Acoustic analysis of the knee joint in the study of osteoarthritis detection during walking

by
Costas Yiallourides

A Thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy of Imperial College

Communications and Signal Processing Research Group
Department of Electrical and Electronic Engineering
Imperial College London
November, 2019

Copyright Declaration

Declaration of Originality

I hereby declare that this thesis and the work herein detailed, was composed and originated by myself, except where appropriately referenced and credited.

Costas Yiallourides

Acknowledgments

Several groups and individuals supported my efforts in the development of this dissertation.

First, I wish to express my sincere gratitude to my supervisor, Prof Patrick A. Naylor, for the advice he has given me over the past years but also for all the comments and suggestions during the preparation of this manuscript. His expertise, encouragement and support made this thesis possible and equipped me with great knowledge and skills.

Second, I would also like to thank my examiners, Prof David Simpson and Dr Tania Stathaki for their critical feedback and suggestions.

Third, as with all the endeavours which I undertake, friends and colleagues provided essential support. It is impossible to list all of them here, but I wish to mention a few: Fani Iakovou, Constantinos Papayiannis, Sina Hafezi and Hamza Javed.

Last but not least, I would like to acknowledge the immeasurable and unconditional support I received from my parents Andreas and Chrystalla, and my two sisters Georgia and Elena. I dedicate my dissertation to them.

Abstract

This thesis investigates the potential of non-invasive detection of knee Osteoarthritis (OA) using the sounds emitted by the knee joint during walking and captured by a single microphone. This is a novel application since, until now, there are no other methods that considered this type of signals. Clinical detection of knee OA relies on imaging techniques such as X-radiology and Magnetic Resonance Imaging. Some of these methods are expensive and impractical while others pose health risks due to radiation. Knee sounds on the other hand may offer a quick, practical and cost-effective alternative for the detection of the disease.

In this thesis, the knee sound signal structure is investigated using signal processing methods for information extraction from the time, frequency, cepstral and modulation domains. Feature representations are obtained and their discriminant properties are studied using statistical methods such as the Bhattacharyya distance and supervised learning techniques such as Support Vector Machine. From this work, a statistical feature parameterisation is proposed and its efficacy for the task of healthy vs OA knee condition classification is investigated using a comprehensive experimental framework proposed in this thesis.

Feature-based representations that incorporate spatiotemporal information using gait pattern variables, were also investigated for classification. Using the waveform characteristics of the acoustic pulse events detected in the signal, such representations are proposed and evaluated. This approach utilised a novel stride detection and segmentation algorithm that is based on dynamic programming and is also proposed in the thesis. This algorithm opens up potential applications in other research fields such as gait analysis.

Contents

List of Figures

List of Tables

Abbreviations

AE Acoustic Emission . . . 36

AUC Area under the curve . . . 16

CART Classification and Regression Tree . . . 82

CDF Cumulative Distribution Function . . . 68

DCT Discrete Cosine Transform . . . 65

DFT Discrete Fourier Transform . . . 48

DP Dynamic Programming . . . 112

FFT Fast Fourier Transform . . . 50

FIR Finite Impulse Response. A filter whose output is a weighted sum of past input values and whose system function contains only zeros and no poles. . . . 57

HPF High-Pass Filter . . . 56

ICA Independent Component Analysis . . . 180

KST Kolmogorov-Smirnov Test . . . 71

LP Linear Parameter . . . 118

LPC Linear Predictive Coding . . . 117

MA Moving Average . . . 56

MCC Matthew's Correlation Coefficient . . . 79

MFCC Mel-frequency Cepstral Coefficients . . . 42

OCA Open Chain Activity . . . 31

PCA Principal Components Analysis . . . 180

PSD Power Spectral Density . . . 50

RBF radial basis function . . . 117

RHS Right-Hand Side 68

RMS Root Mean Square 56

RMSE Root Mean Square Error 56

ROC Receiver Operating Characteristic 91

SNR Signal-to-Noise Ratio 112

STFT Short Time Fourier Transform 50

SVM Support Vector Machine 82

ADL Activity in Daily Living 55

ASDSA Automatic Stride Detection and Segmentation Algorithm 15

CC Cepstral Coefficient 85

CDF Cumulative Distribution Function 68

CI Contact Instant 115

CT Computed Tomography 35

DP Dynamic Programming 112

DR Detection Rate 145

FAR False Alarm Rate 145

FIR Finite Impulse Response 56

GD Group Delay 119

GRF ground reaction force 44

ICI Initial Contact Instant 113

KOOS Knee injury and Osteoarthritis Outcome Score 54

KL Kellgren-Lawrence 34

kNN k-Nearest Neighbors 37

LCI Last Contact Instant 113

LP Linear Prediction 118

LFCC Linear Frequency Cepstral Coefficient 89

MCI Muscle Contraction Interference 50

MEMS microelectromechanical systems 37

MR Miss Rate 145

MRI Magnetic Resonance Imaging 30

MVN multivariate normal 78

NZC negative-going zero crossings 120

TC threshold crossings 174

TCR threshold crossing rate 174

VAG Vibroarthrography 36

PAG Phonoarthrography 35

OA Osteoarthritis 5

OAH Ostearthritis Healthy 43

LDA Linear Discriminant Analysis 78

QDA Quadratic Discriminant Analysis 78

QOL Quality of Life 55

RFE Recursive Feature Elimination 184

List of Symbols

A note on mathematical symbols: Lower case letters are used to define random variables (e.g. f). Bold lower case letters are used to define vectors (e.g. $\boldsymbol{f}$). Bold upper case letters are used to define matrices (e.g. $\boldsymbol{F}$). Slant upper case letters are used to define matrix operators and other general operators (e.g. $\mathcal{F}$). Both lower and upper case letters are used to define constants and equation parameters (e.g. f, F).

$s_i(n)$	Sampled knee acoustic emission signal of the i^{th} knee. 63, 71, 87, 194
n	Discrete time index. 63, 64
τ_s	Window in seconds within which the sounds related to abnormalities appear in a knee acoustic emission signal. 63, 67, 70, 71, 79, 87, 91, 92, 179, 181, 192, 194
$s_{i,j}(n)$	Segment j of the knee acoustic emission signal of the i^{th} knee. 63, 66, 67, 69, 70, 71, 77, 84, 86, 87, 88, 91, 104, 179, 180, 181, 185, 188, 191, 192, 194
l	Size in ms of a general frame of a signal. 63, 71, 88, 92, 93, 94, 95, 96, 97, 98, 100, 101, 103, 104, 105, 106
N_f	Number of frames in a signal. 63, 64, 67, 68, 88, 89
l_n	Size in samples of a general frame of a signal. 63, 64, 88, 89
$\boldsymbol{S}$	Matrix of frames from a sampled knee acoustic emission signal. 63, 88
$\boldsymbol{H}$	Hanning window diagonal matrix. 63

$\boldsymbol{W}$	Vandermonde matrix for the roots of unity. 63, 64
π	The ratio of the circumference of a circle to its diameter (mathematical constant). 64
i	Imaginary unit. 64
k	Frequency index. 64, 120
$\boldsymbol{\Psi_F}$	Magnitude of the first half of the points from the discrete fourier transform of $\boldsymbol{S}$. 64, 65, 70, 71, 73, 76, 77, 85, 88, 89, 198
$\boldsymbol{\Psi_{FF}}$	Magnitude of the modulation domain of $\boldsymbol{S}$. 64, 70, 71, 73, 74, 76, 77, 85, 198
T	Number of acoustic time frames used in computing the modulation magnitude spectrum. 64, 71
N_B	Number of filters in a filter-bank. 65, 71, 77, 89, 91, 92, 93, 100, 106, 107
$\boldsymbol{U}_M$	Filter-bank matrix of triangular filters equally spaced along the mel-frequency axis (in columns). 65
$\boldsymbol{\Psi_D}$	Magnitude of the discrete fourier transform of $\boldsymbol{S}$ multiplied with a filter-bank matrix with triangular band-pass filters equally spaced on the mel-frequency axis. 65, 89, 90, 106
$\boldsymbol{C_M}$	MFCC feature vector of $\boldsymbol{\Psi_D}$. 65, 70, 76, 77, 85, 90
$\mathcal{F}$	Discrete Cosine Transform operator. 65, 89
$s_0(n)$	High pass filtered version of $s(n)$. 66, 67, 69, 173, 174
η_0	Acoustic emission pulse definition amplitude. 66, 67, 68, 69, 70
t_0	Acoustic emission pulse definition duration. 66, 67, 69, 71, 174
r_0	Acoustic emission pulse definition rest time. 66, 67, 69, 71, 174

F_s	Sampling frequency. 69, 71, 101, 104, 133, 138, 147, 149, 161, 163, 166, 174
H_0	Null Hypothesis for statistical tests. 71, 72, 76, 136, 137, 178
B_d	Bhattacharyya distance. 73, 74, 75, 76
B_c	Bhattacharyya coefficient. 73, 74, 75
Γ	Gamma function. 74, 75
$\mathcal{D}$	Operator that acts along the columns of a matrix and extracts 11 statistical parameters. 77, 89, 194
$\boldsymbol{\phi_M}$	Statistical feature vector of $\boldsymbol{C_M}$. 77, 78, 90
$\boldsymbol{\phi_F}$	Statistical feature vector of $\boldsymbol{\Psi_F}$. 77, 78, 90, 97
$\boldsymbol{\phi_{FF}}$	Statistical feature vector of $\boldsymbol{\Psi_{FF}}$. 77, 78
$\boldsymbol{\phi_A}$	Statistical feature vector of the acoustic emission pulse matrix. 77, 78, 194
$\boldsymbol{\Phi_M}$	Matrix comprised of a collection of $\boldsymbol{\phi_M}$ vectors for a number of knee acoustic emission signal segments. 78, 80, 85, 86, 90, 92, 93, 94, 95, 101, 103, 105, 106, 107, 109, 189
$\boldsymbol{\Phi_F}$	Matrix comprised of a collection of $\boldsymbol{\phi_F}$ vectors for a number of knee acoustic emission signal segments. 78, 80, 85, 86, 90, 93, 97, 98, 100, 103, 105, 106
$\boldsymbol{\Phi_{FF}}$	Matrix comprised of a collection of $\boldsymbol{\phi_{FF}}$ vectors for a number of knee acoustic emission signal segments. 78, 80, 85
$\boldsymbol{\Phi_A}$	Matrix comprised of a collection of $\boldsymbol{\phi_A}$ vectors for a number of knee acoustic emission signal segments. 78, 85
E_r	Error rate. 79, 80, 81, 82, 84, 87, 91, 103, 109, 182, 197

$F_{0.5}$ Variation of F_1 which weighs recall lower than precision. 79, 80, 81, 82, 84, 87, 91, 103, 104, 109, 182, 197

F_1 Harmonic mean of precision and recall. 79

S_c Classification metric computed as the average of E_r, $F_{0.5}$ and MCC. 80, 81, 82, 83, 84, 85, 92, 103, 104, 106, 107, 110, 182, 186, 188, 189, 190, 194, 196, 197

θ_e Threshold value for the error rate. 82, 83, 92

$\theta_{0.5}$ Threshold value for the $F_{0.5}$ score. 82, 83, 84, 92

θ_{mcc} Threshold value for the Matthew's Correlation Coefficient. 82, 83, 84, 92

$\boldsymbol{U}_L$ Filter-bank matrix of triangular filters linearly spaced in frequency (in columns). 89

$\boldsymbol{\Psi_E}$ Magnitude of the discrete fourier transform of $\boldsymbol{S}$ multiplied with a filter-bank matrix with triangular band-pass filters linearly spaced in frequency. 89, 90

$\boldsymbol{\phi_E}$ Statistical feature vector of $\boldsymbol{\Psi_E}$. 89, 90, 97

$\boldsymbol{\phi_D}$ Statistical feature vector of $\boldsymbol{\Psi_D}$. 89, 90, 97

$\boldsymbol{C_L}$ LFCC feature vector of $\boldsymbol{\Psi_E}$. 89, 90

$\boldsymbol{\phi_L}$ Statistical feature vector of $\boldsymbol{C_L}$. 89, 90

$\boldsymbol{\Phi_E}$ Matrix comprised of a collection of $\boldsymbol{\phi_E}$ vectors for a number of knee acoustic emission signal segments. 90, 92, 93, 97, 98, 100, 101, 103, 105, 107

$\boldsymbol{\Phi_D}$ Matrix comprised of a collection of $\boldsymbol{\phi_D}$ vectors for a number of knee acoustic emission signal segments. 90, 92, 93, 97, 98, 100, 101, 103, 105, 107

$\boldsymbol{\Phi_L}$ Matrix comprised of a collection of $\boldsymbol{\phi_L}$ vectors for a number of knee acoustic emission signal segments. 90, 92, 93, 94, 96, 101, 103, 104, 105, 107

σ	Standard deviation. 115, 116
$s(n)$	Sampled knee acoustic emission signal. 117, 121, 122, 123, 131, 143, 159, 160, 161, 163, 166, 173, 178
$\hat{s}(n)$	Signal obtained by linearly weighted summation of past samples of $s(n)$. 118
p	Linear prediction order. 118, 124, 131, 146, 148, 149
α_i	Linear predictor coefficient i. 118
E_m	Minimum squared error between $s(n)$ and $\hat{s}(n)$. 118, 119
$w(n)$	General window signal. 118, 123
l_w	Length of the general window signal $w(n)$. 118, 119, 143
$s_w(n)$	Windowed knee acoustic emission signal. 118, 122, 143
$r_{s_w s_w}$	Autocorrelation sequence of a windowed knee acoustic emission signal. 119, 122
$u(r)$	Linear prediction residual signal. 119, 120, 124, 131, 133, 153, 166
$u_r(n)$	Windowed linear prediction residual signal. 120
$\tau_g(k)$	Group delay function. 120
U_r	Fourier transform of the windowed linear prediction residual signal. 120
$\delta(n)$	Unit impulse function. 120
d_{EW}	Energy-weighted group delay. 120, 124, 127, 133
N_w	Window size of the group delay function. 120, 131, 133, 134, 146, 148, 149
$\hat{r}_{s_w s_w}$	Normalised autocorrelation sequence of a windowed knee acoustic emission signal. 122

N_l	Estimated period in seconds of a short segment of $s(n)$. 122, 123
N_1	Lower limit for N_l. 122, 123
N_2	Upper limit for N_l. 122, 123
$\hat{N}_s$	Estimated period in seconds of $s(n)$. 123, 138, 143, 149, 160
$\|u(r)\|$	Amplitude envelope of $u(r)$. 124, 125, 129
C_{sd}	Slope deviation cost. 127, 132, 133, 153, 156
n_z	Sample index of the current contact instant candidate. 127
l_s	Length of slope window in samples. 127, 133, 147
C_{pc}	Period consistency cost. 127
n_{z+1}	Sample index of the next contact instant candidate. 127
n_{z+2}	Sample index of the next next contact instant candidate. 127
$\boldsymbol{\Omega}$	Set of candidate contact instants generated by the negative going zero crossing of the group delay function applied on $u(r)$ 131, 132, 148
$\boldsymbol{\Xi}_I$	Subset of $\boldsymbol{\Omega}$ containing only initial contact instant estimates obtained by ASDSA. 132, 133, 141, 159, 160
$\boldsymbol{\Xi}_L$	Subset of $\boldsymbol{\Omega}$ containing only last contact instant estimates obtained by ASDSA. 132, 133, 141, 159, 160
$\boldsymbol{\Theta}_1$	Vector of the weighting factors of C_{sd}, C_{pd} and C_{er} for the ASDSA cost function. 132
$\boldsymbol{V}_{\boldsymbol{\Xi}_I}$	Cost vector evaluated for the initial contact instant with index z in the set $\boldsymbol{\Xi}_I$. 132, 133
$\boldsymbol{V}_{\boldsymbol{\Xi}_L}$	Cost vector evaluated for the last contact instant with index r in the set $\boldsymbol{\Xi}_L$. 132

θ_{ss}	Weighting factor of the swing to stance ratio cost. 132, 133, 150
C_{ss}	Swing to stance ratio cost. 132, 137, 149, 156, 157
θ_{sd}	Weighting factor of the slope deviation cost. 132, 133, 150
θ_{pd}	Weighting factor of the period deviation cost. 132, 133, 150
θ_{er}	Weighting factor of the energy ratio cost. 132, 133, 150
C_{pd}	Period deviation cost. 132, 134, 149, 157
C_{er}	Energy ratio cost. 132, 136, 157
$\boldsymbol{C}(z,r)$	Overall cost vector evaluated for the initial contact instant with index z in the set $\boldsymbol{\Xi}_I$ and for the last contact instant with index r in the set $\boldsymbol{\Xi}_L$. 132, 156
$\boldsymbol{\Theta}$	Vector of all weighting factors for the ASDSA cost function. 133, 147, 150, 151, 154, 156
q_{z-1}	Sample index of the $z-1$ initial contact instant candidate. 133, 134, 135, 137
q_z	Sample index of the z initial contact instant candidate. 133, 134, 135, 136, 137, 138
q_{z+1}	Sample index of the $z+1$ initial contact instant candidate. 133, 134
q_{r-1}	Sample index of the $r-1$ last contact instant candidate. 133
q_r	Sample index of the r last contact instant candidate. 133, 137, 138
q_{r+1}	Sample index of the $r+1$ last contact instant candidate. 133

α	Control parameter for the period deviation cost function. 134, 147
ϵ	Tolerance factor for the stride duration variability during walking. 134, 135, 147
Δ_p	Period ratio between estimated contact instant candidates. 134
$R_{q_z,q_{z-1}}$	Energy ratio between the current and previous estimated initial contact instant candidates. 135, 136
$E(q_z)$	Energy of signal in a short window centred at q_z. 136
l_e	Length, in samples, of the window in which the signal energy is computated. 136, 147
p_e	Ensemble probability density of the energy ratio of consecutive initial and last contact instants. 136, 137
p_{ss}	Ensemble probability density of the swing to stance ratio of consecutive obtained with temporal ground truth data. 137
T_{min}	Lower limit of the dynamic programming algorithm search window. 138
T_{max}	Higher limit of the dynamic programming algorithm search window. 138
w	Weight factor. 138, 143
$\hat{\tau}_\delta$	Detection time error expressed as a percentage of the stride cycle length. 141, 145, 151, 163
$\mathbf{\Gamma}_I$	Set of reference initial contact instants. 141, 143, 148
$\mathbf{\Gamma}_L$	Set of reference last contact instants. 141, 143, 148
γ_j	Reference initial or last contact instant of the foot to the floor in stride cycle with index j. 141
ζ	Initial or last contact instant estimate. 141, 159, 160

N_s	True period of $s(n)$ obtained from the force plate signal. 143
$RMSE$	Root mean squared error. 143
τ_δ	Detection time error. 145
l_j	Length in samples of stride cycle with index j. 145, 148
ϱ	Detection accuracy. 145, 146, 148, 150, 151, 153, 154, 159, 162, 163, 166, 167, 189, 200
D_{ss}	Evaluation score for the ASDSA. 146, 150, 151, 153, 154, 159, 161, 163, 166, 167
D_{ss}^{ici}	Evaluation score of the initial contact estimates of the ASDSA. 146, 153, 154, 162, 163
D_{ss}^{lci}	Evaluation score of the last contact estimates of the ASDSA. 146, 153, 154, 162, 163, 166, 167
β	Control parameter of the ASDSA evaluation score. 146, 161
C_I	Cost function for assessing the generated initial contact instant candidates. 148
τ_{ref_j}	Sample instance of a reference contact instant in stride cycle with index j. 148
τ_{cl_j}	Sample instance of the closest candidate to the reference contact instant in stride cycle with index j. 148
N_b	Bias parameter. 148, 149
C_L	Cost function for assessing the generated last contact instant candidates. 148
C_{IL}	Cost function for assessing the generated candidates. 148
$\hat{\theta}$	Element of $\hat{\mathbf{\Theta}}$. 150
$\hat{\mathbf{\Theta}}$	Estimate of $\mathbf{\Theta}$ during training of ASDSA. 150

r_s	Cross correlation sequence. 159, 160
$\boldsymbol{X_s}$	Stride-synchronous matrix representation of a knee acoustic emission signal segment 179, 180, 181, 184, 185, 186, 187, 188
$\boldsymbol{X_{st}}$	Stance-phase-synchronous matrix representation of a knee acoustic emission signal segment 179, 180, 181, 184, 185, 186, 187, 188
$\boldsymbol{X_{sw}}$	Swing-phase-synchronous matrix representation of a knee acoustic emission signal segment 179, 180, 181, 184, 185, 186, 187, 188

Chapter 1

Introduction

OSTEOARTHRITIS is the most common disabling and financially burdensome of all musculoskeletal diseases, and prevalence is rising [1, 2]. It occurs most frequently in the knee and commonly affects adults over the age of 45 [3], although younger adults may also develop OA [4]. It is estimated that 100 million people worldwide are affected by knee OA [5]. Several studies have raised the concern regarding the increasing economic and socioeconomic burden of OA [6, 7, 8, 9]. In European countries, the average annual cost of OA per patient ranges from 1,330 EUR to 10,452 EUR which is usually covered by both the country's healthcare system and the patient [7]. The corresponding cost for a patient in the United States of America was estimated at 11,029 USD per year while the national cost was 62.1 billion USD for each year between 2008 and 2011 [10]. Risk of OA is associated with increased mechanical wear, such as through older age and high body weight [11]. Currently, there is no cure and treatments aim to manage symptoms through lifestyle modification, physio- and pharmacological therapy [3]. In severe cases, total knee replacement is required.

Clinical detection of knee OA relies on a combination of patient reported symptoms and medical imaging of cartilage and subchondral bone degradation. Current imaging methods such as X-ray, Magnetic Resonance Imaging (MRI) and ultrasound (see section 1.3.1) have poor sensitivity in early disease and as a result at the time of diagnosis, OA is already at a progressed stage, and understanding of its cause and development is still

limited. Additionally, current imaging techniques provide images of the static anatomical structure of knee joints at a particular posture, as in the example shown in Figure 1.2. Fluoroscopy, on the other hand, is an imaging technique that uses X-rays to obtain real-time moving images of the interior of an object [12]. Since the patient is exposed to a continuous source of X-rays instead of a momentary pulse, as in ordinary radiography, a fluoroscopy procedure generally results to a higher absorbed dose of radiation by the patient [12]. Adding to this, the equipment needed for the examination limits the range of knee and patient movements. Therefore, these techniques are limited in assessing the integrity of the knee during dynamic Open Chain Activity (OCA), for example when the foot leaves and makes contact again with the ground as happens during walking. This is important since OA subjects experience more pain and discomfort when their knee is functional [13]. Although dynamic MRI produces good measurements in the assessment of knee function, it is normally not practical in terms of cost and accessibility [14].

Therefore, there exists the need for a quick, non-invasive, portable and cheaper technique that would ideally be accessible in a non-clinical environment and could be used as a screening tool for the mild disease cases. This technique will facilitate effective intervention toward knee OA management and to improve the selection of patients that need further clinical testing, thus reducing associated costs for both the patient and the healthcare system.

It is well recognized that knees generate sounds during movement, often called crepitus in the medical community, the grinding and creaking sensation when the knee joint moves [15]. These acoustic events can potentially provide information regarding the health of the internal joint surfaces and its structural integrity [16]. In general, the focus of this thesis dissertation is to present the analysis conducted on the sounds emitted by the knee joint during walking and captured by a single microphone attached at the patella, with particular focus on the effects due to osteoarthritis.

In the remainder of the Chapter the knee joint mechanisms and physiology in the human body will be described, followed by an overview of the effects of OA in the knee with a description of the current medical diagnosis methods. A literature review on past

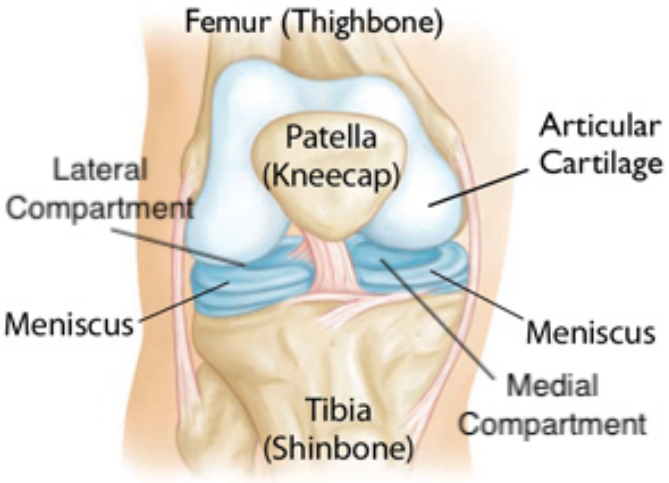

Figure 1.1: Internal structure of a knee joint
(modified from: http://orthoinfo.aaos.org/topic.cfm?topic=a00212).

signal processing approaches for the measurement and characterization of knee joint events will also be presented. Following, the research challenges of the thesis will be defined along with the outcomes of the work.

1.1 Knee joint mechanisms and physiology

The knee joint is one of the most complicated and largest joints in the human body [17]. As shown in Figure 1.1 the knee includes the femur (thighbone) and the tibia (shinbone) that form the main joint and the patella bone (kneecap) that joins with the femur to form the patellofemoral joint [17]. At the meeting points of the three bones there is a smooth, slippery substance called articular cartilage that covers the end of the bones and protects them during motions like knee flexion or extension [17]. The meniscus in particular, are pieces of cartilage that are strong but rubber-like and absorb the shock forces between the femur and the tibia [17]. The patella also protects the knee by mitigating the friction forces between the cartilages and muscles during bending. Another layer of protection is offered by the synovial membrane that surrounds the knee joint and secretes the synovial fluid that lubricates the cartilage [18]. This fluid helps in reducing friction forces.

The movement of the knee joint during OCA (activities in which the foot is free to

move) involves the interaction of muscles, tendons, ligaments and articular surfaces which prevent the joint from moving in certain directions [19]. Hence the knee is not capable of withstanding strong rotational forces that occur most frequently in sport activities and often lead to knee related injuries. In addition to sports injuries, the gradual wear or damage of the cartilage results in the reduction of the protective space between the bones and can lead to degenerative diseases such as OA [11].

1.2 Knee osteoarthritis

Knee OA can occur in the medial and lateral compartments (see Figure 1.1), in the patellofemoral joint as well as in all three of these locations [11, 17]. OA can lead to crepitus and knee pain, particularly during everyday movements like walking or climbing the stairs [20]. Another common knee OA symptom is stiffness and swelling that usually occurs after rest. These symptoms greatly degrade the patients' quality of life [20].

Knee joints generate sounds during movement. When the knee is active, the tibia and the femur bones come into contact. The regions and perhaps the quality of joint surfaces coming into contact are different at each angular position, generating therefore a number of different sounds during movement. In healthy knee joints, the bones have smooth surfaces due to a thin layer of cartilage and are separated by a protective space filled with synovial fluid to reduce friction [18]. They are able to move freely and the level of sound emitted is low. In OA knees this structure is degraded, as can be seen in the example of Figure 1.2 (b), and the protective space and associated lubrication reduce. Consequently, this leads to increased friction which accelerates the wear of cartilage [11]. It is hypothesized that this increased friction makes the knee more noisy during motion.

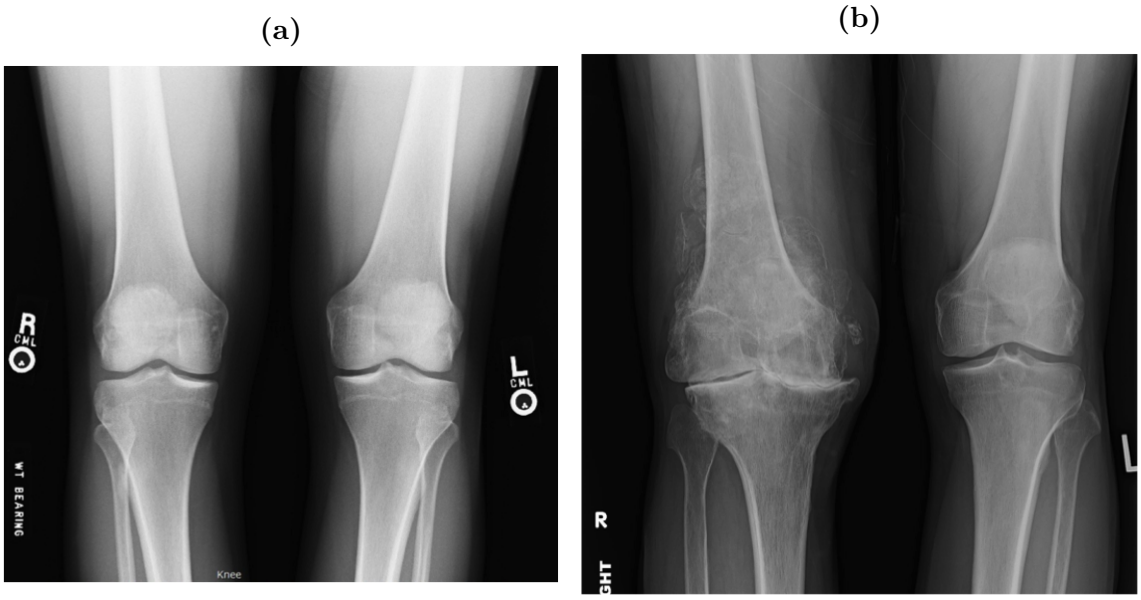

Figure 1.2: X-ray images of (a) subject with healthy knee joints and (b) subject with OA in the right knee.

1.3 Methods of knee joint assessment

1.3.1 Clinical methods

X-radiography is one of the most common methods used by clinicians to diagnose OA by measuring, for example, the spacing between the tibia and the femur bones. Figure 1.2 displays X-ray images of a subject with healthy knee joints and a subject with OA in the right knee where the difference in joint spacing in the second image is clear. In general, a relatively reliable and reproducible analysis using X-ray images can be performed by the radiologist [21]. However, the image interpretation can sometimes vary for different radiologists depending on their skills and experience. Based on the visual inspection of the image, the knee is graded using the Kellgren-Lawrence (KL) radiographic scoring method [22, 23]. The KL score classifies the severity of OA in the knee using five grades, from 0 to 4, where grade 0 indicates no presence of radiographic features of OA and grade 4 is definite bone deformity [23]. In X-ray imaging however, the cartilage is not visible and therefore, its status cannot be determined. MRI on the other hand uses magnetic field and radio frequency pulses and is more sensitive to articular cartilage surface defects than X-radiology [24]. Therefore, it can provide a more complete static picture of the knee

joint [25]. Computed Tomography (CT) and ultrasound are two more clinical imaging methods used for knee OA diagnosis and assessment. A CT scan provides more detailed images than traditional X-rays. The drawbacks are its low soft-tissue contrast and like X-rays, the health risks that it poses due to radiation, especially for pregnant women [26]. Ultrasound is an imaging technique that enables real-time multiplanar imaging at a relatively low cost [26, 27]. Ultrasound can detect inflammatory and structural abnormalities, features that are associated with OA, without the risks of radiation [27]. The primary limitations of this technique are that it is dependent on the skills of the operator and that the penetration capabilities of sound limit the assessment to shallower regions than other imaging techniques (impossible to view through bone) [27]. Further limitations of the clinical methods exist and were discussed in the Chapter's introduction.

1.3.2 Phonoarthrography

The potential for using internal knee joint sounds for diagnostic purposes has been known for many years. Blodgett, in 1902, reported on auscultation of the knee, with attention to sounds of normal joints and their change with repetitive motion, where a relation between an increase of sound activity and age was noted [28]. In 1913, Bircher reported that different types of meniscal injury generate distinctive sound signals [29, 30]. Steindler in 1937, used a system consisting of a cardiophone, an oscilloscope and a recorder to study 397 knees [31]. He found a relation between pathologies and the pitch, amplitude and the sequence of sounds and was able to classify the joints based on these features. It was observed, however, that it was difficult to separate other body sounds such as muscle activity from articular cartilage sounds. In [32] the authors claim that sounds could be detected in rheumatoid arthritis before any changes were observable in an X-ray image but no further work was conducted to confirm this claim.

Phonoarthrography (PAG) utilises acoustic microphones in the audible frequency range to record sounds generated during movement. Important work on PAG by Chu et al. reported that the spectral activity of pathological knees (recorded during knee flexion-extension) spanned the entire audible frequency range and the signals' acoustic power

increased with severity of cartilage damage [33, 34, 35, 36]. The microphones used in this work were large condenser microphones placed at a short distance from the patella [33].

1.3.3 Vibroarthrography

Significant work was directed to the development of Vibroarthrography (VAG) as an alternative to PAG which relies on accelerometer sensors, operating at frequencies below 1 kHz, to pick up mechanical vibrations [37]. Algorithms proposed for classifying the knee VAG signals according to pathological conditions, vary from linear prediction modelling [38,39] to time-frequency analysis [40,41,42] and wavelet matching pursuit decomposition [43,44]. Several features were used for classification, including spectrogram features, waveform variability parameters, statistical features [45], fundamental frequency, mean amplitude of pitch and their jitter and shimmer [46, 47]. Classifiers used range from early neural network architectures [42,45] to maximal posterior probability decision criterion [48], bagging ensemble and multiple classifier system based on adaptive weighted fusion [43]. A thorough description of VAG analysis can be found in [49].

1.3.4 Acoustic emission

The use of Acoustic Emission (AE) at ultrasonic frequencies was explored as a potential biomarker for assessing the knee joint condition. In the same way that AE is used for machine condition monitoring [50], the authors in [51] defined the potential biomarker as elastic waves generated by the joint during movement that satisfy a set of amplitude and time constraints [51, 52]. Piezoelectric contact sensors were used to capture ultrasonic AE signals (50 kHz to 200 kHz) emitted during sit-to-stand movements. It was found that OA knees produce substantially more AE events with higher peak magnitude and average signal level compared to normal knees [52, 53, 54]. It was demonstrated, using Principal Component Analysis, that healthy and OA knees are separable in the feature space [51,54]. It was also shown that ultrasonic AEs are sensitive to age-related degeneration in both OA and normal knees [53]. The movement protocols most often reported in the literature, for OA and other arthritis related studies, are knee flexion-extension and sit-to-stand

movements [49, 51, 55, 56].

In [57] a comparative study was presented on the quality of the knee sound signals captured by a piezoelectric contact microphone, a miniature electret microphone and a microelectromechanical systems (MEMS) based microphone (with a sound port to measure airborne sounds) during knee flexion-extension and sit-to-stand movements. The microphones were attached on the lateral and medial sides of the patella using tape. The authors recommended using air microphones over contact microphones and showed that electret and MEMS microphones performed similarly at detecting joint sounds. It was suggested that the latter are a viable substitute for the more expensive electret-based microphones and was also reported that high amplitude acoustic events were repeatable and occurred at consistent joint angles [57].

AE analysis during knee flexion-extension was also explored in the context of knee injury rehabilitation [55, 58, 59]. In [55], a 64-dimensional feature representation of 200 ms frames of the knee sound signal was used, from which a k-Nearest Neighbors (kNN) graph was constructed. A graph-based metric was then proposed to quantify the homogeneity of the feature matrix without modelling the underlying distribution. Based on this metric, it was concluded that injured knee joints produce more heterogeneous features than healthy knee joints [55]. Although this approach alleviates the need for prior algorithm training, it is only accurate when sound data from a healthy and an injured knee of the same subject is available since the study focused on the intra-subject knee sound differences. Inter-subject sound differences were not considered.

In [60], the intra-subject consistency of knee acoustic events in healthy subjects was explored. The sound signals were captured by a MEMS-based microphone that was attached at the lateral side of the patella. These events were defined as short, high frequency sounds and extracted from the knee signal after filtering with a 1-20 kHz band-pass filter. Consistency was quantified based on the number of such events in ranges of normalized knee angles that were computed by the measurements of two gyroscope-accelerometer pairs. It was reported that knee acoustic events for healthy subjects were consistent between the two knees, but the inter-subject consistency was not explored [60].

In [61] the effect of mechanical loading on the knee during walking was explored using the graph based metric developed in [55]. The analysis however, only considered the swing phase i.e. when the knee is suspended over the floor and simply ignored the sounds generated during the stance while the knee actually experiences loading conditions (foot in contact with the ground). In the context of knee OA detection, it will be shown later in Chapter 5 that both the swing and the stance phases contain information that is relevant to the detection of the disease.

1.3.5 Discussion

The scientific literature is, to some extent, mixed regarding the term used for referring to the sounds emitted by the knee joint. As mentioned in the chapter's introduction, the term crepitus is most commonly used in the medical community. In the signal processing literature, PAG and VAG are used instead, with the term AE gaining more traction in the recent years. The difference amongst these terms is in the type of sensor used. PAG utilises acoustic microphones in the audible frequency range while VAG uses accelerometers to pick up mechanical vibrations at the surface of the skin. The sensors used for picking up sounds at ultrasonic frequencies are commonly called AE sensors and are piezoelectric based contact sensors.

