the EDA measurements of runners during the semi-marathon competition in order to shed more light into their performances and profile. Particularly, we focused on the temporal signature of the EDA at the start and at the finish phases of the race.

We were able to identify “key instants” relative to change points of this physiological measure leading to define EDA signatures of the start and the finish phases. Indeed, we were able to describe the start phase by three parameters (ΔT_{stab} , EDA_{start} , EDA_{stab}) characterizing the “transient regime” well documented in physiology. In the same manner, the finish phase of the race was also characterized through (ΔT_{finish} , EDA_{finish}), a time duration that precedes the end of the competition and when we could observe a specific EDA evolution in a number of subjects.

On the other hand, it is worth to stress that throughout the competition, electrodermal reactions (EDRs) were detected reflecting the runner's high physical effort. The analysis revealed differences between subjects and concluded by the choice of a typical EDA profile for a high level runner.

It is also worth to mention that EDA studies in the domain of sport or physical activity are rare. So, the fore mentioned results constitute a first experience in an effort to begin to characterize features of high level runners through EDA tracking. In future works, we can focus on detection automation of these "key instants" and their correlation with physical performance to better evaluate the latter and to insure the safety of athletes and tracking his/her training and competition performances.

Acknowledgement

The authors are grateful to the COMAR⁷ 2012 and Foulées du Lac⁸ 2012 organizers for their permission to lead the measurements and also the reference runner P6, Safa Aissaoui, who is a vice champion in the 2004 world junior championships in athletics (Grosseto Italia 2004) in 3000 m. We would also thank all participants and their clubs.

4.8 References

- [1] Clochard, M., "L'activité électrodermale, technique pertinente pour l'évaluation des émotions?", 2010-2011.
- [2] Picard, R.W., Sano, A., Kaliouby, R., "Palmar vs. Forearm EDA during Natural Sleep at Home," 50th annual meeting of Society for Psychophysiological Research, Portland, OR, USA, Sep.29-Oct.3, 2010.
- [3] Clarion, A., "Recherche d'indicateurs électrodermaux pour l'analyse de la charge mentale en conduite automobile," Thesis presented in 2009, Claude Bernard university-LyonI
- [4] Boucsein, W., "Electrodermal activity," New York: Plenum Press, 1992
- [5] Ghozi, R., Fraj, O., Khalfa, N., Jaidane, M. and Hussein, F., "Presbycusis and Stress Evaluation in Urban Settings," 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, Barcelone, October 2011
- [6] Storm, H., "Changes in skin conductance as a tool to monitor nociceptive stimulation and pain," *Curr Opin Anaesthesiol.*, 21(6):796-804, 2008.
- [7] Sano, A., Picard, R.W., "Toward a taxonomy of autonomic sleep patterns with electrodermal activity," *Conf Proc IEEE Eng Med Biol Soc*, 2011.

⁷The mediterranean company for insurance and reinsurance, www.comar.com.tn

⁸www.lesfouleesdulac.org

- [8] Cheuvront, S.N., and Haymes, E.M., "Thermoregulation and marathon running: Biological and environmental influences," *Sports Medicine*, 31(10):743-762, 2001.
- [9] Bertrand, P.R., Fhima, M., and Arnaud, G., "Off-Line Detection of Multiple Change Points by the Filtered Derivative with p-Value Method," *Sequential Analysis*, 30: 172206, 2011
- [10] Khalfa, N., Bertrand, P.R., Boudet, G., Chamoux, A. and Billat, V. "Heart Rate Regulation processed through wavelet analysis and change detection. Some case studies," *Acta Biotheoretica*, 60 (1-2): 109-29, 2012.
- [11] Fontana, P., Boutellier, U., Knipfli-Lenzin, C., "Time to exhaustion at maximal lactate steady state is similar for cycling and running in moderately trained subjects," *Eur J Appl Physiol.*, 107:187192 DOI, 2009.
- [12] Machado, M., Junior, O., Marques, A., Colantonio, E., Cyrino, E., De Mello, M.T., "Effect of 12 weeks of training on critical velocity and maximal lactate steady state in swimmers," *European Journal of Sport Science*, 11(3): 165-170, 2011.

Chapter 5

Tonic level and phasic activity extraction and motion artefact and special events detection

Preamble

Electrodermal activity (EDA) offers an objective global view of motivation and emotion namely in sport competition [32, 60]. In studying EDA signals, research literature has identified two levels of temporal changes in an EDA measurement: a slowly varying part that reflects the global tendencies and has been known as the *tonic* part of EDA and a rapidly changing component which is related to instantaneous electrodermal reactions due to the internal or external stimuli [21, 22]. This part is known as the *phasic* component in contrast with the tonic component. This has been detailed in Chapter 3. The tonic part of the EDA is usually extracted via a low pass filtering giving an insight into global behavior and general trend of the EDA.

In the previous chapter, thanks to ambulatory recording techniques, it was possible to assess the EDA of marathon runners during a competition which is a rarely explored application. Key phases of a marathon race were depicted at the start and the end of a marathon. As a result, a profile of an elite runner was highlighted through temporal signatures in EDA measurement.

In this chapter, we go further into the analysis of EDA of this population by exploring the tonic level. For that, we propose a particular Intrinsic Mode Functions aggregation of an Empirical Modal Decomposition (EMD) of an EDA signal in order to better extract its tonic level. Moreover, this allows pseudo-periodic motion artefacts reduction. Besides, we compare this artefact removal method to filtering approaches classically used [87, 91].

As a result this approach presents the EMD as a pre-processing step for affective state

change detection. It is worth to note that the case study P presented in this chapter was designated by P5 in the previous chapter (see Table 4.1, Chapter 4).

Except the last section (5.4), this study is an integral part of a conference paper titled "*EMD for Electrodermal Activity tonic level extraction and motion artefact detection*" co-worked by N. Khalifa, F. Mhamdi¹, R. Ghozi, M. Jaïdane. It has been recently submitted to the 3rd International Congress on Sport Sciences Reaserch and Technology Support (icSPORTS 2015, Lisbon).

5.1 Introduction

The autonomous nervous system (ANS) reacts to all stimuli, no matter what their origin is. The electrodermal activity (EDA) is one of its reactions which is totally governed by the sympathetic branch of the ANS [21, 57, 81]. The EDA offers an objective global view of motivation and emotion namely in sport competition [32, 60]. Technically, the EDA refers to the electrical property of the skin and is sometimes used interchangeably with skin conductance (SC), a form of the EDA. It is measured in microsiemens (μS) [47]. Furthermore, the EDA can be quantified through two components: electrodermal activity level (EDL) or *tonic level* and electrodermal reactions (EDRs) or phasic activity. The EDL reflects the overall physiological arousal over a relatively large period of time. On the other hand, the EDR or phasic activity corresponds to short time fluctuations with a characteristic shape [21, 61] happening generally due to an external stimuli. It lasts several seconds ranging from 2.5 sec to 18 sec after a stimuli. Phasic activity reflects the higher-frequency variability of the signal that is modulated on top of the slower changes in tonic level [47].

In this work, we are interested in the evaluation of the tonic level of the EDA signal via a new approach where the signal is decomposed into several components. Then, an aggregation strategy is proposed for tonic part extraction.

Furthermore, we consider the case of an EDA signal of a marathon runner during a competition of long duration. It is worth to note that the EDA tracking for marathoners² remained restricted and the assessment of heart rate (HR) has for long time been the classical main physiological indicator of athlete performance and well being [63]. This can be due to data recording technology that did not mostly afford wireless and non invasive sensor of EDA recordings. In this work, we used a wireless bio sensor for EDA recording practical in the case of runners and proved useful in many applications [91, 81].

¹Signals and Systems Lab, Ecole Nationale d'Ingénieurs de Tunis, Université de Tunis El Manar

²www.scienceofrunning.com: a website belonging to Mr. Steve Magness, author of the book, The Science of Running and writer for Running Times magazine.

Hence, in the sequel, we propose in section 2 a tonic level extraction method using the Empirical Modal Decomposition (EMD) [102]. The case study of the marathon runner is discussed in the third section. Then, main results and challenges on decomposing EDA signal are given in the conclusion.

5.2 EMD for EDA tonic level extraction

In the following, we stress the importance of the tonic level and its significance. We also propose the EMD approach and algorithm and present an aggregation strategy for the purpose at hand.

5.2.1 EDA tonic level extraction

In Fig.5.1, we can see an EDA signal of a marathon runner (who will be studied in detail in next section) and the corresponding 15 sec-segments averaged EDA signal. In fact, averaging is a way to approximate the tonic level. We can note that it describes slow variations of the EDA signal.

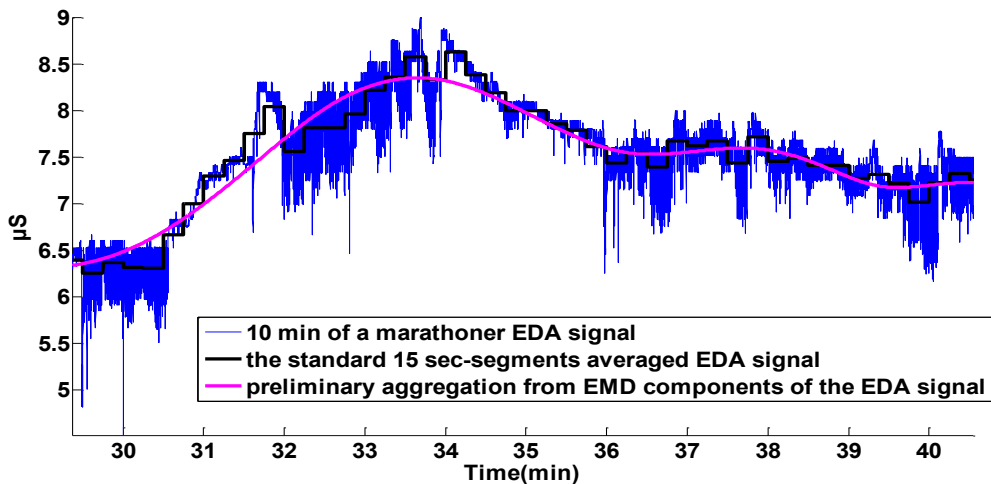


Figure 5.1: 10 minutes of a marathon runner EDA signal during a competition (Blue), the standard 15 sec-segments averaged EDA signal (Black), preliminary aggregation from EMD components of the EDA signal (Pink). More natural EMD result than averaged signal but need ”‘intelligent choice’”

Let us recall that the EDA is the cumulative results due to several (internal and external) stimuli appearing physiologically as an EDR. An EDR can be represented as the response of

5.2. EMD for EDA tonic level extraction

a basic (usually modeled as a first order) system to a simple stimuli $x(t)$. Given the physiological system characteristics (via its impulse response $h(t)$), the EDR is the convolution results of $h(t)$ and $x(t)$, as can be seen in Fig.5.2.

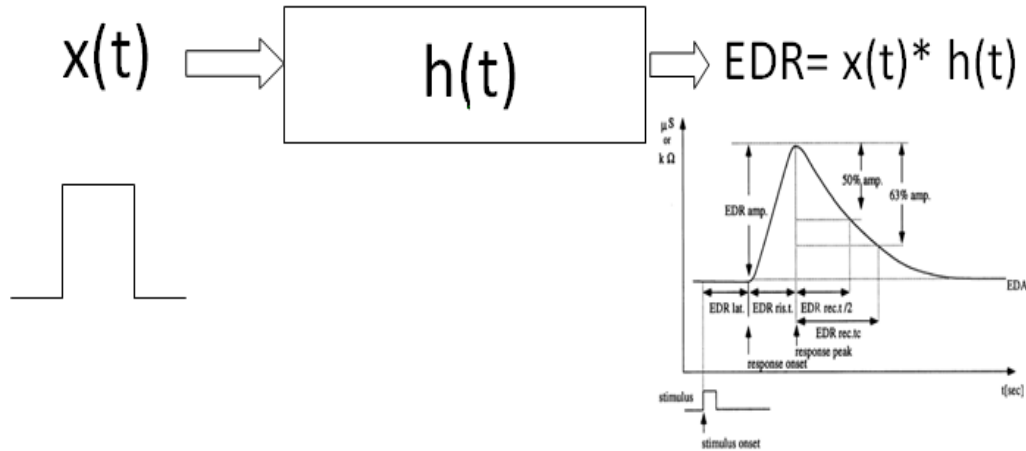


Figure 5.2: Modeling an Electrodermal Reaction (EDR) by a first order system. An EDR is the results of a convolution between the stimulus $x(t)$ and the impulse response $h(t)$. The EDR figure is extracted from [21].

Physiologically, it reflects the level of arousal which can be due to stress, anxiety [31]; it also offers an objective global view of motivation and emotion namely in sport competition [32, 60]. Indeed, a runner need to control his/her stress and motivation so as to re-boost it and achieve better performance.

For tonic level extraction and since slow variations of EDA activity act under 0.5 Hz, low pass filtering is a common approach in recent works as well as for artefact reduction [47]. For instance, the tonic component is quantified by taking the average of EDA values within well determined segments without overlap after removing artefacts. Periods of 10 to 30 seconds are recommended for averaging. For short time durations, the average is affected by EDR (short term variations) and possible numeric artefacts. However, averaging on long observation periods minimizes the effect of these variations [31]. So, averaging presents limitations related to averaging segments durations. On the other hand, the EMD signal analysis allows a multi-level decomposition of non stationary signals without imposing predetermined base as it is the case for multi-scale approaches such as wavelet. Several advantages and applications were reported based on the EMD decomposition like inspecting energy seasonality [73, 74] wherever the signal presents local and global rich dynamics similar to that of EDA signals.

For that, we detail the EMD method in the next paragraph.

5.2.2 EMD approach

The Empirical Mode Decomposition (EMD) has been introduced by Huang [102], as an alternative approach to traditional methods for signals analysis such as wavelets or Fourier methods. The key idea of EMD is to locally decompose a signal $y(t)$ into oscillatory components so-called Intrinsic Mode Functions (IMF). This method is particularly useful to deal with possibly nonstationary and nonlinear components which often characterize time series. It considers the signal as a superposition of oscillatory components, which are extracted from upper and lower envelopes, so-called Intrinsic mode functions. The IMFs are fully data-driven and therefore, the decomposition is signal dependent unlike the Fourier or Wavelet which require a basis for the decomposition.

Therefore, the extraction of IMFs from a given signal $y(t)$ is called sifting and it consists of the following 6 main steps:

- Step 1 Initialize the residue $r_0(t) = y(t)$, set $g_0(t) = r_{k-1}(t)$ and $i = 1$; the index of IMF $k = 1$
- Step 2 Construct the lower minima $Imin_{i-1}$ and the upper maxima $Imax_{i-1}$ envelopes of the signal by the cubic spline method,
- Step 3 Calculate the mean values by averaging the upper envelope and the lower envelope; Set $m_{i-1} = [Imax_{i-1} + Imin_{i-1}]/2$,
- Step 4 Subtract the mean from the original signal: $g_i = g_{i-1} - m_{i-1}$ and $i = i + 1$, and repeat steps (2)-(4) until g_i being an IMF (see below for the definition). In that case, the k^{th} IMF is given by $IMF_k = g_i$,
- Step 5 Update residue $r_k(t) = r_{k-1}(n) - IMF_k(t)$. This residual component is treated as a new data and subjected to the process described above to calculate the next IMF_{k+1} ,
- Step 6 Repeat the steps above until the final residual component $r(t)$ becomes monotone.

It turns out that an IMF satisfy the two following properties. First: the upper and lower envelopes are symmetric and second: the number of zero-crossings and the number of extremum are equal or differ at most by one.

The advantage of the EMD method resides in the fact that the oscillatory modes which are generated, are derived directly from the data, without any reference to a predetermined dictionary of functions.

At the end of this process, the initial time series is decomposed into K IMF's components and $r(t)$ is the final residue:

$$y(t) = \sum_{k=1}^K IMF_k(t) + r(t) \quad (5.1)$$

In Fig.5.1, we superimpose an aggregation of IMFs ($IMF_k, k = K$ to $K - 4$) of the corresponding EMD decomposition of the runner's EDA signal. The aggregation reflects a smooth variation of the overall evolution of the EDA level. It avoids the abrupt changes in the window based average of EDA. So, the question is: which decomposition level fits the best to the tonic level? Hence, comparing the averaged signal to the estimated EMD tonic level could help in deciding of the precision of the decomposition level.

In the following section, we describe an approach for IMF aggregation decision of tonic level extraction based on the comparison with the corresponding averaged signal.

5.2.3 IMF aggregation strategy for estimating EDA tonic component [73, 74]

In order to select appropriate IMF components that constitute the tonic level of an EDA signal denoted by $y(t)$, we compare different aggregations of IMFs to the averaged EDA signal ($y_m(t)$). This last is calculated by processing the mean on successive segments of a duration $\Delta(t)$ such that $\Delta(t) \in [10sec, 30sec]$ without overlap. The comparison between ($y_m(t)$) and ($y_{EMD}(t)$) is made by computing the correlation coefficient between the two approximations. For instance, IMF aggregations corresponds to the following formula:

$$\begin{cases} y_{EMD}(t) = A(i) \\ A(i) = r(t) + \sum_{k=1}^{i-1} IMF_{K-k}(t) \end{cases} \quad (5.2)$$

for $i=1$ to K .

So, for a given m_{EDA} matching with a dt segments period, correlation coefficients for the different aggregations $A(i)$ correspond to :

$$\rho(i) = Cor(y_{EMD}, y_m) \quad (5.3)$$

Selected aggregations correspond to the highest correlation coefficients and are denoted by $A(i^*)$ or $y_{EMD}^*(t)$.

5.3 EDA signal analysis: Case study of a marathon runner

In this section, we consider the case of a marathoner who participated in Foulées du Lac competition³. The participant is a 34 year old woman who ran the semi-marathon (21 km). Her EDA signal, dermal temperature and the movement of her wrist were recorded using a wristband wearable biosensor (the Q sensor)³. Sampling frequency was fixed to 32 Hz.

The participant is denoted by **P** in the sequel. Next, we will see how EMD analysis will not only give an "intelligent" and "intuitive" way of tonic level extraction but also allows EDA artefact motion detection.

5.3.1 EDA tonic level: extraction via EMD components aggregation

The decomposition of the EDA signal of runner P during the competition (race) gives $K = 26$ IMFs and a residue ($r(t)$). In Fig.5.4, we choose to present only the four first IMFs and the three last ones with the residue obtained from the EMD decomposition of the athlete P of his EDA signal.

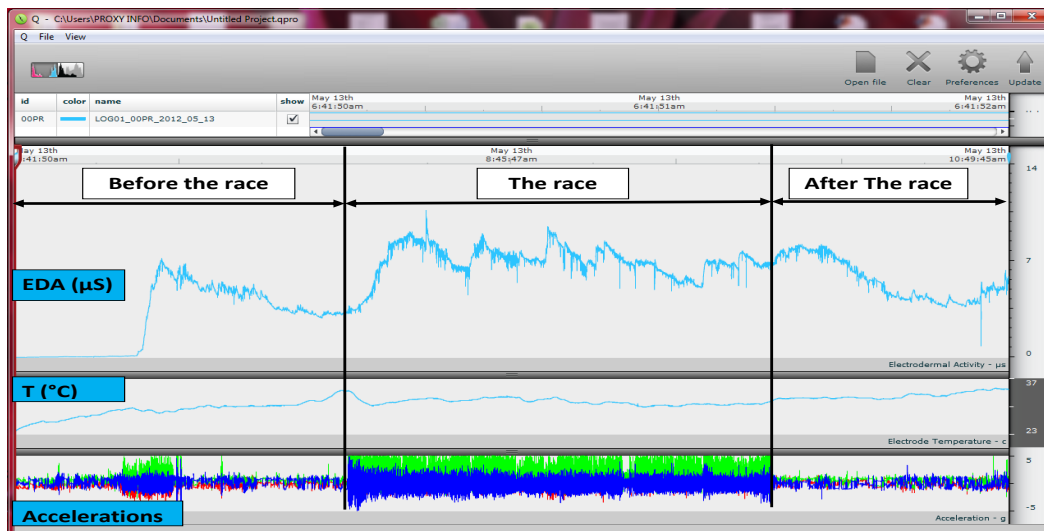


Figure 5.3: Simultaneous measurements of the runner P before, after and during the race (Sensor used on the wrist: The Q sensor). (Top) EDA recording. (Middle) Skin temperature. (Bottom) Wrist accelerations. The race period is well delimited through the accelerations. This study concerns the race period.

³www.lesfouleesdulac.org. Data were collected during the 8th edition on May, 13, 2012.

³The qsensor is designed by Affectiva, a spin-Off of MIT, www.affectiva.com

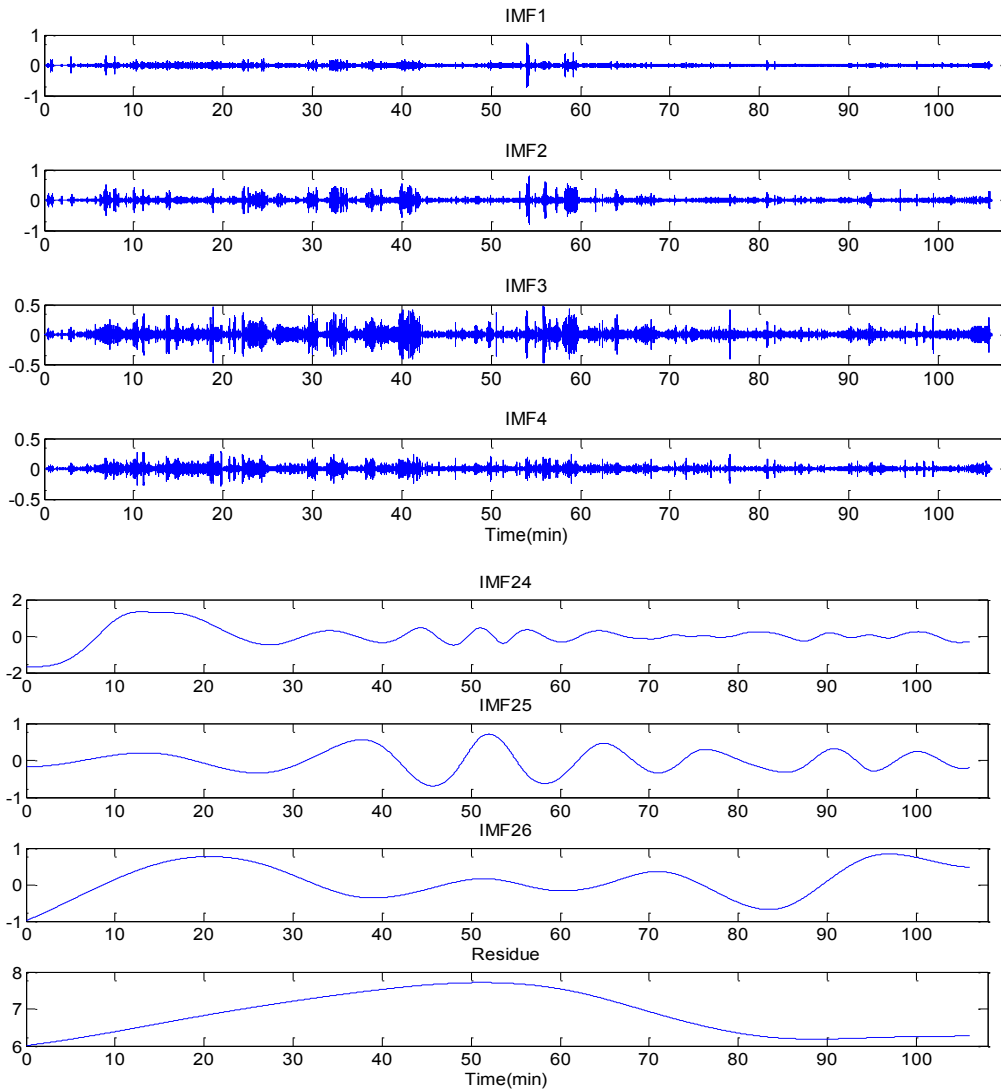


Figure 5.4: Results of the EMD decomposition of the EDA signal of participant P during the race (26 IMFs and a residue $r(t)$). (Top), The first four EMD components (4 IMFs). (Bottom), The last four EMD components (3 IMFs and final residue $r(t)$)

Aggregated tonic component (ATC) for runner P For participant P, we compare different aggregations of IMFs to the averaged EDA signal ($y_m(t)$). For instance, in Fig.5.5, we can see different correlation coefficients computed for $A(i)$ and the 30 sec-segments averaged EDA signal ($y_m(t)$).