As mentioned in Section 1.3.2, Chu et al. is attributed for advancements in PAG signal analysis through a series of papers published in the 1970s. In their work, the authors used condenser microphones to capture the knee sounds. This type of microphones in the 1970s were not as sensitive at low frequencies as were the accelerometers which also allowed better attachment on the knee due to their small size compared to the much bigger condenser microphones. In later studies it was found with the use of accelerometers that there is information at the low frequencies of the knee sound signal that is significant for distinguishing between healthy and unhealthy knee joints (Section 1.3.3). This is also shown in the findings of Chapter 3 where a specific frequency band below 1 kHz was found to be important for classifying OA and healthy knees.

For these reasons, I believe, the research community turned its focus towards the

use of accelerometers and VAG signal analysis and as a consequence the term PAG is now rarely used. The term AE for knee joint signal analysis was introduced in [51] by a research group at the university of Central Lancashire. Perhaps, this term had already been used in the literature but it appears that, after this paper, subsequent publications started to use it more often such as in [55, 57, 60, 62]. The term AE is generally used for an acoustic signal generated by a source and it is more general than PAG or VAG. It is therefore the preferred term in this thesis work.

1.4 Research challenges and thesis structure

Based on the literature review findings and the inability of clinical imaging methods to assess the knee joint condition during dynamic activities, the overarching research question for this thesis can be formulated in the following:

> *Is the underlying structure of the acoustic signal emitted from a knee joint with osteoarthritis during walking and captured using a single microphone useful for the disease detection and is it possible to be effectively and efficiently extracted?*

To the date this thesis was written and to the best knowledge of the author, no prior art exists on the analysis of knee acoustic signals acquired during walking for the detection of OA. Therefore, an exploratory analysis work needs to be performed before attempting to investigate the research question in order to answer some fundamental questions relating to the data at hand. This is presented in Chapter 2 along with a description of the data acquisition system and test protocol.

In investigating the aforementioned overarching question several challenges needed to be addressed. Inter-subject knee sound differences is one of the most challenging aspects of knee joint sound analysis as there is strong variability in the knee sounds amongst individuals [59]. This is likely due to their joints' structural differences. Unlike previous studies that focused on acoustic differences between the knees of the same subject [55, 57, 59], the analysis framework developed and employed in Chapter 3 takes into account the inter-subject variability and investigates the knee sound signals with the aim of identifying

possible features and feature combinations that can discriminate between healthy and OA knee populations. To this respect, a statistical parameterisation of the feature distributions is proposed.

Furthermore, in order to incorporate spatiotemporal information in the analysis, the time instants of the first and last contact of the foot to the floor need to be known. This is a challenging problem given that only single microphone recordings are used. Chapter 4 investigates the development of an automatic algorithm for detecting the first and last contact instances of the foot to the floor. This novel algorithm enables the investigation of the relationship between knee sounds with functional (i.e. gait biomechanical variables) and clinical outcomes, which is currently unknown. This is investigated in Chapter 5. In the concluding chapter of the thesis, Chapter 6, the major findings of the individual Chapters are summarized and integrated, from which fruitful directions for future research are identified.

The work presented in this dissertation focuses on the knee joint but ultimately the research can be potentially extended and applied to, primarily, the other main joints involved in the locomotion system, namely the talocrural joint (ankle joint) and the acetabulofemoral joint (hip joint). Aside these particular joints, the methodologies developed could also potentially help in the research for OA detection in any other joint that causes pain and discomfort, such as in the elbow and shoulder.

Finally, a note about how the thesis dissertation can be read. Each of the chapters 2, 3, 4 and 5 have been written to stand independently and therefore, they can be read in any order. However, each subsequent chapter capitalizes to a certain degree on the findings of the previous one(s). Therefore, it is advisable for the readers who intend to read more than one chapter to do so in the order in which they appear. The aim of writing chapters that stand alone created several challenges, the most important of which was how to avoid unnecessary repetition. To this respect, cross-referencing was used in order to direct the reader to the algorithms and outcomes that were described in previous chapters. However, in order to avoid degrading the flow of reading, important results and conclusions that form the basis for certain parts of the work in subsequent chapters, are restated.

1.5 Thesis outcomes

The following list shows the publications related to the research presented in this thesis:

J1 C. Yiallourides and P. A. Naylor, "Time-frequency analysis and parameterisation of knee sounds for non-invasive detection of osteoarthritis," *IEEE Trans. Biomed. Eng.*, Jul. 2019, Submitted

C1 C. Yiallourides, A. H. Moore, E. Auvinet, C. Van Der Straeten, and P. A. Naylor, "Acoustic analysis and assessment of the knee in osteoarthritis during walking," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 281–285

C2 C. Yiallourides, V. Manning-Eid, A. H. Moore, and P. A. Naylor, "A dynamic programming approach for automatic stride detection and segmentation in acoustic emission from the knee," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 401–405

C3 V. Manning, C. Yiallourides, Brevadt M, A. H. Moore, E. Auvinet, P. A. Naylor, and J. Cobb. "Knee Sounds May Predict Osteoarthritis Severity, Symptoms and Function: Pilot Investigation Toward a Novel Dynamic Imaging System," [abstract]. Arthritis Rheumatol. 2015; 67 (suppl 10). https://acrabstracts.org/abstract/knee-sounds-may-predict-osteoarthritis-severity-symptoms-and-function-pilot-investigation-toward-a-novel-dynamic-imaging-system/

1.5.1 Statement of originality

The following aspects of the thesis are, as far as the author is aware, original contributions:

1. The use of the knee sounds sensed at the patella during dynamic functional activity (walking) for knee joint assessment and condition classification.
2. The development of a comprehensive experimental framework for the investigation of the discriminant capabilities of various features for the supervised classification task of normal (clinically healthy) vs abnormal (clinically OA) knee joint sound signals.

This framework takes into account the possible inter-subject knee sound differences and proposes the parameterisation of the features extracted from the knee sound signals using 11 statistical parameters. (Chapter 3)

3. The use of the modulation magnitude spectrum and the Mel-frequency Cepstral Coefficients (MFCC) are novel in the context of knee joint condition assessment in the presence of osteoarthritis using the sound signal sensed at the patella. In particular, the MFCCs were shown to generate the best classification results amongst all the features considered using the framework above. (Chapter 3)

4. The frequency bands 220 Hz to 420 Hz and 1 kHz to 3.4 kHz were found to contain a collection of features that are important for classification and are derived from the spectrum of the knee sound signal. This finding enables the extraction of more targeted features for future research in OA detection. (Chapter 3)

5. The use of knee joint sounds generated during walking and sensed at the patella for stride detection and segmentation. Furthermore, the development of an automatic, dynamic programming based method for temporal stride detection and segmentation using only the sound signal. This algorithm enabled the study of the knee sounds in relation to spatiotemporal information. In addition, the algorithm permits gait analysis with low time bias error without the need of expensive and bulky equipment as is the case with the state of the art force plate based systems. (Chapter 4)

6. The development of stride and stride-phase based representations of the knee sound signal using short acoustic pulse events detected in the signal and their waveform characteristics. In particular, it was found that the number of threshold crossings, the threshold crossing rate and the energy, together with the number of pulse events are the most important features for classification. Furthermore, the classification accuracy was found to improve when the stride-based representations were used compared to features obtained over a longer time frame. This work establishes the foundations for any future research that uses these pulse events emitted by the knee during walking for characterizing knee health and relating to the main gait phases. (Chapter 5)

Chapter 2

Data Acquisition and Exploration

THE research work carried out for this thesis began as an interdisciplinary project in which a London orthopaedic clinic[1] was responsible for acquiring the raw data which was then given to the author of this thesis for analysis and information extraction. This Chapter describes the detailed settings of the knee joint sound signal acquisition system in Sections 2.1 and 2.2. The text also presents in Section 2.3 the participant demographical characteristics and in Section 2.4 an exploratory analysis conducted on the acquired knee joint sound signals. More specifically, temporal and time-frequency approaches for the signal analysis are described and an investigation for the association of the knee sounds to osteoarthritis symptoms and function is conducted. Ultimately, the purpose of this Chapter is to raise an understanding about the subject demographics and the sound signals and prepare the reader for the work that is presented in the Chapters to follow.

2.1 Subjects and equipment

Adults with clinically healthy and clinically OA knees and reporting no knee pain in the 2 weeks prior to data collection were recruited. Knees were classified by clinicians as: 1) normal (clinically healthy), 2) abnormal (clinically OA), 3) Ostearthritis Healthy

[1]MSk Lab in Charing Cross Hospital, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London

(OAH) (contralateral knee OA). Exclusion criteria were: aged <18 years, previous surgery, unable to provide consent. Knee sound signals were acquired with a sampling frequency of ≥44.1 kHz using a contact microphone with a sound port for capturing airborne sounds and an electret condenser microphone mounted inside a capsule (Basik Pro Schertler, 20 Hz – 20 kHz), attached to the patella of the knee joint. The microphone signal was connected to the input of a digital preamplifier (RME Babyface; PreSonus DigiMax LT). Measurements were made while subjects were walking on a treadmill instrumented with force plates (see Appendix A for technical specifications). Force plates recorded kinetic and spatiotemporal gait characteristics during stance phase (foot in contact with ground), including ground reaction force (GRF) at heel strike, mid-stance and push-off, enabling the calculation of time spent in swing phase. More information regarding gait parameters and their extraction is discussed in Chapter 4. Pictures of the equipment used for data collection can be seen in Figure 2.1.

Precautions were taken to reduce skin friction with the microphone during data collection. The skin area under and near the microphone was shaved and double adhesive tape was used to ensure the firm attachment of the microphone to the patella. However, it must be taken into account that minor movements of the microphone might have still taken place during the recordings.

2.2 Assessment protocol

The assessment commenced with a 5 minute warm-up and acclimatisation to treadmill walking followed by data acquisition at 1) increasing speeds on a flat level until maximum walking speed was achieved (speed increments of 0.5 km/h, maximum walking speed defined as the maximum pain-free speed where one foot was always in contact with the ground), 2) fixed speed up an increasing incline (speed 4 km/h, incline increments of 5%, maximum incline 20%), 3) increasing speeds on a fixed decline until maximum walking speed (starting at 4 km/h, speed increments of 0.5 km/h). The speeds and inclines were subject dependent. Maximum speeds achieved on a flat level ranged from 2.5 to 9 km/h.

Figure 2.1: Experimental equipment: RME Babyface (top left), PreSonus DigiMax LT (bottom left), treadmill instrumented with force plates (right).

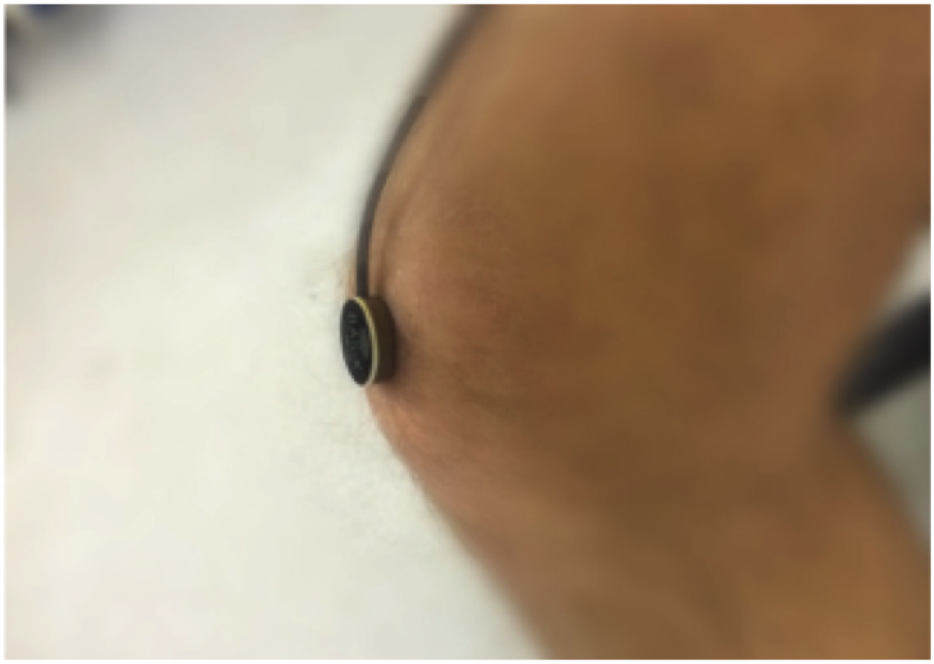

Figure 2.2: Contact microphone (Basik Pro Schertler) attached on the patella.

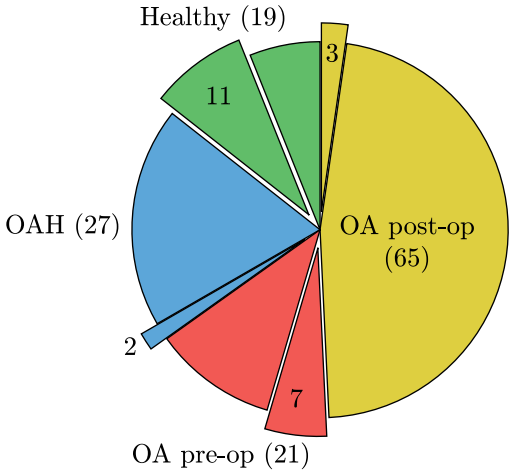

Figure 2.3: Number of knees in the database, per condition.

2.3 Data and subject demographics

Sound data was acquired for 132 knees using the contact microphone attached as shown in Figure 2.2. Force plate data that is time synchronised with the sound data was obtained for 23 of the 132 knees. Even though the sound and force plate data were acquired continuously, the signals that are time synchronised were acquired in 'snapshots' of 20 seconds per assessment stage. Therefore, uninterrupted force plate signals from which temporal ground truth data can be extracted is 20 seconds long. A number of recordings that correspond to the different assessment stages are available for each knee in the database.

Various interesting facts about the recruited subjects can be derived from the collected data. The pie chart in Figure 2.3 for example, shows the number of knees in each category (joint condition). The smaller (chopped) segments within each category indicate the number of knees for which force plate data was also collected. The OA knees are divided to preoperative (OA pre-op) and postoperative (OA post-op). In this thesis the OA postoperative knees are excluded from any analysis because they do not add any value to the investigation of the overarching research question posed in the previous Chapter. The research interest focuses on the detection of OA and the differences between healthy and OA knee joints prior to any surgical modification. The internal structure of the knee changes after surgery and it is unknown how this affects the emitted sounds with respect

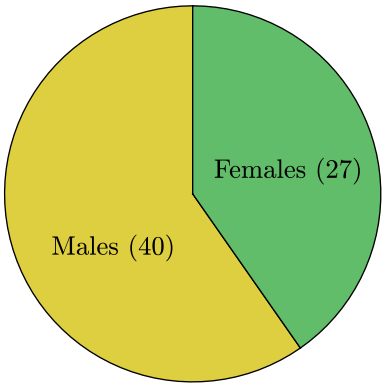

Figure 2.4: Number of knees per gender.

Age group	Knees
20 – 37	20
44 – 59	13
60 – 79	29
≥ 80	5

Table 2.1: Number of knees in each age group.

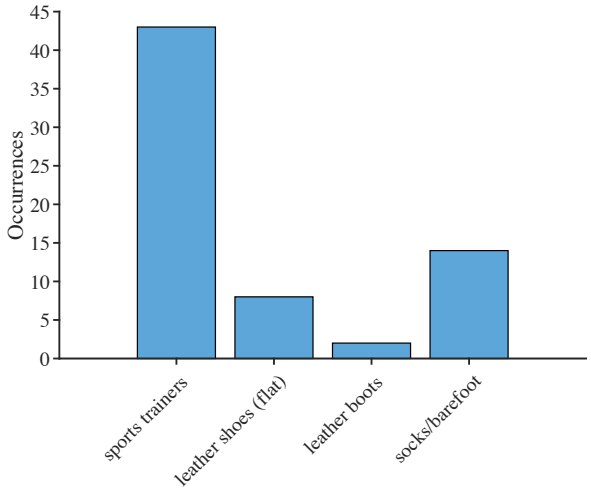

Figure 2.5: Footwear of subjects during data acquisition (occurrences are per leg).

to OA. In the remainder of the thesis, preoperative OA knees are simply referred as OA.

Excluding the postoperative knees, sound data was acquired from 37 males and 21 females resulting in 40 and 27 knees respectively. The difference in numbers is largely due to the exclusion of the aforementioned knee category but also because data was not collected from both knees of every subject. The majority of the subjects were wearing sports training shoes but a large proportion preferred to walk wearing only their socks or even barefoot (further distinction between these two was not made by the researchers at the time of data collection). The above are depicted in Figures 2.4 and 2.5. Table 2.1 divides the 67 knees into different age groups and shows that the majority fall in the early adulthood (20 – 37) and late adulthood (60 – 79).

2.4 Knee joint sound signals

There isn't any known established work in the literature that uses the type of sound signals acquired as described in the previous section. Therefore, this necessitates an exploratory analysis to be carried out, prior to any attempt at investigating the overarching research question outlined in the previous Chapter. Questions such as *"Is the knee sound signal stationary or non-stationary?"* or *"What is a meaningful sampling frequency value to use for processing the signals?"* are the type of questions for which the answer is taken for granted for other type of signals. In the case of the knee sound signals however, the answers are unknown.

Common temporal signal analysis methods include filtering techniques, signal variability measures, segmentation using windows and so on. Signal segmentation can be performed using windows of fixed or adaptive length. In signal processing, segmentation is commonly used to split a non-stationary signal into several locally stationary segments so that conventional techniques such as the Discrete Fourier Transform (DFT) can be applied [66]. Knee sound signals are obtained during an OCA in which the knee performs dynamic movements. Such signals are essentially non-stationary, that is the signals' statistical moments vary with time. In practice, only the first and second order moments are used for showing non-stationarity. Figure 2.6 displays examples of amplitude-normalised knee sound signals recorded from a healthy subject and a patient with OA in the medial compartment of the knee (see Figure 1.1). Figure 2.7 illustrates the variance of these two signals, computed using a sliding window of 120 ms (5292 samples with 44.1 kHz sampling frequency). The variance changes significantly from one portion of the signal to another and therefore, the signals are said to be non-stationary in their second-order statistics.

Due to the rhythmic operation of normal walking (e.g. no tripping), the resulting knee joint sound signal is almost periodic, and may be referred to as a cyclo-stationary signal. Such signals are characterized by the periodicity they exhibit in their mean, correlation, or spectral descriptors [67]. The statistics of the knee sound signal vary within the duration of a cycle (stride) but repeat themselves at regular intervals (gait pattern) as can be seen in Figure 2.7. The cyclic repetition due to walking facilitates ensemble

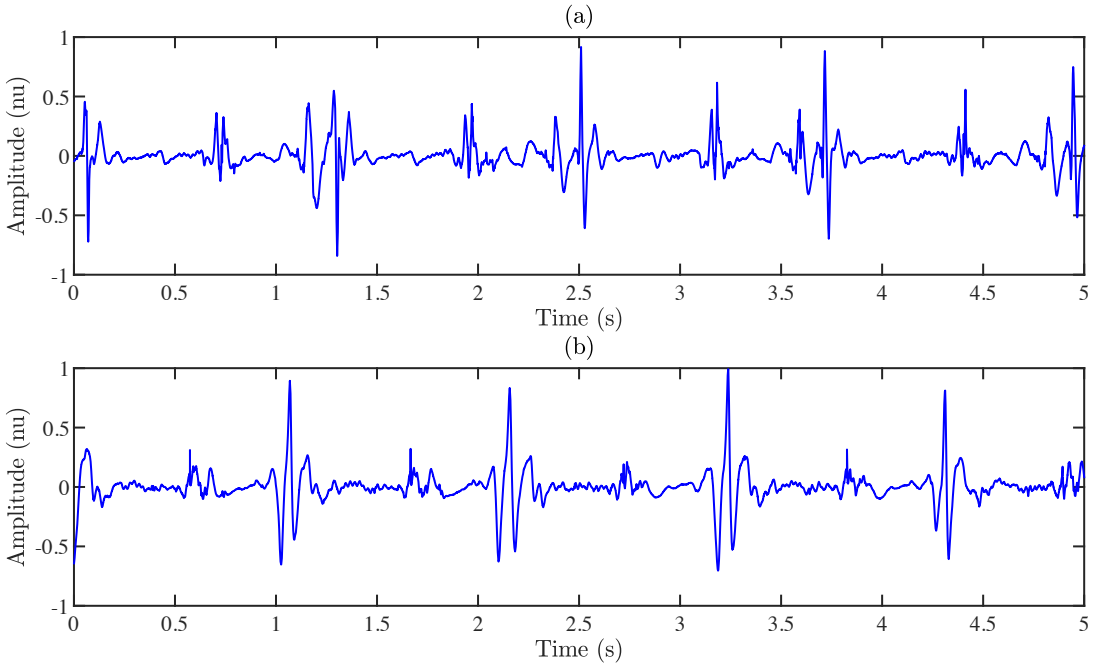

Figure 2.6: (a) The amplitude-normalized knee sound signal of a healthy subject. (b) The amplitude-normalized knee sound signal of a patient with osteoarthritis in the medial compartment of the knee. *nu* : normalised units.

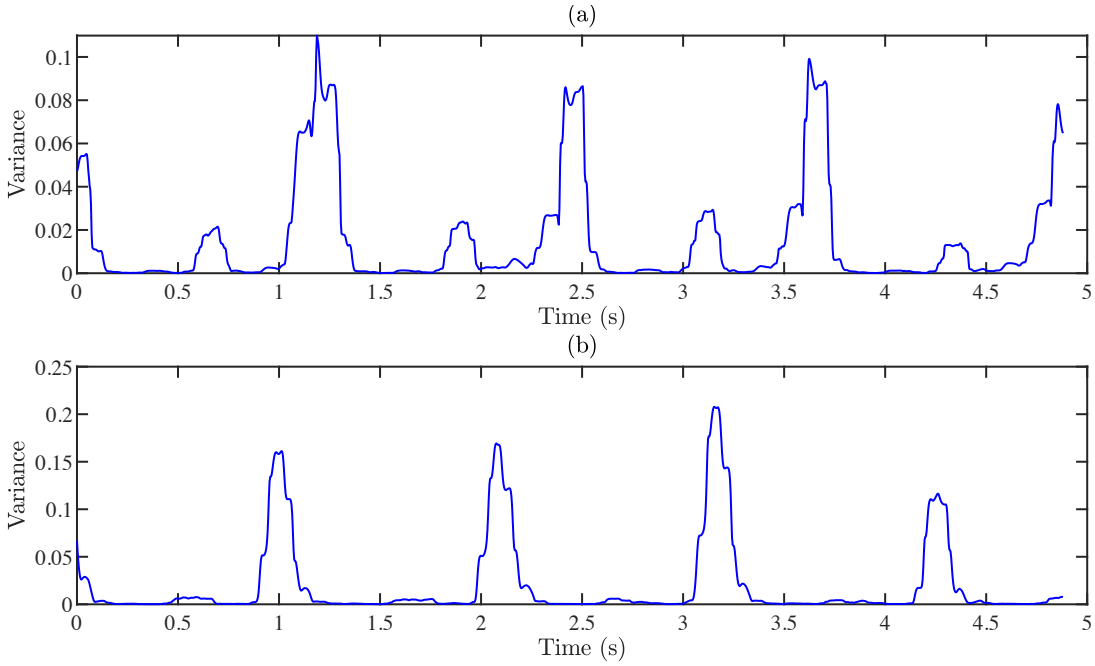

Figure 2.7: Variance of the knee joint sound signals of Figure 2.6, computed using a sliding window of 120 ms (5292 samples with 44.1 kHz sampling frequency).

averaging using events extracted from an observation of the signal over many cycles. The cyclic nature of the knee joint sound signal is exploited for stride-synchronous analysis and will be illustrated in Chapter 5.

The temporal analysis concentrates on the detection of physiological and pathological events in the time scale. Fixed window as well as adaptive window segmentation is investigated in Chapter 3 in the context of feature extraction and classification. Chapter 4 investigates the detection and identification of the walking patterns from the knee sound signal alone. This allows the adaptive segmentation of the signal into strides as well as stride phases from which gait-related features can be extracted.

The information about the inherent periodicity of a given knee sound signal can be expressed and analysed in the frequency domain. The Short Time Fourier Transform (STFT) is a very useful tool for characterizing the frequency distribution of signal segments [66]. Figure 2.8 shows the spectrograms of the non-normalised (in amplitude) knee sound signals of Figure 2.6. The spectrograms were computed using the Fast Fourier Transform (FFT) and a 2048 samples long window with an overlapping segment length of 1024 samples. The window function used is the 4-term Blackman-Harris which is a good general-purpose window with side lobe rejection of more than 90 dB and a moderately wide main lobe [66]. To enhance the visualisation in Figure 2.8 a threshold of 60 dB was set on the Power Spectral Density (PSD) of each segment such that if $10\log(\text{PSD}) \leq 60$ dB was true then the PSD of that segment was set to zero in order to make the spectrogram image sharper. It is evident that the spectral characteristics of the signal vary over its duration. Short temporal events appear at regular intervals and are visible in both spectrograms. Such events arise, for example, from the impact of the heel during walking and from the impact of the foot when it becomes flat. In Figure 2.8 (b) some events appear to be broadband and extend to cover the whole frequency range.

The spectrograms also show the presence of noise and its harmonics, appearing at approximately 3.9 kHz, 7.8 kHz etc. The recording of knee joint sound signals is susceptible to several different types of noise artefacts including Muscle Contraction Interference (MCI), the occasional tremor of the treadmill belts and random noise due to

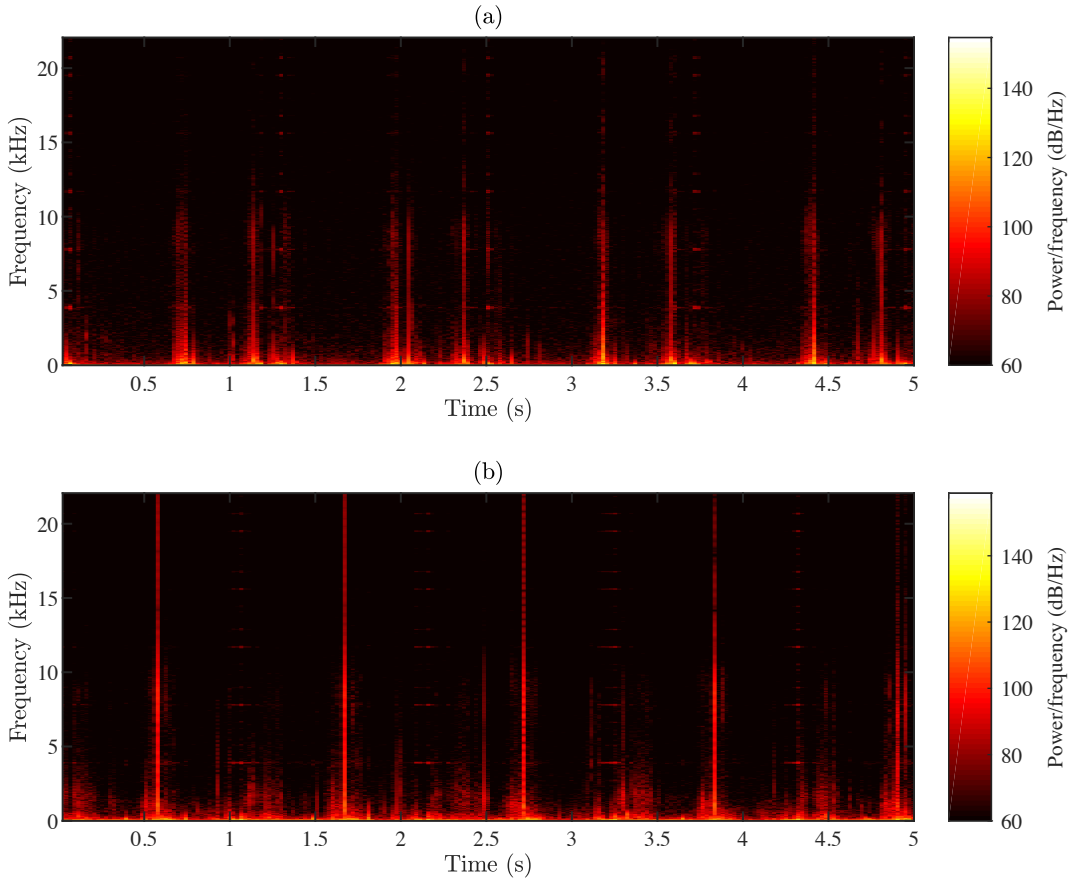

Figure 2.8: Spectrograms of the non-normalised knee joint sound signals in Figure 2.6 computed using the short-time Fourier transform with a Blackmann-Harris window of length 2048 samples and 50% overlap: (a) of a healthy subject; (b) of a patient with osteoarthritis in the medial compartment of the knee. Signals were sampled at 44.1 kHz.

the thermal effect in the cables and amplifiers. In VAG signal analysis it was found that MCI cancellation is not essential and that it could even make the results of VAG signal classification worse compared to not doing the cancellation [39].

From the illustrations of Figure 2.8 it can be observed that the resolution of the STFT time-frequency representations can facilitate feature identification in the knee joint sound signals. However, the trade-off of the window length used in fixed-frame segmentation is a matter of experimentation. This is investigated in Chapter 3 in the context of information extraction from STFT based representations for classification.

From the spectrogram visualisations it can be seen that there is not as much content at high frequencies (> 10 kHz) as there is at lower frequencies. Therefore, the 44.1 KHz

that was used to sample the signals during the recording is not the most appropriate for further signal analysis. Processing at high sample rates will unnecessarily incur high computational costs if the signal does not have information at the highest frequencies. For this reason the suitability of down-sampling the signals for further processing is explored.

It is common in information theory to measure the information of a signal using the Shannon entropy measure [68]. Considering the joint time-frequency function $C(t, f)$ (the spectrogram), the Shannon entropy can be computed as

$$H(C) = -\iint C(t, f) \log_2 C(t, f)\ dtdf \tag{2.1}$$

after following a probabilistic analogy as in [69] where $C(t, f)$ is shown to have properties similar to those of probability densities. Williams et al. [70] proposed a measure of time-frequency information by employing the generalized Rényi entropies [71] as

$$H_R(C) = \frac{1}{1-\alpha} \log_2 \iint \left(\frac{C(t, f)}{\iint C(u, v) du dv} \right)^{\alpha} dtdf \tag{2.2}$$

where α is the Rényi entropy order. The normalisation inside the parenthesis is performed so as to make the information measure invariant to the energy of the signal [70]. Note that the spectrogram has the desired property of non-negativity in all values for t (time) and f (frequency) and hence the integral in (2.2) is well defined. Numerical evidence in [70, 72] and an extensive theoretical study of (2.2) in [69] indicated that H_R, for $\alpha = 3$, is well-defined for a large class of signals and time-frequency representations and is therefore a reasonable selection for measuring the content of a spectrogram such as those in Figure 2.8.

The formulation of Rényi entropy in (2.2) with $\alpha = 3$ is therefore used in the following. A segment of 20 seconds duration is obtained from each of the 67 signals in the database (the post-operated OA knee signals are excluded) and the STFT of each segment is computed using the 4-term Blackmann-Harris window of length 2048 samples (at 44.1 kHz sampling frequency) and 50% overlap. The change in H_R is examined using frequency bands of 10 bins, giving approximately 215 Hz of frequency increase for each additional band. That is, the entropy is computed using all time frames and adding 10

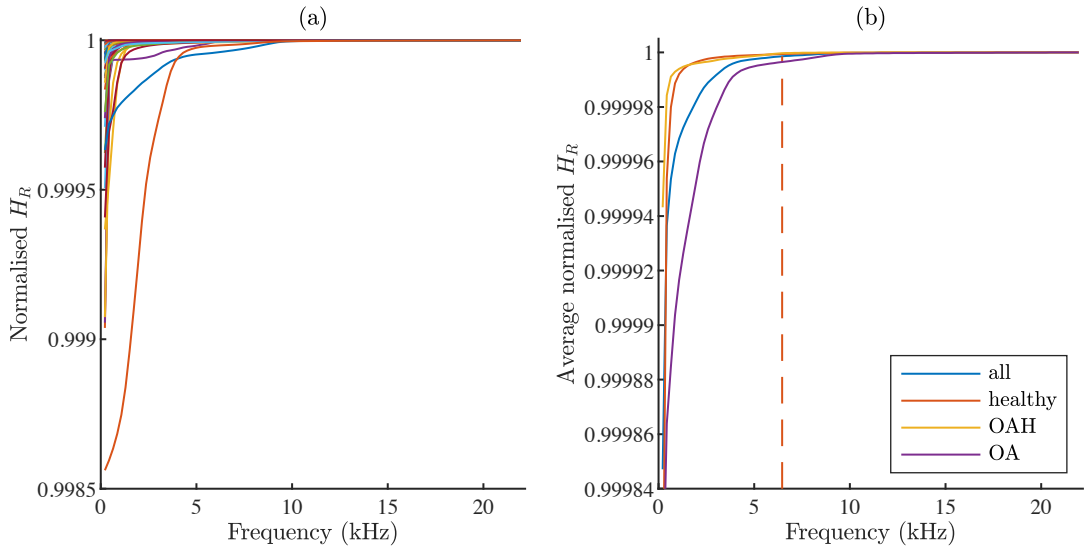

Figure 2.9: For a database of 67 signals: (a) Rényi information entropy per signal, normalised by the maximum H_R of each signal, (b) Average of the normalised Rényi information entropy values per knee condition. Dashed line is drawn at x value of 6.45 kHz.

bins at each iteration.

Figure 2.9 displays the results obtained from running the experiment. Plot (a) shows the Rényi information entropy per signal against frequency. To improve the visual comparison, the H_R values per signal are normalised by the maximum H_R of that signal. In plot (b) the average of the lines in the first plot is displayed (legend name 'all') together with the average of the lines corresponding to the condition class as indicated by the figure legend. It can be observed that the increase in bits of information decreases every time a new frequency band is added to the computation of (2.2). This suggests that most of the energy (spectrogram information) is concentrated at the low frequencies, which is expected as seen in the spectrograms of Figure 2.8. This is more prevalent for the healthy and OAH knees as indicated by Figure 2.9 (b) where the two information-related lines corresponding to these conditions flatten out at a much lower frequency than the OA knees. By setting appropriate threshold values or using more sophisticated approaches can help to decide the point in frequency over which no additional information is present. However, visual inspection of Figure 2.9 (b) suffices the purpose of this exploratory analysis. The increase in the average H_R is minimal above 30 bands as indicated by the dashed line. With 215 Hz per band this means that by 6450 Hz almost all of the information is obtained

from a spectrogram. Based on this value, a suitable choice for the sampling frequency is 16 kHz which is used in the analysis from this point onwards unless explicitly stated otherwise.

Adding fewer than 10 bins at each iteration of computing (2.2) will generate much smoother curves in Figure 2.9 making it more difficult to choose a suitable frequency threshold. On the other hand, adding a lot more than 10 bins would produce a step-like curve with nearly vertical rise in the first couple of bands since most of the information is in the low frequencies as can be seen in Figures 2.8 and 2.9. Grid search can be performed to find a more suitable value for the number of bins to add, by optimizing a metric that measures the change in $H_R(C)$ from one iteration to the next. However, given that a value for the sampling frequency can be justified from the generated results (as explained in the previous paragraph) and that it is computationally expensive to perform grid search for this task, this experiment was not conducted.

2.4.1 Correlation with Osteoarthritis symptoms and function

The diagnosis of knee OA relies heavily on a combination of characteristic structural features from radiological images and OA symptoms [73]. However, x-ray features correlate relatively poorly with symptoms [13, 74]. As described in [13], 21 studies examined the relation of MRI findings in OA to symptoms and 13 of these demonstrated a statistically significant correlation ($P < 0.05$). This indicates the inconsistency between the studies in determining associations of structural features to symptoms. The relationship between knee OA symptoms and gait changes remains to be clarified and a better understanding of this link could advance the treatment and prevention of disease progression. Unlike static imaging, knee sounds captured during dynamic activity may correlate with symptoms and function. The study presented in this section relates to the following publication [C3] and aims at exploring the association between knee sounds and knee OA symptoms and function.

Knee symptoms and function were assessed, at the time of sound data collection, via self-reported Knee injury and Osteoarthritis Outcome Score (KOOS) questionnaire [75].

	Healthy	OA
Participants	19	17
Females/males	5/14	9/8
Age (years)	40.1 ± 18.3	60.5 ± 14.6
BMI (kg/m^2)	23.7 ± 2.9	28.5 ± 6.6
KOOS		
Symptoms	94.7 ± 4.8, (82.1 to 100)	66.9 ± 20.0, (35.7 to 100)
Pain	94.1 ± 7.3, (77.8 to 100)	69.6 ± 19.5, (16.7 to 100)
ADL	97.2 ± 4.0, (87.5 to 100)	77.2 ± 22.3, (18.3 to 100)
Sport/rec	88.1 ± 14.1, (56.3 to 100)	51.4 ± 25.7, (5.0 to 100)
QOL	84.2 ± 21.0, (25.0 to 100)	47.7 ± 26.3, (6.3 to 100)

Table 2.2: Demographical characteristics of study participants. All numbers reported are mean $\pm$ standard deviation. The range of KOOS values is given inside the parenthesis.