By varying the Δt segments duration of averaging from 10 to 30 sec, we note in Table 5.1 best aggregations ($A(i^*)$) corresponding to the highest correlation coefficients as specified in Section 5.2.3. We can notice that aggregation 8 (8 last IMFs of the EMD decomposition plus the residue $r(t)$) is selected for the different Δt segments-duration averaged EDA signal. Hence, we choose this IMFs aggregation, as the aggregated tonic component (ATC). So,

where $ATC = A(8)$.

Averaging window Δt (sec)	Selected aggregations $A(i^*)$
10	A(8), A(9), A(10)
15	A(8), A(9)
20	A(7), A(8), A(9)
25	A(7), A(8)
30	A(7), A(8)

Table 5.1: Selected IMF aggregations of the residue $r(t)$ and the i^* last IMFs for different averaging windows dt (sec). Selected aggregations correspond to highest correlation coefficients and are denoted by $A(i^*)$. Multiple value of i^* are possible.

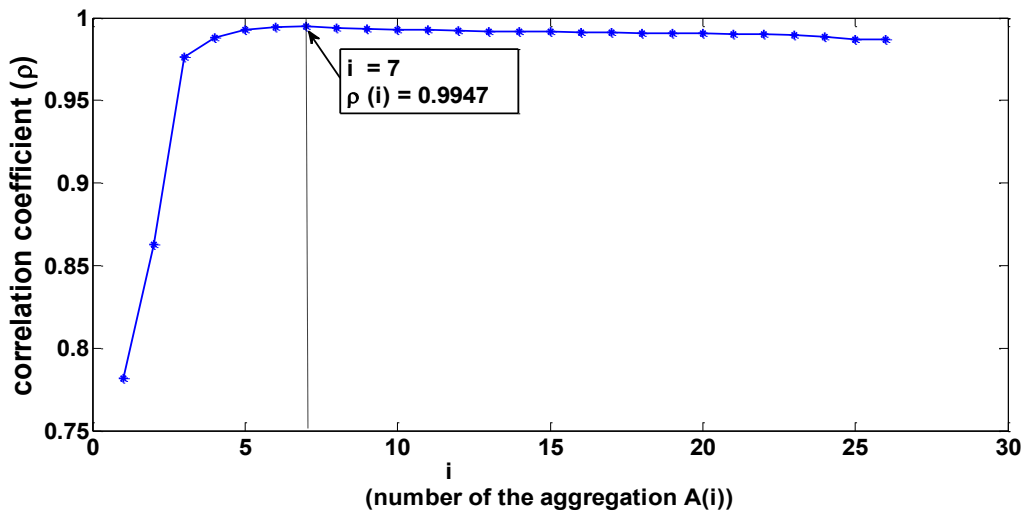


Figure 5.5: Correlation coefficient $\rho(i)$ between different aggregations $A(i)$ and the 30 sec-segments averaged EDA signal $y_m(t)$ for strategic aggregation for the extraction of EDA tonic level. The result is the aggregated tonic component (ATC). Maximum value of correlation with 30 sec-averaged EDA signal is obtained for $A(7)$ and $A(8)$, $i^* = 7, 8$

In Fig.5.6, we superimpose the averaged EDA signal m_{EDA} and the Aggregated Tonic Component (ATC) obtained for the best aggregation. ATC describes the slow variations of the EDA signal without discontinuities and without the delay caused by averaging.

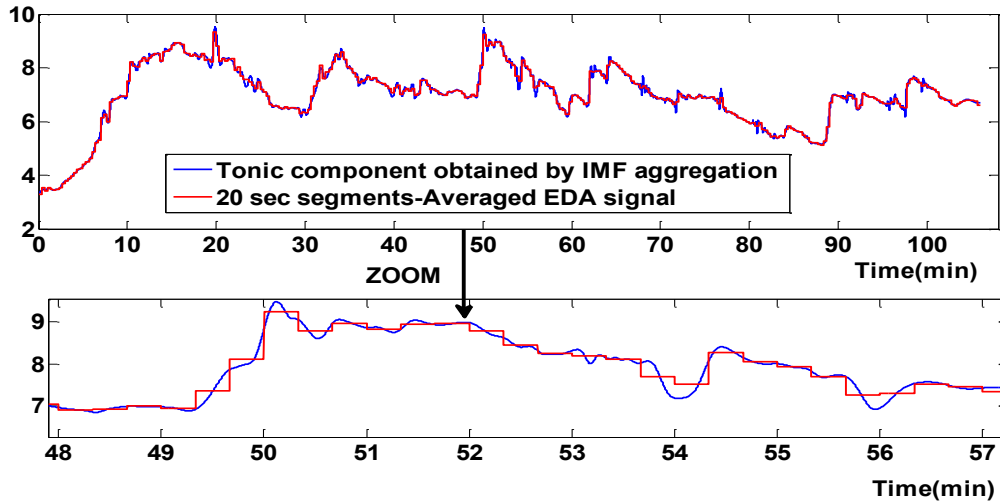


Figure 5.6: Superposition of the Aggregated Tonic Component (ATC) obtained with strategic aggregation of IMFs ($IMF_k, k = 1$ to 8) and the 20 sec-segments averaged of the EDA signal of the runner P

In Fig.5.6, we can note that the aggregation corresponding to the highest correlation, has a maximum overshoot of order 5% of the mean value (around time instant = 50 min) and a maximum undershoot of 3% of the same mean value (around time instant = 56 min).

In the following, we show how to detect athlete motion artefacts via EMD components.

5.3.2 Pseudo periodic artefact detection via EMD components

Pseudo periodic artefact identification We reconsider the EDA data of the same athlete previously mentioned. The signal variations and key moments of the race at the start and finish phases are given in [61]. In particular, we superimposed 8 seconds of the EDA of athlete P before the beginning of the race and 8 seconds after as illustrated in Fig.5.7 (Middle).

5.3. EDA signal analysis: Case study of a marathon runner

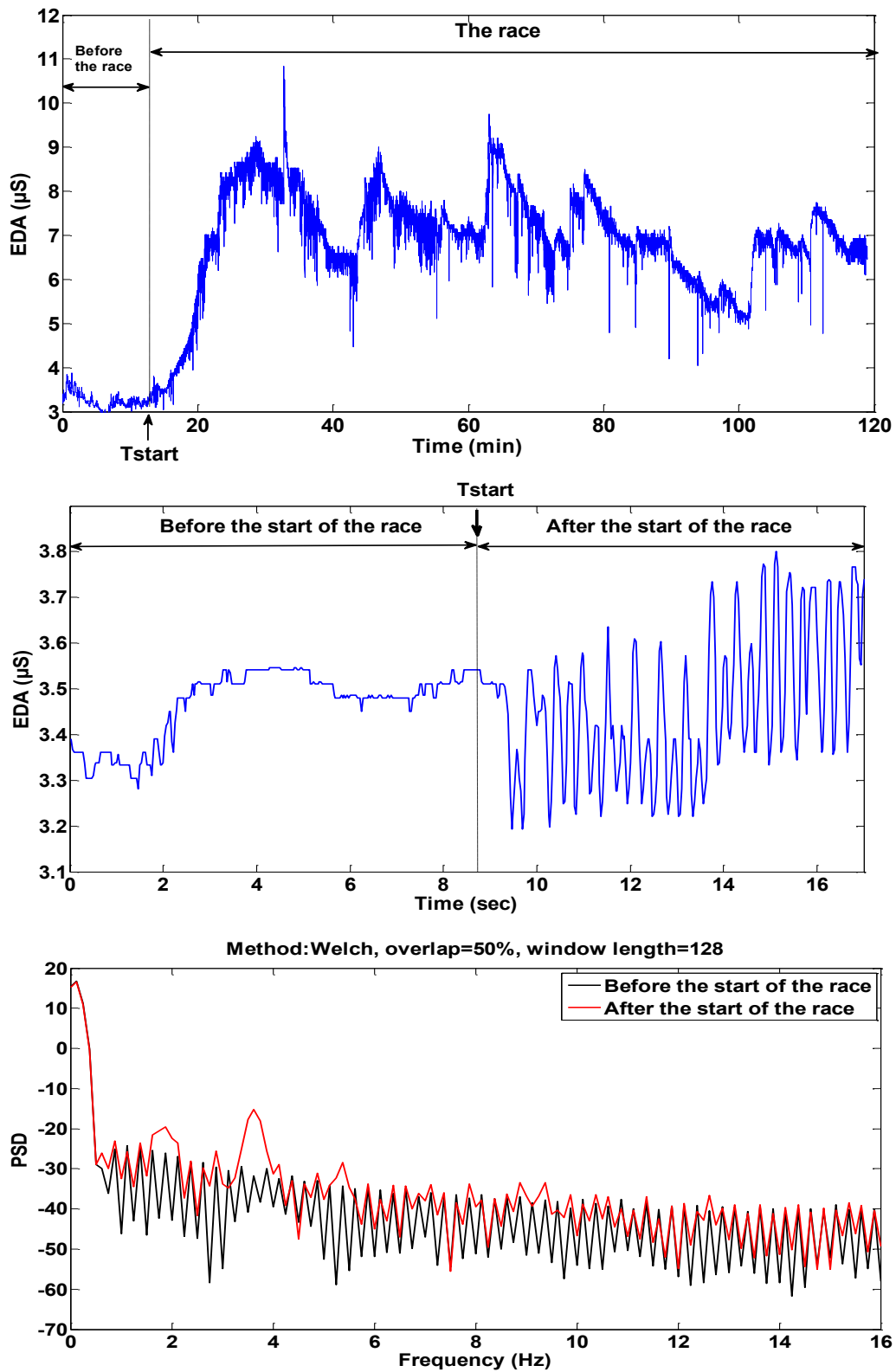


Figure 5.7: (Top), EDA signal of participant P: 13 minutes before the start and during the race. (Middle), Zoom on the 8 first seconds of the race and 8 seconds after the start. (Bottom), Power spectrum density Presence of frequency components due to the oscillations of hand motion

5.3. EDA signal analysis: Case study of a marathon runner

We can notice the presence of oscillations in EDA signal after the race that are not present before. Indeed, in the power spectrum density plot (Fig.5.7 (Bottom)), the spectral analysis confirm such artefacts due to periodic hand movement.

We suspected the movement of the sensor wristband. Hence, we inspected the origin of these oscillations by simulating the movement of the wrist during a race. For that, we carried an experience within which we simulated hand movement similar to the runner's one. The EDA is recorded and shown in Fig.5.8 (Top). We can note the same oscillations as for the runner after the race start in Fig.5.7 (Middle). We also explore the frequency content of the EDA recorded signal in Fig.5.8 (Bottom). Harmonics can be observed on the PSD which refers to oscillation of the hand motion in (Top).