The KOOS questionnaire is an evaluative instrument that is used to assess the patient's opinion about their knee and associated problems. It is intended for evaluating the changes induced by treatment for OA. The KOOS consists of five sub-scales that are separately scored from 0 to 100: Pain, other Symptoms, Activity in Daily Living (ADL), Function in Sport and Recreation (Sport/rec), and knee-related Quality of Life (QOL) [75]. A score of 0 indicates extreme problems and a score of 100 indicates no problems. Table 2.2 presents the demographical characteristics of the participants used in this study with the associated KOOS values for each sub-scale.

Features were extracted from the knee sound recordings using spectro-temporal and cepstral analysis. More specifically, the STFT of a knee sound signal was obtained using fixed segmentation with 32 ms hamming windows with 50% overlap, giving 257 frequency bins. Subsequently, 11 summary statistics were obtained for each frequency bin by considering all the time frames (i.e. a horizontal slice through the spectrogram). The computed statistics are the mean, variance, skewness, kurtosis, max, min and the 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles. These statistical moments aim at capturing different aspects of the frequency variation across time. In addition to the spectrogram features, the log energy and 12 MFCC coefficients, computed as in [76], together with their corresponding first and second time derivatives (delta and delta-delta coefficients) were also extracted and the 11 statistics were subsequently computed per coefficient. In total, a 1×3256 feature

vector was used to represent each knee recording in the database.

The correlations between each individual feature and the KOOS were examined using the Pearson's correlation coefficient r [77]. Features producing $|r| \geq \theta$ were used to train binary regression decision trees, [78], in order to identify feature combinations predictive of KOOS. The threshold values, θ, were tested at $0.3, 0.35, \ldots, 0.65$. The limit cases were chosen so that combinations can be formed with features that are at least weakly correlated ($|r| \geq 0.3$) with the response variable and that $|r| \geq 0.7$ was observed only for 2 sub-scales (Pain, ADL). A 5-fold cross-validation training procedure, [79], was employed in which the data was randomly split into 5 folds (groups) where 4 folds were used for training a regression tree model and the left out fold was used to evaluate the model. This was repeated until all 5 folds were evaluated. Prior to this, the training data was scaled by subtracting the mean and normalising by the variance. The same normalization values were then applied to the test set. This cross-validation procedure was executed 100 times in order to reduce the variance of the estimator and the final predicted KOOS values were averaged. Pearson's correlation coefficient and Root Mean Square Error (RMSE) were used to assess the association between predicted and true values. The RMSE was computed per trial for each KOOS sub-scale and the final score was averaged over the number of trials (100).

Several techniques were used for conditioning the signal prior to feature extraction: (a) Root Mean Square (RMS) normalisation so that all signals have equal loudness, (b) High-Pass Filter (HPF) using the Parks-McClellan optimal Finite Impulse Response (FIR) filter design, [80], with a filter cut-off frequency of 250 Hz, (c) Moving Average (MA) filter with a 1.5 ms window, (d) Mean subtraction, achieved by splitting the signal into 1.5 ms frames with 50% overlap, finding the mean amplitude value per frame and subsequently subtracting, from each signal's sample amplitude, the value found by shape-preserving piecewise cubic interpolation, [81], using the per frame mean amplitude values as query points and (e) Mean subtraction and variance scaling: the signal from (d) is divided by the standard deviation of its sample amplitude values to achieve unit variance. Moreover, combinations of these procedures were also used. For technique

(b) the filter is optimal in the sense that the maximum error between the desired and the actual frequency response is minimized. Technique (c) is employed in order to reduce random noise. Techniques (b) and (d) are employed in order to remove the baseline signal and the 250 Hz cut-off frequency in (b) was chosen to attenuate the motion artifact noise from the microphone's contact with the skin. In [61] the authors used a Kaiser-window Finite Impulse Response (FIR) bandpass filter with a lower cut-off frequency of 400 Hz in order to attenuate this type of noise and in [60] the same window type was used with 1 kHz lower threshold. This inconsistency is due to the lack of studies focused on measuring the frequency response of the noise due to skin rubbing and motion artifacts.

The prediction results are presented and compared in Table 2.3. The table displays the lowest average RMSE values (averaged over 100 trials and rounded to 1 decimal place) between the predicted and the true values per KOOS sub-scale, obtained amongst all correlation threshold values tested for each combination of the preprocessing techniques. The techniques are represented in the table with the list symbol used to describe them in the previous paragraph and N/A in the table indicates that no preprocessing is applied to the signals. Bold values indicate the lowest RMSE achieved per sub-scale.

The sensitivity analysis performed for the correlation threshold showed that in the case of predicting the individual KOOS scores, the lowest RMSE is obtained with $\theta = 0.5$ and $\theta = 0.45$ in the majority of the cases (41 out of 75), as indicated in Table 2.3. Decreasing θ lower than 0.4, results in higher RMSE values. One factor that causes the reduced performance is overfitting that is caused by the inclusion of more and more features in the training stage. Another factor is the large number of features which overshadows the patterns that are present in the data, as Guyon et. al. in [82] describes. The features are often not all equally informative and the presence of additional redundant features does not always result in improved findings. On the contrary, these features can mask the hidden natural patterns in the data [82]. Regression trees trained using features extracted from the knee joint sound signals achieve an RMSE for symptoms, pain and ADL that is lower than Sport/rec and QOL. The scatter plots in Figure 2.10 display the predicted vs true values per KOOS sub-scale for the bold cases of Table 2.3. The red lines indicate

Preproc.	Symptoms		Pain		ADL		Sport/rec		QOL	
	RMSE	θ	RMSE	θ	RMSE	θ	RMSE	θ	RMSE	θ
N/A	19.2	0.5	20.5	0.5	18.4	0.6	25.7	0.45	26.9	0.45
(a)	20.9	0.5	19.9	0.65	19.0	0.6	31.4	0.5	31.7	0.45
(b)	**16.4**	0.45	21.1	0.55	17.4	0.6	29.8	0.45	30.0	0.5
(a),(b)	16.8	0.5	19.2	0.65	21.4	0.4	32.4	0.5	29.7	0.5
(c)	20.8	0.45	20.0	0.45	19.0	0.45	30.1	0.45	32.6	0.4
(a),(c)	19.9	0.45	20.1	0.45	19.3	0.65	31.0	0.5	34.1	0.45
(a),(b),(c)	18.8	0.5	20.1	0.65	19.6	0.5	30.6	0.5	30.7	0.45
(d)	21.3	0.5	20.2	0.45	**17.2**	0.45	28.5	0.55	30.8	0.4
(e)	19.8	0.6	19.5	0.65	18.5	0.45	26.1	0.65	**26.5**	0.5
(a),(d)	20.6	0.5	18.8	0.65	19.3	0.5	26.8	0.5	30.9	0.45
(b),(d)	17.2	0.65	**18.2**	0.55	18.1	0.5	27.9	0.5	33.6	0.45
(a),(e)	20.1	0.6	19.2	0.65	18.7	0.45	**25.7**	0.6	26.6	0.5
(b),(e)	17.0	0.65	21.0	0.6	19.4	0.45	28.7	0.55	31.3	0.5
(a),(b),(d)	17.2	0.65	21.2	0.6	21.8	0.65	29.0	0.5	29.9	0.5
(a),(b),(e)	17.2	0.65	21.1	0.6	19.3	0.45	29.2	0.55	31.1	0.5

Table 2.3: Lowest RMSE (averaged over 100 trials) between the predicted and the true values per KOOS sub-scale, obtained amongst all correlation threshold values, θ, used for each combination of preprocessing techniques (N/A indicates that no preprocessing is applied to the signals). Bold values indicate the lowest RMSE achieved per sub-scale.

the line of best fit through the points which is obtained by a simple linear model found using least-squares [83]. The Pearson's correlation coefficient (r) between the true and the predicted values is also displayed in the title of each plot. Figure 2.10 shows that regression trees trained using the selected features are generally effective at discriminating between high and low KOOS and within the bounds expected given the subjective nature of the scores. The predictions are seen to correlate strongly with the true scores for Symptoms ($r = 0.67$), QOL ($r = 0.65$) and Function in Sport and Recreation ($r = 0.6$) and less for ADL ($r = 0.55$) and pain ($r = 0.52$). In general, the RMSE values are high but this is expected given that the scoring range of the available data is very large (see Table 2.2) and that there isn't enough data to adequately cover this range for both healthy and OA knees. Nevertheless, the findings of this section of work indicate that knee sounds are predictive of OA symptoms and function ($r \geq 0.6$) and could be used as a screening tool for the mild OA cases, avoiding the frequent need for X-radiology or MRI and their associated health and economic risks. Whilst encouraging, further investigation is required among a larger cohort with KOOS values that cover the entire range (0 to 100). Such further investigation is however outside the scope of this thesis.

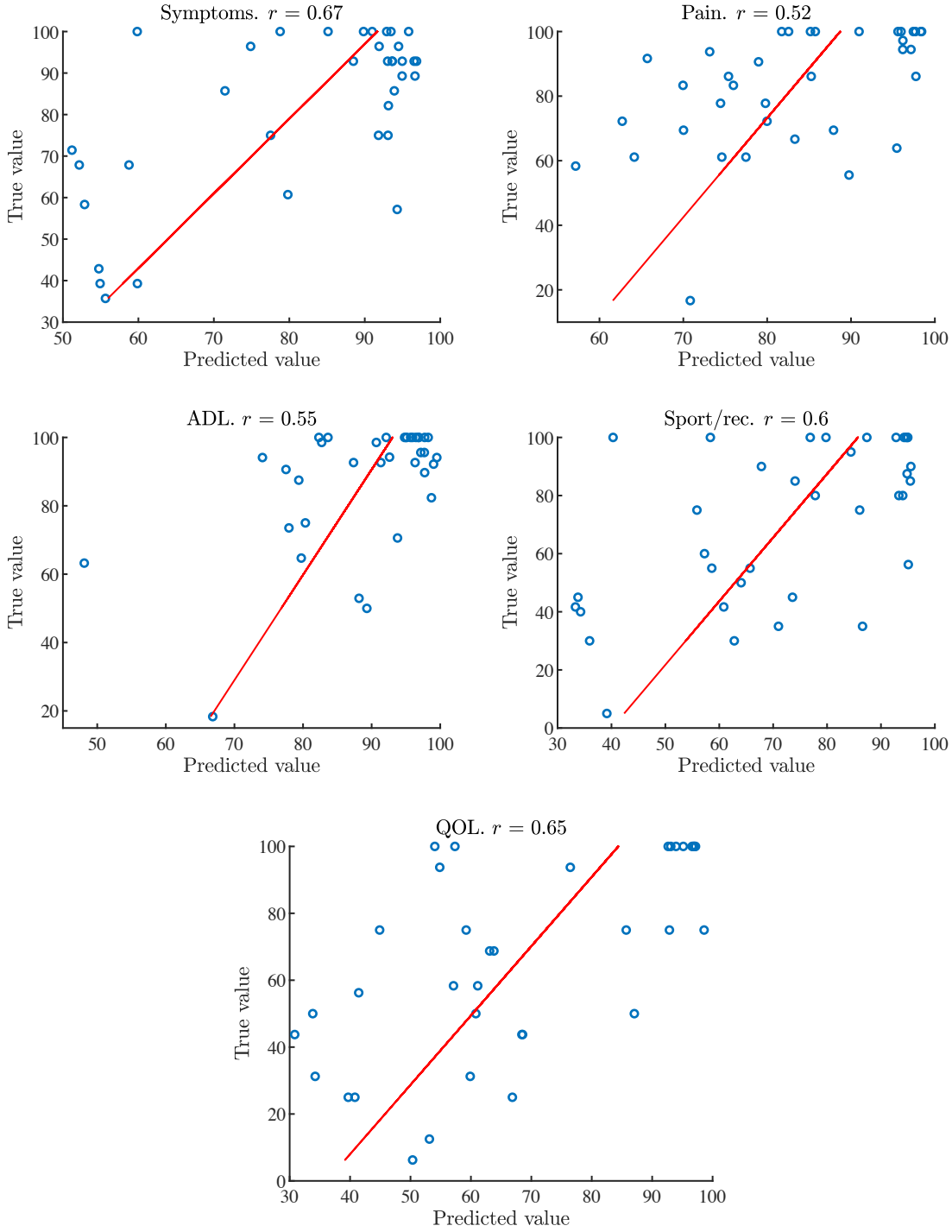

Figure 2.10: Scatter plots show the true vs predicted values per KOOS sub-scale. Predictions were performed using a 5-fold cross-validation training procedure using regression trees and features extracted from the knee joint sound signals only. Red lines indicate the line of best fit through the points and r is the Pearson's correlation coefficient between true and predicted values.

2.5 Conclusion

This Chapter presented the data acquisition system, test protocol and the demographical characteristics of the subjects recruited for sound data acquisition. Exploratory analysis was conducted on the knee joint sound signals using temporal and time-frequency methods. The generalised Rényi entropy was employed to investigate the spectrogram content and showed that OA knees generate sounds with more high frequency content than healthy or OAH knees (Figure 2.9). Correlation and prediction of the KOOS was then explored. MFCC and STFT features extracted from the sound signal and represented with 11 statistical parameters were used with regression trees. In the following Chapter, the 11 statistical parameters are further explored and will be mathematically formulated in the context of effective parameterisation and classification of the knee joint sounds.

Chapter 3

Segment-based Analysis and Classification of Knee Sounds

THE work presented in this Chapter investigates various feature-based descriptions for the sound signals emitted from the knee. In particular, discriminative features are sought that are relevant to the analysis and classification of normal (clinically healthy) and abnormal knee joints (clinically OA). The Chapter is divided into two parts. The first part, Section 3.2, develops a comprehensive experimental framework for the comparative analysis of several features and presents an evaluation of their discriminant capabilities. The second part, presented in Section 3.3, uses the experimental framework developed in the previous section and studies the discriminant properties of time-frequency representations of the knee sound signals. Finally, important results and conclusions are summarized at the end of the Chapter. Section 3.2 relates to conference publication [C1]. The work presented in Section 3.3 is submitted for publication to IEEE Transactions on Biomedical Engineering [J1].

3.1 Introduction

The work presented throughout this thesis focuses on the study of the acquisition and analysis of sounds generated by the knee during walking with particular focus on the effects

due to OA. In this context, Chapter 3 examines the relation between the knee condition and this type of sounds and investigates their potential for non-invasive detection of knee OA. The research question of this Chapter is therefore formulated in the following:

> *Can we find structure in the sound of the knees that will enable us to identify acoustic signatures that are indicative of Osteoarthritis?*

In line to the above question, a hypothesis is formed which states that *"During walking, the sounds emitted by knees with osteoarthritis are statistically different from those emitted by clinically healthy knees"*. The aim of this Chapter is to present an analytical study towards testing this hypothesis and answering the research question with evidence from the data. More specifically, it explores various features extracted from the knee sound signals and examines their discriminant power for the binary classification task of normal (healthy) vs abnormal (OA). Ultimately, the aim is to obtain insights into the nature and the fundamental differences in sounds between clinically healthy and clinically OA knees. The novel contributions of the work in this Chapter are:

1. The use of dynamic functional activity for knee joint assessment and condition classification using the knee sounds sensed at the patella.
2. The development of a comprehensive experimental framework for the investigation of the discriminant capabilities of various features for the supervised classification task of normal vs abnormal knee joint sound signals. This framework takes into account the possible inter-subject knee sound differences and proposes the parameterisation of the features extracted from the knee sound signals using 11 statistical parameters.
3. An efficient algorithm for the detection and extraction of the acoustic events characterised by short duration and high frequency.
4. The use of the modulation magnitude spectrum and the Mel-frequency Cepstral Coefficients are novel in the context of knee joint condition assessment in the presence of osteoarthritis using the sound signal sensed at the patella.
5. The frequency bands 220 Hz to 420 Hz and 1 kHz to 3.4 kHz were found to contain

a collection of features that are important for classifying such a signal into normal or abnormal and are derived from the spectrum of the knee sound signal.

In what follows, the work that lead to the findings outlined above will be presented with experiments and results. Most importantly, it will be shown why the answer to the research question is 'yes'.

3.2 Acoustic analysis and assessment of the knee

3.2.1 Pre-processing and feature extraction

Acoustic signals are sensed over the patella using a contact microphone, as was described in the previous Chapter. Let $s_i(n)$ denote the signal at discrete time index n captured by the patella microphone for the i^{th} knee in the data-set, where $i = 1, 2, \ldots, I$ for I knees in total. Prior to feature extraction, all recorded signals are normalised to have equal RMS level to allow the comparison of level dependent features from signals obtained at different recording levels. It is assumed that sounds related to abnormalities appear within time periods of τ_s seconds. Accordingly, $s_i(n)$ is divided into non-overlapping segments of length τ_s, denoted as $s_{i,j}(n)$ for $j = 1, 2, \ldots, J_i$ segments. Each segment is then labelled for classification according to the condition of the knee from which it was obtained.

I. Spectral and cepstral features

The signal segment $s_{i,j}(n)$ is further divided into frames of length l ms with 50% overlap. This creates an $N_f \times l_n$ matrix $\boldsymbol{S}$ where N_f is the number of frames and l_n is the frame length in samples. Considering a hanning window of length l_n transformed into the diagonal squared matrix $\boldsymbol{H}$, the DFT of $\boldsymbol{S}$ is computed as

$$\boldsymbol{\Psi_f} = (\boldsymbol{SH})\boldsymbol{W}_{l_n} \tag{3.1}$$

where

$$\boldsymbol{W}_{l_n} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-\frac{2\pi \mathrm{i}}{l_n}} & e^{-\frac{4\pi \mathrm{i}}{l_n}} & \dots & e^{-\frac{2\pi \mathrm{i}(l_n-1)}{l_n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\frac{2\pi \mathrm{i}(l_n-1)}{l_n}} & e^{-\frac{4\pi \mathrm{i}(l_n-1)}{l_n}} & \dots & e^{-\frac{2\pi \mathrm{i}(l_n-1)(l_n-1)}{l_n}} \end{bmatrix} \tag{3.2}$$

is the Vandermonde matrix for the roots of unity, otherwise known as the DFT matrix in this context. Each element of $\boldsymbol{W}$ is given by $e^{-\frac{2\pi \mathrm{i} nk}{l_n}}$ where for each row $n = 0, 1, \dots, l_n - 1$ and for each column $k = 0, 1, \dots, l_n - 1$ where k is the frequency index. By taking the magnitude of each element in $\boldsymbol{\Psi_f}$ and retaining only the first $K = \lfloor 1 + l_n/2 \rfloor$ columns, the matrix $\boldsymbol{\Psi_F}$ is constructed.

Additionally to $\boldsymbol{\Psi_F}$, the dynamics of the STFT magnitude spectrum also contain information, as can be seen for example in the spectrograms of Figure 2.8. To examine whether this information is relevant to the classification task at hand, the modulation magnitude spectrum $\boldsymbol{\Psi_{FF}}$ is obtained as in [84]. That is, the DFT of the acoustic magnitude spectrum $\boldsymbol{\Psi_F}$ is computed at each frequency bin using a window of T acoustic time frames and without overlap, resulting to the matrix $\boldsymbol{\Psi_{FF}}$ with N_f/T rows and K columns.

Extraction procedures inspired by human auditory perception are widely used in many applications. MFCCs are a common choice of features that are successfully used in speech recognition and music genre classification. MFCC have been previously used for the analysis of VAG signals, [85,86], but have not so far been used for OA detection from the analysis of acoustic signals emitted from the knee and sensed at the patella. Their extraction process involves mapping the power of the STFT spectrum using triangular overlapping windows onto the mel scale which is designed to approximate the human auditory system's response. The aim is to exploit the property of the mel scale and apply it to the knee signals. In particular, for the sounds heard as pops, clicks, grindings etc. during knee motion. The use of the mel scale for knee signals is motivated by the fact that these sounds can be distinguished in a recording by even the untrained human ear with minimal effort, so that a perceptually inspired approach is reasonable, and that they

are sounds that are likely generated by the friction between the tibia and the femur bones which in turn is caused by the effects of OA in the knee.

A filter-bank with N_B triangular band-pass filters equally spaced along the mel-frequency axis which is defined as in [87], is used to construct the matrix

$$\boldsymbol{U}_M = \begin{bmatrix} U_1(0) & U_2(0) & \dots & U_{N_B}(0) \\ U_1(\frac{2\pi}{K}) & U_2(\frac{2\pi}{K}) & \dots & U_{N_B}(\frac{2\pi}{K}) \\ \vdots & \vdots & \ddots & \vdots \\ U_1(\frac{2\pi(K-1)}{K}) & U_2(\frac{2\pi(K-1)}{K}) & \dots & U_{N_B}(\frac{2\pi(K-1)}{K}) \end{bmatrix} \tag{3.3}$$

where each element is the magnitude of the bandwidth of a single filter at a single frequency bin $k = 0, 1, \dots, K-1$. A compact spectrum representation can then be obtained as

$$\boldsymbol{\Psi_D} = \boldsymbol{\Psi_F} \boldsymbol{U}_M \ . \tag{3.4}$$

Hence, MFCC are computed as

$$\boldsymbol{C_M} = \mathcal{F}(\log(\boldsymbol{\Psi_D})) \tag{3.5}$$

where $\mathcal{F}(\cdot)$ is the Discrete Cosine Transform (DCT) operator [88]. For each element in $\boldsymbol{C_M}$, defined as a static coefficient and denoted as a_t for time frame t, the trajectories in time are computed as the first derivatives using

$$d_t = \frac{\sum_{u=1}^{U} (a_{t+u} - a_{t-u}) u}{2 \sum_{u=1}^{U} u^2} \tag{3.6}$$

where $U = 4$ and d_t is called the delta coefficient from frame t computed using the static coefficient of that frame. The second derivatives are also computed using (3.6) but replacing a_t with d_t and setting $U = 1$ which makes it a simple difference equation. The delta coefficients and then the delta-delta coefficients are appended, column-wise, at the end of $\boldsymbol{C_M}$, which results to increasing its column rank to $3N_B$. The choice of $U = 1$ and $U = 4$ are adopted from speech recognition.

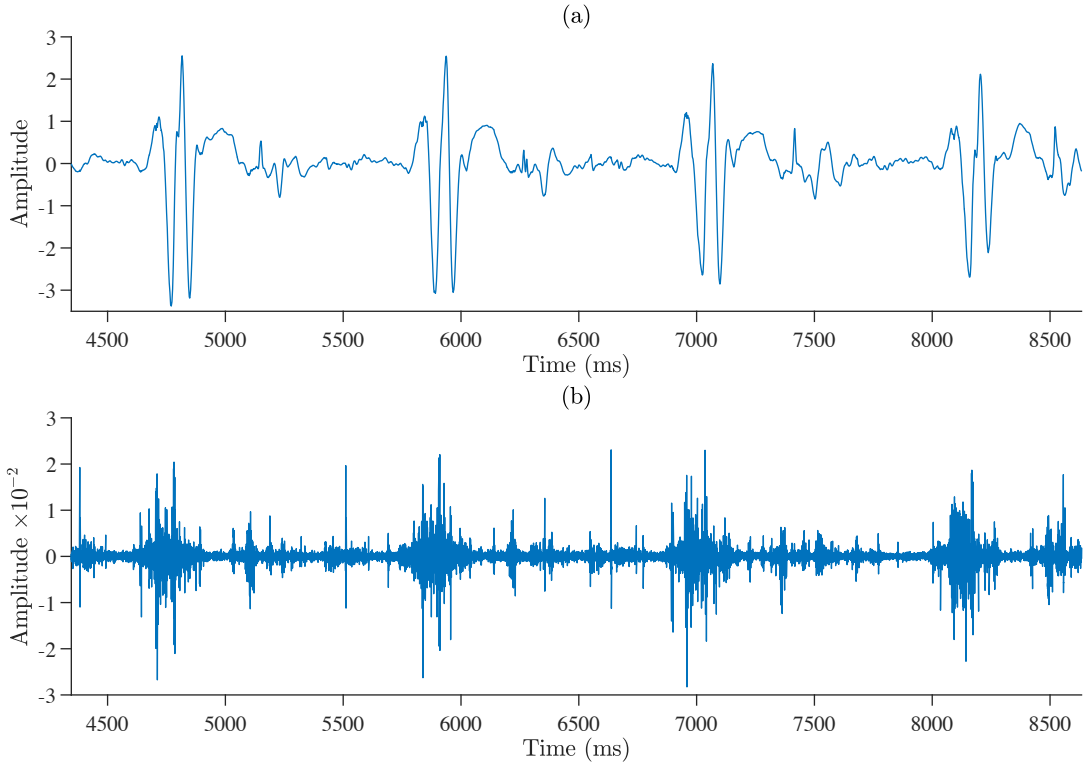

Figure 3.1: (a) Example signal $s_{i,j}(n)$ from an OA knee and (b) its equivalent high pass filtered version, $s_0(n)$, obtained using an FIR filter with cut-off frequency 250 Hz.

II. Acoustic emission pulse waveform elements

In the study of knee sound signals, occasional short acoustic pulses were observed. These pulses are identified as high frequency waveform elements appearing along the slowly varying baseline signal. Figure 3.1 (a) shows an example signal $s_{i,j}(n)$ and in plot (b) its equivalent high pass filtered version, $s_0(n)$, obtained using the Parks-McClellan optimal FIR filter design, [80], with a filter cut-off frequency of 250 Hz. The filter was employed in order to remove the baseline signal from $s_{i,j}(n)$ and the 250 Hz high-pass cut-off was chosen to attenuate the motion artifact noise from the microphone's contact with the skin. The explicit dependency of $s_0(n)$ to i and j in $s_{i,j}(n)$ is omitted for notational simplicity. It can be clearly seen that there is strong activity in the signal above 250 Hz.

To extract the AE pulses, a set of parameters that characterise them is defined as $\{\eta_0, t_0, r_0\}$. The amplitude definition parameter η_0 is used to distinguish potential pulses from spurious variations in $s_0(n)$. For the same reason the duration definition parameter

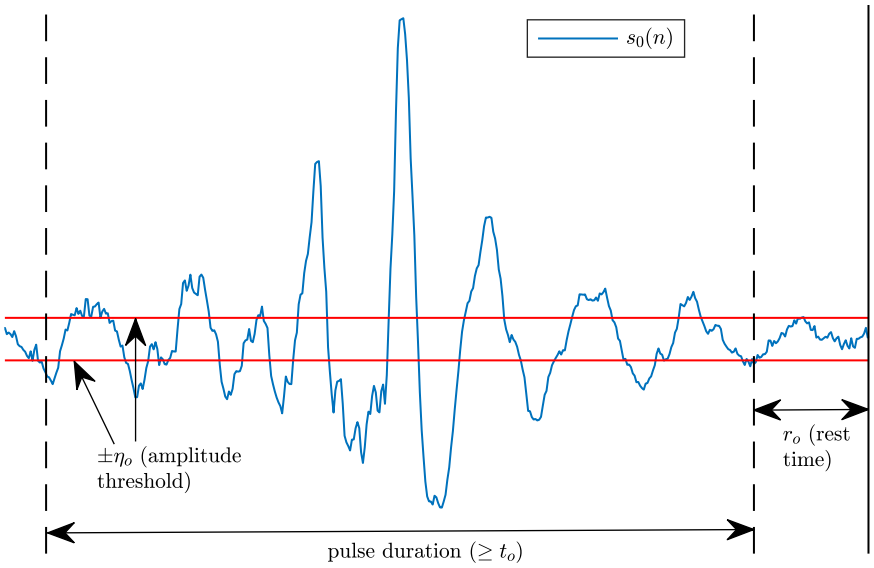

Figure 3.2: Acoustic emission pulse with definition parameters.

t_0 is also used and defines the minimum pulse duration. The third definition parameter, r_0, is the rest time and is a window within which the pulse waveform amplitude is below η_0 and determines the end of the pulse. Both t_0 and r_0 are defined in number of samples. Figure 3.2 shows $\{\eta_0, t_0, r_0\}$ on an example AE pulse that was obtained using the method described in the paragraphs to follow which was also used to extract all AE pulses from $s_{i,j}(n)$.

After high-pass filtering the signal segment at 250 Hz, $s_0(n)$ is split into 50 ms frames with 50% overlap which gives $N_f = \lfloor \tau_s/0.025 - 1 \rfloor$ frames. The variance of the sample amplitude values in each frame is then estimated. Let $\boldsymbol{y}$ be the vector formed by the collection of all the variances and let $\text{sort}(\cdot)\colon \mathbb{R} \to \mathbb{R}$ be the function defined using the general vector $\boldsymbol{x} \in \mathbb{R}$ as

$$\text{sort}(\boldsymbol{x}) := [x_1, x_2, x_3, \ldots]$$

that satisfies the following condition:

$$x_i \leq x_{i+1} \quad \forall\ i = 1, 2, 3, \ldots$$

where i indexes the elements of $\boldsymbol{x}$ and is not to be confused with the i that was used at

the beginning of this section to index the knees in the data-set. Then

$$\text{sort}(\boldsymbol{y}) \equiv \left[100(^{0.5}/_{N_f}), 100(^{1.5}/_{N_f}), \ldots, 100(^{(N_f-0.5)}/_{N_f})\right] \tag{3.7}$$

where the elements of the Right-Hand Side (RHS) vector are considered as the percentiles, and their associated value is the value of the corresponding element in sort($\boldsymbol{y}$). If a percentile value is not given by any of the elements in the RHS vector, it is found by linear interpolation. Following, the frames with sample amplitude variance less than the 10th percentile value are identified as noise segments. The percentile number (10th) was chosen to be as low as possible in order to collect frames with AE pulses of low amplitude but at the same time avoid the noise frames which are expected to have a much lower sample amplitude variance than the acoustic events of interest. The sample amplitudes of the noise segments form a Gaussian distribution, an example of which is shown in Figure 3.3. The distribution parameters are estimated using maximum likelihood, [89], and the inverse Cumulative Distribution Function (CDF) is subsequently obtained. The value of the amplitude definition parameter η_0 is computed from the CDF at a probability of $1 - \eta$ where η is a control variable. It was empirically found that $\eta = 10^{-14}$ is a good

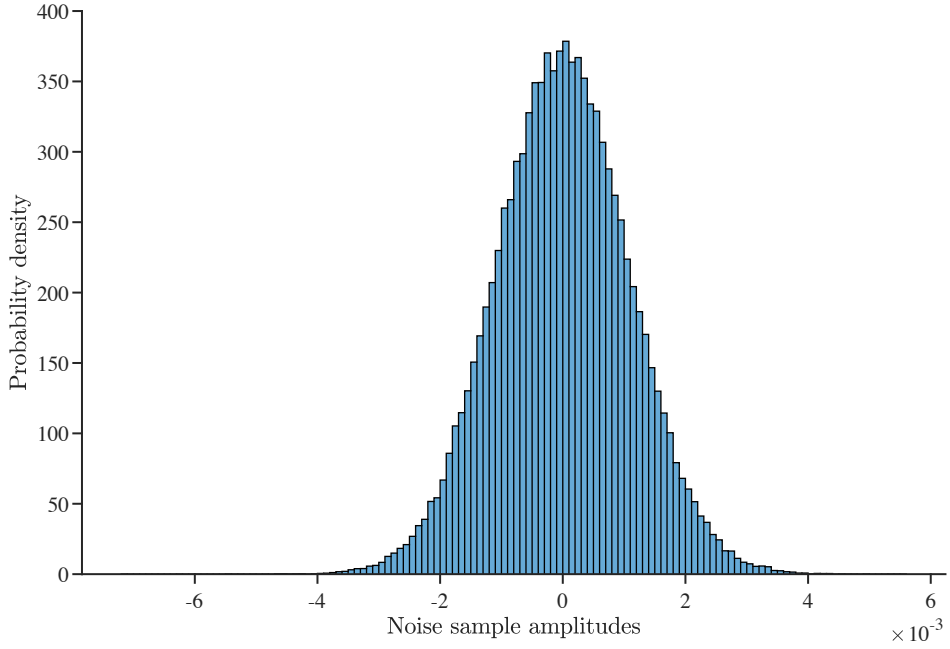

Figure 3.3: Probability density of noise samples amplitude.

choice which produces an η_0 that discards subtle and uninteresting variations in $s_0(n)$, as shown in the example of Figure 3.4. Changing any of the $\{\eta_0, t_0, r_0\}$ values will alter the number of pulses detected. The results in the following sections use the values displayed in Table 3.1.

Given η_0, the peaks in $|s_0(n)|$ that exceed $|\eta_0|$ are identified using standard peak finding techniques such as [90, 91]. If the width of a peak is not at least 2 samples long and with both samples located at amplitudes above $|\eta_0|$ it is discarded in order to avoid spurious threshold crossings. The width of 2 samples is empirically determined for a full Nyquist bandwidth signal sampled at $F_s = 16$ kHz. Suppose that the selected peaks occur at sample locations defined by the set $\{p_1, p_2, \ldots, p_c\}$ for c peaks. Let the unknowns $\{b_{ii}(j_1), v_{ii}(j_2)\}$ denote the pulse's waveform start and stop sample numbers for peak location p_{ii} with j_1 and j_2 denoting the repetition indices and $\{b_{ii}(0), v_{ii}(0) = p_{ii}\}$. To find the unknowns for each peak index ii, the following is repeated

$$b_{ii}(j_1) = \big[\min(b_{ii}(j_1 - 1) - n);\ n = 0, 1, \ldots, t_0;\ j_1 = 1, 2, 3, \ldots\big] \tag{3.8}$$

$$v_{ii}(j_2) = \big[\max(v_{ii}(j_2 - 1) + n);\ n = 0, 1, \ldots, t_0;\ j_2 = 1, 2, 3, \ldots\big] \tag{3.9}$$

subject to

$$\big|s_0\big(b_{ii}(j_1 - 1) - n\big)\big| > \big|\eta_0\big| \tag{3.10}$$

$$\big|s_0\big(v_{ii}(j_2 - 1) + n : v_{ii}(j_2 - 1) + n + 1\big)\big| > \big|\eta_0\big| \tag{3.11}$$

until the inequalities (3.10) and (3.11) do not hold for any n at the current j_1 and j_2 values. When this occurs, the repetition stops and the value r_0 is added to $v_{ii}(j_2)$. When the above is executed for all peaks, any overlapping pulses that might occur are merged into one. Figure 3.4 shows a snapshot of $s_0(n)$ with identified AE pulses. Clearly, there are many pulses within a single signal segment and can occur at various waveforms. For subsequent data analysis the peak-to-peak amplitude, duration and energy of these waveforms are extracted. Let the matrix $\boldsymbol{A}$ denote the collection of all the pulse features obtained from a single $s_{i,j}(n)$, where each row holds the values of the 3 features, for a single AE pulse, in the order: peak-to-peak amplitude, duration and energy.

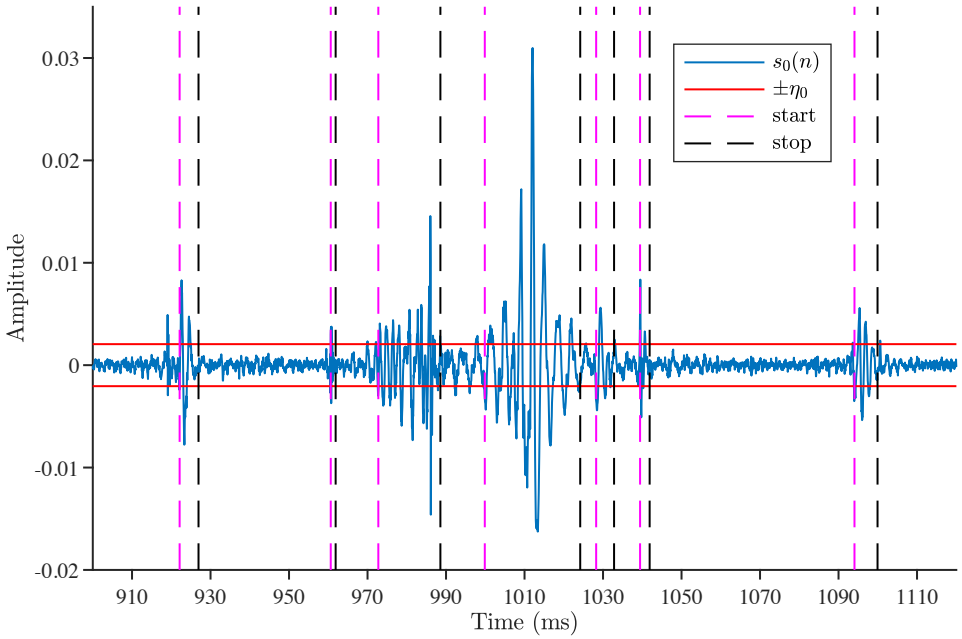

Figure 3.4: High-pass filtered signal $s_0(n)$ with the identified pulses indicated by the start (magenta) and stop (black) dashed lines. The pulses were obtained based on the algorithm values displayed in Table 3.1.

III. Synopsis

In the next sections, the discriminant capabilities of $\boldsymbol{\Psi_F}, \boldsymbol{\Psi_{FF}}, \boldsymbol{C_M}$ and $\boldsymbol{A}$ for the task of normal vs abnormal segment classification is studied. A statistical analysis on the separability of the feature distributions will be conducted first, followed by the multivariate discriminant analysis in Section 3.2.3. The aim of the analyses is to obtain insights into the nature and the fundamental differences between normal and abnormal knee signals.

3.2.2 Statistical analysis of feature distributions

Let $\mathbf{F}$ denote the set of the feature matrices obtained from all signal segments $s_{i,j}(n)$ in the order of $\{\boldsymbol{C_M}, \boldsymbol{\Psi_F}, \boldsymbol{\Psi_{FF}}, \boldsymbol{A}\}$. The parameter values used to extract these features are summarised in Table 3.1 where τ_s is set to 20 s which results to 249 normal and 297 abnormal acoustic signal segments to be used in the classification experiments. The acoustic emission pulse definition amplitude, η_0, is not shown in the table because its value is different for each $s_{i,j}(n)$ as it depends on the amplitude distribution of the noise

Parameter	Description	Value
F_s	Sampling frequency (in Hz)	16000
τ_s	Window size (in seconds) used to segment $s_i(n)$ into $s_{i,j}(n)$	20
l	Frame-size (in ms) used to segment $s_{i,j}(n)$ into frames	32
K	Number of frequency bins in $\boldsymbol{\Psi_F}$	257
T	Number of acoustic time frames used to compute $\boldsymbol{\Psi_{FF}}$	6
N_B	Number of filters in the mel filter-bank	20
t_0	Acoustic emission pulse definition duration (in samples)	$F_s/10^3$
r_0	Acoustic emission pulse definition rest time (in samples)	$F_s/10^3$

Table 3.1: Parameter values used in the feature extraction step.

samples which are identified per $s_{i,j}(n)$. Given the values in the table, the total number of features extracted per $s_{i,j}(n)$ are $N = 574$ from which 57 MFCC, where the zeroth MFCC and its delta and delta-delta coefficients were discarded, 257 magnitude STFT values, 257 modulation magnitude values and 3 AE pulse features.

As stated in the introduction of the Chapter, the aim is to discriminate between normal and abnormal signal segments $s_{i,j}(n)$. Therefore, let $\mathbf{F}_{he}$ and $\mathbf{F}_{oa}$ denote the feature sets obtained from normal and abnormal segments respectively, where $\mathbf{F}_{he}, \mathbf{F}_{oa} \subset \mathbf{F}$ and

$$\mathbf{F}_{he} = \{\boldsymbol{x}_1^{d_1}, \boldsymbol{x}_2^{d_2}, \ldots, \boldsymbol{x}_N^{d_N}\}$$
$$\mathbf{F}_{oa} = \{\boldsymbol{y}_1^{m_1}, \boldsymbol{y}_2^{m_2}, \ldots, \boldsymbol{y}_N^{m_N}\}$$

where the meaning of each element in both sets is the same as in $\mathbf{F}$ and the constants d_i and m_i denote each element's dimension. For example, $\boldsymbol{x}_1^{d_1}$ denotes the values of the first MFCC where d_1 is equal to the total number of frames obtained from all healthy $s_{i,j}(n)$ and similarly, $\boldsymbol{y}_1^{m_1}$ denotes the values of the first MFCC where m_1 is equal to the total number of frames obtained from all OA $s_{i,j}(n)$.

It is interesting to investigate whether $\boldsymbol{x}_n^{d_n}$ and $\boldsymbol{y}_n^{m_n}$ for $n = 1, 2, \ldots, N$ generate dissimilar sample distributions such that a classifier with low error rate could be designed. Let the null hypothesis H_0 state that $\boldsymbol{x}_n^{d_n}$ and $\boldsymbol{y}_n^{m_n}$ originate from the same continuous distributions $\forall n$. To test H_0, the two-sample Kolmogorov-Smirnov Test (KST) is used, which calculates the maximum absolute difference between the empirical CDFs [92], and

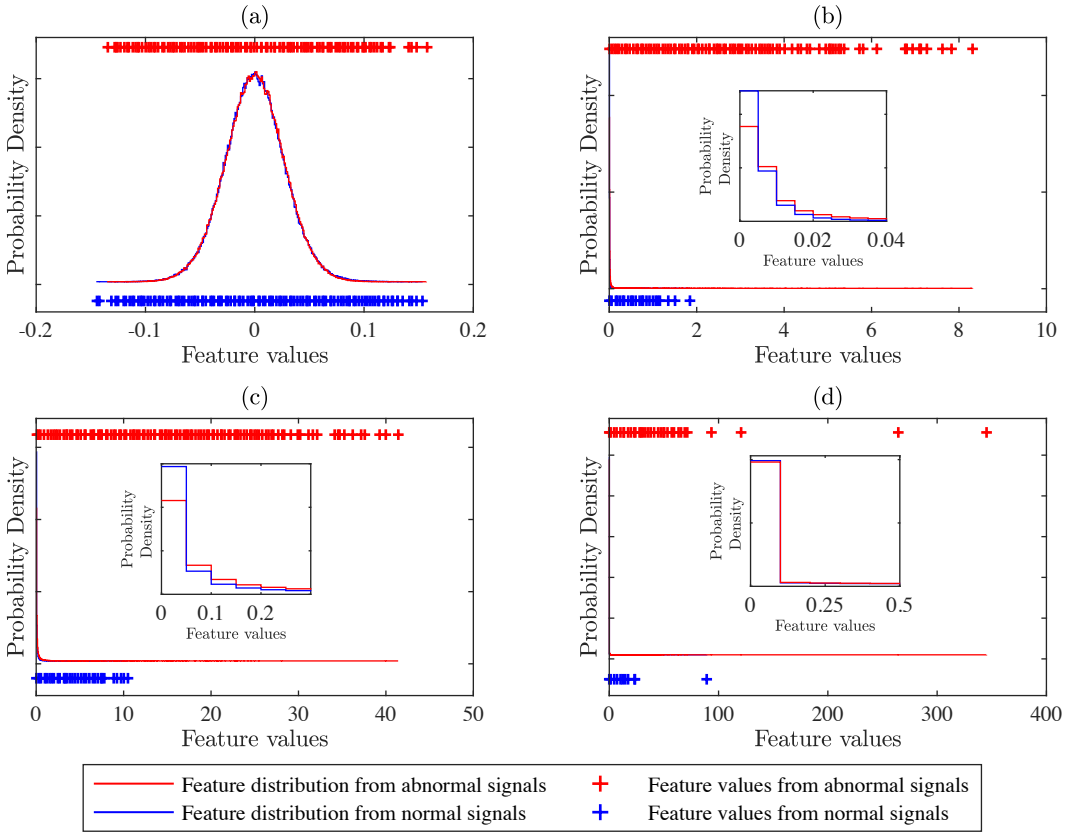

Figure 3.5: Distribution of feature values obtained from all normal signals (healthy; blue) and abnormal signals (OA; red) for the (a) first derivative of the 15th MFCC, (b) 94th STFT frequency bin, (c) 37th modulation index and (d) AE pulse energy, obtained using the parameters in Table 3.1. Scatter points indicate the individual feature values.

is one of the most standard tests to apply for such tasks. Test results at 5% significance level showed that H_0 was rejected for 565 features implying that their statistical differences are significant. H_0 was accepted for only 9 out of 574 features and these are the second derivatives of the 9th, 10th, 12th−16th and the 19th MFCC and the first derivative of the 15th MFCC. Figure 3.5 shows four examples of such $x_n^{d_n}$ and $y_n^{m_n}$ distributions where for those in plot (a) the null hypothesis was accepted whereas for the distributions in plots (b), (c) and (d) it was rejected. The smaller plots within (b), (c) and (d) are the enlarged versions of the respective distribution plots, showing the distributions close to the origin in order to enhance the visual comparison. It is clear that in plot (a) there is significant overlap between the distributions which results to H_0 being accepted. In plots (b), (c) and (d) the tails of the feature distributions obtained from the abnormal signals extend

to much larger values than the corresponding obtained from the normal signals. However, the majority of the feature values are clustered close to zero.

To further compare the distributions, the Bhattacharyya distance is used to measure the distance between two distributions [93]. It takes into account both the separation of means and the separation due to covariances (or variances for univariate models) [93]. It is defined as

$$B_d = -\log_e(B_c) \tag{3.12}$$

where B_c is the Bhattacharyya coefficient given by

$$B_c = \int_{-\infty}^{+\infty} \sqrt{p_1(x)p_2(x)}\, dx \tag{3.13}$$

where $p_1(x)$ and $p_2(x)$ are the two continuous distributions in question. The coefficient value varies from 0 to 1, where $B_c = 1$ indicates two identical distributions w.r.t. the mean and covariance and $B_c = 0$ indicates that there is no overlap at all. MFCC and their first and second derivatives are assumed to follow normal distributions, as shown in the example of Figure 3.5 (a), in which case (3.13) can be easily simplified [79]. That is, for two normal distributions, $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, B_c is given by

$$B_c = \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} e^{-\frac{1}{4}\frac{(\mu_2-\mu_1)^2}{\sigma_1^2+\sigma_2^2}} \tag{3.14}$$

and the corresponding distance is computed as

$$B_d = \frac{1}{4}\frac{(\mu_2 - \mu_1)^2}{\sigma_1^2 + \sigma_2^2} - \frac{1}{2}\log_e\Big[\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}\Big]\ . \tag{3.15}$$

The first term of (3.15) gives the class separability due to the means while the second term gives the separability due to the variances. The normality assumption however, does not hold for the features in $\boldsymbol{\Psi_F}$, $\boldsymbol{\Psi_{FF}}$ and $\boldsymbol{A}$, as can be seen in the examples of Figure 3.5 and therefore various distribution fits, obtained using maximum likelihood estimation, were explored (gamma, exponential, weibull, inverse gaussian, rayleigh). The final choice was made based on visual inspection of the histograms with the distributions overlaid. Gamma was observed to be a better approximation for the features in $\boldsymbol{\Psi_F}$ and $\boldsymbol{A}$ while exponential

was for the features in $\boldsymbol{\Psi_{FF}}$. The gamma distribution is given by, [94],

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \tag{3.16}$$

for $x > 0,\ \alpha > 0$ and $\beta > 0$ where the latter two are the distribution parameters and $\Gamma(\cdot)$ is the gamma function defined as

$$\Gamma(c) = \int_0^{+\infty} e^{-t} t^{c-1}\ dt \tag{3.17}$$

for a positive real number c [94]. The expressions for B_c and B_d between two gamma distributions are not readily available [95]. To derive them, (3.16) is substituted in (3.13) for gamma distributions with parameters β_1, α_1 and β_2, α_2 respectively, giving

$$\begin{aligned} B_c &= \int_0^{\infty} \Big[\frac{1}{\beta_1^{\alpha_1}\Gamma(\alpha_1)} x^{\alpha_1-1} e^{\frac{-x}{\beta_1}} \frac{1}{\beta_2^{\alpha_2}\Gamma(\alpha_2)} x^{\alpha_2-1} e^{\frac{-x}{\beta_2}} \Big]^{\frac{1}{2}} dx \\ &= \frac{1}{\big[\Gamma(\alpha_1)\Gamma(\alpha_2)(\beta_1^{\alpha_1}\beta_2^{\alpha_2})\big]^{\frac{1}{2}}} \int_0^{\infty} x^{\frac{\alpha_1+\alpha_2}{2}-1} e^{-(\frac{1}{2\beta_1}+\frac{1}{2\beta_2})x} dx\ . \end{aligned} \tag{3.18}$$

An expression for the integral in the equation above can be found by first solving (3.17) in which the exponent of e is multiplied with a positive real number b:

$$\int_0^{+\infty} e^{-bt} t^{c-1}\ dt = \tag{3.19}$$

$$= \underbrace{\big[-\frac{t^{c-1}e^{-bt}}{b} \big]_0^{+\infty}}_{0} + \int_0^{+\infty} \frac{(c-1)}{b} t^{c-2} e^{-bt}\ dt$$

$$= \frac{(c-1)}{b} \left[\underbrace{\big[-\frac{(c-2)t^{c-2}e^{-bt}}{b} \big]_0^{+\infty}}_{0} + \int_0^{+\infty} \frac{(c-2)}{b} t^{c-3} e^{-bt}\ dt \right]$$

$$= \frac{(c-1)(c-2)}{b^2} \int_0^{+\infty} t^{c-3} e^{-bt}\ dt$$

$$= \underbrace{\cdot\cdot\cdot\cdot\cdot\cdot}_{c-3 \text{ times}}$$

$$= \frac{(c-1)!}{b^c}$$

$$= \frac{\Gamma(c)}{b^c} \ . \tag{3.20}$$

The last step can be easily shown if (3.17) is solved in the same way as above. Recognizing that the integral in (3.18) is equivalent to (3.19) with $c = \frac{\alpha_1+\alpha_2}{2}$ and $b = \frac{1}{2\beta_1} + \frac{1}{2\beta_2}$ and using (3.20), gives the expression for the Bhattacharyya coefficient and distance between two gamma distributions:

$$B_c = \frac{1}{\left[\Gamma(\alpha_1)\Gamma(\alpha_2)(\beta_1^{\alpha_1}\beta_2^{\alpha_2})\right]^{\frac{1}{2}}} \frac{\Gamma(\frac{\alpha_1+\alpha_2}{2})}{(\frac{1}{2\beta_1} + \frac{1}{2\beta_2})^{\frac{\alpha_1+\alpha_2}{2}}}$$

$$= \frac{\Gamma(\frac{\alpha_1+\alpha_2}{2})}{\left[\Gamma(\alpha_1)\Gamma(\alpha_2)(\beta_1^{\alpha_1}\beta_2^{\alpha_2})(\frac{1}{2\beta_1} + \frac{1}{2\beta_2})^{\alpha_1+\alpha_2}\right]^{\frac{1}{2}}} \tag{3.21}$$

$$\implies B_d = -\log_e \left[\frac{\Gamma(\frac{\alpha_1+\alpha_2}{2})}{\left[\Gamma(\alpha_1)\Gamma(\alpha_2)(\beta_1^{\alpha_1}\beta_2^{\alpha_2})(\frac{1}{2\beta_1} + \frac{1}{2\beta_2})^{\alpha_1+\alpha_2}\right]^{\frac{1}{2}}} \right] . \tag{3.22}$$

The exponential distribution is a special case of gamma and is obtained by setting $\alpha = 1$ in (3.16) and noting from (3.17) that $\Gamma(1) = 1$. The expression for the B_d between two exponential distributions can be derived in the same way as done for gamma or by simply setting $\alpha_1 = 1$ and $\alpha_2 = 1$ in (3.21) to obtain

$$B_c = \frac{2\sqrt{\beta_1\beta_2}}{\beta_1 + \beta_2} \tag{3.23}$$

$$\implies B_c = -\log_e \left[\frac{2\sqrt{\beta_1\beta_2}}{\beta_1 + \beta_2} \right] . \tag{3.24}$$

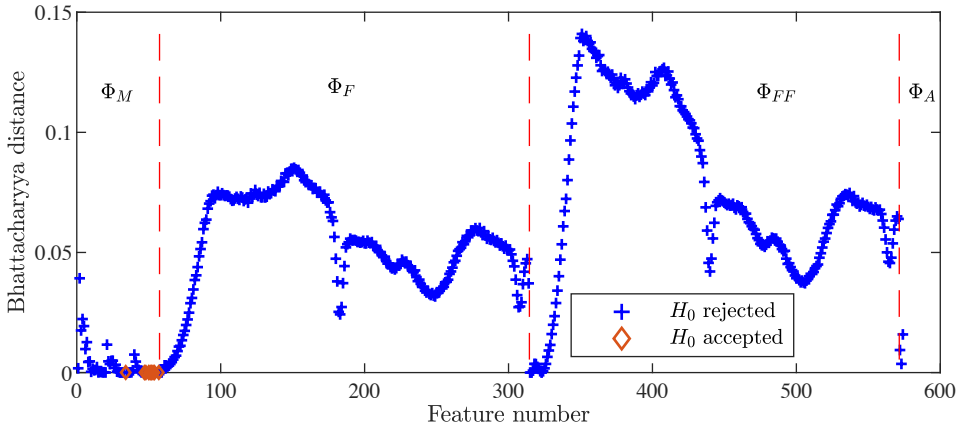

Figure 3.6: Bhattacharyya distance between the distributions formed by the corresponding features in $\mathbf{F}_{he}$ and $\mathbf{F}_{oa}$. The result of the two sample Kolmogorov-Smirnov Test is indicated by the marker type. Dashed lines separate the feature types.

Figure 3.6 shows the Bhattacharyya distances, B_d, computed between each corresponding feature distribution from $\mathbf{F}_{oa}$ and $\mathbf{F}_{he}$ using the appropriate equations as indicated earlier. The model parameters were estimated from the sample population. The results shown in the figure support the KST outcome (showing minimal B_d for the features where H_0 was accepted) but also indicate that the magnitude modulation features, $\boldsymbol{\Psi_{FF}}$, corresponding to frequencies in the range 0.97 to 3.30 kHz exhibit the largest separations suggesting that they have higher discriminant power than the rest. For $\boldsymbol{\Psi_F}$, the feature distributions corresponding to frequencies in the range 1.09 to 3.67 kHz exhibit the largest separations. The B_d for $\boldsymbol{A}$ and $\boldsymbol{C_M}$ are amongst the lowest obtained which indicates significant overlap between the distributions. Column 2 of $\boldsymbol{C_M}$ (i.e. the 2$^{\text{nd}}$ MFCC) and column 3 of $\boldsymbol{A}$ (i.e. the energy of AE pulses) produce the most distinctive feature distributions within their respective feature sets. In general, B_d values are low which implies significant overlap between the feature distributions obtained from normal and abnormal signals respectively. However, a closer look at these distributions reveals that despite the overlap, the shapes and tails can be different, as can be seen in the example of Figure 3.7 and those of Figure 3.5 (b), (c) and (d). Using the whole distribution is hindering certain class differences. Therefore, by using the mean, variance, kurtosis, skewness, max, min and the 10th, 25th, 50th, 75th, 90th percentiles to represent each feature distribution, the

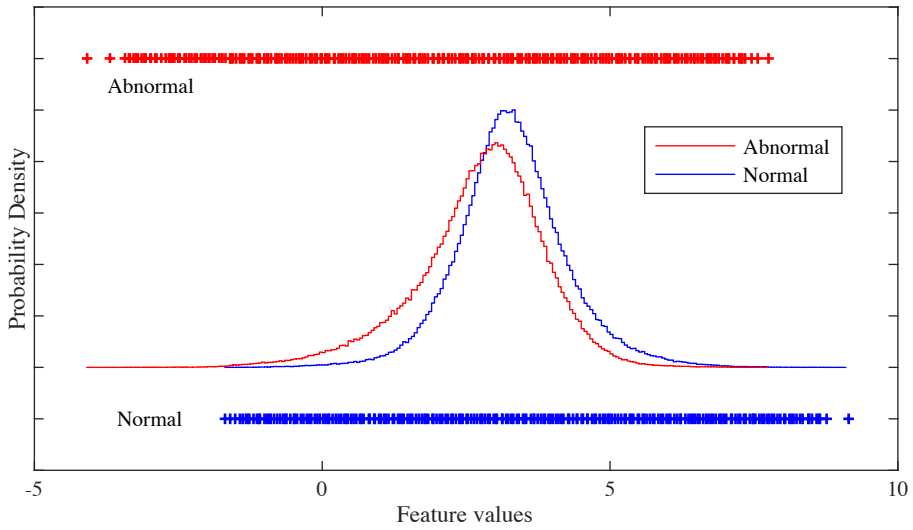

Figure 3.7: Distribution of feature values obtained from all normal signals (healthy; blue) and abnormal signals (OA; red) for the 2nd MFCC. Scatter points indicate the individual feature values.

aim is to capture certain attributes that will instead highlight these class differences.

Let $\mathcal{D}(\cdot)$ be a matrix operator such that when it is applied to an $L \times M$ matrix, for example, it returns 11 statistical moments of each column in a new $11 \times M$ matrix. These 11 moments are the mean, kurtosis, variance, skewness, max, min and the 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles. Let $\boldsymbol{f}_1^F$ denote the 11-dimensional statistical feature vector representation of the first column of $\boldsymbol{\Psi_F}$. Using this notation, the feature matrices representing a single signal segment $s_{i,j}(n)$ are defined as

$$\begin{aligned}
\boldsymbol{\phi_M} &= \mathcal{D}(\boldsymbol{C_M}) = [\boldsymbol{f}_1^M, \boldsymbol{f}_2^M, \ldots, \boldsymbol{f}_{N_B}^M]^T \\
\boldsymbol{\phi_F} &= \mathcal{D}(\boldsymbol{\Psi_F}) = [\boldsymbol{f}_1^F, \boldsymbol{f}_2^F, \ldots, \boldsymbol{f}_K^F]^T \\
\boldsymbol{\phi_{FF}} &= \mathcal{D}(\boldsymbol{\Psi_{FF}}) = [\boldsymbol{f}_1^{FF}, \boldsymbol{f}_2^{FF}, \ldots, \boldsymbol{f}_K^{FF}]^T \\
\boldsymbol{\phi_A} &= \mathcal{D}(\boldsymbol{A}) = [\boldsymbol{f}_1^A, \boldsymbol{f}_2^A, \boldsymbol{f}_3^A]^T \ .
\end{aligned} \tag{3.25}$$

Following similar notation, the feature sets representing all signal segments are defined as

$$\begin{aligned}\boldsymbol{\Phi_M} &= [\boldsymbol{\phi_{M1}}, \boldsymbol{\phi_{M2}}, \ldots, \boldsymbol{\phi_{MC}}] \\ \boldsymbol{\Phi_F} &= [\boldsymbol{\phi_{F1}}, \boldsymbol{\phi_{F2}}, \ldots, \boldsymbol{\phi_{FC}}] \\ \boldsymbol{\Phi_{FF}} &= [\boldsymbol{\phi_{FF1}}, \boldsymbol{\phi_{FF2}}, \ldots, \boldsymbol{\phi_{FFC}}] \\ \boldsymbol{\Phi_A} &= [\boldsymbol{\phi_{A1}}, \boldsymbol{\phi_{A2}}, \ldots, \boldsymbol{\phi_{AC}}] \end{aligned} \tag{3.26}$$

for $C = \sum_{i=1}^{I} J_i$ total segments. The effectiveness of this parameterisation is examined in the discriminant analysis described in the following section.

3.2.3 Multivariate feature analysis

Bayes' rule for minimum error states that a vector $\boldsymbol{y} = \{y_1, \ldots, y_p\}$ is assigned to class ω_j if $p(\omega_j|\boldsymbol{y}) > p(\omega_c|\boldsymbol{y})$ for all classes c where $c \neq j$ [79]. Using Bayes' theorem the following decision rule is obtained

$$p(\boldsymbol{y}|\omega_j)p(\omega_j) > p(\boldsymbol{y}|\omega_c)p(\omega_c) \ . \tag{3.27}$$

For Linear Discriminant Analysis (LDA) it is assumed that the classes follow multivariate normal (MVN) distributions with equal covariances Σ. Therefore, $p(\boldsymbol{y}|\omega_c)$ is given by, [79],

$$\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}\exp\Big[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu_c})^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu_c})\Big] \tag{3.28}$$

where p is the order of the multivariate distribution. The goal is to maximize over c the discriminant function $h_c(\boldsymbol{y})$ which is obtained by substituting (3.28) into (3.27), taking the logarithm and ignoring constant terms across classes, [96],

$$h_c(\boldsymbol{y}) = \log(p(\omega_c)) - \frac{1}{2}\boldsymbol{\mu_c}^T\Sigma^{-1}\boldsymbol{\mu_c} + \boldsymbol{y}^T\Sigma^{-1}\boldsymbol{\mu_c} \ . \tag{3.29}$$

It has been shown that (3.29) is robust to deviations from the covariance equality assumption [97]. Quadratic Discriminant Analysis (QDA) on the other hand does not assume equal class covariances and is also used in the experiments of this section in order to search for non-linear discriminants. LDA and QDA are attractive classifiers because they

have closed-form solutions that can be easily computed, are inherently multi-class, have proven to work well in practice, and have no hyperparameters to tune [96, 98].

The MVN assumption was tested using the Henze-Zirkler test, [99], on each row of the matrices in (3.26). Using (3.28) with $p = 11$ (the number of statistical moments used) and 5% significance value, the p-values obtained from the test were less than 0.05 for all features, rejecting therefore the assumption. However, based on previous findings that LDA performs robustly for certain tasks even when the data is not MVN [96, 98, 100], it is employed in the following.

The parameters $\{\Sigma, \boldsymbol{\mu_c}, p(\omega_c)\}$ are calculated from the sample data for each class c over a training set. In this work, the task is normal vs abnormal classification, as stated in Section 3.1, and therefore, $c = 1, 2$. For LDA, Σ is computed as the unbiased estimate of the pooled within-class covariance matrix. New observations $\boldsymbol{y}$ are assigned to the class for which $h_c(\boldsymbol{y})$ is largest. For QDA the appropriate equation derived from (3.27) is available in many textbooks, such as in [79, 98].

Let x refer to any of the symbols $\{M, F, FF, A\}$. The aim is to evaluate the discriminant power of each $\boldsymbol{f}_i^x \; \forall \; i = 1, 2, \ldots, N$ independently, where $N = N_B$ for $x := M$, $N = K$ for $x := F, FF$ and $N = 3$ for $x := A$. The classification performance of each feature is assessed based on several metrics. Relying only on the error rate (E_r) is often not enough to deduce safe conclusions since potential classification errors other than the number of misclassified observations are not captured by the error rate. Hence, the $\mathrm{F}_{0.5}$ measure (a variation of F_1) and Matthew's Correlation Coefficient (MCC) are also used. Both are computed from the confusion matrix [101]. From a clinical perspective, the false prediction of abnormal segments as normal is worse than the contrary. $\mathrm{F}_{0.5}$ emphasises this error type more than F_1 and is thus preferred. MCC is a balanced measure ranging from -1 (prediction totally different from observation), to 1 (perfect prediction), with 0 stating no better than random prediction [102].

The knee database used in these tests is comprised of 19 normal knees and 21 abnormal and the parameter values used in the feature extraction stage are the same as in Table 3.1. Given that $\tau_s = 20$ s, it gives 249 normal and 297 abnormal acoustic signal

segments for the classification experiments. A cross-validation procedure is employed using 5 groups, randomly constructed from the database, with a normal to abnormal knees ratio of 3:5, 3:5, 3:5, 5:3, 5:3 for each group which are then made up with the segments of their constituting knees. In this way the problem of having segments of a knee signal in more than one group is avoided. Some variability in the group sizes exists given that some knee recordings are longer than others and thus have more segments. Four groups are used for training and the group left out is used for testing. This is repeated until all 5 groups are evaluated. The above procedure is executed 100 times in order to reduce the variance of the estimator and the results are averaged at the end. To aid the comparison, an average score is computed using the average metric values as

$$\mathrm{S_c} = \frac{1}{3}\left[(1 - \mathrm{E}_r) + \mathrm{F}_{0.5} + \mathrm{MCC}\right] . \tag{3.30}$$

$\mathrm{S_c}$ can vary between 0 and 1 (where $\mathrm{S_c} = 1$ indicates perfect prediction).

The results of the top 9 performing features, obtained as the maximum $\mathrm{S_c}$ from both classifiers (LDA and QDA), are reported in Table 3.2. Figure 3.8 shows the performance of all features. It was found that non-linear boundaries are more suitable, over linear, for 517 features as these generated higher $\mathrm{S_c}$ scores with QDA. LDA on the other hand, performed better for 31 features from the $\boldsymbol{\Phi_M}$ set, 13 features at frequencies $\leq$ 435 Hz and 5 features at frequencies $\geq$ 7.85 kHz from the $\boldsymbol{\Phi_F}$ set and 8 features from the $\boldsymbol{\Phi_{FF}}$ set which correspond to frequencies $\leq$ 1.34 kHz. Evidently, the most discriminant features are predominantly at low frequencies coming from the cepstral and STFT domains, as shown in Table 3.2. In general, the performance of the features from $\boldsymbol{\Phi_M}$ varies from very poor ($\mathrm{S_c} = 0.188$) to the best ($\mathrm{S_c} = 0.734$) and there is no evidence of increased discrimination in the delta and delta-delta coefficients over the static MFCC, as seen in Figure 3.8. Features from $\boldsymbol{\Phi_F}$ and $\boldsymbol{\Phi_{FF}}$ sets gave results that have similar structure in $\mathrm{S_c}$. The maximum $\mathrm{S_c}$ of 0.680 for the former set is obtained with the feature corresponding to approximately 342 Hz. For the latter set that maximum $\mathrm{S_c} = 0.574$ obtained with the feature corresponding to approximately 2.11 kHz. Pulse duration performs the best among the pulse features with $\mathrm{S_c} = 0.476$ but its performance is mediocre overall.

Feature	E_r		$F_{0.5}$		MCC		S_c
	mean	std	mean	std	mean	std	
$\boldsymbol{f}_2^M$	0.194	0.0247	0.789	0.0277	0.608	0.0495	0.734
$\boldsymbol{f}_{11}^F$	0.233	0.0265	0.741	0.0279	0.533	0.0535	0.680
$\boldsymbol{f}_{10}^F$	0.249	0.0256	0.724	0.0267	0.501	0.0522	0.659
$\boldsymbol{f}_{12}^F$	0.253	0.0287	0.721	0.0297	0.491	0.0590	0.653
$\boldsymbol{f}_{13}^F$	0.265	0.0261	0.707	0.0268	0.467	0.0541	0.637
$\boldsymbol{f}_9^F$	0.267	0.0327	0.705	0.0350	0.463	0.0679	0.633
$\boldsymbol{f}_8^F$	0.272	0.0367	0.699	0.0406	0.454	0.0766	0.627
$\boldsymbol{f}_{14}^F$	0.285	0.0274	0.686	0.0298	0.426	0.0569	0.609
$\boldsymbol{f}_7^F$	0.286	0.0300	0.688	0.0346	0.422	0.0621	0.608

Table 3.2: Average cross-validation results for the best features ranked by S_c.

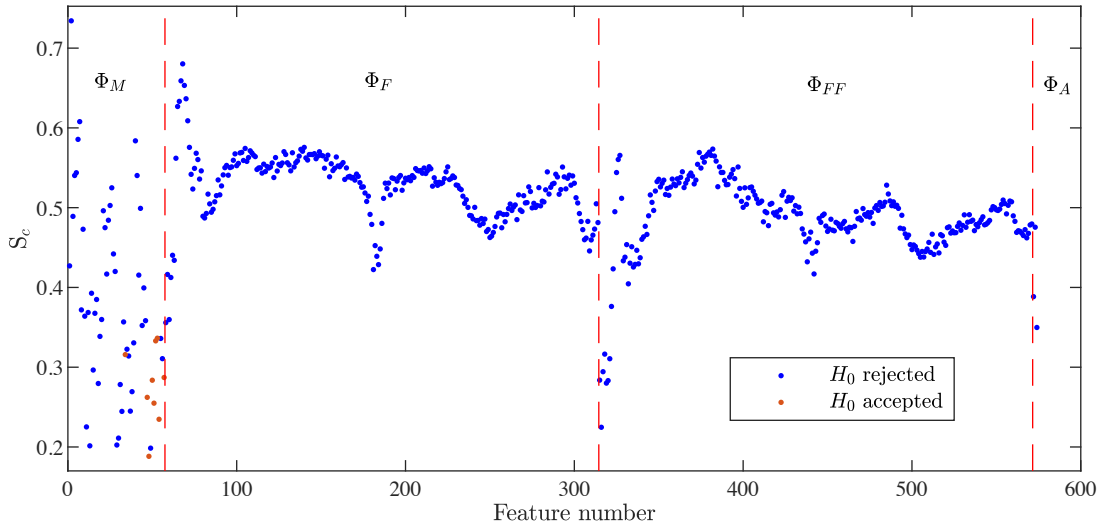

Figure 3.8: Performance of all features in **F** as measured by S_c. H_0 refers to the KST results.

Given that the MVN assumption is not valid, the approximation to (3.27) is not optimal and therefore, LDA and QDA are not the most suitable classifiers to use with this data. Nevertheless, the results in Table 3.2 and Figure 3.8 show that linear and non-linear hyperplanes exist and are captured by LDA and QDA with good performance. The results of Table 3.2 strongly suggest that designing a classifier for normal vs abnormal knee sound signal classification can be achieved with good performance even using information from a single feature represented by the proposed 11-dimensional statistical feature vector.

3.2.4 Feature selection and classification experiments

In this section, classification experiments are performed using several classifiers with different subsets of $\mathbf{F}$ that are constructed by setting thresholds on the performance metrics $\mathrm{E}_r, \mathrm{F}_{0.5}$ and MCC. The classifiers employed for these experiments are the (a) LDA, (b) Classification and Regression Tree (CART) [78], (c) Support Vector Machine (SVM) with a linear kernel, [103], denoted as SVM_l and (d) SVM with a Gaussian kernel, denoted as SVM_g, in order to look for more complex and non-linear boundaries in the feature space [103]. Classification results are evaluated using $\mathrm{S_c}$ that was defined in (3.30).

Even though QDA was found to provide good classification results in the individual feature tests of the previous section, it is not used in the experiments to follow. In this section, a feature selection method is employed in which more than 1 feature will be included in the feature subsets to be tested. The increased number of features will in turn make the number of parameters to train for QDA very large. This can cause ill-conditioned covariance when the ratio of the number of observations to the number of features is very small. The effect of this is that the covariance matrix becomes singular because the amount of data is insufficient fir fitting Gaussian models. To avoid this singularity issue the Moore-Penrose pseudo-inverse was used. In spite of this, the classification results were found to be poor. The classifier consistently overfitted the training data which resulted in very bad performance with the test sets.

The feature selection method employed is a hybrid of a filter and a wrapper approach. First, the features are ranked in each of the 3 metric categories (the filter step) and with the application of thresholds, $\left[\theta_e, \theta_{0.5}, \theta_{mcc}\right]$ for $\left[\mathrm{E}_r, \mathrm{F}_{0.5}, \mathrm{MCC}\right]$ respectively, the best N features are selected, where N is dictated by those that satisfy all thresholds. Secondly, by allowing θ_e, $\theta_{0.5}$ and θ_{mcc} to vary in the range $\left[0, 1\right]$ with discrete steps of size w, the entire feature space is searched and all possible subsets are constructed subject to these constraints (the wrapper step). The three metrics are bounded in a continuous range and therefore defining a discrete set of constraints is necessary in order to make the search space tractable as it is not practical to test all possible combinations of features as

an exhaustive search. This feature selection method forms nested subset of features,

$$
\begin{aligned}
S_1^x &= \{\boldsymbol{f}_q^x, \ldots, \boldsymbol{f}_{q+N_1}^x\} \textbf{ s.t. } \theta_e^1 \geq \{J(\boldsymbol{f}_q^x), \ldots, J(\boldsymbol{f}_{q+N_1}^x)\} \geq \{\theta_{0.5}^1, \theta_{mcc}^1\} \\
S_2^x &= \{\boldsymbol{f}_q^x, \ldots, \boldsymbol{f}_{q+N_2}^x\} \textbf{ s.t. } \theta_e^2 \geq \{J(\boldsymbol{f}_q^x), \ldots, J(\boldsymbol{f}_{q+N_2}^x)\} \geq \{\theta_{0.5}^2, \theta_{mcc}^2\} \\
&\vdots \\
S_r^x &= \{\boldsymbol{f}_q^x, \ldots, \boldsymbol{f}_{q+N_r}^x\} \textbf{ s.t. } \theta_e^r \geq \{J(\boldsymbol{f}_q^x), \ldots, J(\boldsymbol{f}_{q+N_r}^x)\} \geq \{\theta_{0.5}^r, \theta_{mcc}^r\}
\end{aligned}
$$

where

$$
\theta_e^1 \geq \theta_e^2 \geq \cdots \geq \theta_e^r, \quad \theta_{0.5}^1 \leq \theta_{0.5}^2 \leq \cdots \leq \theta_{0.5}^r, \quad \theta_{mcc}^1 \leq \theta_{mcc}^2 \leq \cdots \leq \theta_{mcc}^r
$$

with $N_1 \leq N_2 \leq \cdots \leq N_r$, $J(.)$ is any of $\left[\mathrm{E}_r, \mathrm{F}_{0.5}, \mathrm{MCC}\right]$, evaluated against the corresponding threshold and q is an index that takes integer values in the range 1 to N where $N = N_B + 2K + 3$. Each feature subset is used for training and testing the classifier by employing the cross-validation procedure with 5 groups as described earlier (Section 3.2.3). Prior to this, the training data (4 groups) is scaled by subtracting the mean and normalising the variance. The same normalization values are applied to the test data (the group left out). This data standardisation will not have any effect on LDA but is necessary for SVM and CART in order to produce meaningful results [96]. The classification performance is evaluated using $\mathrm{S_c}$ and the subset that gives the highest $\mathrm{S_c}$ is chosen.