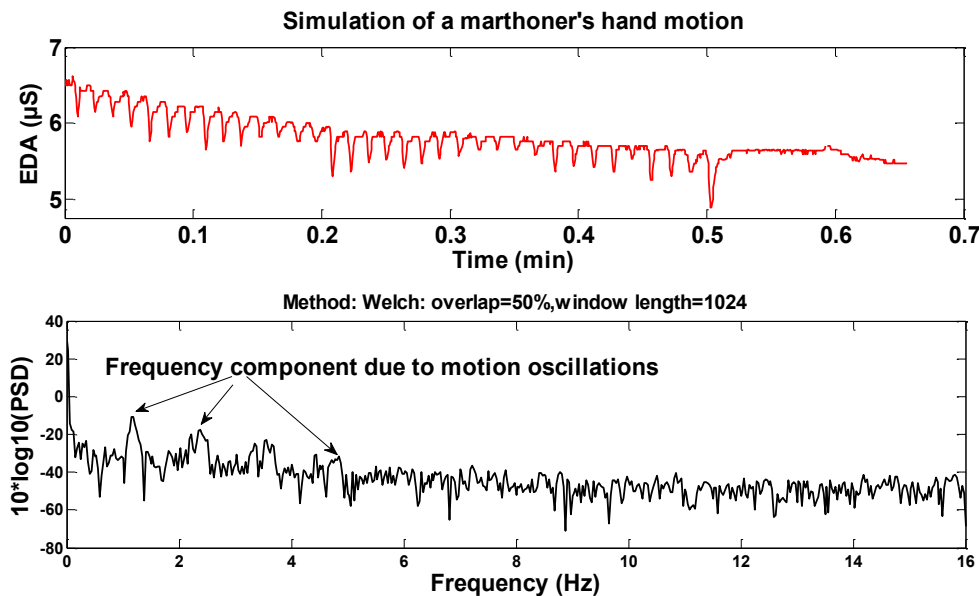


Figure 5.8: (Top), EDA signal during a simulation (of a marathoner's hand motion). (Bottom), PSD of the corresponding EDA signal. Presence of frequency components due to motion oscillations

Hence, these oscillations are not the "natural" variations of a marathoner's EDA during a race and are the most probably due to the marathoner's hand movement. These specific artefacts have to be removed in order to get valuable interpretations.

Now, that we have demonstrated via simulation in real EDA measurement that the regular pseudo-periodic movement induce an EDA artefact, we proceed detection via an EMD. So far, we have presented a way of extracting the tonic part of the EDA from carefully selected IMF aggregation. Let us now examine the content of the remaining EDA's IMF components:

$$X_r(t) = EDA(t) - ATC(t) \quad (5.4)$$

X_r is plotted in Fig.5.9(Top). The power spectrum density of the aggregation of its first fourth IMF components (Fig.5.9(Middle)) show frequency components corresponding to periodic oscillations (non sinusoidal) which refer to oscillating hand movement of the marathoner as we explain previously. The aggregation of the remaining IMFs present a frequency content relatively close to phasic component. We note that motion artefact was not completely removed. This has to be confirmed through further studies.

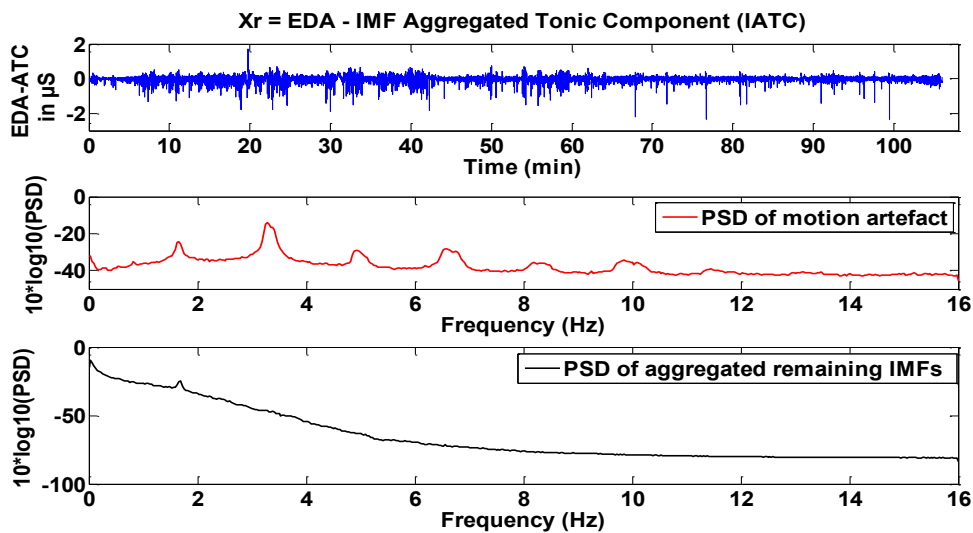


Figure 5.9: (Top), EDA signal of participant P without Aggregated Tonic Component (ATC) obtained by strategic aggregation of IMFs. (Middle), PSD of motion artefact (Aggregation of the first four IMF components, ($IMF_k, k = 1$ to 4)). (Bottom), PSD of the aggregated remaining IMF components ($IMF_k, k = 18$ to 5)

5.4 Change point detection on IMF components of the EDA signal

In this section, we propose to detect changes on the EDA signal of the subject P during the race. We have already noted that the phasic component of the EDA makes change detection with the FDpV method more difficult (see chapter 2). To overcome this issue, filtering as a preprocessing step for change detection seem to be a valuable solution. As far as we are concerned, we propose to process change detection on specific aggregations of the IMF components of the EDA signal.

So, for a duration of 1 h 40 min of a semi-marathon race; we suppose that about 20 state changes can occur. And a state change can occur every 4 minutes. Consequently, the FDpV

5.4. Change point detection on IMF components of the EDA signal

parameters can be fixed as follows:

- The maximum number of change points $K_{max} = 20$
- The sliding window $A = 2 \text{ min}$
- We choose a relatively "low" p-value equal to 0.1

According to sport physiologists, the 20th minute of a race represents a physiologically important event for a runner. In the original EDA signal, see Fig.5.10 (Top), we can verify that at the 20th an EDR appears. So, detecting it automatically would confirm the normal evolution of a runner during a race. For that, we processed FDpV change detection using the same parameters on different forms of the signal: the original EDA signal, the filtered signal (with a 32 order low pass FIR filter and a cut-off frequency of 0.4 Hz as recommended in [81]) and IMF aggregations of the EDA signal namely A(3) having few details and describing the overall EDA variations, A(5) with more details and A(8)); let us recall that A(8) is the IMF aggregation chosen as the tonic level of the EDA signal of the runner P (see Sect.5.3).

Results of FDpV change detection on the original EDA signal, the filtered signal and aggregations A(8) and A(5) are shown in Fig.5.10 below. For A(3), dates of change points are different from the other signals, see 5.11. This can be explained by the lack of precision due to the few details in A(3).

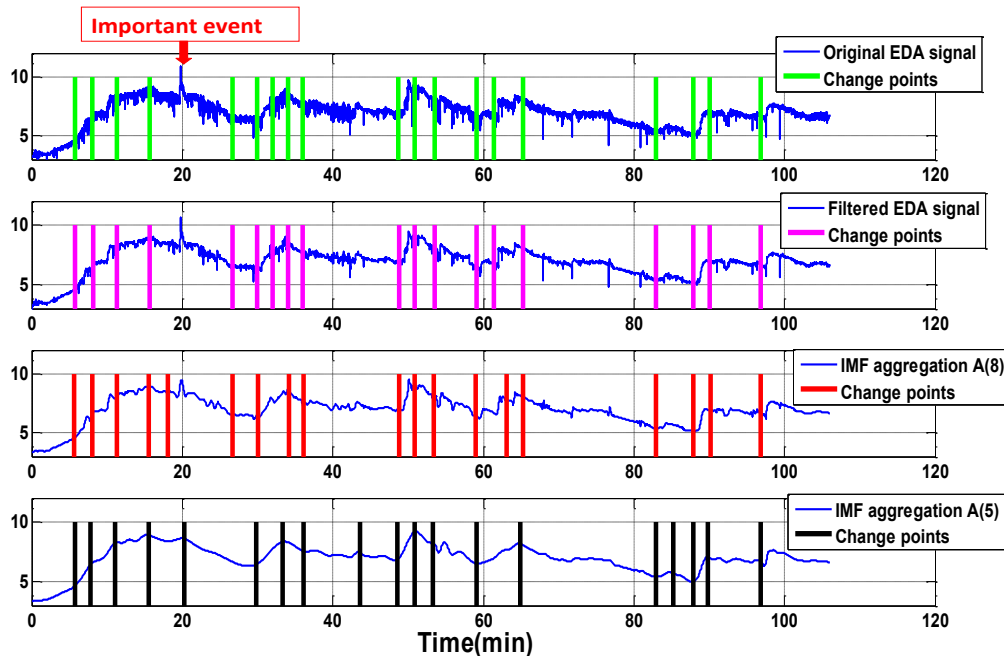


Figure 5.10: Change points detection with FDpV for different signals: original signal, filtered signal, ATC (A(8)) and A(5). FDpV parameters: $A = 2 \text{ min}$, $K_{max} = 20$, $p - \text{value} = 0.1$

5.4. Change point detection on IMF components of the EDA signal

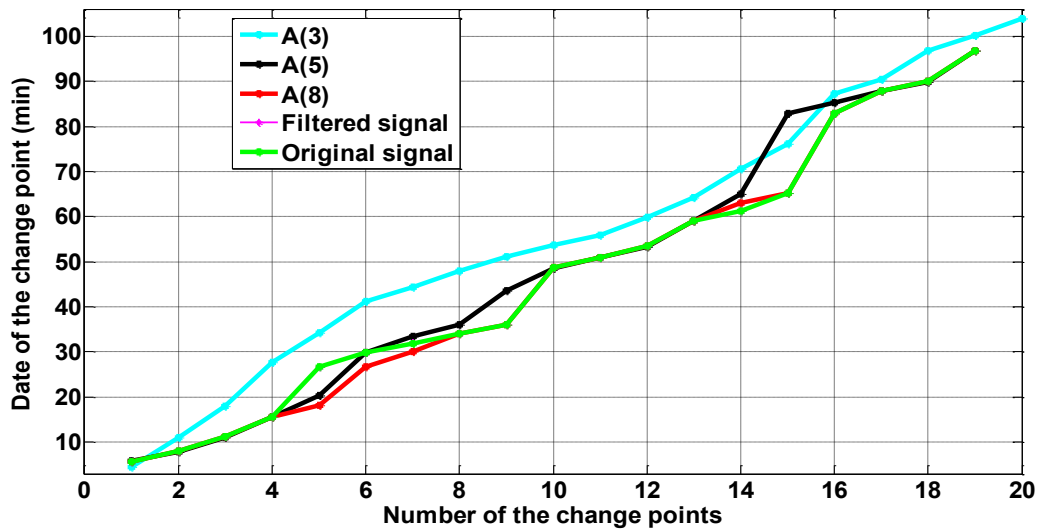


Figure 5.11: Dates of change points obtained by FDpV for different signals: original signal, filtered signal, ATC (A(8)), A(5) and A(3). FDpV parameters: $A = 2min$, $K_{max} = 20$, $p - value = 0.1$

Original Vs filtered signal We can note that the result obtained for the original signal and the filtered signal are very similar. Indeed, in Fig.5.12 (Top), we can see that their lines of change points are mingled. On the bottom of Fig.5.12, the maximum of difference in date is about 17 sec. So, the result of FDpV change detection on low-pass filtered signal is slightly different from that obtained by processing change detection on the original signal.