The parameter values used in the experiment are again those depicted in Table 3.1 and therefore, $N = 574$. Additionally, the error rate threshold θ_e was fixed at 0.456 which is the error rate obtained when the predicted class is always the largest $(1 - \frac{297}{546})$. This is the error rate attributed to random guessing and hence anything worse than this would mean that the classifier performs very poorly. By keeping θ_e constant, the values of $\theta_{0.5}$ and θ_{mcc} are varied in the range $[0, 1]$ with step size $w = 0.05$ and the possible feature subsets are constructed and subsequently fed to the classifier.

Table 3.3 presents the best results (based on $\mathrm{S_c}$) obtained with each classifier and Table 3.4 shows the threshold values along with the selected features that were used to generate the results in the former table. Comparing to the results of Table 3.2 and

Classifier	E_r		$F_{0.5}$		MCC		S_c
	Mean	Std	Mean	Std	Mean	Std	
LDA	0.157	0.0215	0.817	0.0227	0.688	0.0435	0.783
SVM_l	0.139	0.0189	0.841	0.0213	0.722	0.0378	0.808
SVM_g	0.247	0.0309	0.719	0.0316	0.515	0.0607	0.662
CART	0.178	0.0244	0.796	0.0279	0.646	0.0478	0.755

Table 3.3: Best cross-validation performance per classifier.

Classifier	$\theta_{0.5}$	θ_{mcc}	Selected features
LDA	0.70	0.10	$\{\boldsymbol{f}_2^M, \boldsymbol{f}_9^F, \ldots, \boldsymbol{f}_{13}^F\}$
SVM_l	0.70	0.45	$\{\boldsymbol{f}_2^M, \boldsymbol{f}_9^F, \ldots, \boldsymbol{f}_{13}^F\}$
SVM_g	0.65	0.55	$\boldsymbol{f}_2^M$
CART	0.75	0.25	$\boldsymbol{f}_2^M$

Table 3.4: Metric thresholds used to select the feature subset that gave the results in Table 3.3. The features that comprise the subset which produced the results, as indicated, are also shown.

Figure 3.8, the classification performance is improved except for the case of SVM_g where $S_c = 0.662$ was obtained. On average, SVM_l performs better and with less variability, achieving $S_c = 0.808$ using $\boldsymbol{f}_2^M$ and $\boldsymbol{f}_9^F, \ldots, \boldsymbol{f}_{13}^F$ which correspond to frequencies 280 Hz to 405 Hz. The class of each signal segment $s_{i,j}(n)$ (normal and abnormal) as predicted by each classifier in Table 3.3 can be used to predict the condition of the knee that emitted these signals. This is done by majority vote, that is, the knee joint condition class is either normal or abnormal depending on which class was predicted most frequently for the knee's $s_{i,j}(n)$. Performing this on all 40 knees of the database generates $S_c = 0.777$ with average $\big[E_r, F_{0.5}, \text{MCC}\big]$ equal to $\big[0.175, 0.849, 0.658\big]$, obtained by the linear kernel SVM.

It is observed that the best results per classifier were achieved with only a few features from the total of 574 and in two cases with only $\boldsymbol{f}_2^M$. Even CART, which uses an inherent feature selection method tied to the classifier model, performs best with a single feature. Given that one feature adds 11 variables, this outcome is attributed to the small dataset since by adding more features the feature space quickly becomes sparse. With too many variables the classifier is likely to overfit the training data and fail to generalise to new data resulting in poor performance as evaluated using the cross-validation procedure.

3.2.5 Synopsis

Section 3.2 addressed the discriminant analysis of various features for the task of classifying the sound signals emitted from knee joints during walking into normal (clinically healthy) and abnormal (clinically OA). The features explored are the magnitude of the STFT ($\boldsymbol{\Psi_F}$), MFCC ($\boldsymbol{C_M}$), modulation magnitude ($\boldsymbol{\Psi_{FF}}$) and AE pulses from which the energy, duration and peak-to-peak amplitude were extracted ($\boldsymbol{A}$). The AE pulses and the use of MFCC and modulation magnitude spectrum are novel in the context of knee joint condition classification using the sound signal sensed at the patella. The feature distributions were statistically evaluated using the two-sample Kolmogorov-Smirnov test and the Bhattacharyya distance which was extended for univariate gamma distributions in (3.22). The results of these experiments, shown in Figure 3.6, as well as the observations derived (Figures 3.5 and 3.7), motivated the use of the 11 statistical parameters to represent the feature distributions. Subsequent classification experiments were conducted on a database of 249 normal and 297 abnormal acoustic signals obtained from 40 different knees and using the proposed statistical parameterisation of the features ($\boldsymbol{\Phi_M}, \boldsymbol{\Phi_F}, \boldsymbol{\Phi_{FF}}, \boldsymbol{\Phi_A}$). A number of off-the-shelf classifiers were compared and the results showed that SVM with linear kernel achieves the best performance giving $\left[\mathrm{E}_r, \mathrm{F}_{0.5}, \mathrm{MCC}\right] = \left[0.139, 0.841, 0.722\right]$. From the analysis of the results it was found that the most discriminant features are the low order MFCC ($\boldsymbol{f}_2^M$ and $\boldsymbol{f}_7^M$; where the 0th Cepstral Coefficient (CC) was not considered as a feature) and STFT features ($\boldsymbol{f}_7^F, \ldots, \boldsymbol{f}_{14}^F$) that correspond to the frequencies in the range of 250 Hz to 435 Hz. These 16 features individually achieve the highest $\mathrm{S_c}$ scores (0.608 to 0.734) from all the features considered. In summary, the results of the experiments carried out in this section have validated the hypothesis stated at the beginning of the Chapter: Feature representations of the sound signals emitted from abnormal (OA) knees are statistically different from the corresponding representations of the sound signals emitted from normal (healthy) knees (Figure 3.6). Moreover, the results have also shown that there is structure (information) in the knee sound signals which can be captured by $\{\boldsymbol{\Phi_M}, \boldsymbol{\Phi_F}, \boldsymbol{\Phi_{FF}}, \boldsymbol{\Phi_A}\}$ and enables the detection of OA in the knee with high classification performance (Tables 3.2 and 3.3).

In addition to the results above, it is also interesting to mention the following observations regarding the AE pulses, extracted as described in Section 3.2.1 II, even though these observations are not immediately relevant to the analysis of this or the following section. Using the same database as before (249 normal and 297 abnormal segments) it was found that the number of AE pulses observed in a signal segment obtained from a knee with OA are 529, on average, which is 50.7% more than the 351 pulses observed, on average, in a same duration segment (20 s) obtained from a healthy knee. Adding to this, the average energy of the pulses in OA knee sound signals was found to be approximately 142% larger than the corresponding pulses from healthy knees. This is in line with the observations derived from the work on ultrasonic AE detection from normal and OA knees where the authors defined a hit as a short burst of energy and found that OA knees produce significantly more hits than normal knees [51]. The sounds in that study however, were captured using a piezoelectric sensor that is sensitive in frequencies over 50 kHz and while the knee was performing flexion-extension. For these reasons and given that no load is active on the knee during flexion-extension movements, the observations of the aforementioned study are not immediately applicable on the sound signals of this thesis work. Therefore, the observations described in this paragraph are considered to be novel. Further work on AE pulse events will be presented in Chapter 5.

3.3 Time-frequency analysis and effective parameterisation of knee sounds

The work presented in this section builds on the results of the previous section where it was shown that spectrum and cepstrum based features carry significantly discriminant information for the task of normal vs abnormal classification (see Table 3.2). These features were represented by the sets $\boldsymbol{\Phi_F}$ and $\boldsymbol{\Phi_M}$ respectively, using the proposed parameterisation with the 11 statistical parameters. Section 3.3 investigates further the discriminant properties of these feature sets, explores other time-frequency representations of the knee sound signals $s_{i,j}(n)$ and further examines the classification effectiveness of the proposed parameterisation. For these, the comprehensive experimental framework introduced in the

previous section is redefined in Section 3.3.1 and employed in the experiments presented in the following sections. Based on this framework, the impact of the feature extraction parameters on the classification performance is investigated using Classification and Regression Tree (CART), Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) classifiers but the classifiers' optimisation is beyond the scope of this Chapter. The main aim of this work is to answer the following four questions:

(a) Does the classification performance improve when the DFT spectrum is compressed using triangular filter-banks?

(b) Does the classification performance improve when the natural logarithm and DCT are also used instead of just DFT?

(c) Is the classification performance better when using uniform or non-uniform frequency spacing for the triangular filters in (a)?

(d) Which frequency ranges of knee sound signals contain more discriminative information and hence are important for OA classification?

In answering these questions an insight will be obtained into which features best characterise healthy and OA knees.

3.3.1 Methodology

The experimental framework developed in the previous section is elaborated further here and is utilised for the experiments to follow. Figure 3.9 shows a schematic of this framework and outlines the whole process of feature extraction, analysis, selection and classification. In summary, the signal $s_i(n)$ of the i^{th} knee is RMS level-normalised and is then divided into non-overlapping segments, $s_{i,j}(n)$, of length τ_s. Each segment is treated as a separate observation for the classification task and is further divided into frames from which certain features are extracted. Each of these features is then used to classify the signal segments into normal and abnormal and the performance is measured by $\mathrm{E}_r, \mathrm{F}_{0.5}$ and MCC, as previously discussed. The results are then used to select subsets of features for subsequent classification which is used to measure the final performance of the feature set.

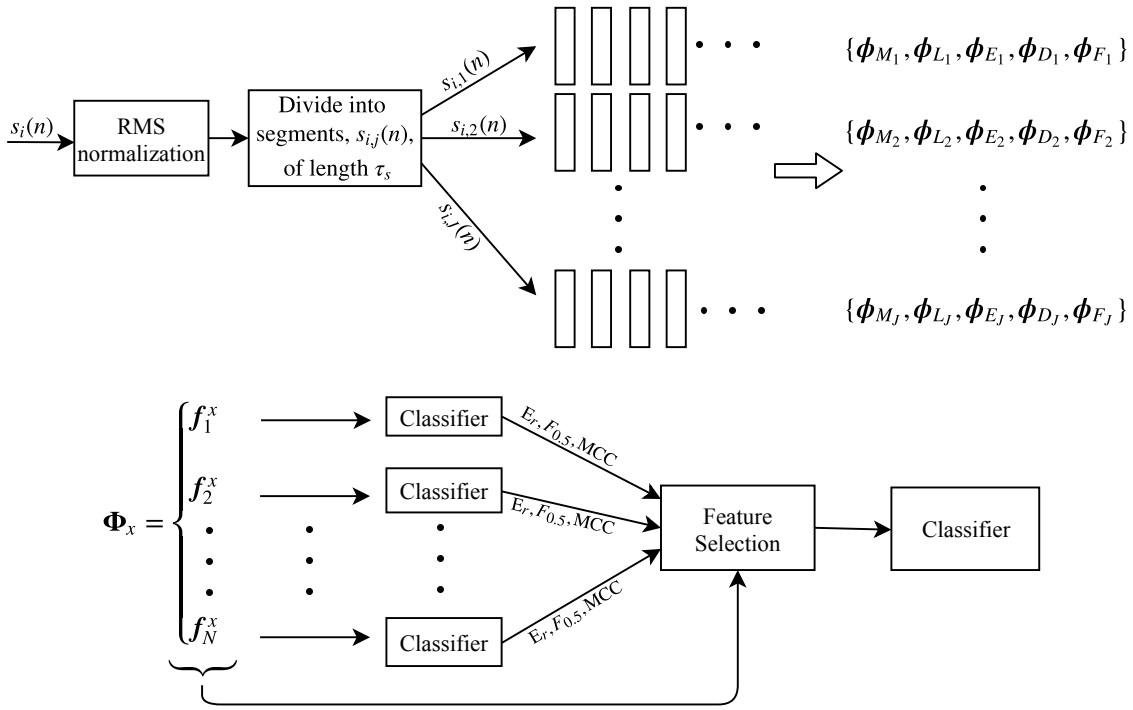

Figure 3.9: Top: Feature extraction process for a single $s_i(n)$ that is divided into J segments. **Bottom:** Feature analysis and subset selection process with subsequent classification presented for feature set $\mathbf{\Phi}_x$ where x denotes any of the symbols $\{D, E, F, L, M\}$ and $N = K$ for $x := F$ or $N = N_B$ otherwise.

The feature extraction, analysis and selection steps indicated in the schematic of Figure 3.9 are described in the following Section. The series of experiments in Section 3.3.3 examine the effectiveness of the feature parameterisation with the 11 statistics and study their discriminant power for the task of normal vs abnormal signal segment classification.

3.3.2 Feature extraction, analysis and selection

Extraction

Section 3.2.1 introduced the matrix representation, $\boldsymbol{S}$, of the j^{th} signal segment of the i^{th} knee, $s_{i,j}(n)$, that was obtained by dividing $s_{i,j}(n)$ into frames of length l ms with 50% overlap. The number of rows and columns of $\boldsymbol{S}$ are N_f and l_n respectively, where N_f is the number of frames and l_n is the frame length in samples. The DFT of $\boldsymbol{S}$ was then computed in (3.1) as $\mathbf{\Psi_f}$ from which the feature matrix $\mathbf{\Psi_F}$ was obtained by taking the

magnitude of each element in $\boldsymbol{\Psi_f}$ and retaining only the first $K = \lfloor 1 + l_n/2 \rfloor$ columns.

Similarly to (3.3), a filter-bank with N_B triangular band-pass filters linearly spaced in frequency is used to construct the matrix

$$\boldsymbol{U}_L = \begin{bmatrix} U_1(0) & U_2(0) & \dots & U_{N_B}(0) \\ U_1(\frac{2\pi}{K}) & U_2(\frac{2\pi}{K}) & \dots & U_{N_B}(\frac{2\pi}{K}) \\ \vdots & \vdots & \ddots & \vdots \\ U_1(\frac{2\pi(K-1)}{K}) & U_2(\frac{2\pi(K-1)}{K}) & \dots & U_{N_B}(\frac{2\pi(K-1)}{K}) \end{bmatrix}$$

where each element is the magnitude of the bandwidth of a single filter at a single frequency bin $k = 0, 1, \dots, K-1$. A compact spectrum representation is then computed as

$$\boldsymbol{\Psi_E} = \boldsymbol{\Psi_F} \boldsymbol{U}_L \ . \tag{3.31}$$

The matrices $\boldsymbol{\Psi_E}$ and $\boldsymbol{\Psi_D}$, from (3.4), are $N_f \times N_B$ and in this way the dimensionality reduction is achieved. In this section the discriminant capabilities of $\boldsymbol{\Psi_E}$ and $\boldsymbol{\Psi_D}$ will be explored and therefore, as was done in (3.25) for $\boldsymbol{\Psi_F}$, their statistical parameterisation is obtained as

$$\begin{aligned} \boldsymbol{\phi_E} &= \mathcal{D}(\boldsymbol{\Psi_E}) = [\boldsymbol{f}_1^E, \boldsymbol{f}_2^E, \dots, \boldsymbol{f}_{N_B}^E]^T \\ \boldsymbol{\phi_D} &= \mathcal{D}(\boldsymbol{\Psi_D}) = [\boldsymbol{f}_1^D, \boldsymbol{f}_2^D, \dots, \boldsymbol{f}_{N_B}^D]^T \ . \end{aligned} \tag{3.32}$$

A set of Linear Frequency Cepstral Coefficients (LFCCs) can be computed using $\boldsymbol{\Psi_E}$ as

$$\boldsymbol{C_L} = \mathcal{F}(\log(\boldsymbol{\Psi_E})) \tag{3.33}$$

and the corresponding statistically parameterised matrix is therefore given by

$$\boldsymbol{\phi_L} = \mathcal{D}(\boldsymbol{C_L}) = [\boldsymbol{f}_1^L, \boldsymbol{f}_2^L, \dots, \boldsymbol{f}_{N_B}^L]^T \ . \tag{3.34}$$

In the extraction of $\boldsymbol{\Psi_E}$ and hence of $\boldsymbol{C_L}$, triangular shaped band-pass filters were used as described earlier. The choice for the shape was made based on the fact that it is

normally used in the extraction of MFCC. The filter shape was chosen to be the same in all of the relevant feature sets so that it gets excluded from the reasons of any potential differences in the classification performance between the sets.

Given the fixed frame segmentation process employed using short time frames, it is likely that a knee sound related to OA might extend to more than one frame. By taking the time derivatives of the coefficients, the aim is to extract the information present in the evolution of these sounds across multiple frames. Therefore, for each element in $\boldsymbol{\Psi_D}$, $\boldsymbol{\Psi_E}$ and $\boldsymbol{C_L}$ the trajectories in time are computed as the first and second derivatives using (3.6), in the same way that they were computed for $\boldsymbol{C_M}$ in Section 3.2.1.

All the values of the first derivative obtained from a single feature in either $\boldsymbol{C_M}, \boldsymbol{C_L}, \boldsymbol{\Psi_E}$ and $\boldsymbol{\Psi_D}$, (i.e. cepstral coefficient or frequency band), can be considered to form a distribution from which the 11-dimensional vector $\boldsymbol{f}$ is extracted. The same is performed for all the features in these 4 matrices for their first and second derivatives and the new vectors are appended to the appropriate $\boldsymbol{\phi}$ feature matrix. Performing the above process for all signal segments produces 5 feature sets that will be used for the classification experiments. These are

$$\begin{aligned}
\boldsymbol{\Phi_E} &= [\boldsymbol{\phi_{E}}_1, \boldsymbol{\phi_{E}}_2, \ldots, \boldsymbol{\phi_{E}}_C] \\
\boldsymbol{\Phi_D} &= [\boldsymbol{\phi_{D}}_1, \boldsymbol{\phi_{D}}_2, \ldots, \boldsymbol{\phi_{D}}_C] \\
\boldsymbol{\Phi_F} &= [\boldsymbol{\phi_{F}}_1, \boldsymbol{\phi_{F}}_2, \ldots, \boldsymbol{\phi_{F}}_C] \\
\boldsymbol{\Phi_L} &= [\boldsymbol{\phi_{L}}_1, \boldsymbol{\phi_{L}}_2, \ldots, \boldsymbol{\phi_{L}}_C] \\
\boldsymbol{\Phi_M} &= [\boldsymbol{\phi_{M}}_1, \boldsymbol{\phi_{M}}_2, \ldots, \boldsymbol{\phi_{M}}_C]
\end{aligned} \tag{3.35}$$

for $C = \sum_{i=1}^{I} J_i$ total segments. The matrices $\boldsymbol{\Phi_F}$ and $\boldsymbol{\Phi_M}$ above are the same as in (3.26) but are redefined here for completeness.

Analysis

Let x refer to any of the symbols $\{D, E, F, L, M\}$. The aim is to evaluate the discriminant power of each $\boldsymbol{f}_i^x \ \forall \ i = 1, \ldots, N$ in (3.35) independently, where $N = K$ for $x := F$ or

$N = N_B$ otherwise. The classifier employed for this purpose is SVM because it is efficient for small training data and avoids making any assumptions on the underlying data distribution [103]. This makes it a suitable choice since the distribution is unknown and was shown in Section 3.2.3 that it does not follow a MVN distribution as was the assumption for LDA. The linear kernel SVM is used, SVM_l, as it is the simplest form of kernel, it is less prone to overfitting than other more complex kernels, [103], and it was shown to achieve the highest classification results in the experiments of Section 3.2.4. For a comprehensive analysis of SVM classifiers the reader may consult the following excellent reference [103] but also the original paper by Cortes and Vapnik [104]. Furthermore, a brief technical review of SVM classifier using kernels is presented in Section 5.2.

The knee database used in the experiments of this section is the same as in Section 3.2, that is, 19 normal knees and 21 abnormal from which the signal segments $s_{i,j}(n)$ are obtained using a window size of τ_s seconds. Moreover, the same cross-validation procedure is employed in which the database is divided into 5 randomly constructed groups and in the same ratio of normal to abnormal knees as before, that is, 3:5, 3:5, 3:5, 5:3, 5:3. Four groups are used for training the SVM model which is then tested on the group left out. This is repeated until all 5 groups are evaluated. Prior to this, the training data is scaled by subtracting the mean and normalising by the variance. The same normalization values are then applied to the test set. The above procedure is executed 100 times in order to reduce the variance of the estimator and the results are averaged at the end. The performance of each feature is assessed based on $\text{E}_r, \text{F}_{0.5}$ and MCC.

Selection

Following the pre-processing with feature extraction and analysis steps, the selection of feature subsets for subsequent classification and analysis is performed. The selection method used is described in Section 3.2.4. Each subset is subsequently used for training and testing the SVM classifier by employing the cross-validation procedure described earlier. The subset's classification performance is in this case evaluated using the Area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve, [105], instead of

using S_c as in the previous section's experiments. AUC applies to binary classifiers that use decision thresholds, like for example SVM and logistic regression [79]. Unlike accuracy, AUC is not testing the quality of a particular choice of the classifier's decision threshold. It is instead an evaluation of the classifier as the threshold varies over all possible values, measuring the trade off between true positive rate and false positive rate [105]. Therefore, it is suitable for comparing the performance obtained with different feature subsets and classifiers. The subset that gives the highest AUC is chosen. Experiments and results are discussed in the next section.

3.3.3 Effect of feature extraction parameters on classification

The experimental framework is based on a systematic approach that aims to (a) find the best frame length l for extracting the 5 alternative feature sets in (3.35), (b) examine the effect of the number of filters N_B on the classification performance when using either one of $\boldsymbol{\Phi_D}$, $\boldsymbol{\Phi_E}$, $\boldsymbol{\Phi_L}$ or $\boldsymbol{\Phi_M}$ and (c) obtain insights into the time-frequency information of normal and abnormal signals and their differences.

An investigation of the effect of frame length values based on a deterministic approach was initially conducted by defining a suitable range and quantifying the classifier performance in order to choose the best l. Subsequent Monte Carlo simulations were performed to test the suitability of the choice and to identify performance trends of the feature sets in a larger range. Finally, experiments were conducted varying the number of filters N_B.

In all the experiments that follow, the signal segment length τ_s was set to 20 s. Other time periods that do not violate the assumption outlined at the beginning of Section 3.2.1 could also be used but would affect the total number of segments obtained. Additionally, the error rate threshold θ_e was fixed at 0.456 as described in Section 3.2.4. By keeping θ_e constant, the values of $\theta_{0.5}$ and θ_{mcc} are varied in the range $[0, 1]$ with step size $w = 0.05$ and the possible feature subsets are constructed and subsequently fed to the classifier.

Data used in the following experiments originates from 40 knees, of which 19 are normal (from 15 patients) and 21 are abnormal (from 18 patients). Following the segmen-

tation process described in Section 3.2.1, 249 normal and 297 abnormal segments of 20 s are obtained that is, 83 minutes of sound data from healthy knees and 99 minutes from OA knees is used.

I. Deterministic search in a specified range of frame lengths

The frame length l is tested for the values $\left[20, 24, 28, \ldots, 100\right]$ ms. The limiting values were chosen so that 20 ms is a short enough window to allow good time resolution in the time-frequency representations for the sounds (clicks, pops, grindings) heard during walking and 100 ms is a large enough window to capture the two major events in a single stride, namely, the heel strike and the push off responses as captured by the patella microphone. This information was extracted from the ground reaction force signals obtained from the treadmill's force plates from which can be extracted the timings of each event in the gait cycle. More details about this process are described in Section 4.5.2.

The experimental framework developed in the previous Sections is applied and for each frame length l the features are extracted, analysed individually and all the possible subsets are constructed for evaluation using the SVM_l classifier. In all cases $N_B = 20$ is chosen, giving 60 coefficients (including the 0^{th} cepstral coefficient) for all the feature sets except $\boldsymbol{\Phi_F}$. For $\boldsymbol{\Phi_F}$, the size depends on l since the DFT length used is equal to the frame length, as shown in Section 3.2.1. For the analysis and interpretation of the results, a top-down approach is followed. The overall results are summarised in Figure 3.10. Each point on a line represents the highest AUC, averaged over 100 trials, obtained by any subset of the corresponding frame-size and of the particular feature set.

It is evident from Figure 3.10 that classifying using features obtained from the $\boldsymbol{\Phi_M}$ set scores consistently higher than any other set, for any l, with its best performance occurring with $l = 48$ ms (AUC $= 0.915$). On the other hand, $\boldsymbol{\Phi_F}$, $\boldsymbol{\Phi_L}$, $\boldsymbol{\Phi_E}$ and $\boldsymbol{\Phi_D}$ give highest AUC at 24 ms, 20 ms, 88 ms and 20 ms respectively. Initial observations suggest that the optimal frame-size is different for each set and that reducing the dimensionality of the spectrum with triangular shaped filters generates comparable results with the full spectrum features and in some cases ($l \geq 72$ ms) improves the classification performance.

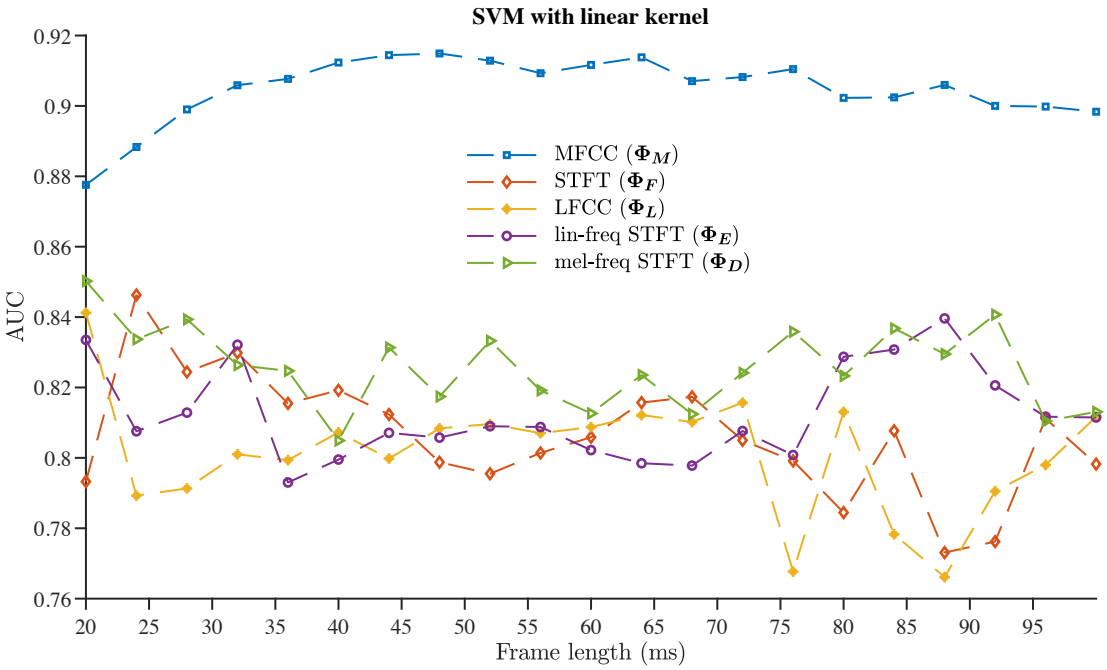

Figure 3.10: AUC against frame length for SVM (linear kernel). The points are connected with dashed lines to aid the visualisation.

Comparison of MFCC and LFCC results

It is seen in Figure 3.10 that AUC values range from 0.878 to 0.915 and from 0.77 to 0.84 for $\mathbf{\Phi_M}$ and $\mathbf{\Phi_L}$ respectively. Hence, the performance for $\mathbf{\Phi_M}$ is only weakly sensitive to the choice of frame-size and in fact, above 32 ms the variance of the metric value drops approximately to 1/3 of the variance obtained when including the AUC for $l < 32$ ms. The fluctuation is small because, for any l, the final selected subset that gave the highest AUC consisted of only $\boldsymbol{f}_3^M$ which contains the statistical parameters of the distribution of the 2$^{\text{nd}}$ MFCC. All other subsets considered from this set resulted in worse performances. For $l \geq 88$ ms and $l \leq 32$ ms, the AUC of $\mathbf{\Phi_M}$ follows a downward trend suggesting that, even to a small degree, the performance drops for frame-sizes outside of this range. This hypothesis is tested in Section III to follow. Higher variability is observed in the results of $\mathbf{\Phi_L}$ where more than one feature vector is chosen in the final subset for 10 out of 21 frame lengths. The highest AUC for this set occurs at 20 ms which is obtained with a subset of 17 feature vectors. The classification rate for this case, averaged over the 100 cross-validation iterations, is 75.07% compared to 85.25% obtained from the $\mathbf{\Phi_M}$ feature

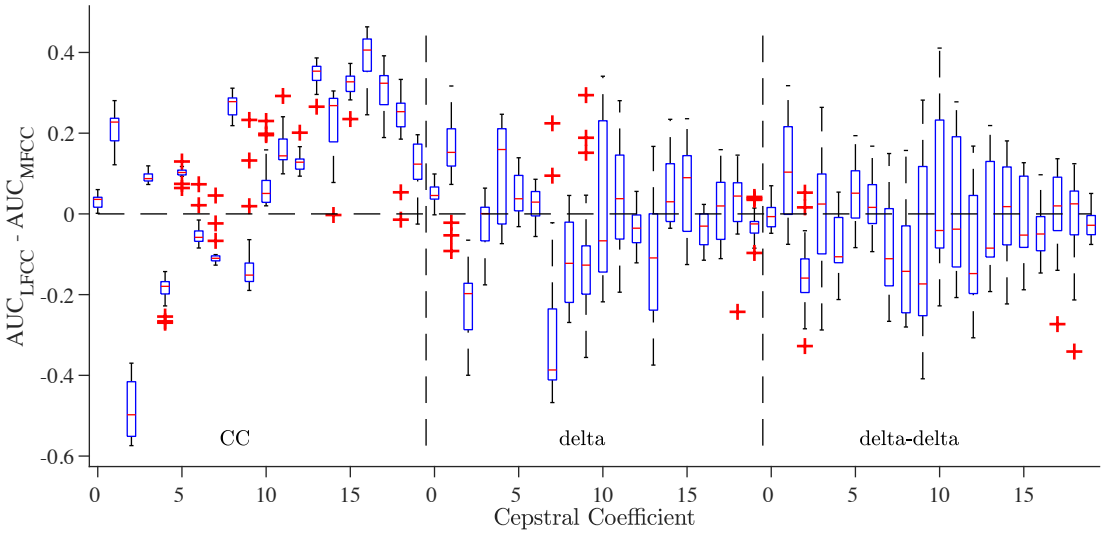

Figure 3.11: Difference in AUC values of LFCC and MFCC for all frame lengths.

set with $l = 48$ ms. This translates into classifying correctly, 55 more segments (from a total of 546 segments) when using the $\mathbf{\Phi_M}$ set.

Figure 3.11 shows a box plot of the differences in AUC for all frame lengths per Cepstral Coefficient (CC). It is evident that many CC obtained with a linear-frequency filter-bank generated better classification results than their mel-frequency filter-bank counterparts. More precisely, the feature vectors corresponding to high order (10 to 19) as well as low order coefficients $(0, 1, 3, 5, 8)$ consistently give higher AUC values in any of the frame length cases tested. Exceptions occur in each of the 14^{th} and 18^{th} coefficient at 20 ms and one in the 19^{th} at 28 ms but can be considered as outliers because of their very small value. The feature vector representing the 2^{nd} MFCC, which was found to be the most important for the $\mathbf{\Phi_M}$ set, clearly outperforms the 2^{nd} LFCC with a difference that reaches -0.57 in the AUC result (absolute median value is 0.5). The same holds for the 4^{th} and 9^{th} CC but with a smaller absolute median value of 0.18 and 0.15 respectively. For the trajectory coefficients however, there is no apparent advantage of one feature set over the other that persists with changing frame length, except in very few cases (at the 2^{nd} and 7^{th} delta coefficients).

Figure 3.12 compares the static coefficients with their time derivatives for MFCC in plot (a) and LFCC in plot (b). The top performing coefficient is found, per feature

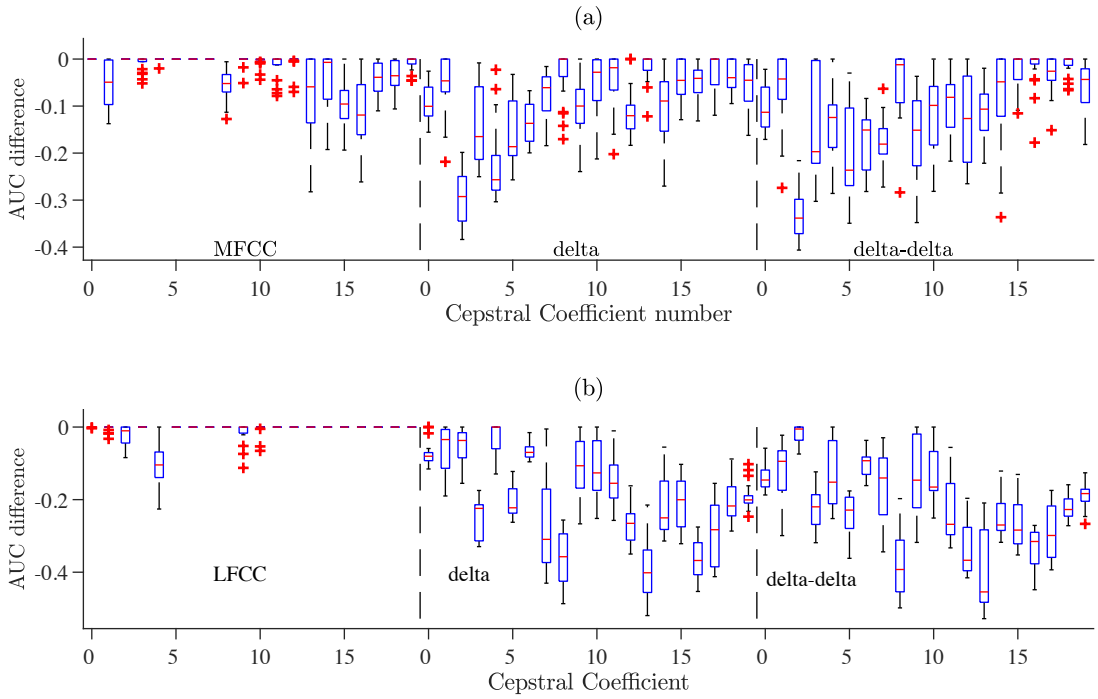

Figure 3.12: Comparison of AUC output for the static, delta and delta-delta cepstral coefficients for MFCC (a) and LFCC (b), in all frame lengths.

set and for each index (0 to 19), by comparing the AUC scores of the static coefficient and its delta and delta-delta. Each box in the figure consists of the values obtained, for every frame length, by subtracting the AUC of the coefficient indicated by the index on the horizontal axis from the top performing one found as before. The MFCC static coefficients $0, 2$ to 7 and 9 to 12 exhibit better classification performance in over 65% of the frame lengths with the 0^{th}, 2^{nd} and 5^{th} to 7^{th} scoring consistently higher than their trajectory coefficients in any frame length. This is different for the 1^{st}, 8^{th} and 13^{th} to 18^{th} static CCs where they score lower than either the delta or delta-delta coefficients in over 50% of the frame lengths. The feature vector representing the 19^{th} MFCC scores similarly with the vectors corresponding to the coefficient's derivatives but its performance degrades for $l > 84$ ms. For the $\boldsymbol{\Phi_L}$ feature set, the overwhelming majority of the static CCs score consistently higher than their delta and delta-delta counterparts. More specifically, the feature vectors $\boldsymbol{f}_i^L$ for $i = 4, \ldots, 20, i \neq 5, 10, 11$ score higher for any l.

Overall, the results show that the higher order LFCC outperform the corresponding MFCC (Figure 3.11) but Figure 3.10 suggests that, from a classification point of view, it

is better to use mel-scaled instead of linear-scaled filters to extract the CC. However, to justify this, further tests are needed in order to investigate the effect on the performance, the number of filters in the filter-bank has. This is explored in Section IV to follow.