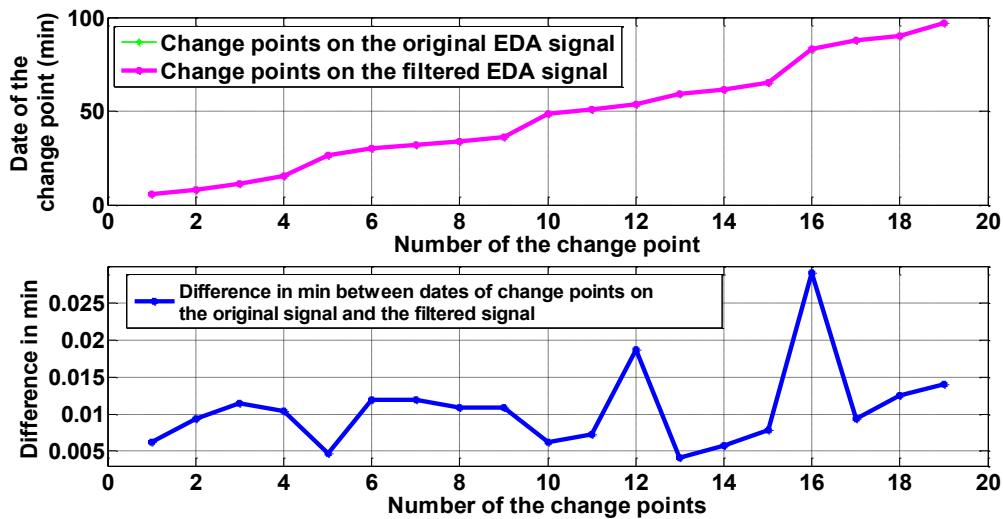


Figure 5.12: Similarity of change points on the original EDA signal and the filtered signal. FDpV parameters for the trend change detection: $A = 2min$, $K_{max} = 20$, $p - value = 0.1$

The 20th minute is detected more precisely in A(5) than in A(8) or the original or filtered signal. The rest of change points are obtained similarly for the different signals.

Consequently, with the same FDpV parameters, we can detect the important dates on A(5) which is not possible to get with the original signal or the classically de-noised signal (with low pass filtering).

Hence the IMF aggregation can represent a pre-processing step for FDpV change detection method. In Fig.5.13, we superimpose the original EDA signal, A(5) and the corresponding change points. The result represents well the main changes of the signal.

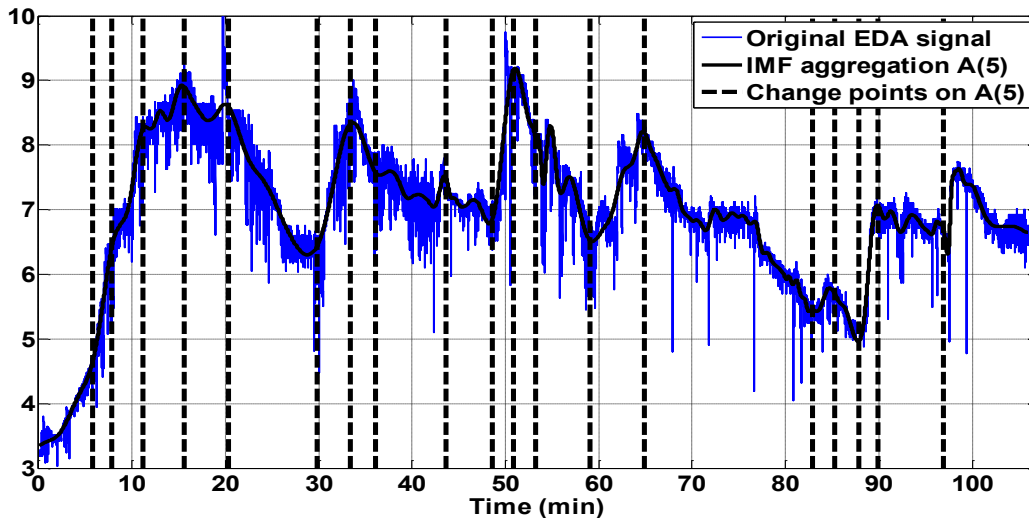


Figure 5.13: Superposition of the Original EDA signal (Blue) and the change points detected on aggregation A(5).

5.5 Conclusion

In this work, we proposed an EMD analysis for the tonic component extraction of the EDA signal. It was built with an aggregation of selected IMFs in accordance with the highest correlation with the global EDA averaging approach. The extracted tonic component has the advantage to be smooth and describes more naturally the overall level of EDA than the corresponding averaged EDA signal.

The other EMD components of the EDA signal revealed also EDA artefact specific to athlete motion. On the other hand, we show that specific IMFs aggregations represent a pre-processing way to obtain "important" change points with the FDpV change detection method. As a perspective, we propose to investigate the Ensemble EMD (EEMD) known for its performance to deal with noisy data [73, 102]. A comparison with this approach and other multiscale resolution techniques would also be of value to the presented work.

Conclusion and perspectives

The main objective of this multidisciplinary thesis was to develop a protocol allowing analysis and detection of physiological state changes in marathon runners based in in vivo settings. Specifically, we consider two signals: a classical one the heart rate (HR) and a rarely explored signal the electrodermal activity (EDA). Such an objective was tackled via the method "Filtered Derivative p-Value" (FDpV) which has the advantage of dynamic tracking of signal statistics, coupled with signal preprocessing.

After presenting an overview of the state of the art on HRV analysis based on non dynamic approaches, we opted for a probabilistic approach for modelling heartbeat series by a piece-wise locally stationary Gaussian process, the main hypothesis behind the FDpV method. Since the Heart Rate Variability (HRV) is usually evaluated from the ration of Low to High Frequency rate changes (LF/HF), we have relied on the wavelet transform in evaluating those spectral energies. In a first stage, our work focused on the application of the FDpV approach for change detection on the spectral density of heart rate time series of seven marathon runners ⁴; the algorithm was also tested and validated on a cohort of five postal service shift workers.

As a result, the first chapter of this work not only introduced the potential of the FDpV method (via illustrative results for HRV signal change point detection), but also raised the importance of proper experimental protocol definition for physiological data acquisition. Moreover, the need to explore another physiological indicator that reflects the motivational and stress level of athletes an addition to the HRV analysis seemed necessary. We also stressed that the HRV gives an insight of the heart functioning while the EDA covers different organs functioning. For that we have opted for the Electrodermal Activity (EDA), confirmed for its ability to measure the arousal levels due to cognitive, physical and psychological states. The interest of the EDA measurement has already been demonstrated [21] but relative embedded sensors recently appeared. So, we could finally explore the contribution of the EDA measurement on marathon runners. As mentioned, there is no analysis to our knowledge on this population but rather those on different application such as autism [21],

⁴Paris marathon 2006

cognition and emotions [87, 32] or sleep [92].

The need to adjust the FDpV algorithm to appropriately detect EDA changes, were an important considerations as a consequence.

In chapter 2, a close investigation of the change detection method (FDpV) via academic sample signals and real case studies (HRV and EDA), was presented. For that, the two main stages in the algorithm were detailed, namely: the Filtered derivative (FD) stage and the p-value evaluation for data mean change detection. One of the main attractiveness of the FDpV algorithm lies in its low computational complexity and parametrization flexibility. Indeed, the FDpV change detection can be processed on the mean, the slope and variance. This allows the possibility to adjust the parameters according to the nature of the signal variances. For instance, the nature of EDA temporal evolution displays clear changes in tendencies, and therefore the EDA change point detection via the FDpV approach had to be performed based on slope changes rather than the mean values (which was the case for the HRV data). The Gaussian and independence assumptions could be asymptotically justified for large enough datasets.

Hence, the original EDA series and the filtered one are processed by FDpV method on the slope. The linear piece-wise model and the T-statistic for slope comparison are then adopted.

In chapter 3, we introduce the EDA signal with its two components: tonic level and phasic activity and their physiological significance. We show that some properties of the phasic activity namely "the storm" feature can alterate change detection result as well as other measurement artefact that can be due to the misuse of the sensor. Linear interpolation and low pass filtering are proposed to tackle these issues. We showed that this preliminary preprocessing must be associated with specific preprocessing of artifacts related to the studied population, namely motion artefacts. Furthermore, we stress the importance of the measurement protocol for EDA assessment within invivo settings in the cohort of marathon runners. Indeed, a pre-test experience was held in Foulées du Lac semi-marathon for controlling the start and finish phases of a competition. After facing the different issues of such an experience, we propose a measurement protocol using a wireless wearable EDA sensor. The second experience was held later during the Comar marathon. This last experience has allowed us to analyse the evolution and structure of a race and compare different runners between them. Besides, it gave us the possibility to identify marathoners' specific artefacts and to process them. This is what we could explain in the following chapters.

Hence, in Chapter 4, we assess the reliability of the electrodermal activity (EDA) to characterize the performance of few amateur and professional Tunisian runners during an annual semi-marathon. We focused on the start and finish phases of the competition. So, we

examine the EDA temporal signature during those key phases. We note that the overall EDA performance tends to be similar for all subjects during the starting phase of the competition. However, the end phase seems to differentiate among them: specifically, we note that with a "reference" subject (winner of the semi-marathon) there is a better management of stress level. Other measurements on this "reference" runner would confirm our observations and other measurements on different elite marathoners can confirm and generalize our results.

More precisely, in chapter 5, we are interested in the tonic part of the EDA since it gives insight on the overall trend and variations which is the concern of the physiologist. Furthermore, the FDpV performances get better when we process it on the tonic component than on the original signal. Low pass filtering was often adopted in extracting the tonic component of EDA signal. Here, we propose a multi-level approach to constructing the tonic part of an EDA signal by adequate strategy based on Empirical Mode (EMD) approach inspired from two references [73, 102]. In particular, we generate the Intrinsic Mode Functions (IMF) components of EDA signal then select the optimal (minimum) number of IMF that is most correlated to the EDA window-based averaged level construction. The correlation coefficient between the last and the different aggregations decides for the selected tonic level. For illustration, we have used the EDA signal of an elite marathon runner who is a vice champion in the 2004 world junior championships in athletics (Grosseto Italia 2004) in 3000 m. The proposed EMD based analysis of EDA signal did not only offered a strategy of tonic level extraction but uncovered with an EDA artefact related to regular athlete motion. Furthermore, the strategy for selecting the tonic component can be reconsidered. Indeed, comparison of the different IMFs' aggregations with the tonic and phasic component obtained from other techniques than averaging can be considered. In fact, we can cite for instance the deconvolution method as described in [13] and that allows to separate both components of an EDA signal. Moreover, we can explore an other extraction technique for tonic and phasic components that is the Ensemble EMD (EEMD) known for its performance to deal with noisy data [73, 102]. Besides, a comparison with this approach and other multiscale resolution techniques would be useful.

In this work, the FDpV segmentation technique was used for change detection state. It offers fast computation and low memory allocation which makes it very practical for embedded sensing. Meanwhile, assumptions of gaussianity, independence of observation and stationarity needed are not always satisfied.

On the other hand, the applicative novelty of this work is dealing with *invivo* situations using wireless technological solutions for physiological assessment. Indeed, elsewhere, studies are generally held in centers or laboratories under controlled conditions, see [50, 52]. This

eliminates the occurrence of spontaneous events and responses and consequently the "real" side of physiology and systems' behavior.

So, in this work different and multidisciplinary aspects of design and analysis of in vivo protocol for runners are treated: measurement protocol and instrumentation, artefact signal identification and processing, change state detection. Besides, applicative study for marathon runners is proposed.

During the thesis work, another application was tackled in the frame of the ANR project "Design of Well Being Monitoring Systems (Do Well B.)" ⁵. It is not deeply discussed in this report for a purpose of coherence.

So, in the following, we give three possible and future perspectives of this work.

- **Study of another population: autistic population**

It is estimated that around 1% of all children is diagnosed with autism, see [64] and [72]. Autism spectrum disorders (ASD) are a family of heterogeneous neuro-developmental disorders characterized by deficits in (a) social interaction, (b) communication, and (c) unusual repetitive behaviors [45] and [58]. It seems quite obvious that the way people with autism process sensory information, sound information in particular, will be among determinants of their quality of life. As many people with an ASD, in auditory stressing situations, are not able to say or explain properly what do they feel and why, it is then relevant to examine physiological parameters involved in usual reaction at stress inducing situations (stressors). The protocol dedicated for marathoners has been generalized to the evaluation of the response to everyday events situations in autistic children. It is approved by Ethics Committee of Clermont- Ferrand (South-East I), France and recently published, see [26]

So, physiological parameters should lead to a better understanding of the way people with autism perceive and react to their environment, and more precisely how they perceive their sound environment.

To illustrate the importance of EDA analysis by means of an application other than marathon runners, in a first experience, we compared physiological reactions of an autistic subject and a control one in real life situation. Indeed, we assessed the EDA of an 18 years old autistic subject S_A during a daylong in his normal activity. S_A was accompanied with his neuphysiologically normal friend S_N of the same age, see the figure below.