Comparison of STFT and compressed STFT feature sets

All three feature sets, $\boldsymbol{\Phi_F}$, $\boldsymbol{\Phi_E}$ and $\boldsymbol{\Phi_D}$, achieve comparable maximum performances in AUC, with values 0.846 ($l = 24$ ms), 0.839 ($l = 88$ ms) and 0.850 ($l = 20$ ms) respectively (Figure 3.10). At $l = 24$ ms the subset selected from $\boldsymbol{\Phi_F}$ that gave the highest AUC, only consists of features that fall within the range 0.29 to 2 kHz. Comparing the performance of the individual feature vectors (i.e. the 11-dimensional elements of $\boldsymbol{\phi_E}$, $\boldsymbol{\phi_D}$ and $\boldsymbol{\phi_F}$) within this range, shows that the feature vectors in $\boldsymbol{\phi_F}$ perform better than the corresponding from $\boldsymbol{\phi_E}$ and $\boldsymbol{\phi_D}$. This is depicted in Figure 3.13 that shows the average AUC scores per feature vector, against frequency. At approximately 3.5 kHz however, the performance starts to drop rapidly, becoming comparable to, or worse than, the corresponding features from the $\boldsymbol{\Phi_E}$ and $\boldsymbol{\Phi_D}$ sets. At around 6 kHz the performance starts to improve again, reaching a maximum near 7.2 kHz at a value of 0.75, comparable to that of the lower frequency features ($\leq$ 3.5 kHz). Above 4.5 kHz it is observed that the classification performance actually improves when using the static coefficients or their second derivatives from the $\boldsymbol{\Phi_E}$ and $\boldsymbol{\Phi_D}$ sets. The line plots in Figure 3.13 are seen to follow similar trends.

Adding to the above, the static coefficients from both $\boldsymbol{\Phi_E}$ and $\boldsymbol{\Phi_D}$ perform better than their corresponding derivatives for frequencies up to 2 kHz. The delta coefficients peak in the range 2.2 to 2.6 kHz but then begin to drop rapidly whereas for higher frequencies ($\geq$ 5 kHz) the delta-delta coefficients perform better overall. This shows that additional information exists in the dynamics of the spectrum for mid to high frequencies. The above observations can also be derived from the results at other, higher values of l. From the plots of Figure 3.14 it becomes apparent that, with increasing frame length, the delta coefficients capture more information than their static and delta-delta in the frequency range 1.6 to 3 kHz.

With increasing frame length, the frequency resolution improves and it becomes

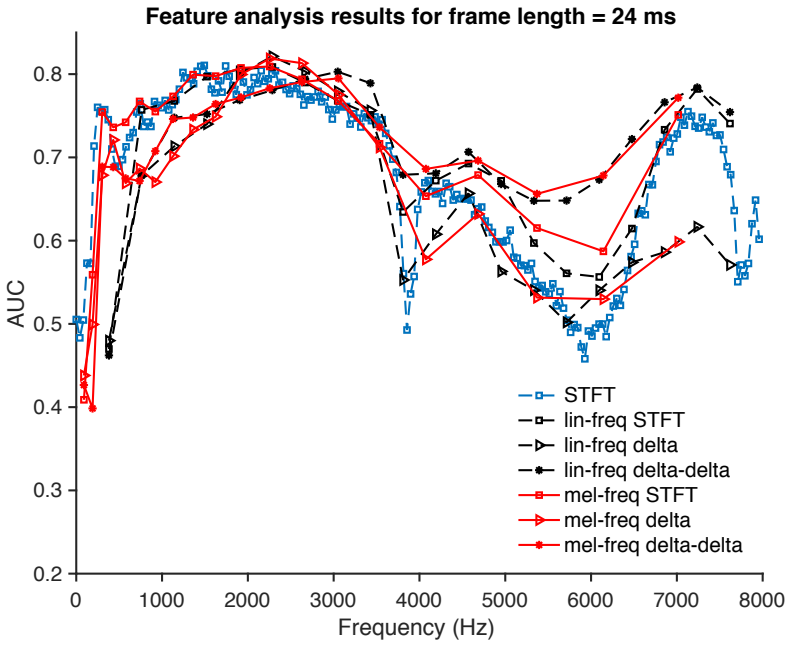

Figure 3.13: Individual performance of STFT based feature vectors at $l = 24$ ms with $N_B = 20$.

immediately apparent from Figure 3.14 that two regions of high performing features exist, one in the range $[220, 420]$ Hz, denoted as F_{r_1}, and another in the range $[1, 3.4]$ kHz, denoted as F_{r_2}. These frequency regions show a collection of features that individually score higher than the rest in their respective set (e.g. ≥ 0.726 AUC in the $\mathbf{\Phi_F}$ set). It is also worth noticing that for $l \leq 36$ ms there exists another frequency band ($6.6 - 7.6$ kHz), where some feature vectors that fall within the band, from all three sets ($\mathbf{\Phi_F}$, $\mathbf{\Phi_E}$, $\mathbf{\Phi_D}$), also scored high AUC (≥ 0.75). However, in the three examples of Figure 3.14 it is seen that their performance gradually drops with increasing frame length to at, or below, AUC $= 0.5$. This means that the classifier randomly assigns observations to classes (AUC $= 0.5$), given the input features, or the classifier failed to apply the information at hand correctly (AUC < 0.5). For the second case one may reverse the classifier's decisions and obtain a ROC curve that would give AUC > 0.5, as long as the classifier consistently produces results falling in the lower right part of the ROC space as described in [106].

Feature vectors from $\mathbf{\Phi_F}$ that fall within F_{r_1} consistently perform better than the corresponding vectors from both $\mathbf{\Phi_D}$ and $\mathbf{\Phi_E}$ in all frame-sizes tested. On the contrary, the performance of the feature vectors from the latter two sets, at frequencies that fall

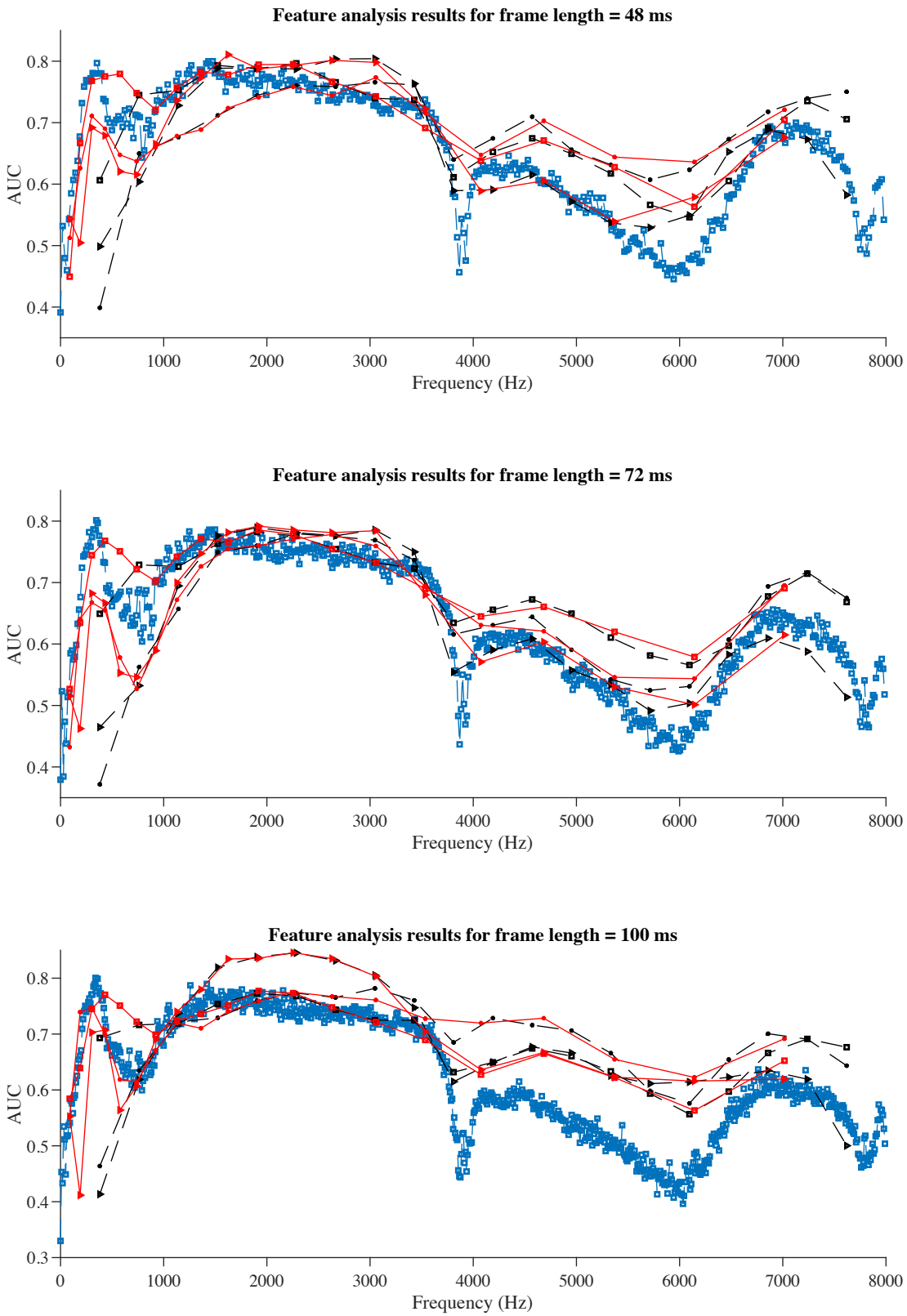

Figure 3.14: Individual performance of STFT based feature sets at (a) $l = 48$ ms, (b) $l = 72$ ms and (c) $l = 100$ ms with $N_B = 20$. Line styles and colours are the same as in Figure 3.13 .

within F_{r_2}, increases with increasing l and becomes similar to or even better than those of the $\mathbf{\Phi_F}$ set. As can be seen in Figure 3.14 (c), for frequencies between 1.5 kHz and 3 kHz (within F_{r2}), the delta coefficients from both $\mathbf{\Phi_E}$ and $\mathbf{\Phi_D}$ sets outperform the rest, achieving AUC ≥ 0.8. This effect is reflected in the performance curve (Figure 3.10) of the $\mathbf{\Phi_E}$ and $\mathbf{\Phi_D}$ feature sets as a slightly upward trend compared to the degrading performance with the $\mathbf{\Phi_F}$ set.

The feature subsets from the $\mathbf{\Phi_F}$ set, selected based on the method described in Section 3.3.2, contain only features that fall within F_{r_1} or F_{r_2} for $l \geq 24$ and only from F_{r_1} for $l \geq 32$. For the $\mathbf{\Phi_D}$ and $\mathbf{\Phi_E}$ sets, the majority of the feature vectors included in the selected subsets fall within the two frequency bands for any l tested. Therefore, these observations highlight the importance of the spectrum features that fall within the two identified bands of frequencies and show that these features have a strong impact on the classification performance.

In the work presented in Section 3.2 it was observed that the top performing features obtained from the magnitude spectrum are primarily at the low frequencies. This is further supported here where in addition, specific bands that carry significantly discriminant information were also identified.

From the above results and observations it can be concluded that the information in the time-frequency spectrum that enables SVM to discriminate between the two classes (normal/healthy vs abnormal/OA) is contained in a range of frequencies but in specific bands that depend on l. It can be deduced that the frame-size, and hence the DFT length, is a classification performance trade-off. Small l is preferred in order to achieve suitable frequency resolution for the knee sounds occurring in the range of 0.7 to 3.5 kHz and at frequencies ≥ 6 kHz but a larger l is preferred to capture the finer details of the spectrum in the 220 to 420 Hz band. For the sets $\mathbf{\Phi_D}$ and $\mathbf{\Phi_E}$, the latter two hold true. However, for the range 0.7 to 3.5 kHz the choice of l does not significantly affect the maximum AUC achieved but it has an effect in the performance of the individual coefficients.

The suitability of the choice $N_B = 20$ is tested in Section IV, even though reducing the dimensionality of the STFT spectrum using 20 filters scaled either linearly or non-

linearly (mel) in frequency improves the AUC in 18 out of the 21 l values tested. An adaptable filter-bank with narrower filters at the low (< 1 kHz) and highest frequencies (≥ 6 kHz) and broader at the mid to high frequencies could capture, with fewer coefficients than the full resolution spectrum, the information needed to discriminate the normal and abnormal signals with higher than 0.85 in AUC. This is further supported by the results showing that the $\boldsymbol{\Phi_D}$ low frequency features (up to 500 Hz or 1.6 kHz depending on l) consistently perform better than the corresponding $\boldsymbol{\Phi_E}$ whereas at high frequencies the difference in AUC is diminished and sometimes reversed. These observations are further investigated and validated in Section IV.

II. Local search in the vicinity of the best frame length

The l values that gave the maximum AUC, denoted as $l_{b,x}$ for the frame length b of the x feature set, were found from the experiments in the previous Section in which the time step used was $t = 4$ ms. The existence of local maxima in the vicinity of these points can be tested by defining a grid of 6 values with a time step of 1 ms centred at $l_{b,x}$ i.e. $l_{b,x} - t$ for $t = 3, 2, \ldots, -3, t \neq 0$. Steps shorter than 1 ms will not have a large impact in the result given that, at $F_s = 16$ kHz, the sample difference would be less than 16 samples.

The same experimental framework is followed and the training and test sets are standardized as before. In this experiment, 3 more classifiers are used to evaluate the subsets created at the feature selection step. This is done in order to assess how well the subsets generalise with different classifiers. The classifiers used are the (a) LDA classifier that finds linear hyperplanes in the feature space which separate the two classes, (b) CART and (c) SVM with a Gaussian kernel (SVM_g). Classification results are again evaluated using AUC.

The results of the experiments are presented in Figure 3.15 where for comparison the AUC at $l_{b,x}$ is also shown. The results show that for the sets $\boldsymbol{\Phi_M}$ and $\boldsymbol{\Phi_L}$, SVM_l and LDA generated comparable results at approximately AUC = 0.9 and 0.8 in the respective sets, with SVM_l being slightly better by 1% for $\boldsymbol{\Phi_M}$ and 0.5% to 7% for $\boldsymbol{\Phi_L}$. SVM_g and CART achieved AUC scores that ranged from 9% to 55% lower than the maximum.

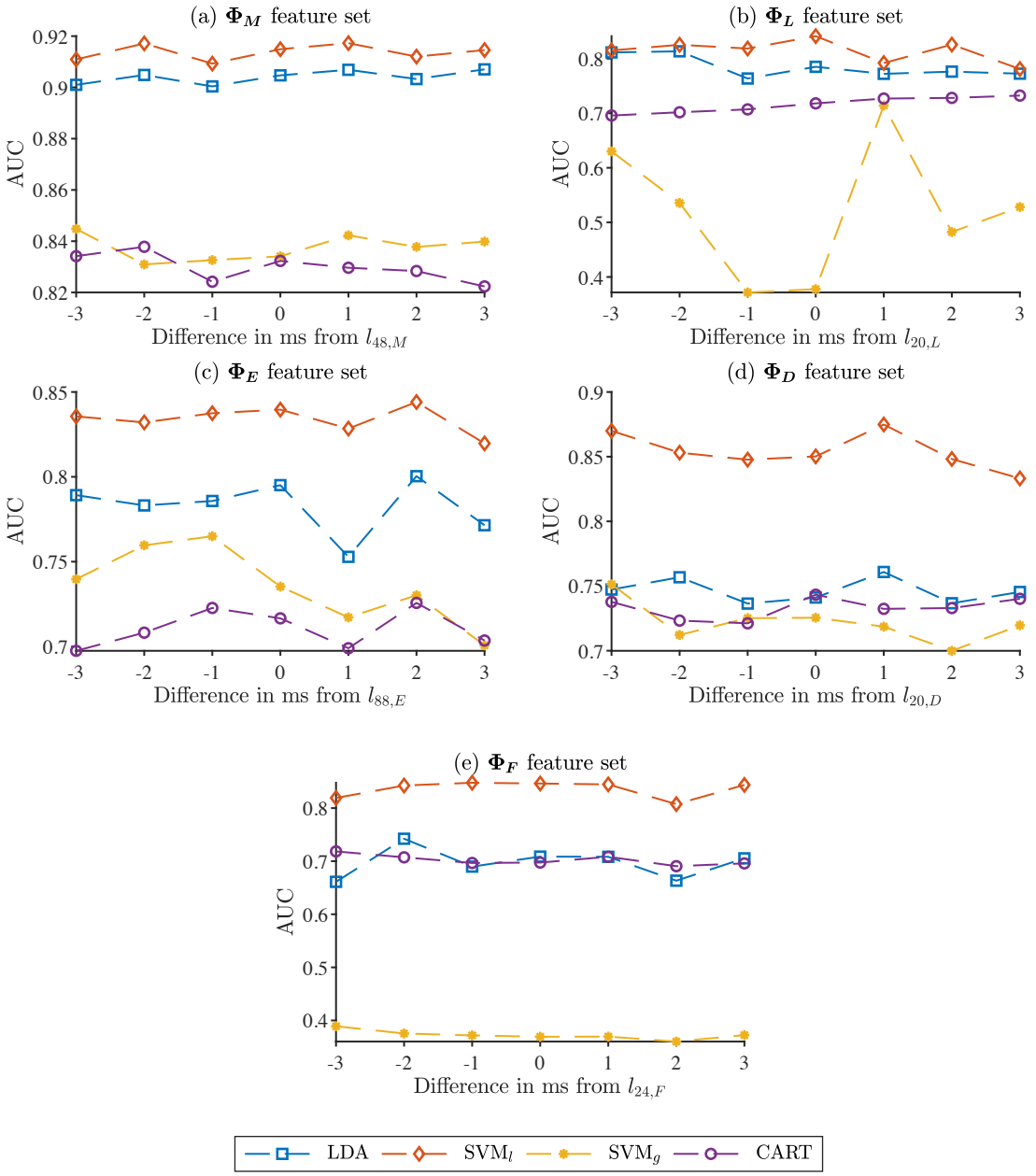

Figure 3.15: AUC per feature set against frame length evaluated with four different classifiers.

Feature set	AUC	E_r	$F_{0.5}$	MCC	S_c	l (ms)	Feature vectors used
$\boldsymbol{\Phi_M}$	0.917	0.147	0.853	0.705	0.804	49	$\boldsymbol{f}_3^M$
$\boldsymbol{\Phi_L}$	0.841	0.249	0.723	0.501	0.658	20	$\boldsymbol{f}_i^L$ for $i = 1, 2, 4, 6, \ldots, 9, 13, \ldots, 18, 22, 28, 30, 31$
$\boldsymbol{\Phi_F}$	0.848	0.218	0.756	0.564	0.701	23	$\boldsymbol{f}_i^F$ for $i = 7, 18, 30, 34, 35, 41, 49, 51, 52, 53, 56$
$\boldsymbol{\Phi_D}$	0.875	0.195	0.780	0.611	0.732	21	$\boldsymbol{f}_i^D$ for $i = 3, 6, 8, \ldots, 13, 31, \ldots, 34, 48, \ldots, 54, 60$
$\boldsymbol{\Phi_E}$	0.844	0.239	0.723	0.536	0.673	90	$\boldsymbol{f}_{26}^E, \boldsymbol{f}_{27}^E$

Table 3.5: Average cross-validation results of the best subsets per feature set using linear kernel SVM. The features that comprise the subset which produced the results, as indicated, are also shown.

For both $\boldsymbol{\Phi_E}$ and $\boldsymbol{\Phi_D}$ sets, SVM_l outperformed by at least 5.2% and 10.5% respectively the other classifiers which generated comparable scores (for $\boldsymbol{\Phi_D}$ only). For all $l_{b,x} - t$ in $\boldsymbol{\Phi_F}$ and 4 out of the 6 values of t in $\boldsymbol{\Phi_L}$, SVM_g performed significantly worse than the other classifiers by at least 70% and 31% respectively, producing results as low as AUC = 0.36. The power of the SVM classifier with a Gaussian kernel is limited in this case by the relatively large number of features compared to the small training set size, which increases the risk of overfitting when the data is transformed to a high dimensional feature space. Therefore, for the binary classification task of this work and based on the amount of available data, the performance advantage of linear classifiers suggests that the two classes can be linearly separated in the feature spaces explored.

As seen, SVM_l achieves higher AUC with all of the feature sets. However, the features were selected based on results obtained from training and testing this specific type of classifier. Therefore, the subsets created at the feature selection step are tailored to work better with this and similar classifiers. Nevertheless, the work in section 3.3 is not concerned with finding the best classifier to use with the available data. It is rather focused on the efficacy of the features in question to separate the two classes and capture information that would eventually lead to finding the specific abnormality (OA) signatures

in the signals and the various classification techniques are employed to help illuminate this question.

The experiments showed that the average AUC values for the frame-sizes $l_{b,x}-t$, for $t = 3, 2, \ldots, -3, t \neq 0$, were close to those of $l_{b,x}$. The results are improved for all feature sets except for $\boldsymbol{\Phi_L}$. Table 3.5 reports the frame length value that gave the highest AUC per feature set, found using SVM_l. The corresponding S_c score computed as in (3.30) is also used and consists of the MCC and $\text{F}_{0.5}$ measures which capture different attributes of the classification result than the AUC and would therefore be useful in the interpretation. It is shown in Table 3.5 that S_c ranks the top performance of the feature sets in the same way as AUC.

III. Monte Carlo experiments

In the previous experiments, the l values to be tested were defined in a deterministic approach. In this Section, a stochastic approach is followed in which the values are randomly chosen from a range. In this way the optimality of the result in the previous Section is assessed and the effect of increasing or decreasing the frame length even further is examined.

Firstly, plausible limits for l need to be set. Using frames larger than a single stride will cause overlap of the sounds from two strides resulting in poor modelling of those sounds. In addition, applying the DFT operation will become inappropriate because the signal in a single frame will be non-stationary. It was shown in Chapter 2 that the maximum recorded speed is 9 km/h giving an average stride duration of 0.7 s. Given that a single frame length value is applied at the feature extraction stage for all $s_{i,j}(n)$, the upper bound is set at 0.7 s. For the lower bound 2 ms is chosen which gives a time resolution that allows fine localisation of the sounds and approximately 470 Hz frequency resolution at $F_s = 16$ kHz for the DFT.

Monte Carlo simulations were performed by randomly assigning 20 values to l within the range 2 to 700 ms, excluding the range 20 to 100 ms, since it was examined in the previous experiments (Sections I and II). The same experimental framework is executed

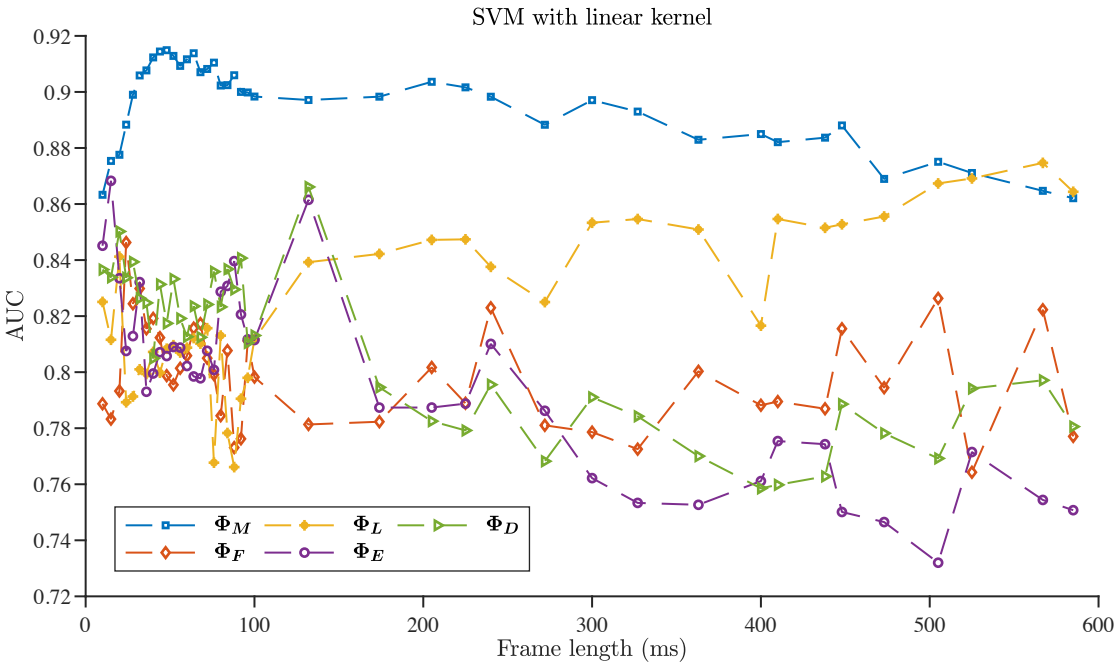

Figure 3.16: AUC per feature set against frame length for SVM (linear kernel) - including the Monte Carlo results.

as before using only SVM_l and the results are shown in Figure 3.16. In comparison to Table 3.5, an improvement of 2.8% is observed for the $\boldsymbol{\Phi_E}$ set at 15 ms, giving an AUC of 0.868 compared to the previous 0.844. With this experimental framework and using fixed frame segmentation, only the $\boldsymbol{\Phi_M}$ feature set generates a classification performance that varies smoothly over the entire range and has a global maximum. The $\boldsymbol{\Phi_L}$ feature set on the other hand generates a clear local maximum in the range of 20 to 100 ms. However, for $l \geq 205$ ms (except at 272 and 400 ms) the AUC scores are higher than the previously best one obtained at $l = 20$ ms. The maximum is achieved at $l = 567$ ms giving AUC $= 0.875$ which is even higher than that achieved with $\boldsymbol{\Phi_M}$. The $\boldsymbol{\Phi_D}$ and $\boldsymbol{\Phi_E}$ feature sets achieve an AUC of 0.86, both at 132 ms, whereas for $l \geq 174$ ms, AUC values less than 0.81 are achieved with both sets. The $\boldsymbol{\Phi_F}$ set on the other hand exhibits a more variable behaviour, with the AUC ranging between 0.76 and 0.85. However, it shows a clear global maximum at $l = 23$ ms.

IV. Experiments on the number of filters

The effect of the N_B parameter on the classification performance of all the feature sets except $\boldsymbol{\Phi_F}$ is examined in this section. N_B is varied from 10 to 75 while l is kept fixed at the values of Table 3.5. The outcome is evaluated using AUC and S_c. The experimental outcomes are summarised in Figure 3.17 for each set, showing the highest values obtained by any feature subset in each N_B case based on the plots' metric. As expected, the specific values of the two performance metrics are classifier depended. However, the general outcomes and the observations derived are similar across the different classifiers.

Figure 3.17 (a) and (b) show an overall negative trend with a small variance in the final value in both metrics as N_B increases. It can be deduced that there is a strong indication that a small number of filters (≤ 20) is more suitable for the extraction of $\boldsymbol{\Psi_D}$. The two metrics are shown to improve slightly for $N_B > 50$ only with CART but later drop and never exceed the highest score obtained with $N_B = 14$. With SVM_l and CART, the subsets that generated the highest results for any N_B always included $\boldsymbol{f}_3^M$. In fact, $\forall N_B \neq 10, 11, 13$, this particular feature vector was the only one selected in the final subset, one of which generated the overall highest result (AUC = 0.921 for $N_B = 33$ and SVM_l). For LDA, the selected feature subsets included $\boldsymbol{f}_3^M$ together with at most 2 more features (both static and delta coefficients) for $N_B = 11, \ldots, 18$ whereas for all other cases $\boldsymbol{f}_3^M$ was the only feature vector in the subset.

The maximum classification performance measured in both AUC and S_c is achieved when using only $\boldsymbol{f}_3^M$. However, if an exhaustive search is performed through the entire $\boldsymbol{\Phi_M}$ feature set and all possible combinations are used with a classifier, it is likely that a subset containing more vectors than only $\boldsymbol{f}_3^M$ would generate better performance. For this experiment however, such a method is computationally very costly as it would generate $\sum_{n=1}^{3N_B} \binom{3N_B}{n}$ possible combinations per classifier. Given that there is only a limited amount of data, the classifier results would become meaningless when n becomes larger than a certain value because the feature space will eventually become very sparse. When this happens, the classifier's decision boundaries will be formed due to the sparseness of the feature space and not due to the information captured by the features which can also lead

to overfitting, among other problems, [79]. For these reasons, a suboptimal search method, like the one employed in this work, is favoured and was found to generate significantly high results as shown in Figure 3.17 (a) and (b).

Compared to $\boldsymbol{\Phi_M}$, the results of the $\boldsymbol{\Phi_L}$ feature set are more variable as shown by plots (a) to (d) in Figure 3.17. High classification scores are obtained for $13 \leq N_B \leq 50$ (depending on the classifier) and the best result for this set is obtained with SVM_l and $N_B = 17$ giving an AUC of 0.853. A slightly upward trend for $N_B \geq 51$ is observed, which is more noticeable with SVM_l. From Figure 3.17 (e) and (f), it is clear that, when using LDA with the $\boldsymbol{\Phi_E}$ set, the classification performance in terms of AUC and S_c is higher for $N_B \leq 20$. With CART, an increase in the values of both metrics is observed for $N_B \geq 46$ which exceeds that achieved with LDA for those N_B values. This jump is attributed to the inclusion of features falling in the frequency band F_{r_1} compared to only using features that fall within F_{r_2} for $N_B < 32$. This effect is less obvious with SVM_l where the scores achieved in the range of the N_B values tested are comparable. For the final set, $\boldsymbol{\Phi_D}$, the plots are similar amongst the classifiers and suggest that, for best results, a small number of filters ($N_B \leq 20$) is also preferable. It is observed that the selected subsets with which the classifiers scored the highest AUC and S_c only contain feature vectors that fall in either F_{r_1} or F_{r_2} or both. This further supports the observations and results discussed in Section I, stressing the importance of these frequency bands, as well as the importance of the information in the dynamics of the spectrum for classifying normal and abnormal knee signals.

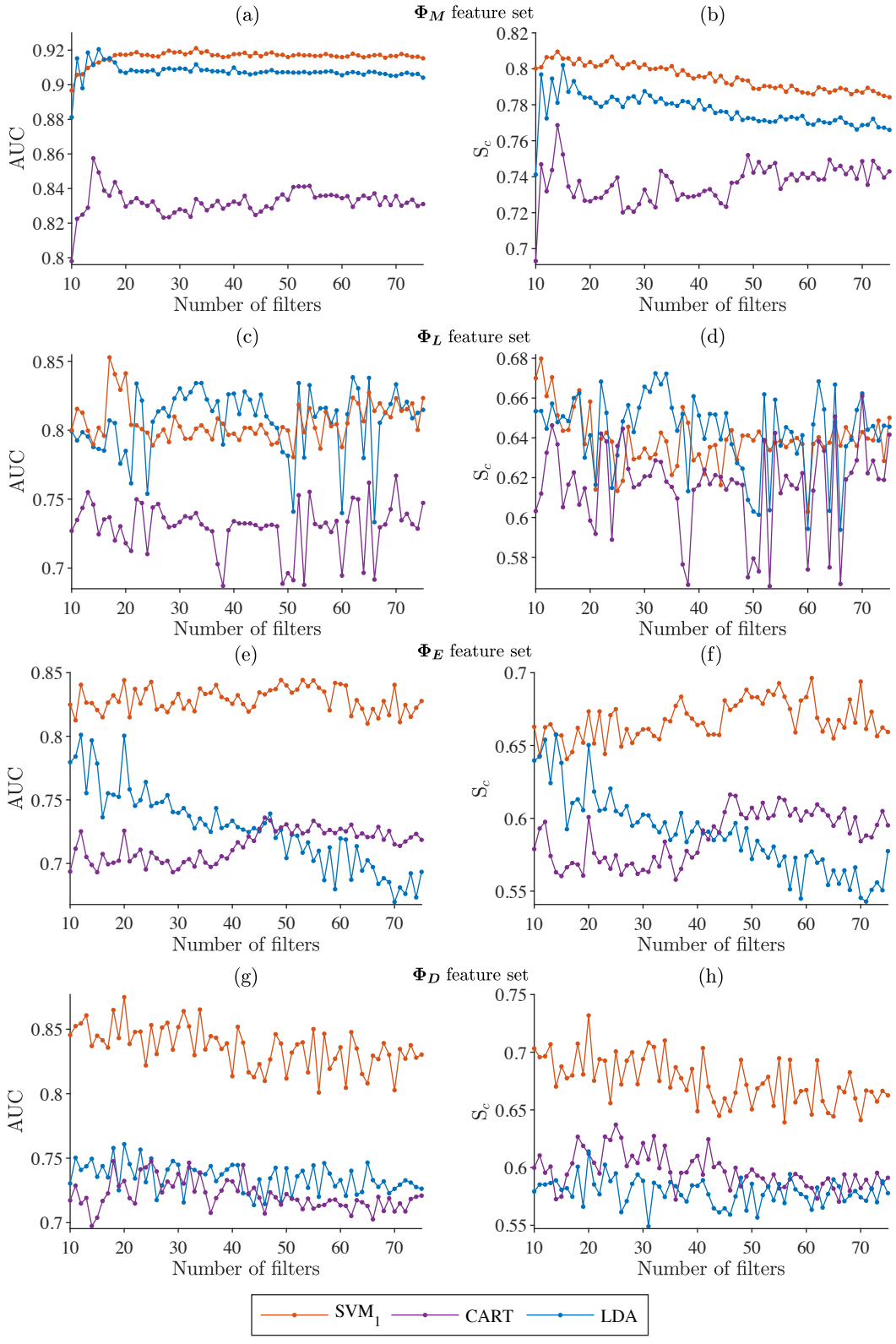

Figure 3.17: Effect of the number of filters, N_B, on classification performance per feature set, using 3 different classifiers.

3.4 Conclusion

This Chapter proposed an experimental framework to investigate the discriminant power of time-frequency and other features, for the task of classifying the knee condition, and explored effective parameterisations of the knee sounds collected from healthy and OA knees during walking. The efficacy of knee condition classification was evaluated by qualitative and quantitative analysis using the AUC of the ROC curve and metrics derived from the confusion matrix (E_r, $F_{0.5}$, MCC). Additionally, a study of the effect of varying the values of the feature extraction parameters of frame length and number of filters in the filter-bank was presented in order to examine their impact on the classification performance. In addition to the conclusions of Section 3.2.5, the results of this work enable the extraction, from AE signals, of spectrum features and CC focused on specific frequency bands that were shown to carry significantly discriminant information.

The results show that reducing the dimensionality of the STFT spectrum using a mel-spaced triangular filter-bank improves the classification performance compared to using the full resolution spectrum for the case of LDA, CART and SVM with linear or Gaussian kernel (see Figure 3.15 and Table 3.5). Furthermore, the analysis signifies that taking the natural logarithm of the STFT spectrum and subsequently computing the DCT also improves the performance. In fact, using CART, LDA and SVM as the tools for classification, the findings demonstrate that low order coefficients from the $\boldsymbol{\Phi_M}$ feature set (especially the $\boldsymbol{f}_3^M$ feature vector) can distinguish between healthy and OA knee sound signals with an AUC of 0.92 which is the highest value obtained amongst the 5 feature sets examined.

In the work presented in Section 3.2 it was found that the top performing features (in classification) obtained from the magnitude spectrum are primarily at the low frequencies. The results of Section 3.3 further supported this finding where in addition, specific bands that carry significantly discriminant information were also identified. The experiments conducted in order to investigate the effect of the frame length and the number of filters, revealed two frequency regions, namely 220 Hz to 420 Hz and 1 kHz to 3.4 kHz. These regions contain a collection of features (both static and derivative coefficients) that

individually score higher classification results, measured by AUC and S_c, than the rest in their respective feature sets. The analysis performed highlighted the importance for classification performance of the spectrum features within these two frequency bands.

Contrary to other studies that focus on sit-to-stand movements and similar variants (e.g. knee flexion and extension), this study analysed signals obtained from knees performing an OCA. The outcomes presented in this Chapter suggest that the analysis of such signals could potentially lead to non-invasive detection of knee OA with high accuracy and could also potentially extend the range of available tools for the assessment of the disease as a more practical and cost effective method without requiring clinical setups.

Chapter 4

Stride Detection and Segmentation

THIS Chapter presents the research work carried out towards the development of an algorithm that can detect the instants of first and last contact of the foot to the floor, from the knee sound signal alone. The technique is designed for acoustic signals sensed at the patella of the knee and picked up by a single contact microphone. The Chapter begins with a review of gait analysis methods and continues with an introduction to Linear Predictive Coding and Group Delay which are important techniques for understanding the material presented later. Initial investigation in stride detection and segmentation is presented in Section 4.3 with results and discussion. Section 4.4 presents the development of a Dynamic Programming based algorithm that automatically estimates both the initial and last contact instants. A series of tests and experiments for evaluating this algorithm against ground truth data are presented in Section 4.5, along with discussion and interpretation of the results. Following this, in Section 4.6, a method that improves the temporal location of the estimates obtained from the Dynamic Programming algorithm is presented. Finally, important results and conclusions are summarized at the end of the Chapter. Part of the work in this Chapter relates to conference publication [C2].

4.1 Introduction

The work presented throughout this thesis focuses on the study of the acquisition and analysis of sounds generated by the knee during walking with particular focus on the effects due to OA. In this context, reliable stride detection is fundamental for the acoustic analysis that will enable intermittent sounds to be discriminated from those that occur in every stride (stride-synchronous analysis). Stride synchronization also enables inter-stride signal averaging to improve detection of quasi-periodic acoustic events with poor Signal-to-Noise Ratio (SNR) as well as the extraction of features per stride, for knee condition classification.