⁵ <http://math.univ-bpclermont.fr/DoWellB/index-fr.html>

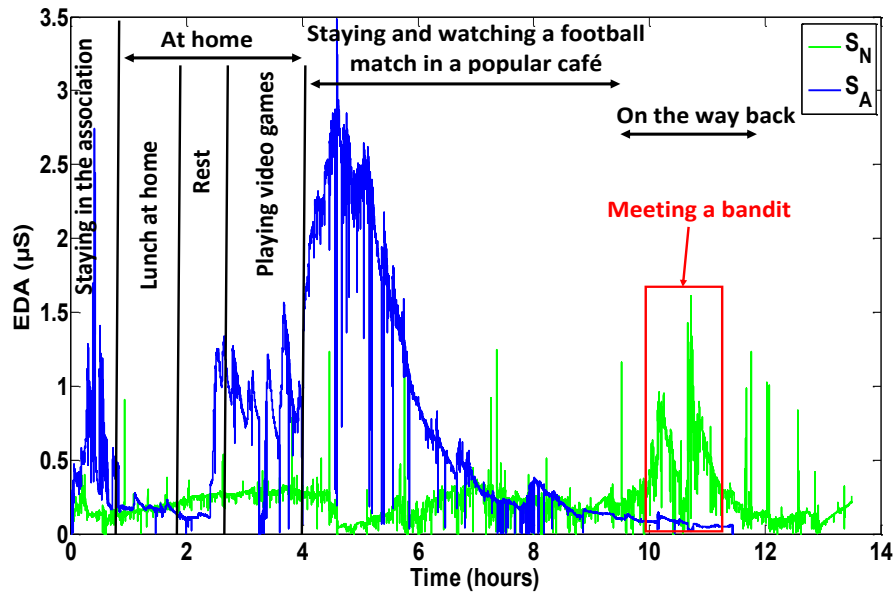


Figure. A daylong simultaneous EDA measurement of an autistic subject S_A and the neuphysiologically normal friend S_N

We can note change points on the mean and the tendency of the EDA signal of S_A high EDA level and variability during the stay in the association, the video game play and while watching the football match in comparison with the control subject S_N , however no behavioral reaction was noted.

Interestingly, that day was marked by “an event”. Indeed, S_N and S_A met a bandit while returning home. We can notice a reaction during this “event” in the EDA signal of S_N .

However, in the EDA recording of the autistic subject S_A , we do not observe this change. We interpret that by the fact that the last did not acknowledge the bandit while subject S_N do and reacted emotionally. Change detection on the slope and the mean is then particularly interesting. Besides, the related artefacts and their pre-processing will also be different. A first oral communication presenting these first results was made ⁶

Lately, we get the idea to focus on the EDA reactions to everyday auditory stimuli in autistic subjects within their learning environment (classroom). Experiences were held in “Centre Cordoue d’éducation spécialisée” in Testour at El Kef which receives children with handicaps namely autism. First results have also been presented ⁷. Besides, further work is being processed and will be published soon.

⁶During the Société de Mathématiques Appliquées et Industrielles (SMAI) 2013 conference, Seignosse Le Penon, Landes, France.

⁷Poster presentation that got the innovation prize in “Colloque Autisme: Prise en charge et intégration

5.5. Conclusion

- For a deeper physiological understanding, multivariate analysis based on the control of other physiological signals (such as EEG or breathing) measured simultaneously with the heart rate (HR) and the EDA is possible. A physiological index summarizing the different information could be deduced . Besides,the pre- processing and change detection algorithms as well as real time optimization of the FDpV parameters can be integrated to the sensor and allow the athlete or his/her coach better monitoring and instant reaction
- In a more general way , the classification of stress and motivation states as well as change detection on physiological signals (EDA, HR and others) can be done in a personalized way and within in vivo settings. This can involve many important applications as important as the autistic children learning faculties or the high-level athletes performance

Appendices

Appendix A

Conception of an experimental protocol

In the field of experimental science, experimentation means scientific testing. It aims to resolve a scientific problem and gives a response through testing an explicative hypothesis. Fig.A.1 Fig.A.2 show the principles of experimentation: it is based on varying what we call “factor(s)” that can act on one or many “parameter(s)” of the studied scientific phenomenon. It tests their causal relationship.

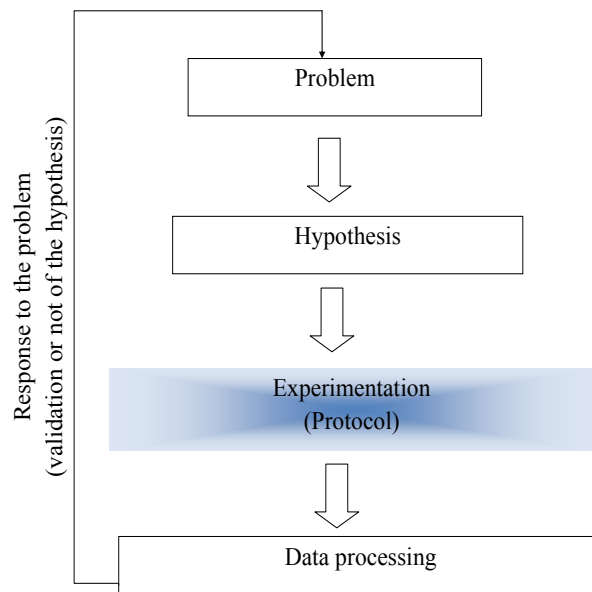


Figure A.1: Resolution of a scientific problem [40]

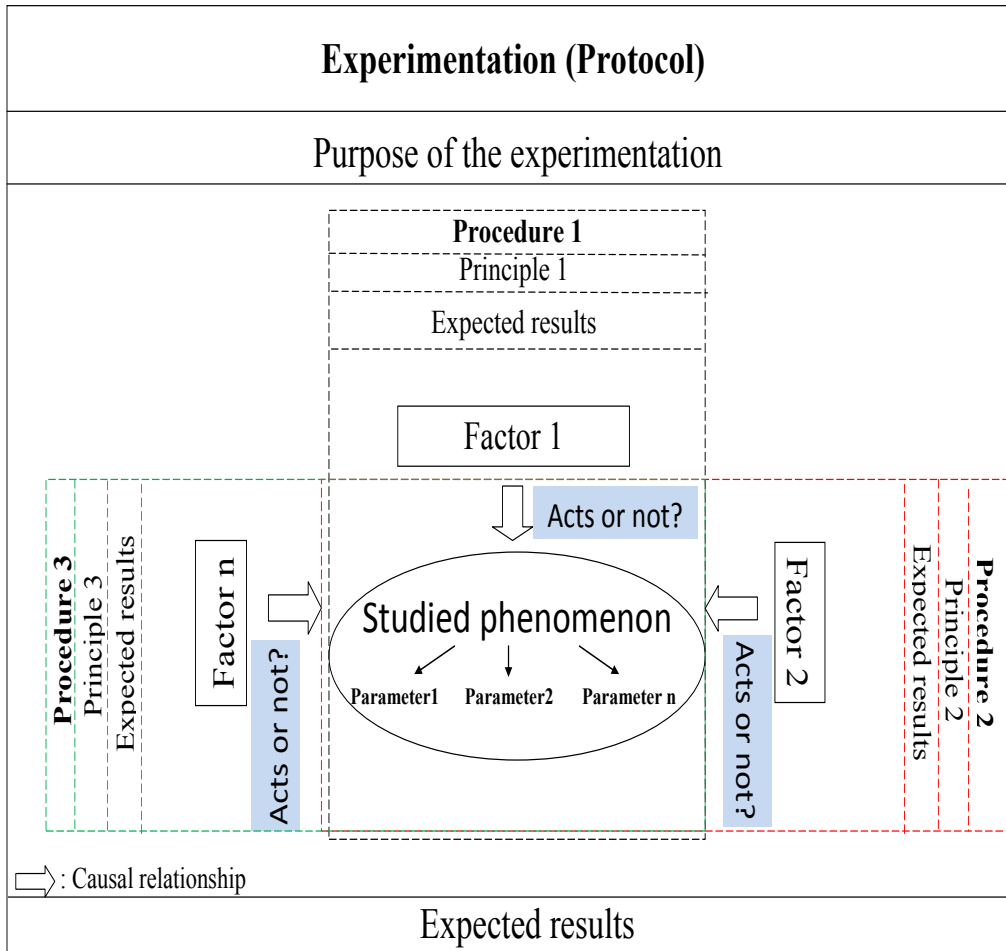


Figure A.2: Basics of an experimentation: does selected factors act on the studied phenomenon?

So, to lead an experimentation, different aspects have to be specified such as factors, equipment, timing or sampling. The formulation of the different steps dealing with all aspects is designated as a *protocol* or a *protocol of data acquisition*.

Definition of a protocol

After our bibliographic research on the subject, we can cite two definitions:

- Definition 1: According to [39], a protocol constitutes the link between hypothesis and the experiment and describes it as a confrontation. On the other side, it helps to attribute a meaning to the experiment: “The activity of designing a protocol is likely to produce a confrontation between theory, problem and experiment. This confrontation is an important way to encourage students to construct meaning and to acquire

knowledge that can be reinvested in practical work.”

- Definition 2: Knowing that a procedure is a series of actions to be executed, an experimental protocol is a procedure or a number of procedures that allows the execution of the experiment [40].

So, in the following, we deal with experimental protocol: its conception (principle of the experimentation, principle of each procedure constituting the experimentation and experience execution) and the ways of its evaluation.

We explain these points in the following paragraphs. Most details are illustrated in FigA.2.

Principle of the experimentation

First, we have to determine *the purpose of the experimentation, the studied phenomenon* and the supposed *causal relation between factors and the phenomenon*. It has to be concise and classed in ascending order. It is a crucial step in the protocol design.

Principle of each experience/procedure

Second, each causal relation to be tested induces an experience. So, for one experience, factors to vary, factors to keep constant and parameters (level, frequency, etc.) to be assessed have to be specified as well as the control case if it exists. The expected result of the possible variation of the observed parameter is described.

Procedure/ experience execution

After getting “theoretical” specification about each procedure, we have to specify *tools* to:

- vary some factors and keep others constant
- assess the observable parameter. The principle of measurement is mentioned.

We have also to mention:

- values of varying factor and the time moments it shift
- values of constant factors
- the duration of the experience
- measurement frequency of the observed parameter
- security measures

Presentation of the protocol

It consists on a schematical presentation of the chosen tool to assess the observable parameter

Evaluation of a protocol

According to [78], the evaluation of a protocol can be done following tasks/actions mentioned in the protocol and parameters throughout four axes:

1. Relevance

It is about evaluation of tasks and theoretical knowledge. We can distinguish:

- External relevance: it means to verify if the theoretical hypothesis is coherent with recorded data.
- Internal relevance: once the parameter to be recorded is chosen, does the strategy adopted in the protocol lead to that parameter?
- Validity domain: adequacy of the protocol with the population in interest.

Error in relevance induce the absence of data.

2. Quality

It concerns the quality of recorded data. It can be assessed through:

- Accuracy: evaluates if the recorded data is close to the correct one
- Fidelity/repeatability

Error in quality induces wrong data.

3. Executability

We can mention two main purposes of executability :

- Evaluation of the description of tasks and parameters in the protocol
- Assessment of the temporal and logical order of the protocol tasks

4. Communicability

It verifies adaptability of the protocol to the person that will execute it through:

- The level of explanation
- Organization of the information
- Type of information (theoretical...)

Appendix B

Participant form

Participant form	Date:
	Operator:
	Event:

Details of the participant:

Reference:

First name:

Last name:

Gender:

Age:

Contact (e-mail/phone):

Details of the sensor(s) (that the participant wears)

Denomination of the sensor(s):

Reference(s):

Appendix C

Sensor form

Sensor form	Date:
	Operator:
	Event:

Details of the sensor

Denomination of the sensor:

Reference:

Details of the participant (who wears the sensor):

Reference:

Last name:

First name:

Bibliography

- [1] Anthony, J., (1958). An experimental approach to the psychopathology of childhood: Autism, *British Journal of Medical Psychology*, Vol. 31, pp 211-225
- [2] Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D., (2003). Self-similarity and long-range dependence through the wavelet lens, in *Long-range Dependence: Theory and Applications*, P. Doukhan, G. Oppenheim and M.S. Taqqu editors, Birkhäuser
- [3] Taylor, A.H., Schatz, S., (2011). Measuring the effectiveness of stress prevention programs in military personnel, *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems Lecture Notes in Computer Science*, Vol. 6780, pp 636-646
- [4] Ayache, A. and Bertrand, P.R., (2011). Discretization error of wavelet coefficient for fractal like process, *Advances in Pure and Applied Mathematics.*, Vol. 2, No. 2, pp 297-321
- [5] Bahri,H., Ghozi, R., Hussein, F. and Malouche, D., (June 2014). Identification et caractérisation des états de stress par une segmentation probabiliste des signaux de l'activité électrodermale, *46ème Journées de Statistique*, Rennes
- [6] Bardet, J.M. and Bertrand, P.R. (2010). A nonparametric estimator of the spectral density of a continuous-time Gaussian process observed at random times, *Scand. J. of Statistics*, Vol. 37, pp 458-476
- [7] Bardet, J.M. and Kammoun, I., (2008). Detecting abrupt changes of the long-range dependence for the self-similarity of a Gaussian processes, *C. R. Math. Acad. Sci. Paris*, Vol. 346, No. 13-14, pp 789-794
- [8] Bardet, J.M. and Kammoun, I. (2008). Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes, *IEEE Trans. Inform. Theory*, Vol. 54, No. 5, pp 2041-2052
- [9] Bardet, J.M. and Surgailis,D., (2011). Measuring the roughness of random paths by increment ratios, *Bernoulli*, Vol. 17, pp 749-780

- [10] Bai, J., and Perron, P., (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, Vol. 66, No. 1 , pp 47-78
- [11] Basseville, M. and Nikiforov, I., (1993). Detection of Abrupt Changes: Theory and Application, *Prentice Hall, Englewood Cliffs, NJ*
- [12] Basseville M. and Nikiforov I., (1993). Detection of Abrupt Changes: Theory and Application,
- [13] Benedek M. and Kaernbach C., (2010). Decomposition of skin conductance data by means of nonnegative deconvolution, *Psychophysiology*, vol. 454, No. 47(4), pp 647-658
- [14] Benveniste, A. and Basseville M., (1984). Analysis and Optimization of Systems, Chapter Detection of abrupt changes in signals and dynamical systems : Some statistical aspects, *Lecture Notes in Control and Information Sciences*, 62, pp 143-155.
- [15] Bertrand, P.R., Fhima, M. and Guillin, A., (2011). Off-line detection of multiple change points with the Filtered Derivative with p-Value method, *Sequential Analysis*, 30 (2), pp 172-206
- [16] Bertrand, P.R., Fhima, M. and A. Guillin (2011). Fast change point analysis on the Hurst index of piecewise fractional Brownian motion, *Proceeding of the 43ème Journées de Statistiques, Tunis 2011*
- [17] Bertrand, P.R, Teyssière, G., Boudet, G. and Chamoux, A., (2009). Detection of Change-Points in the Spectral Density. With Applications to ECG Data, proceedings, EGC'09, Strasbourg, 28-30 janvier 2009. <http://hal.archives-ouvertes.fr/hal-00354416/fr/>.
- [18] Billat, V., Mille-Hamard, L., Meyer, Y. and Wesfreid, E., (2009). Scaling in heart rate and speed dynamics of runners in the marathon race, *Physica A*
- [19] BioHarness BT User Guide, http://www.zephyr-technology.com/media/pdf/BH_UG_P-BioHarnessBTUserGuide_20130121_V03.pdf
- [20] Birgé, L. and Massart, P., (2007). Minimal penalties for Gaussian model selection, *Probab. Theory Related Fields* 138, pp 33-73
- [21] Boucsein, W., (2012). Electrodermal activity, Second Edition, Springer
- [22] Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., Roth, W.T., Dawson, M.E., Filion, D.L., (2012). Publication recommendations for electrodermal measurements, *Psychophysiology* , 49(8), pp 1017-34

- [23] Boudet, G., Albuissou, E., Bedu, M. and Chamoux, A., (2004). Heart rate running speed relationships during exhaustive bouts in the laboratory, *Can. J. Appl. Physiol.* 29(6), pp 731-742
- [24] Brodsky, B. E. and Darkhovsky, B. S., (1993). Nonparametric methods in change-point problems, *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, Vol. 243
- [25] Burch, N., (1965). Data processing of psychophysiological recordings, *Symposium on the Analysis of Central Nervous System and Cardiovascular Data Using Computer Methods*, pp 165-180
- [26] Dutheil, F., Chambres, P., Hufnagel, C., Auxiette, C., Chausse, P., Ghazi, R., Paugam, G., Boudet, G., Khalifa, N., Naughton, G., Chamoux, A., Mermillod, M., Bertrand, P.R., 'Do Well B.': Design Of WELL Being monitoring systems. A study protocol for the application in autism, *BMJ Open* 2015;5:e007716.doi:10.1136/bmjopen-2015-007716
- [27] Chamoux, A., (1984). Le système Holter en pratique, *Médecine du Sport* 58, 43-273, 54-284
- [28] Chen X., Mukkamala R., (2007). Selective Quantification of the Cardiac Sympathetic and Parasympathetic Nervous Systems by Multi-Signal Analysis of Cardio-Respiratory Variability, *Am J Physiol Heart Circ Physiol*, 294, pp 362-371
- [29] Chevront, S.N., and Haymes, E.M., (2001). Thermoregulation and marathon running: Biological and environmental influences, *Sports Medicine*, 31(10), pp 743-762
- [30] Ching E. S. C., Tsang Y.K., (2007). Multifractality and scale invariance in human heartbeat dynamics *Physical Review* E76
- [31] Clarion, A. Recherche d'indicateurs électrodermiques pour l'analyse de la charge mentale en conduite automobile, Thesis presented in 2009, Claude Bernard university-Lyon I
- [32] Clochard, M.,(2011). L'activité électrodermale, technique pertinente pour l'évaluation des émotions? Master Biologie Gestion, Université de Rennes 1
- [33] Cottin, F., Leprêtre, P.M., Lopes, P., Papelier, Y., Médigue, C. and Billat, V., (2006). Assessment of ventilatory thresholds from heart rate variability in well-trained subjects during cycling, *Int. J. Sports Med.*, 27, pp 959-967
- [34] Cramér, H. and Leadbetter, M. R., (1967). *Stationary and related stochastic processes. Sample function properties and their applications*. Wiley and Sons.

- [35] Csörgo, M., Horváth, L., (1997). *Limit Theorem in Change-Point Analysis*, Wiley and Sons.
- [36] Cysarz D., von Bonin, D., Brachmann, P., Buetler, S., Edeläuser, F., Laederach-Hofmann, K. and Heusser P., (2008). Day-to-night time differences in the relationship between cardiorespiratory coordination and heart rate variability. *Physiol. Meas*, 29, pp 1281-1291
- [37] D'Addio, G., Pinna G.D., Maestri, R., Corbi, G., Ferrara, N., Rengo, F., (2004 IEEE Conferences). Quantitative Poincare Plots Analysis Contains Relevant Information Related to Heart Rate Variability Dynamics of Normal and Pathological Subjects, *Computers in Cardiology* , 31, pp 457-460
- [38] Dahlhaus, R., (1997). Fitting time series models to nonstationary processes, *Ann. Stat.*, 25
- [39] D'Ham, C., de Vries, E., Girault, I. and Marzin, P., (December 2004). Exploiting distance technology to foster experimental design as a neglected learning objective in labwork in chemistry, *Journal of Science Education and Technology*, Vol. 13, No. 4
- [40] D'Ham,C., (2009). La construction de protocole expérimental : objet et moyen d'apprentissage, *CRAP-Cahiers Pédagogiques, Dossier "Faire des sciences physiques et chimiques"*, No. 469
- [41] Desobry, F., Davy, M. and Doncarli, C., (2005). An online kernel change detection algorithm, *IEEE Transactions on Signal Processing*, 53 (8), pp 2961-2974
- [42] Douglas, PS., O'Toole ,ML., Hiller, WD., Hackney, K., Reichek, N., (1987). Cardiac fatigue after prolonged exercise. *Circulation* 76, pp 1206-1213
- [43] Djaziri Larbi, S., Ben Ali, F., Jaïdane, M.,(2010). Stationarity Assumption and Frame Segmentation in Objective Quality Evaluation Systems: a Language Dependency. *Audio Engineering Society (AES) 38th int. Conference on Sound Quality Evaluation, Pitea, Sweden*
- [44] Djaziri Larbi, S., Jaïdane-Saïdane, M.,(2005) Audio watermarking: a way to stationarize audio signals *Signal Processing, IEEE Transactions on* , 53(2),pp 816-823
- [45] DSM-IV-R., (1994). American Psychiatric Association,
- [46] Fontana, P., Boutellier, U., Knöpfli-Lenzin, C., (2009). Time to exhaustion at maximal lactate steady state is similar for cycling and running in moderately trained subjects, *Eur J Appl Physiol*, 107, pp 187-192 DOI

- [47] Figner, B. and Murphy, R. O., (2011). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, and R. Ranyard (Eds.), *A handbook of process tracing methods for decision research*, New York, NY: Psychology Press., pp 163-184
- [48] Flandrin, P., (1992). Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. on Inform. Theory*, 38, pp 910-917
- [49] Ghazi, R., Fraj, O., Khalfa, N., Jaidane, M. and Hussein, F., (October 2011). Presbycusis and Stress Evaluation in Urban Settings, *4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, Barcelone*
- [50] Goldfarb, W., (1956). Receptor preferences in schizophrenic children, *Archives of Neurology and Psychiatry*, 76, pp 643-652
- [51] Hogue, CW Jr., Domitrovich, PP., Stein, PK., Despotis, GD., Re L., Schuessler, RB., Kleiger, RE., Rottman, JN., (1998). RR interval dynamics before atrial fibrillation in patients after coronary artery bypass graft surgery, *Circulation*, 98, pp 429-434
- [52] Huikuri, H.V. , Valkama, J.O. , Airaksinen, K.E., Seppanen, T. , Kessler, K.M. , Takkunen, J.T. and Myerburg, R.J., (1993). Frequency domain measures of heart rate variability before the onset of nonsustained and sustained ventricular tachycardia in patients with coronary artery disease, *Circulation* 4, pp 1220-1228
- [53] Huikuri, H.V., Perkiömäki, J.S., Maestri, R., Pinna, G.D.,(2009). Clinical impact of evaluation of cardiovascular control by novel methods of heart rate dynamics, *Philos Transact A Math Phys Eng Sci*, 367, p 1223-38
- [54] Hanley, B., (2015). Pacing profiles and pack running at the IAAF World Half Marathon Championships, *Journal of sports sciences*33 (11).
- [55] Hanley, B. (2013). An analysis of pacing profiles of world-class racewalkers. *International journal of sports physiology and performance*,8 (4), 435-441.
- [56] Haouas, N. and Bertrand, P. R., (2013). Wind farm power forecasting, *Mathematical-Problems in Engineering*
- [57] Hussein, F., Ghazi, R., Bahri, H., (June 2013). Electro-Dermal Activity analysis for a presbycusis population in different urban traffic configurations, *Handicap 2014*, 8 ème Édition, Les technologies d'assistance: de la compensation à l'autonomie
- [58] ICD-10, World Health Organization [WHO], 1992