Adding to the motivation above, this Chapter's work can be directly applied to, and enable further, gait analysis. For example, the methods developed here can also benefit people with movement related injuries or with knee pathologies other than OA at the initial or developed disease stages for rehabilitation and monitoring purposes. The Dynamic Programming (DP) based algorithm developed in this Chapter could potentially also benefit athletes, as its reliable estimation method can be used for correcting and improving their gait.

The aim of this Chapter is to present an analytical study towards a novel method for automatic temporal stride detection and segmentation that does not require bulky and expensive equipment or patient data collection in a controlled environment as is the case for other methods presented in the literature review. The novel contributions of this work are, first of all, the use of knee joint sounds generated during walking and sensed at the patella for stride detection and segmentation. This type of Acoustic Emission (AE) has not been used before for this task. Furthermore, a novel automatic, dynamic programming based method for temporal stride detection and segmentation using only the sound signal mentioned before is presented.

4.2 Technical background and literature review

4.2.1 Gait analysis

The term gait is used to describe the way of walking and consists of consecutive cycles. During each gait cycle a sequence of events take place that mark the transition from one gait phase to another. Gait analysis is an important clinical assessment tool since changes in gait may reflect changes in general health. It is a useful measure of overall health and is a predictor for cognitive decline, quality of life and longevity [107]. Gait analysis is a broad term that covers many aspects of the human walking pattern, ranging from motion and posture analysis to stride and energy spending analysis [108]. For example, gait analysis is used to assess and treat individuals with conditions affecting their ability to walk [109]. It is also commonly used in sports medicine to help athletes run more efficiently by correcting their gait and posture [110, 111]. In medicine, gait analysis is used to assess the stage of diseases besides OA, such as Parkinson's [112, 113, 114, 115] and Alzheimer's [116, 117].

With respect to Figure 4.1, the gait cycle begins with the heel strike, defined as Initial Contact Instant (ICI) in this thesis and is the time instant when the heel of one foot makes contact with the floor, and ends when the same foot, after having executed one step, makes contact again. The instant when the toes leave the floor is usually called "toe off" but in this thesis it is referred to as Last Contact Instant (LCI). The cycle begins

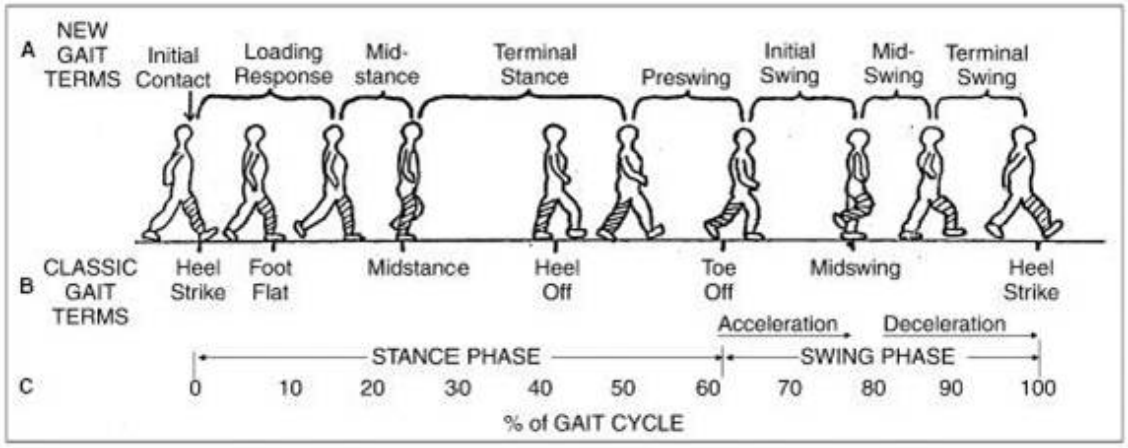

Figure 4.1: Phases of the gait cycle (source: http://physio-pedia.com/Gait).

with the stance phase which is the contact duration (between ICI and LCI) and is followed by the swing phase where the foot is suspended over the floor and moves to complete the cycle (one stride). The stance to swing phase ratio for normal gait is around 3 : 2, that is, 60% and 40% respectively of the total cycle duration [118]. During walking, two temporal events must be detected in order to extract the main gait characteristics such as cadence, duration of double support and the duration of the stance and swing phases. These are the ICI and the LCI which are the time instants that enable temporal stride detection and segmentation and will be the focus of this Chapter.

The development of algorithms for gait event detection using various sensing methods has been an active area of research for years [119, 120]. Reliable identification of ICIs and LCIs can be performed using the ground reaction force (GRF) signals obtained from surface mounted force plates (see Section 4.5.2). This method has become the benchmark to which all other methods compare. In the literature, many studies that investigate temporal parameters in walking or running used force plates or contact mats [118, 121, 122, 123, 124]. Pressure sensitive foot switches have also been used to identify ICIs and LCIs by detecting when a load is placed on or removed from the foot [125, 126]. Foot switches, however, lack mechanical robustness, require the modification of the subject's footwear and the detection accuracy depends on the appropriate and accurate placement of the sensors [127].

Gait analysis based on kinematics, i.e. considering the movement without taking into account the forces that cause it, can be performed by indirect measurements using cameras or other tracking systems. Such methods use reflective markers usually placed at the heel and the toe of the foot to collect positional data and then extract the velocity and acceleration as features to detect ICIs and LCIs [128, 129, 130, 131]. The performance of these methods however, is affected by the walking speed and the kinematics of the joint [132, 133]. The LCI events especially, obtained from such data, were found to be positively correlated with the variability in walking speed, in a comparative study by Hendershot et al. [134]. In [132], data was obtained during overground walking and walking on treadmill, from 18 subjects with different conditions (healthy, with multiple sclerosis,

suffered a stroke) using a six camera motion analysis system with heel and toe markers. It was reported that for treadmill walking and depending on the subject's condition, the estimated ICIs and LCIs had an absolute average time error in the range of 17.4 ms to 37.1 ms and 12.7 ms to 34.7 ms respectively, relative to GRF reference Contact Instants (CIs).

Different algorithms demonstrated a wide range of performance with absolute average time errors ranging from 8 to 55 ms for ICI and 28 to 112.8 ms for LCI, in comparison to vertical GRF [135,136,137]. In these studies, three dimensional kinematic data was obtained using multiple cameras. The algorithm presented in [138] was compared to ground truth data from force plates and the standard deviation σ of the time errors was reported to be 23 ms for ICI and 25 ms for LCI with means of 1 ms and -2 ms, on a database of 40 steps from children with cerebral palsy disorder. The same algorithm was evaluated on a smaller database of just 16 steps obtained from healthy adults, with a reported $\sigma = 19$ ms for ICIs and $\sigma = 12$ ms for LCIs [138]. The respective mean time error values were 27 ms and -14 ms. In the same study two different methods, [128, 139], were evaluated on the same data, achieving ICI time errors with $\sigma = 100$ ms for the method in [128] and $\sigma = 73$ ms for the method in [139] with mean values 31 ms and -8 ms respectively. Similarly for LCIs, the results were $\sigma = 13$ ms and $\sigma = 30$ ms with mean values -28 ms and 21 ms. These results correspond to data from healthy adults only.

Accelerometer based methods are also widely used for gait event detection. Such methods utilize the direct measurement of linear and angular displacements provided by joint angles or accelerations [119]. Accelerometer based ICI/LCI detection received a high interest from the research community [140, 141, 142, 143]. This is primarily because accelerometers are miniature, inexpensive and low-powered [119, 144]. An ICI detection method was developed in [145] using multiple accelerometers placed at various body parts. This method exploited the fundamental spectral relationships during normal gait between the movement of the different body parts. It was evaluated on four accelerometer based datasets, none of which however, had reference ICIs from force plate GRF signals. In another study, the potential of discriminating walking on level ground from walking on

a stairway was investigated using a triaxial accelerometer attached on the waist [146]. ICIs in this work were defined as the maximum acceleration in the vertical direction. ICI and LCI detection was also investigated in [147], considering healthy subjects and transtibial amputees during overground walking. Two uniaxial accelerometers were placed just below each knee and GRF recordings from force plates were used as the reference measurements. The algorithm was evaluated on a database of approximately 325 ICIs and 325 LCIs, obtained at different walking speeds on flat level, from 10 healthy subjects and 8 transtibial amputees. For the healthy subjects data, an average ICI time error of 34 ms with $\sigma = 25$ ms was reported, whereas the corresponding values for LCIs were 19 ms and 36 ms. For transtibial amputees, the average time errors were 33 ms and 13 ms for ICI and LCI detection respectively, with corresponding σ values of 41 ms and 38 ms.

In many of the studies mentioned in the literature, the patients walked in a very controlled manner in order to hit the floor mounted force plate accurately and with the appropriate foot such as in [143, 147, 148]. This might have affected the natural gait pattern of the patients and the fact that the reference strides obtained are not continual could have introduced a bias in the results. This bias might occur due to the data used to train and validate or test the reported algorithms which is derived from a distribution that is different from the real-world data. That is, events that depend on past events (step after step). This type of continual sequence of events is therefore not present in the training and validation/test sets. Testing these algorithms that were developed on such data that poorly reflects the real-world will very likely result in poor performance. The use of a specialised treadmill mounted with force plates, such as the one used in this thesis study, avoids these drawbacks. Furthermore, accelerometers suffer heavily from noise due to mechanical vibrations and their usage requires the compensation of gravity influence [119]. The attachment of sensors is another source of imprecision due to muscle movement during walking, which appears as a high frequency error according to [119,127].

Although force plates are accepted as the gold standard for temporal event detection in walking, they suffer from several limitations [118]. Their lack of portability and their setup complexity, requiring the purchase of expensive complementary equipment and

software that requires a trained expert, restrict their use to only laboratory experiments. This is a major drawback given the in-field nature of walking. Given these limitations there is a clear motivation for an accessible and inexpensive, yet accurate method, for CI estimation. This will enable gait pattern detection with subsequent analysis.

Audio has been used to study footstep sounds and gait patterns, not necessarily from a medical perspective but usually for audio event detection applications [149] and person identification [150, 151, 152]. Little attention has been given to temporal gait event detection for gait analysis purposes [153, 154, 155]. Altaf et al. in [154] explored footstep detection from audio captured by 16 microphones placed in pairs and on the walls around a room. The squared energy estimate, Hilbert transform and Teager-Kaiser energy operator were explored as extraction methods for spatio-temporal gait parameters. It was reported that the Teager-Kaiser energy operator provided the most reliable gait parameter estimates that are used in standardized gait assessments. In [155], an ICI detection system was developed, consisting of a wearable device with a single omni-directional microphone attached at each ankle. The microphones faced towards the ground and captured the footstep sounds. The developed algorithm follows a data driven approach in which 36 features in total were used to extract information from the audio signal. The features were the sub-band energies in 0 to 4 kHz range (10 sub-bands), zero crossing rate, 12 linear prediction cepstral coefficients and 12 MFCC. Additionally to these features, the final detection makes use of self-defined thresholds. It was reported that with an SVM classifier and using a radial basis function (RBF) kernel, the ICI detection rate achieved was 94.52%. However, for these results the reference instants used were manually labelled by people after listening to the audio recordings. This is a major limitation as labelling can be subjective.

4.2.2 Linear predictive coding

Linear Predictive Coding (LPC) is a technique that is widely used in audio and speech signal processing. Suppose that sample n from $s(n)$ can be predicted from linear combi-

nations of past samples weighted by some predictor coefficients

$$\hat{s}(n) = \sum_{i=1}^{p} \alpha_i s(n-i) \tag{4.1}$$

where p is the order of prediction i.e. the number of previous samples used to 'predict' the current sample. The α_i are the Linear Prediction (LP) coefficients (predictors) chosen so as to minimise the squared error E, defined as

$$\begin{aligned} E &= \sum_{m} \big(s(m) - \hat{s}(m)\big)^2 \\ &= \sum_{m} \Big(s(m) - \sum_{i=1}^{p} \alpha_i s(m-i)\Big)^2 . \end{aligned} \tag{4.2}$$

The range of summation m is important and depends on the method used to find the predictors. Equation (4.2) is minimised by setting

$$\frac{\partial E}{\partial \alpha_k} = 0 \tag{4.3}$$

for $1 \leq k \leq p$, which leads to a set of p equations, known as normal equations, with p unknowns, [156],

$$\sum_{i=1}^{p} \alpha_i \sum_{m} \big(s(m-i)s(m-k)\big) = -\sum_{m} s(m)s(m-k). \tag{4.4}$$

Assuming that E is minimized over the infinite duration $-\infty < m < \infty$, by expanding (4.2) and then substituting (4.4) gives the minimum squared error E_m as

$$E_m = \sum_{m=-\infty}^{\infty} s(m)s(m) + \sum_{i=1}^{p} \Big(\alpha_i \sum_{m=-\infty}^{\infty} s(m)s(m-i)\Big) \tag{4.5}$$

In practice however, only a finite portion of the signal is of interest. This can be obtained by multiplying $s(n)$ with a window signal $w(n)$ that is zero outside some interval $0 \leq n \leq l_w$ (e.g. hamming window [157])

$$s_w(n) = \begin{cases} s(n)w(n), & 0 \leq n \leq l_w \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

Tapering the signal segment at the edges to zero helps to avoid predicting the signal from zero valued samples outside the window range (n near 0) and also predicting zero valued samples (outside the window range) from non-zero samples (near $n = l_w$). Considering $s_w(n)$ and $s_w(m)$ as equivalent, the minimum squared error can be shown to be, [158],

$$E_m = r_{s_w s_w}(0) + \sum_{i=1}^{p} \alpha_i r_{s_w s_w}(i) \tag{4.7}$$

where $r_{s_w s_w}(i)$ is the autocorrelation function of the signal $s_w(m)$. The above equation can be expressed in matrix form as

$$\begin{bmatrix} r_{s_w s_w}(0) & r_{s_w s_w}(1) & \dots & r_{s_w s_w}(p-1) \\ r_{s_w s_w}(1) & r_{s_w s_w}(0) & \dots & r_{s_w s_w}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{s_w s_w}(p-1) & r_{s_w s_w}(p-2) & \dots & r_{s_w s_w}(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} r_{s_w s_w}(1) \\ r_{s_w s_w}(2) \\ \vdots \\ r_{s_w s_w}(p) \end{bmatrix}. \tag{4.8}$$

The first matrix is a $p \times p$ toeplitz matrix and the equation can be solved efficiently using the Levinson-Durbin algorithm [159]. The predictors can be used to form an inverse filter which can then be applied to $s(n)$ [158]. This yields an impulsive 'excitation' signal, the LP residual signal $u(r)$, an example of which is shown in Figure 4.22 (c).

To put this in context of the work in this Chapter, the Autocorrelation LPC method described in this section, is a building block that underpins the Automatic Stride Detection and Segmentation Algorithm, proposed and presented in Section 4.4. It forms the basis for generating a set of candidates using $u(r)$, from which the time instants of the foot's first and last contact with the floor will be selected.

4.2.3 Group delay

As will be shown later in the Chapter, the impulsive events in $u(r)$ indicate particular points in the gait cycle that are to some extent 'unpredictable' from previous samples. Their location can be estimated using the Group Delay (GD) within an analysis window. In general, GD is the negative derivative of the unwrapped phase function of the STFT [160].

For a signal $u_r(n)$, the GD is computed as

$$\tau_g(k) = -\frac{d\ \arg(U_r(k))}{d\omega} \tag{4.9}$$

where $U_r(k)$ is the Fourier transform of $u_r(n)$ at frequency $\omega = \frac{2\pi k}{N}$, given by

$$U_r(k) = \sum_{n=0}^{N-1} u_w(n) e^{-2j\pi nk/N} \tag{4.10}$$

in which k varies continuously and $u_r(n) = w(n)u(n+r)$ is an N-sample windowed segment of the residual signal $u(r)$ beginning at sample r, for $n = 0, 1, \ldots, N-1$. From the definition of (4.9) it follows that for a unit impulse function $\delta(n)$ and $u_w(r) = \delta(n-n_0)$, we have $\tau_g \equiv n_0$ for any k. However, if the signal is corrupted by noise, as in $u(r)$, then $\tau_g(k)$ will not be constant. In this case, in order to find an estimate of the delay from the start of the window to the impulsive event, some kind of averaging over k needs to be performed.

A thorough comparison of various GD estimation methods that deal with this problem is given in [161] which puts them in the context of identifying the glottal closure instants in voiced speech. In the algorithm described in Section 4.4, the energy weighted formulation of the GD presented in [161], is used due to its higher performance and substantially lower computational cost compared to other methods. In its simplified form it is given by

$$d_{EW}(r) = \frac{\sum_{n=0}^{N_w-1} n u_w^2(n)}{\sum_{n=0}^{N_w-1} u_w^2(n)} \tag{4.11}$$

for a window size of N_w. The instants where the d_{EW} gives negative-going zero crossings (NZC) correspond to major excitations in an impulsive signal with a gradient of -1 near the zero crossing for a clear impulse train [161, 162]. These instants will be used as candidates for ICIs and LCIs and is further discussed in Sections 4.3.2 and 4.4.1.

4.3 Stride detection - An investigation

The following sections present an investigation carried out for stride detection from the knee AE signal and which led to the development of the DP approach in Section 4.4. Figure 4.2 displays examples of normalised knee sound signals recorded from a healthy subject and a patient with OA in the medial compartment of the knee (see Figure 1.1). Strong peaks that represent loud noises, compared to the low energy regions in between, can be clearly seen. These occur as quasi-cyclic instants as a result of the walking pattern of the subjects. The waveform in the vicinity of these peaks is fairly similar from one stride to the next in Figure 4.2 (b), but can also vary as in Figure 4.2 (a). An early attempt in segmenting the AE signal considered finding the strongest peaks in $s(n)$, under the assumption that these peaks represent the ICI. The quasi-periodicity of $s(n)$ was exploited in order to introduce constraints in the decision process.

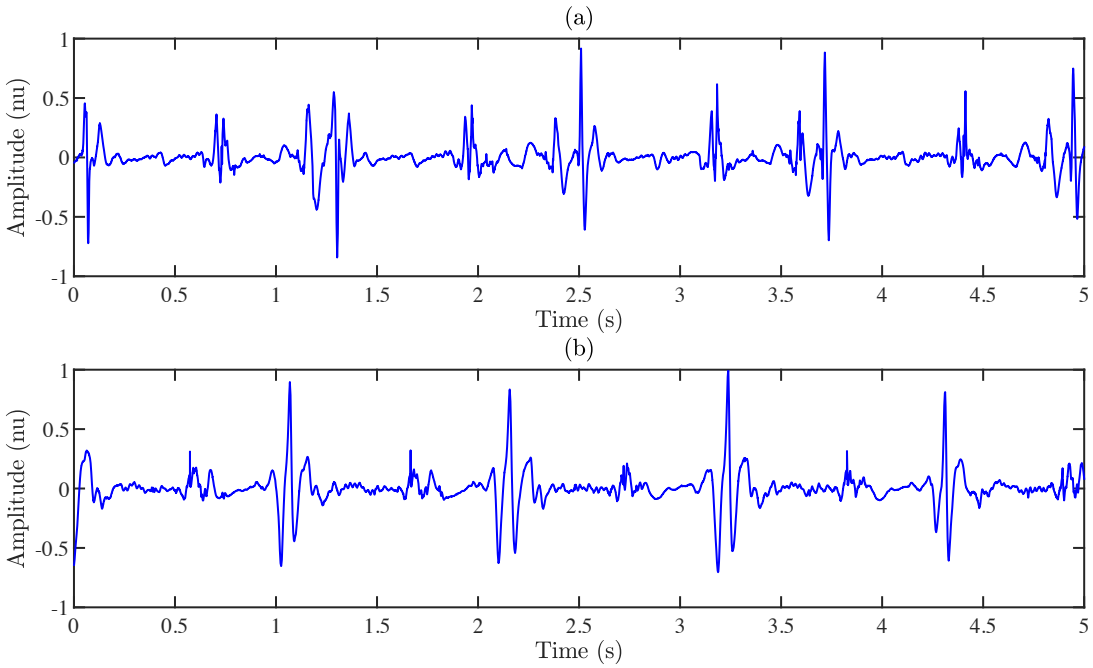

Figure 4.2: (a) The amplitude-normalized knee sound signal of a healthy subject. (b) The amplitude-normalized knee sound signal of a patient with osteoarthritis in the medial compartment. *nu* : normalised units.

4.3.1 Period estimation and peak detection

Given the inherent cyclic nature of the AE signal $s(n)$ due to normal walking, an important parameter that is useful in segmenting the signal into individual strides is the period. This can be determined either from the periodicity in the time domain or from regularly spaced harmonics in the frequency domain. For speech signals the most widely used time domain pitch estimation method is the autocorrelation approach [163, 164]. In this work, this approach is adopted as speech signals and $s(n)$ share similarities in the time domain as both types of signals have quasi-periodic time structure. Moreover, walking is usually performed smoothly with no sudden changes in speed and, assuming no tripping occurs, as is the case for all the signals in the database, autocorrelation is a suitable approach for estimating the period despite the need for a relatively large duration window for segmentation of $s(n)$.

The autocorrelation function has a maximum of 1 at a time lag of zero and in the presence of noise its amplitude envelope decreases exponentially. It is known that the periodicity in the autocorrelation function indicates periodicity in the signal [165, 166]. Given the quasi-periodic signal $s(n)$, the normalised autocorrelation of a short segment $s_w(n)$, obtained as in (4.6) using a rectangular window, is found as

$$\hat{r}_{s_w s_w}(m) = \frac{r_{s_w s_w}(m)}{\sqrt{r_{s_w s_w}(0)}} \ . \tag{4.12}$$

The normalization ensures that the autocorrelation sequence value at zero lag is equal to unity. The local segment period can then be found as

$$N_l = \frac{1}{F_s}\big[\operatorname*{argmax}_m \ \hat{r}_{s_w s_w}(m)\big] \text{ seconds} \tag{4.13}$$

for $m > 0$, subject to

$$N_1 \leq N_l \leq N_2 \tag{4.14}$$

where N_1 and N_2 are limit values. In addition, $s_w(n)$ must satisfy the condition that it contains at least two strides in order for the calculation to be accurate. It is known that estimating the period using the autocorrelation approach is prone to doubling or

halving errors, i.e. estimating double or half the true period. To avoid this type of error, a constraint is set on the local estimate as indicated in (4.14), that is, the accepted estimate must fall within a predefined interval. Taking the median value in (4.15) also helps in moderating this error.

Given that the lowest and highest walking speeds in the database are 2.5 km/h and 9 km/h and according to the findings in [167] regarding the relation between walking speed and stride duration, suitable values for the limits are $N_1 = 0.4$ and $N_2 = 1.7$ in seconds. Moreover, the length of the window $w(n)$ must be chosen accordingly. An experiment was conducted for this purpose from which the sliding window length chosen was 2.4 s with a step-size of 0.24 s. The experiment details can be found in Section 4.5.2.

The entire signal $s(n)$ is thus processed in segments and in each such segment, equation (4.13) is used to calculate the local period. However, the true stride duration can vary slightly even between consecutive strides, hence, small variations in the local period values are observed. Therefore the period of $s(n)$, used throughout this Chapter, is obtained as the median value of the collection of these local period values:

$$\hat{N}_s = \text{median}(\ N_l(i)\ )\ . \tag{4.15}$$

An evaluation of the accuracy of this approach against ground truth data is presented in Section 4.5.2. The mean value could have also been taken in (4.15) instead of the median but it is not preferred as it is more sensitive to double and half period errors.

An initial simple approach in determining the ICIs was considered and is based on identifying the highest peaks in $s(n)$. The highest peak is found first and then simultaneously scanning forward and backward (in time) to find the next peaks subject to the constraints that they must be within $\pm\alpha\hat{N}_s$ from the immediate previous one, where α is a constant slightly larger than 1, and with an amplitude not smaller than $^2/_3$ of the highest peak. This is based on the assumption that in normal walking the heel strike response picked up by the microphone is relatively constant as can be seen in the examples of Figure 4.2.

In the following Section it will be shown that the assumption on the equivalence of the ICIs to the strongest peaks in the signal is not true in all of the recordings. Therefore, a different approach that is not based on this assumption is investigated next.

4.3.2 Forming groups and pruning

Certain regions of high sample energy can be observed in Figure 4.2. Plotting the instants of initial contact derived from the ground truth data (details in Section 4.5.2) shows that they fall inside or in the close vicinity of these regions but do not correspond to the highest peak within the region as assumed in the previous Section. A different approach to peak detection for finding better estimates of ICIs and to also detect LCIs is thus motivated based on the above observation.

The investigation in this Section develops the methodology that is later improved and incorporated into ASDSA. In short, the methodology consists of two stages. In the first stage a number of possible initial or last contact instants are generated and, in the second stage, this set is pruned subject to certain constraints.

As discussed in Section 4.2.3, the LP residual signal $u(r)$ contains identifiable impulsive events that indicate particular points in the gait cycle that are to some extent 'unpredictable' from the previous samples and whose location can be reliably estimated using GD. This characteristic is exploited and (4.11) is used to obtain the set of candidates defined as the NZC of d_{EW}. Given $s(n)$, the residual signal $u(r)$ is obtained using analysis frames of size 50 ms with 50% overlap. The choice for the order p and the GD window size is made after experimenting with different values and selecting the combination that yielded the highest Detection Rate, that is, the percentage of strides where exactly one ICI and one LCI is detected. This metric is defined in Section 4.5.3. In the cases of equal Detection Rate, the combination with the lowest GD window size was chosen for reduced computational complexity.

The amplitude envelope of $u(r)$, denoted as $|u(r)|$, is obtained as described in [168] and is shown in Figure 4.5(a). Various distribution fits were fitted to the collection of the sample amplitude values of $|u(r)|$ from which an inverse Gaussian distribution was chosen

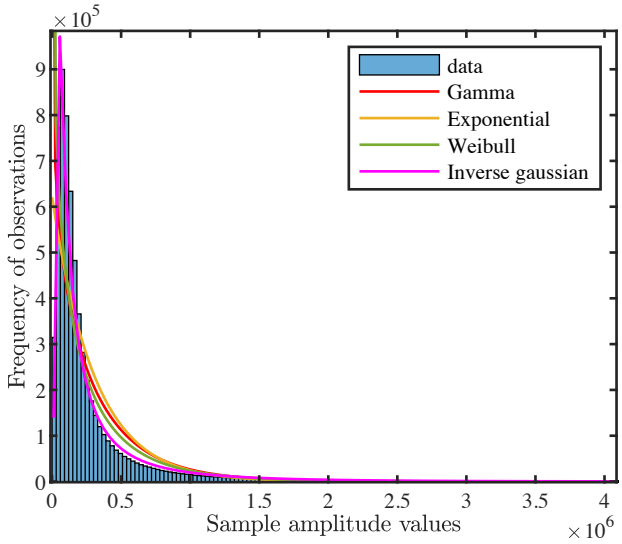

Figure 4.3: Histogram of the amplitude values of $|u(r)|$ from 20 knee acoustic emission signals with various distribution fits. The histogram is zoomed in at the lower amplitude values.

as can be seen in the histogram of Figure 4.3 which was obtained using 20 knee AE signals. In order to form regions of interest, a threshold is found based on the inverse CDF of $|u(r)|$, computed at an input of 0.8 (empirically determined), and is used to separate the high amplitude samples from the rest. A sliding window of 40 ms duration with a 20 ms step increment is applied on $|u(r)|$ and a region of size equal to the window duration is formed if $2/3$ of the samples in the window have an amplitude larger than the threshold. The $2/3$ was chosen on the basis that regions of high signal variability are of interest and therefore, in such regions the majority of the samples should exceed the amplitude threshold as can be seen in Figure 4.5. Finally, formed regions that overlap or are 1 sample apart are merged together.

Assuming two regions per stride and knowing that ICIs and LCIs must alternate and that the stance phase is always longer in duration than the swing phase [169,170], the formed regions can be separated into two groups, one corresponding to possible ICIs and one to possible LCIs. To achieve this, three consecutive regions are chosen at random and the sample differences between the midpoint of the two outer regions and the midpoint of the middle one are computed. The larger difference of the two denotes the stance phase (i.e. between an ICI and an LCI) and the smaller denotes the swing phase (i.e. between

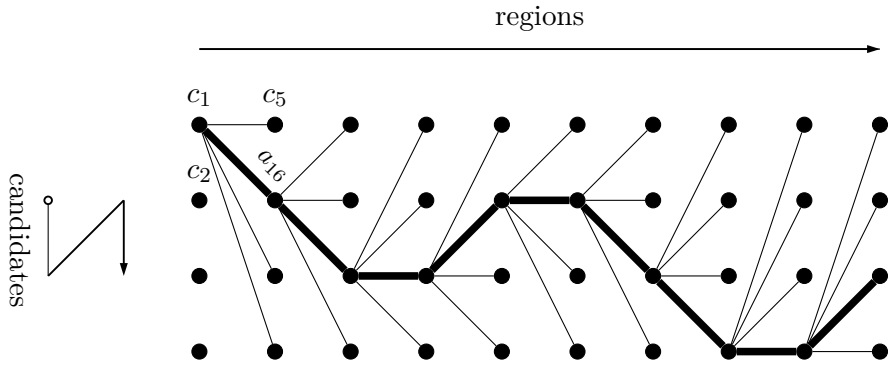

Figure 4.4: Example lattice formed by the regions and their candidates showing one full path.

an LCI and an ICI). Once the labels are identified, all the other regions are marked appropriately by alternating the labels.

The benefit of forming regions of interest is that the number of generated candidates is reduced and therefore, the problem of choosing the best candidates becomes computationally much easier which in turn allows the use of exhaustive search. Having defined the two groups of regions that contain the initial and last CI candidates respectively, the next step is to prune each set down to 2 selections per stride (1 from each group) by performing an exhaustive search subject to certain constraints. The selection algorithm prunes each group separately and in the following paragraphs only the pruning of one group is discussed as the same principles apply for the other.

It is assumed for simplicity, but without loss of generality, that there are 4 candidates per region with 10 regions in total within the group. The problem of pruning can be reformulated into a problem of finding the best path through the lattice formed by the regions and their candidates as shown in Figure 4.4, subject to the constraints that the path starting and ending points as well as the cost of the transition from one node to the next are known and do not vary with time. A further constraint is that only one candidate per node can be chosen and only forward transitions of one step are valid i.e. self-transitions are not possible and no node is skipped. The transition cost is computed based on two components:

1) slope cost of the current candidate:

$$C_{sd} = 1 + m(n_z) \tag{4.16}$$

with

$$m(n_z) = \frac{1}{l_s}\left[d_{EW}(n_z + \frac{l_s}{2}) - d_{EW}(n_z - \frac{l_s}{2})\right] \tag{4.17}$$

where n_z is the sample instance of the current candidate and l_s is the length of the window centred at n_z and is set equal to $^{20}/_{F_s}$ seconds.

2) Period consistency cost, C_{pc}, defined as

$$C_{pc} = 1 - \frac{\min[(n_{z+1} - n_z), (n_{z+2} - n_{z+1})]}{\max[(n_{z+1} - n_z), (n_{z+2} - n_{z+1})]} \tag{4.18}$$

where n_{z+1} and n_{z+2} correspond to the next and the next-next candidates under consideration. Hence, the total cost of going from candidate i to candidate j is given by $a_{ij} = C_{sd} + C_{pc}$.

With the transition cost defined, a path through the lattice starting at point c_1 as shown in Figure 4.4, can now be found. The process entails computing the transition costs a_{1j} to each candidate in the second region (j = 5-8) and selecting the candidate with the smallest cost. This now acts as the starting point for the next search and the same process is continued until a candidate is chosen from the last region. Tracing back the choices made in every region gives the chosen path starting from c_1. This is indicated by the bold line in Figure 4.4. Following the same procedure a minimum cost path is obtained for every possible starting point in the first region. In the end, the overall best path is selected as the one that has the lowest total accumulated transition cost.

Figure 4.5 shows a snapshot of an example signal used to test the pruning algorithm. Magenta squared crosses indicate the candidates selected by the algorithm as the ICI estimates and the black ones are the LCI estimates. Clearly, a large number of candidates are made redundant during the region formation step due to the threshold and the analysis frame-size used. This greatly reduces the size of the lattice from which the set of candidates that satisfy the constraints is found. Given the lattice, the algorithm accurately selects

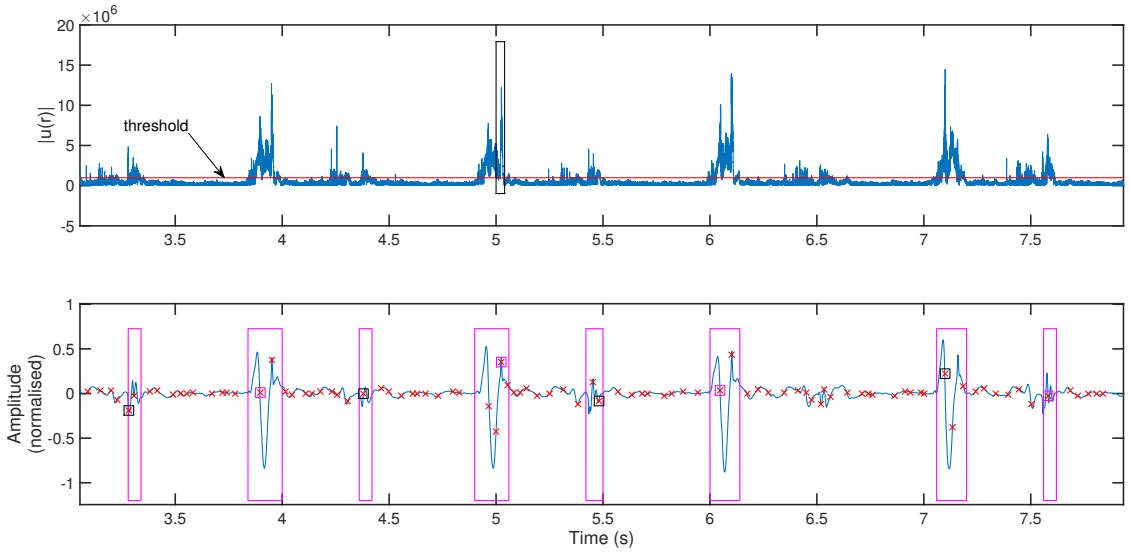

Figure 4.5: Result of forming groups and pruning algorithm. (a) Amplitude envelope of the linear prediction residual signal, $u(r)$, with analysis frame (black rectangle), (b) knee acoustic emission signal with formed regions (magenta squares), candidates (red crosses) and selections made by the algorithm (squared crosses).

one candidate per region and correctly alternates between ICIs and LCIs. However, at approximately 6.5 s it fails to identify a region of high sample energy and consequently the candidate that likely corresponds to a true CI in that region is discarded. However, this can be avoided by appropriately tuning the analysis frame-size and the amplitude threshold value.

4.3.3 Limitations

The line of approach described in the previous two Sections is limited in some important points. First of all, using many different constraints and thresholds that need to be tuned makes it harder to generalise. A set of hard thresholds that work on some signals would likely need some alteration in order to work on others. Therefore, any algorithm that uses predefined thresholds is likely to be less robust and adaptable than another one that does not depend on such thresholds. Moreover, it is difficult to explain why specific threshold values work for some signals but for others different values work better. A more preferable solution is one that can be applied to any possible signal without having to greatly modify it and is as much automated as possible.

For the example shown in Figure 4.5, the algorithm works almost perfectly in identifying two regions in almost every stride. However, there are cases in which the algorithm breaks down. For example, some signals have large variability in the sample amplitude which results in regions of continuously large amplitudes in $|u(r)|$. Therefore, the chosen threshold ($^2/_3$ of the samples in the window) fails to produce two distinct regions per stride. In some other cases however, more than two regions per stride are obtained because the amplitude variability is in shorter durations thus requiring a different threshold value and analysis frame-size. Therefore, the algorithm breaks down again since it is not developed in a way to accommodate the change in the number of groups formed.

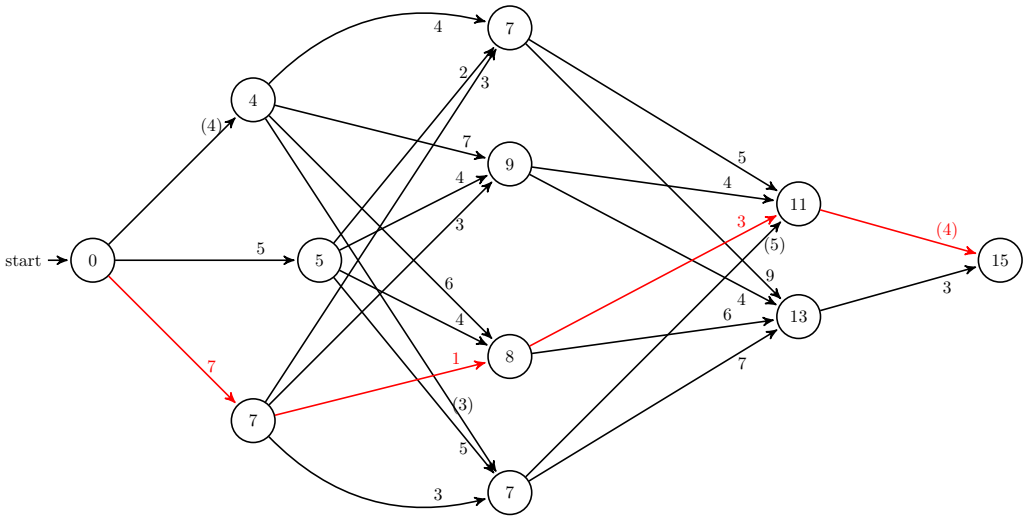

Figure 4.6: Classical example of finding the cheapest path.