- [59] Ivanov, C., Amaral, L., Goldberger, A.L., Havlin, S., Rosenblum, M.G., Struzik, Z. R., Stanley, H. E.,(1999). Multifractality in human heartbeat dynamics, *Nature* 399
- [60] Kerr, J. H., (2014) Motivation and Emotion in Sport: Reversal Theory *Psychology Press*
- [61] Khalfa, N., Drissi, S., Ghazi, R.and Jaïdane, M., (May, 2013). Temporal signatures of electrodermal activity for the evaluation of runners' performance: start and finish phases, *The 9 th International Workshop on Systems, Signal Processing and their applications (WOSSPA)*, doi: 10.1109/WoSSPA.2013.6602345
- [62] Khalfa, N. Some techniques for processing heartbeat time series
- [63] Khalfa, N., Bertrand,P., Boudet, G., Chamoux, A. and Billat, V., (2012). Heart Rate Regulation processed through Wavelet Analysis and Change Detection. Some case studies, *Acta Biotheoretica*, Vol. 60, pp 109-129
- [64] Kogan, M. D., Blumberg, S. J., Schieve, L. A., Boyle, C. A., Perrin, J. M., Ghandour, R. M., (2009). Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007, *Pediatrics*, 124, pp 1395-1403
- [65] Kwiatkowski, D., Phillips, P. C. B. , Schmidt, P. and Shin, Y., (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics*, 54, pp 159-178
- [66] Ludeña, C. and Lavielle, M., (2000). The multiple change-points problem for the spectral distribution, *Bernoulli*, 6, pp 845-869
- [67] Lavielle, M. and Moulines, E., (2000). Least-squares estimation of an unknown number of shifts in a time series, *J. Time Ser. Anal*, 21, 1 , pp 33-59
- [68] Laurent, H., and Doncarli, C., (1998). Stationarity index for abrupt changes detection in the time-frequency plane, *IEEE signal processing letters*, Vol. 5, No. 2, , pp 43-45
- [69] Machado, M., Junior, O., Marques, A., Colantonio, E., Cyrino, E., De Mello, M.T., (2011). Effect of 12 weeks of training on critical velocity and maximal lactate steady state in swimmers, *European Journal of Sport Science*, 11(3), pp 165-170
- [70] Mallat, S., (1998). A wavelet tour of signal processing, *Academic Press*
- [71] Maron, M.B., Horvath, S.M., Wilkerson, J.E., Gliner, J.A., (1976). Oxygen uptake measurements during competitive marathon running. *J Appl Physiol* 40, pp 836-838

- [72] Matson, J. L. and Kozlowski, A. M., (2011). The increasing prevalence of autism spectrum disorders, *Research in Autism Spectrum Disorders*, 5, pp 418-425
- [73] Mhamdi, F., Poggi, J.-M. and Jaidane, M., (2011). Trend Extraction for Seasonal Time Series using Ensemble Empirical Mode Decomposition, *Advances in Adaptive Data Analysis*, vol. 3, No. 3, pp 363-383
- [74] Mohamed, O.M., Mhamdi, F. and Jaidane, M., (2009). Multi-scale analysis of the daily peak load based on the empirical mode decomposition, *IEEE PowerTech, Romania, 2009*, pp 1-6
- [75] Montgomery, D., (1997). *Introduction to Statistical Quality Control*, 3rd Edition. John Wiley& Sons, New York.
- [76] Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J., (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a "default mode" of brain function, *Neuroimage*, 22, pp 243-251
- [77] , Nocua, R., (2009). Conception et développement d'un système ambulatoire pour la mesure de l'activité du Système Nerveux Autonome pour la surveillance de personnes âgées à domicile, *Ph.D. thesis, Université Joseph Fourier, Grenoble I*
- [78] Nziengue, F.B., (2007). Critères d'évaluation des protocoles d'expérimentation, *Cas d'un TP de chimie. Mémoire de Master 2 Recherche-Didactique Education et Médiation des sciences*
- [79] Ornitz, E., and Ritvo, E. R., (1976). Medical Assessment. In E. R. Ritvo (Ed.), Autism. Diagnosis, current research and management, *New York: Spectrum Publications*, pp 7-23
- [80] Peng, C.K., Buldyrev, S., Havlin, S.V., Simons, M., Stanley, H.E., Goldberger, A.L., (1994). Mosaic organization of DNA nucleotides, *Phys. Rev, E* 49, pp 1685-1689
- [81] Picard, R.W., Sano, A., Kaliouby, R.el.,(2010). Palmar vs. Forearm EDA during Natural Sleep at Home, *50th annual meeting of Society for Psychophysiological Research, Portland, OR, USA, Sep.29 - Oct.3*
- [82] Picard, F., S. Robin, M. Lavielle, C. Vaisse, and Daudin, J. J., (2005). A statistical approach for array cgh data analysis, *BMC bioinformatics*, 6, pp 26-7
- [83] Polar RS 800CX user guide, http://www.polar.com/support_files/ca-fr/8525736E0043E095C225754300554A40/Polar_RS300X_user_manual_Francais.pdf

- [84] Pichot, V., Gaspoz, J.M, Molliex, S., Antoniadis, A., Busso, T., Roche, F., Costes, F., Quintin, L., Lacour, J.R. and Barthélémy, J.C., (1999). Wavelet transform to quantify heart rate variability and to assess its instantaneous changes, *Journal of Applied Physiology*, 86, pp 1081-1091
- [85] Pikkujämsä, S.M., Mäkikallio, T.H., Airaksinen, K.E., Huikuri, H.V., (2001). Determinants and interindividual variation of R-R interval dynamics in healthy middle-aged subjects. *Am. J. Physiol. Heart Circ. Physiol.* 280(3), pp 1400-1406
- [86] Pitzalis, M.V., Mastropasqua, F., Massari, F., Forleo, C., Di Maggio, M., Passantino, A., Colombo, R., Di Biase, M. and Rizzon, P., (1996). Short- and long term reproducibility of time and frequency domain heart rate variability measurements in normal subjects. *Cardiovascular research* 32, pp 226-233
- [87] Poh, M.Z., Swenson, N.C. and Picard, R.W., (May 2010). A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity, *IEEE transactions on biomedical engineering*, Vol.57, No. 5
- [88] Quintana-Murci, L., Alcais, A., Abel, L. and Casanova, J. L., (2007) Immunology in natura: Clinical, epidemiological and evolutionary genetics of infectious diseases. *Nature Immunology* 8(11), pp 1165-1171 , doi:10.1038/ni1535.
- [89] Q sensor user guide, http://www.qsensortech.com/assets/Q_User_Manual_20130329.pdf
- [90] Sandercock, G.R.H., Bromley, P., Brodie, D.A., (2004). Reliability of three commercially available heart rate variability instruments using short-term (5-min) recording. *Clin Physiol Funct Imaging* 24, pp 359-367
- [91] Sano, A., Picard, R.W., (2011). Toward a taxonomy of autonomic sleep patterns with electrodermal activity, *Conf Proc IEEE Eng Med Biol Soc*
- [92] Sano, A., Picard, R.W., Stickgold, R., (2014). Quantitative analysis of wrist electrodermal activity during sleep, *International Journal of Psychophysiology*, Vol. 94, pp 382-389
- [93] Singh, N., Mironov, D., Armstrong, P W., Ross, A M., Lange, R A., (1996). Heart Rate Variability Assessment Early After Acute Myocardial Infarction, Pathophysiological and Prognostic Correlates *Circulation*, 93, pp 1388-1395
- [94] Surgailis, D., Teyssière, G. and Vaiciulis, M., (2008). The increment ratio statistic, *Journal of multivariate analysis*, Vol 99, pp 510-541

- [95] Talih, M. and Hengartner, N., (2005). Structural learning with time varying components: tracking the cross section of financial time series, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (3), pp 321-341
- [96] Tartakovsky, A., Rozovskii, B., Blazek, R. and Kim, H., (2006). A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods, *IEEE Transactions on Signal Processing*, 54 (9), pp 3372-3382
- [97] Task force of the European Soc. Cardiology and the North American Society of Pacing and Electrophysiology (1996), Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Circulation* 93 (5), pp 1043-1065
- [98] Tsuboi, K., Deguchi, A., and Hagiwara, H.,(2010). Relationship between Heart Rate Variability using Lorenz Plot and Sleep Level, *32nd Annual International Conference of the IEEE EMBS*
- [99] West, J.B., (2013). A strategy for in-flight measurements of physiology of pilots of high-performance fighter aircraft, *J Appl Physiol*, 1, 115(1), pp 145-9
- [100] Wilhelm, F. H.and Roth, W. T., (1996).
Ambulatory assessment of clinical anxiety. In J. Fahrenberg & M. Myrtek (Eds.), Ambulatory assessment: Computer-assisted psychological and psychophysiological methods in monitoring and field studies, *Göttingen:Hogrefe*, pp 317-345
- [101] Yaglom, A.M., (1958). Correlation theory of processes with stationary random increments of order n , *Trans. A.M.S.*, 8, pp 87-141
- [102] Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N., Tung, C.C. and Liu, H.H.,(1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis, *Royal Society London*, vol.454, No. 1971, pp 903-995

Résumé - Abstract

Détection de ruptures de signaux physiologiques en situation *in vivo* via la méthode FDpV :

Cas de la fréquence cardiaque et de l'activité électrodermale de marathoniens

Résumé : Cette thèse a été réalisée dans un cadre pluridisciplinaire alliant protocole expérimental, instrumentation dans des conditions de mesures *in vivo*, détection de ruptures associées à des changements d'états physiologiques et identification et prétraitement d'artefacts de mesures chez des coureurs de marathons. Nous avons considéré l'analyse de la variabilité du rythme cardiaque (VRC) et l'analyse de l'activité électrodermale (AED) pendant une course et lors des phases qui la précèdent et la suivent. La détection de ruptures de moyenne et de tendance est effectuée par la méthode Filtered Derivative with p-Value (FDpV) tout au long de cette thèse. La méthode FDpV est basée sur une analyse dynamique faisant appel à un modèle stationnaire par morceaux. Elle a permis en particulier d'introduire un indice de régulation cardiaque pour les coureurs. Un suivi des changements des états d'éveil et de motivation à travers l'AED par la détection de ruptures de la tendance pendant la course d'un semi-marathonien est également proposé. Ceci a permis de définir des signatures de début et de fin de course. Une attention particulière a été apportée à la composante tonique de l'AED reflétant le niveau d'activation affectif. Nous avons comparé trois méthodes d'extraction du niveau tonique en tenant compte des artefacts potentiels présents. Ce travail concerne des études de cas ; il peut être étendu à une cohorte et englober plus de paramètres physiologiques (VO_2 , EEG,...). Ainsi, une classification des états d'éveil et de motivation peut être envisagée et représente des éléments significatifs de caractérisation des données physiologiques *in vivo* pour l'optimisation des performances sportives.

Mots clés : Marathonien, Protocole expérimental *in vivo*, Rythme cardiaque, Activité électrodermale, Détection de ruptures par la méthode FDpV, Motivation, Performance

Abstract: This thesis was carried out in a multidisciplinary approach that combines experimental protocol, instrumentation, *in vivo* measurements, physiological change detection instants and identification and preprocessing of measurement artefacts for marathon runners. We considered the analysis of the heart rate variability (HRV) and the electrodermal activity (EDA) recorded during a semi-marathon including pre and post competition periods. A study of the HRV and EDA change detection was carried based on the mean and the trend using the Filtered Derivative with pValue method (FDpV) throughout this thesis. This segmentation method is based on a dynamical approach using a piece-wise stationary model. As a result, it allowed to introduce an index of cardiac regulation for semi-marathon runners. Physiological state changes tracking of affective dimension i.e. "stress" and motivation via the EDA by change detection on its tonic component which reflects the EDA general trend throughout a semi-marathon was also proposed. This enabled us to characterize start and finish phases of a race which are key elements in any competition. A special attention was given to the tonic component of the EDA reflecting the overall level of affective activation. We compared three methods of tonic level extraction by taking into account the present potential artefacts. This work focused on case studies; It can be generalized over a cohort and include more physiological parameters such that VO_2 or EEG. Hence, a classification of stress states may also be considered and represent other significant features for characterizing *in vivo* physiological data for sport performance optimization.

Key words: Marathoners, In vivo experimental protocol, Heart Rate, Electrodermal activity, Change detection by FDpV method, Motivation, Performance