Therefore, an algorithm that avoids the use of hard thresholds and automatically estimates the CIs using the sound signal alone is needed. In this context, it is better to delay making any decisions until the end of the process because it might not lead to the optimal solution. To explain this statement, the classical example of finding the cheapest path from the starting node to the rightmost node as illustrated in Figure 4.6, is useful. The number above the arrows indicates the transition cost and the numbers inside the circles indicate the cheapest accumulated costs of the path from the starting point and up to that node found by DP [171]. If the principle used in Figure 4.4 is applied to this example and at each step the lowest transition cost is chosen (indicated by the numbers

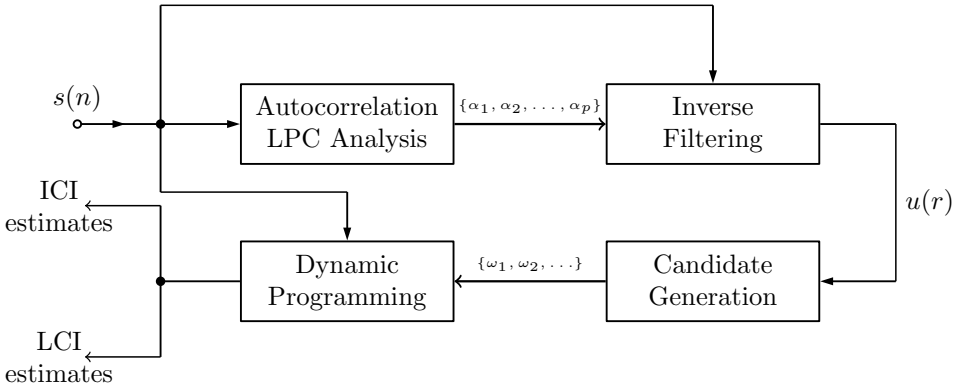

Figure 4.7: Block diagram of the proposed Automatic Stride Detection and Segmentation Algorithm (ASDSA).

inside the parenthesis) it will lead to a path with total cost 16. Selecting at each step the node that gives the lowest transition cost does not necessarily give the optimal path. There are efficient algorithms for these kind of problems that guarantee finding the optimal path (indicated by the red arrows) [172].

4.4 Automatic detection and segmentation using dynamic programming

The investigation in the previous Section led to the development of the Automatic Stride Detection and Segmentation Algorithm (ASDSA) that overcomes the limitations previously described. ASDSA is an automatic technique for estimating the CIs from the knee sound signal alone, for a knee that performs an OCA. The proposed algorithm operates in two parts. In the first part a set of candidate instances is generated using the NZC of the energy weighted GD (4.11). The second part employs DP to select the true instants of first and last contact from the generated set by minimising the cost function defined in (4.20). A block diagram showing the process of the overall system employing ASDSA is shown in Fig. 4.7. In this section the Candidate Generation and the Dynamic Programming blocks will be discussed as the first two blocks have already been described in Section 4.2.2.

4.4.1 Candidate generation

The energy-weighted formulation of the GD defined in (4.11) is used to generate a set of candidates from which the final set of estimates will be selected. By applying (4.11) on $u(r)$, the set of CI candidates is constructed and is defined as

$$\mathbf{\Omega} = \{\omega_1, \omega_2, \ldots, \omega_N\} \tag{4.19}$$

where each element is obtained as the sample location of each NZC of the energy weighted GD waveform for N such zero crossings.

It has been shown in [161] that the choice of the group delay window size N_w is a compromise. In the context of gait analysis, if N_w spans the entire stride cycle then there would be a single NZC per stride corresponding to the strongest excitation event in the signal. In this case, inaccurate detection occurs since the highest peak in any stride, as represented in either $u(r)$ or $s(n)$, does not always correspond to only the ICI or the LCI; it can be both, even for the same patient. On the contrary, if N_w is larger than one stride cycle, then more than one event may fall within the window, resulting to an NZC at the wrong sample position. Detection of both ICIs and LCIs requires a smaller window but when it is much smaller than the stride length it is likely that some windows will contain no impulsive events, giving rise to spurious NZC. However, a large number of CI candidates will not be problematic if there exists an algorithm that can penalise these spurious candidates and select the more suitable ones. This is achieved by defining a cost function for a DP based algorithm that can automatically identify unsuitable and spurious candidates by intentionally assigning a high cost to them. This is discussed in the following Section. In addition, experiments for finding a suitable value for N_w and the LPC order p are presented in Section 4.5.4.

4.4.2 Candidate selection using dynamic programming

Given the set of CIs ($\mathbf{\Omega}$), the next task is to retain only those which, taken together, best fit the constraints of the system and the prior understanding of the structure of the

LP residual signal. Each constraint is incorporated by assigning an associated cost to each candidate. The factors used in the construction of the cost function in (4.20) are based on the attributes of the GD method and the gait cycle characteristics together with the quasi-periodic behaviour of the knee sound signal. The sets $\boldsymbol{\Xi}_I$ and $\boldsymbol{\Xi}_L$ are subsets of ICIs and LCIs respectively, selected from $\boldsymbol{\Omega}$ and jointly minimise the total cost which is solved by employing DP [171]. The solution based on DP requires a starting point and in this case it is given by a dummy state where the transition cost to valid states (i.e. candidates) is given by the slope cost in (4.25) as it does not depend on any other state choice. The choice of the end state is discussed in the algorithm's evaluation, Section 4.5.5. The minimization problem is defined as

$$\underset{\{\boldsymbol{\Xi}_I,\boldsymbol{\Xi}_L\}}{\operatorname{argmin}} \sum_{z=1}^{||\boldsymbol{\Xi}_I||} \left[{\boldsymbol{\Theta}_1}^T \boldsymbol{V}_{\boldsymbol{\Xi}_I}(z) + \sum_{r=1}^{||\boldsymbol{\Xi}_L||} \left({\boldsymbol{\Theta}_1}^T \boldsymbol{V}_{\boldsymbol{\Xi}_L}(r) + \theta_{ss} C_{ss}(z,r) \right) \right] . \tag{4.20}$$

$\boldsymbol{\Theta}_1 = [\theta_{sd}, \theta_{pd}, \theta_{er}]^T$ is a vector of weight factors associated with the individual cost components defined later in the text, z and r index the elements of $\boldsymbol{\Xi}_I$ and $\boldsymbol{\Xi}_L$ respectively, $||\boldsymbol{\Xi}_I||$ and $||\boldsymbol{\Xi}_L||$ are the number of ICI and LCI selections respectively, $C_{ss}(z,r)$ is the swing-to-stance ratio cost evaluated for the z-th ICI and the r-th LCI and is defined in (4.32) with its associated weight factor θ_{ss} and

$$\boldsymbol{V}_{\boldsymbol{\Xi}_I}(z) = [C_{sd}(z), C_{pd}(z), C_{er}(z)]^T \tag{4.21}$$

$$\boldsymbol{V}_{\boldsymbol{\Xi}_L}(r) = [C_{sd}(r), C_{pd}(r), C_{er}(r)]^T \tag{4.22}$$

are the cost vectors evaluated for the ICI with index z in the set $\boldsymbol{\Xi}_I$ and for the LCI with index r in the set $\boldsymbol{\Xi}_L$ respectively. The overall vector of costs is denoted as

$$\boldsymbol{C}(z,r) = \left[C_{sd}(z/r), C_{pd}(z/r), C_{er}(z/r), C_{ss}(z,r) \right] \tag{4.23}$$

where (z/r) denotes that either z or r is used and the explicit dependency of $\boldsymbol{C}(z,r)$ on $\boldsymbol{\Xi}_I$ and $\boldsymbol{\Xi}_L$ is omitted for clarity of notation. Similarly, the vector containing all four weight

factors is defined as

$$\boldsymbol{\Theta} = [\theta_{sd}, \theta_{pd}, \theta_{er}, \theta_{ss}] \ . \tag{4.24}$$

Additionally, the sample instances q_{z-1}, q_z and q_{z+1} are defined for the ICI candidates $z-1, z$ and $z+1$ respectively which depend on $\boldsymbol{\Xi}_I$. Similarly, q_{r-1}, q_r and q_{r+1} are defined for the respective LCI candidates which depend on $\boldsymbol{\Xi}_L$. The composing cost terms have a range of values from 0 to 1 and are described in the equations to follow using only the set $\boldsymbol{V}_{\boldsymbol{\Xi}_I}$ as they are the same for $\boldsymbol{V}_{\boldsymbol{\Xi}_I}$ except the swing-to-stance ratio cost which is dependent on candidates from both sets.

1) ***Slope deviation cost:*** This cost was presented earlier in Section 4.3.2 but is defined more elaborately here and is put in the context of the DP algorithm. The slope in the vicinity of the NZC in d_{EW} for a true impulse is -1 by definition [161]. The events in $u(r)$ are not true impulses and hence the slope will deviate from unity. However, it was observed that candidates corresponding to true CIs have a slope much closer to -1 than other candidates. This provides a way of discriminating true events from spurious NZC where the slope is nearly flat. The slope deviation cost is thus defined as

$$C_{sd}(z) = 1 + m(q_z) \tag{4.25}$$

where

$$m(q_z) = \frac{1}{l_s}\left[d_{EW}(q_z + \frac{l_s}{2}) - d_{EW}(q_z - \frac{l_s}{2})\right] \tag{4.26}$$

is the slope in a window of length l_s, in samples, and centred at q_z. By manually checking the d_{EW} waveform near the locations of the reference CIs, derived as described in Section 4.5.2, it was observed that the nearby NZC are more sparse compared to other locations in the stride. Therefore, to reliably compute (4.26) for the former, more important, cases l_s needs to be sufficiently large. The l_s value also depends on the choice for the energy-weighted GD window size, N_w, which is a more important hyperparameter than l_s and is tuned later in Section 4.5.4. Given N_w and that $F_s = 16$ kHz, $l_s = 0.003F_s$ was found to be good choice and one which does not cause overlapping between neighbouring candidates ($0.004F_s$ and $0.005F_s$ were also examined).

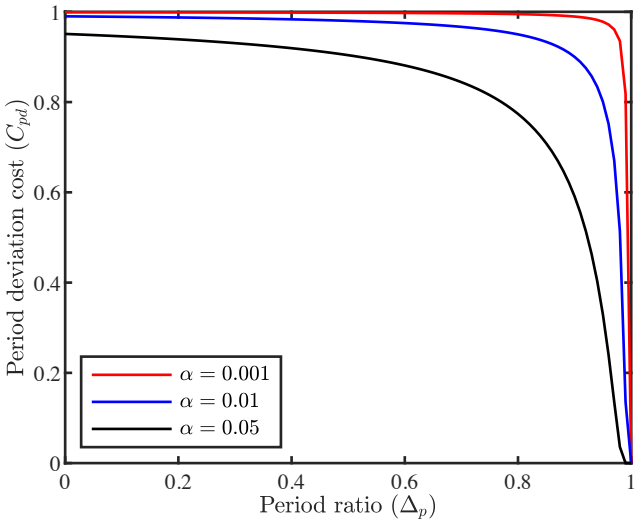

Figure 4.8: Stride period deviation cost as a function of the period ratio Δ_p for different α values and $\epsilon = 0.995$.

2) ***Stride period deviation cost:*** It is based on the assumption of relatively constant stride period and is a function of the previous, the current and the next ICI candidates under consideration by the DP algorithm. It is defined as

$$C_{pd}(z) = \begin{cases} \exp\left(-\frac{\alpha}{(\epsilon - \Delta_p)}\right) & \text{if } \Delta_p < \epsilon \\ 0 & \text{if } \Delta_p \geq \epsilon \end{cases} \tag{4.27}$$

with

$$\Delta_p = \frac{\min[(q_z - q_{z-1}), (q_{z+1} - q_z)]}{\max[(q_z - q_{z-1}), (q_{z+1} - q_z)]} \,. \tag{4.28}$$

The cost increases non-linearly with decreasing period ratio Δ_p and the parameter α controls the rate of increase of the cost: the smaller its value is, the faster the increment. Figure 4.8 shows how C_{pd} varies for changing Δ_p and for different α values. For the algorithm evaluation in Section 4.5 α was set to 0.001. Given that small values for N_w are better (see Section 4.5.4) it means that there will be many candidates to consider in the optimization problem of (4.20). The large number of candidates increases the chances of having spurious candidates that appear periodically in the signal with a period similar to the one modelled by equation (4.27). Therefore, by setting $\alpha = 0.001$ the deviations from (4.28) are aggressively penalised. This is compromised by the tolerance factor ϵ in

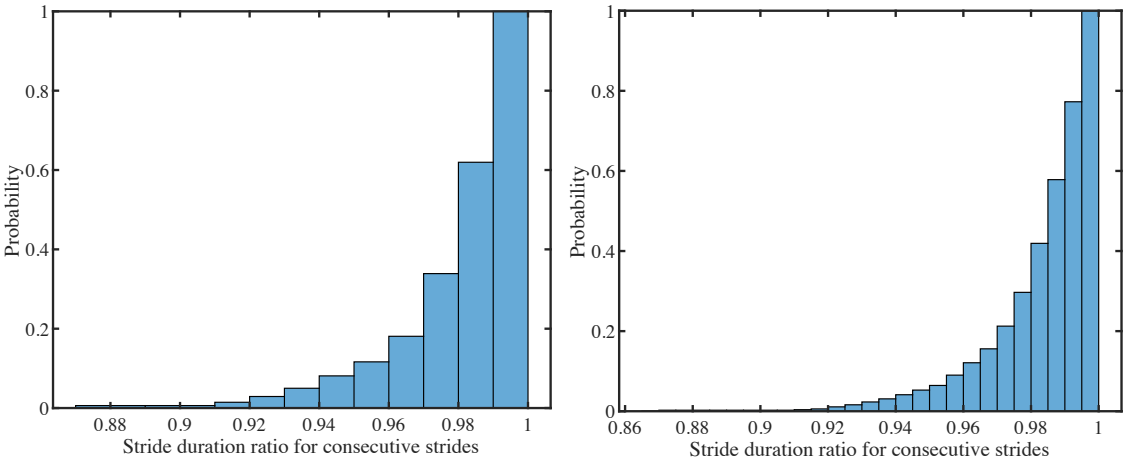

Figure 4.9: Cumulative Distribution of period ratio values of consecutive strides from ground truth data of (a) 36 OA patella recordings (b) 109 patella recordings (OA and healthy)

order to avoid putting a high cost to candidates that are not spurious. ϵ is defined as the variability in stride duration during normal walking.

According to a study by Kaufman et al. [173], OA patients modify their gait pattern so as to protect their knee joint from loading and compensate for the resultant pain. However, the CDF of the period ratio as calculated based on the reference instants from the force plate data of the OA knee patella recordings (Figure 4.9 (a)) shows that the stride period is relatively constant. This validates the assumption made at the beginning and that this cost component can distinguish between periodic candidates (CIs) and candidates that randomly occur within a stride. The value of ϵ is determined from the CDF of all the patella recordings of the database that have ground truth data (see Figure 4.9 (b)) and the choice is made based on the percentage of ground truth strides to be included by ϵ. For the algorithm evaluation, ϵ was set to 0.995, that is, 95% of the ground truth strides are included under this threshold value. The extraction of the reference strides will be described later in Section 4.5.2.

3) ***Energy ratio cost:*** Let the energy ratio between the current and the previous ICI candidates to be

$$R_{q_z,q_{z-1}} = \frac{\min[E(q_z), E(q_{z-1})]}{\max[E(q_z), E(q_{z-1})]} \tag{4.29}$$

with

$$E(q_z) = \sum_{n=-\frac{l_e}{2}}^{\frac{l_e}{2}-1} s^2(q_z + n) \ . \tag{4.30}$$

The energy ratio cost is then defined as

$$C_{er}(z) = 1 - p_e(R_{q_z,q_{z-1}}) \tag{4.31}$$

where l_e is the size in samples of a window centred at the candidate and p_e is the normalised ensemble probability density of the energy ratio of consecutive ICIs as well as that of LCIs, trained using ground truth CIs obtained from the force plate signals. The probability density is normalised by the maximum bin value of the histogram in order to assign a minimum cost of 0 to the candidate that has an energy variability similar to the training ground truth data. An example of a pre-normalised p_e is shown in Figure 4.10. This cost effectively means that candidates that do not have consistent energies get penalised as they are likely not to be true CIs. The rationale is that the candidates that fall in a low probability density region are less likely to occur and hence more likely to be false alarms.

The energy ratio cost is motivated by the observations derived from the exploratory analysis performed on the recordings of both healthy and OA knees. It was found that the signal waveform in the vicinity of consecutive ICIs is similar. The same was observed for LCIs. This waveform similarity is aimed to be captured here by the signal energy within a short window centred at the CIs. In the algorithm's evaluation it was found that $l_e = 0.2F_s$ gives good energy consistency i.e. a skewed histogram towards $R_{q_z,q_{z-1}} = 1$, like in Figure 4.10, compared to uniform-like histograms obtained with other l_e values.

The ensemble distribution p_e is used because the energy ratio of consecutive CIs of the same type is not a discriminant factor between ICIs and LCIs. This can be justified by performing a two-sample KST [92]. The test's null hypothesis H_0 states that the probability density formed by the energy ratios between consecutive ground truth ICIs, as defined in (4.29), and the equivalently defined probability density formed by the ground truth LCIs, originate from the same continuous distribution. At 5% significance level H_0 was accepted, implying that there are no significant statistical differences between the two

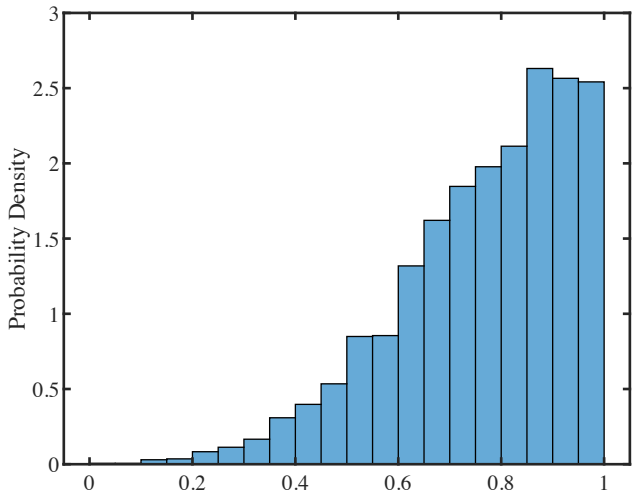

Figure 4.10: Histogram of the signal energy ratios for consecutive ICIs and LCIs obtained from 109 patella recordings.

distributions to reject H_0 and therefore the choice of combining the energy ratios into a single probability density is validated.

4) ***Swing-to-stance ratio cost:*** It is a function of the previous ICI candidate and the current ICI and LCI candidates. It is defined as

$$C_{ss}(z,r) = 1 - p_{ss}\Big(\frac{q_z - q_r}{q_r - q_{z-1}}\Big) \tag{4.32}$$

where p_{ss} is the probability density of the swing to stance distribution trained using ground truth CIs and is normalised by the maximum bin value of the histogram as was done for p_e in (4.31). An example of p_{ss} without normalisation is shown in Figure 4.11. This cost penalises the initial and last CI pairs that are not adequately separated, based on ground truth data. In the algorithm's implementation, the pairs which gave a swing to stance ratio over 1 were given a C_{ss} of 1 as this indicates improper pairs.

The solution to the minimization problem of (4.20) is solved efficiently and fast using DP. However, when the size of the candidate set increases, the number of computations required by the algorithm also increase which in turn results to longer computation times. The knowledge of the stride period can be exploited to avoid processing all combinations by limiting the range of possible sample positions for the previous and next candidates for

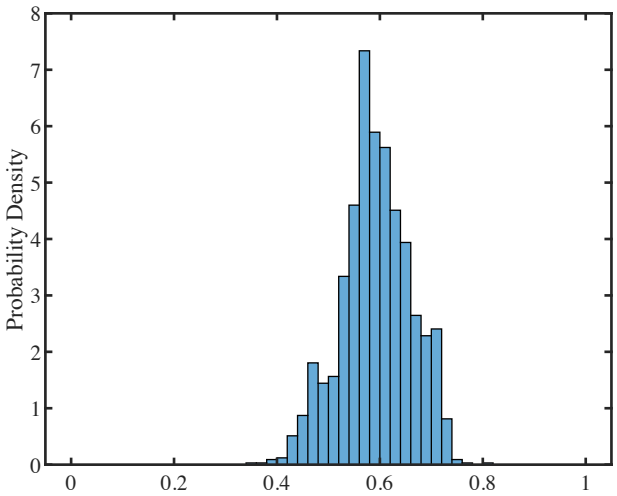

Figure 4.11: Probability density of the swing to stance ratios for consecutive strides, obtained from 109 patella recordings.

each q_z and q_r to a smaller set. Given that the CIs are cyclic points, this can be achieved by introducing two new parameters, namely, T_{min} and T_{max} denoting the edges of the search interval and expressed in number of samples. These are defined as

$$\begin{aligned} T_{min} &= \left(\hat{N}_s - w\hat{N}_s\right)F_s \\ T_{max} &= \left(\hat{N}_s + w\hat{N}_s\right)F_s \end{aligned} \tag{4.33}$$

where F_s is the sampling frequency, $\hat{N}_s$ is obtained as in (4.15) and w is a weight factor which is used to ensure that the true candidate is within the range under consideration. In the evaluation, $w = 1/4$ was found to be sufficient when compared to ground truth data, provided that $\hat{N}_s$ is a true estimate of the stride period. The search intervals for q_z are then defined as $[q_z - T_{min}, q_z - T_{max}]$ and $[q_z + T_{min}, q_z + T_{max}]$. Figure 4.12 shows a snapshot of a force plate signal with the sample locations of the candidate set. It is clear that with the introduction of T_{min} and T_{max} the search space is constrained to a more meaningful range, indicated by the double headed arrows, and therefore the computation time is greatly reduced.

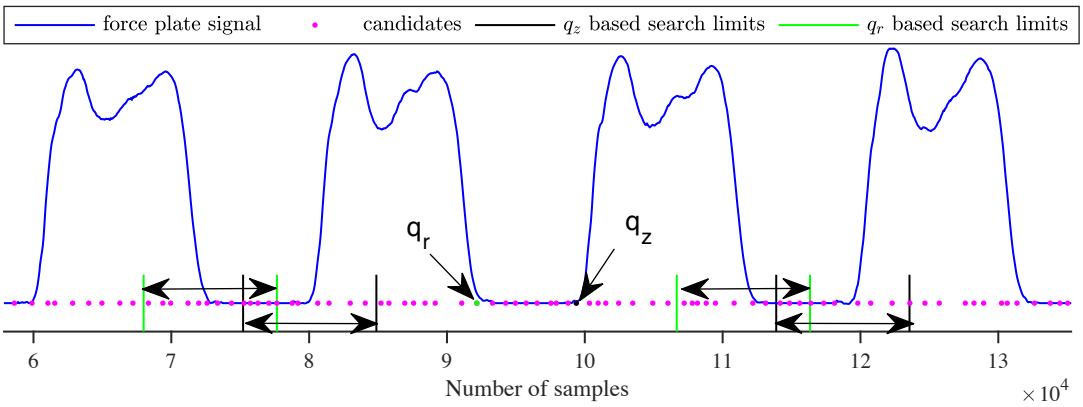

Figure 4.12: Force plate signal with CI candidates and the identified search windows for the previous and next viable candidates for each of q_z and q_r.

4.5 Evaluation

4.5.1 Experimental setup

In the experiments to follow, a database of 109 patella recordings was used, obtained from 11 healthy, 7 OA and 2 OAH knees with 63, 36 and 10 recordings respectively. This is the data for which temporal ground truth was available and could be extracted. Sound data obtained at various walking speeds and treadmill inclination levels, both positive and negative, was included. Table 4.1 shows the details of the knee database at the time of data collection. Even though more recordings were available for some knees it wasn't possible to include them in the analysis because of unsuitable force plate signals and/or due to clipping in the audio signals. For this reason and given that the walking speeds and inclination levels are patient dependent as seen in Chapter 2, there are more recordings available for some knees over others. However, this does not affect the analysis and the algorithm's performance evaluation.

In Section 4.5.4 the optimization of the system's hyperparameters is performed using a small training subset of the database, comprised of 2 knees, 1 healthy and 1 OA having 11 and 8 recordings respectively resulting to approximately 17.4% of the database. In number of CIs, it is 333 ICIs and 333 LCIs. The training data was subsequently excluded from all testing of ASDSA, whose evaluation was conducted on the remaining 18

Gender	Age	BMI[1]	Footware[2]	Knee/condition	Number of recordings
Male	28.3	22.7	1	right/healthy	5
Male	34.3	24.2	1	right/healthy	3
Male	29.7	27.6	1	right/healthy	6
Male	21.3	22.2	1	right/healthy	5
Female	30.2	19.2	1	right/healthy	7
Male	35.9	22.8	1	right/healthy	5
Female	33.4	20.9	1	right/healthy	6
Male	36.3	25.1	1	right/healthy	7
Female	25.4	22.1	1	right/healthy	6
Male	27.6	23.3	1	right/healthy	2
Male	33.9	24.4	2	right/healthy	11
Female	59.7	21.7	2	right/OA	12
Female	48.4	21.0	1	both/OA	10
Male	80.4	25.0	2	right/OA	6
Female	76.0	31.1	3	both/OA	4
Female	68.0	36.6	2	right/OA	4
Male	80.4	25.0	2	left/OAH	5
Male	66.1	37.1	4	left/OAH	5

[1]Body Mass Index

[2]Footware: 1=no shoes/socks, 2=sports trainers, 3=leather boots, 4=flat leather shoes

Table 4.1: Database with details at the time of collection.

knees (90 recordings). The probability densities in (4.31) and (4.32) were obtained from the training data. For the test data-set, the number of CIs is 1444 ICIs and 1444 LCIs, identified from the force plate signals. The data was split in this way in order to have more data available for the algorithm evaluation to support the results on its performance as well as the findings on the gait differences between OA and healthy patients.

4.5.2 Temporal ground truth extraction

The experiments for the algorithm evaluation described throughout Section 4.5 are performed against temporal ground truth data obtained from the synchronised force plate signals. For more details on the acquisition and nature of these signals the reader may refer to Chapter 2. Here, the method for extracting the reference CIs is described.

There is no gold standard for determining the precise instants of first and last contact of the foot to the floor from the force plate signal [174]. The definition of an

instant of contact varies in the literature. A CI can be considered to be the first sample at which the vertical component of the GRF (Figure 4.13) exceeds or drops below a threshold that is commonly specified at 5 and 10 Newtons, [175], or 20 Newtons, [132, 176], or even the first non-zero sample [122, 129]. In this work the reference instants are defined using a threshold of 10 Newtons. Choosing a different threshold value would impact the detection time error ($\hat{\tau}_\delta$ defined in equation 4.37). The other evaluation metrics (see Section 4.5.3) would slightly change. The difference would be due to the possible selections made by the DP algorithm that fall near the boundaries of the regions defined in Figure 4.16.

Occasionally, the signals generated due to walking on the force plates were faulty or the collection of these signals was not suitable, producing therefore corrupted data as in Figure 4.13 (b) and (c). In order to avoid extracting the instants of incorrect reference strides, a threshold is imposed on the stance to stride ratio where both the ICI and LCI of a stride are rejected if this ratio is outside the range $[0.55 - 0.8]$. These limits are chosen based on previous studies done in [177, 178] and after experimental validation on the data at hand. The set of reference ICIs and LCIs obtained as above are defined as $\mathbf{\Gamma}_I = [\gamma_1, \gamma_2, \ldots, \gamma_M]$ and $\mathbf{\Gamma}_L = [\gamma_1, \gamma_2, \ldots, \gamma_M]$ respectively, for M strides with γ_j denoting the sample position of a reference CI in stride cycle j.

In the experiments to follow, an estimated CI denoted as ζ, from either $\mathbf{\Xi}_I$ or $\mathbf{\Xi}_L$ set, is assigned to the j-th stride if it lies in the interval defined as

$$\frac{(\gamma_{j-1} + \gamma_j)}{2} \leq \zeta < \frac{(\gamma_j + \gamma_{j+1})}{2} \tag{4.34}$$

where $j = 2, 3, \ldots, M-1$ for M strides. For $j = 1$, $j = M$ and in the cases of having discontinuities in the force plate signal, the definition of the stride cycle edges is more involved. Initially all the edges are computed as in (4.34) with

$$\begin{aligned} \gamma_1 &= \gamma_1 - \frac{(\gamma_1 + \gamma_2)}{2} \\ \gamma_m &= \gamma_m + \frac{(\gamma_{m-1} + \gamma_m)}{2}. \end{aligned} \tag{4.35}$$

The difference in number of samples between consecutive edges is obtained. A threshold

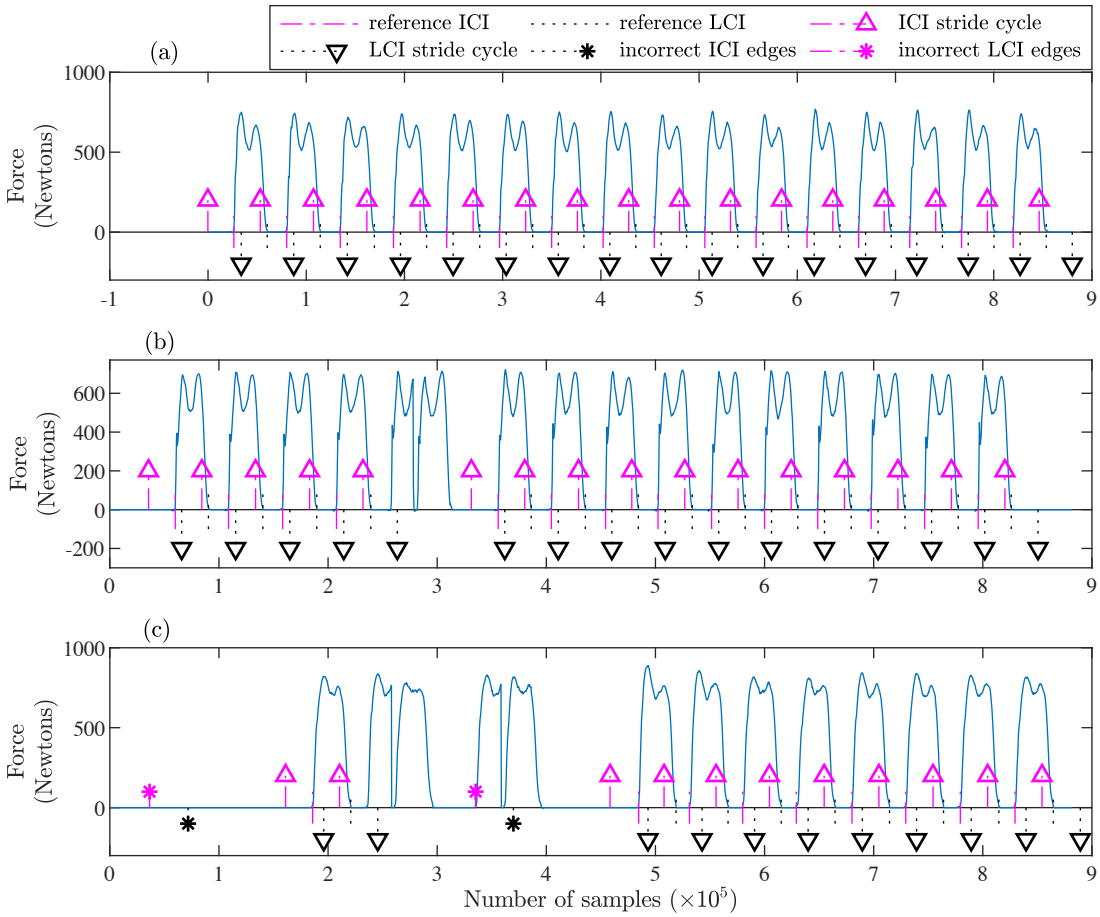

Figure 4.13: Force plate signals with identified contact instants and stride cycle edges.

is then computed as the median value of the collection of all the differences, multiplied by a constant value of 1.25 which was found, after experimentation, to be a good choice. This was done by individually checking for all the signals that had discontinuities. The 25% increase in the median duration is used to capture all the variability in the reference CIs. If the interval size at $j = 1$ and/or $j = M$ is more than the threshold value, then it is replaced by the average value of the intervals which do not exceed the threshold. This is achieved by moving the leftmost or rightmost edge respectively by an appropriate number of samples and recomputing the threshold value (to get a better estimate of the median). A similar process is then followed to replace the initial estimates of the interval edges of improper stride cycles without updating the threshold again. Figure 4.13 shows the force plate signals along with the identified CIs and the stride intervals used for evaluating the ASDSA output estimates. The initial incorrect estimates of the stride cycle edges

are shown only in (c) for comparison. As a side note, for cases as in plots (b) and (c), the incorrect stride cycles were found by time aligning the audio signal waveforms of the strides using the reference CIs and identified as those that do not match. This works since the majority of the strides were correct as in the plots shown in Figure 4.13 (b) and (c).

The identified instants in the sets $\mathbf{\Gamma}_I$ and $\mathbf{\Gamma}_L$ can be used to obtain the stance and swing durations of each stride as well as the true stride period N_s of the associated knee sound signal. An experiment is performed to measure the accuracy of the period estimation approach (Section 4.3.1) against N_s as well as to test the effect of the size l_w of the window used to obtain the signal segment $s_w(n)$ in (4.6) which is then used to estimate the period from (4.13) and (4.15). The window is then slid across the entire signal $s(n)$ and the value for the amount of overlap is also tested. The root mean square error $RMSE$ is used to measure the accuracy of the period estimate and is computed, for N signals, as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}({N_s}^i - \hat{N}_s^i)^2} \tag{4.36}$$

where ${N_s}^i$ and $\hat{N}_s^i$ are the true and the estimated stride period of the signal with index i. Only the data for which $s(n)$ is time synchronised with the force plate signal is included in this experiment. The range of window sizes tested is $[1.5, 1.6, \ldots, 3.5]$, in seconds, with step sizes of $[0, 10, \ldots, 90]$, as a percentage of l_w. The results are shown in Figure 4.14 where a region of low error values can be clearly seen for window sizes over 2.1 s. For smaller windows the error is higher because less than two complete stride cycles fall within the window and therefore the autocorrelation approach produces inaccurate results. This effect is also clearly seen in Figure 4.14 where the minimum error obtained per window is shown. The minimum error occurs at an $RMSE$ of 4.3 ms with $l_w = 2.4$ s and 90% overlap i.e. sliding window step size of 0.24 s. The effect of the amount of overlap is illustrated in Figure 4.15. The small error indicates the high accuracy of the period estimation approach. $\hat{N}_s$ affects the size of the DP search window defined by the pair of equations (4.33) in which the value chosen for the weight factor w suffices to correct for this error as it is small enough.

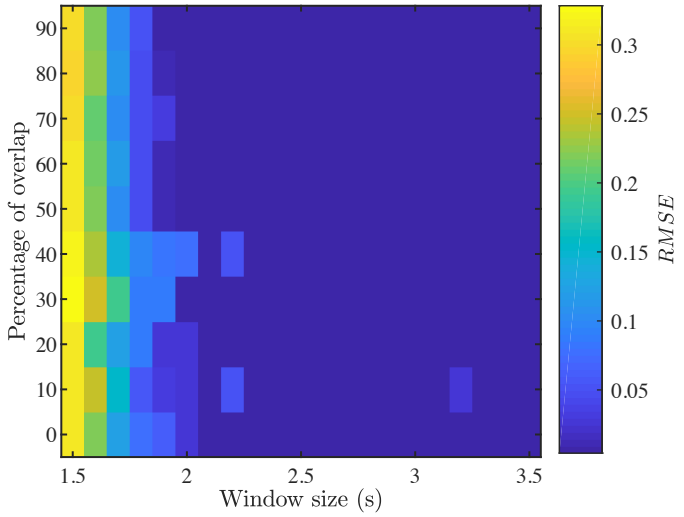

Figure 4.14: Root mean square error (in seconds) between the true and estimated stride rates for 109 recordings, each of 20 seconds duration.

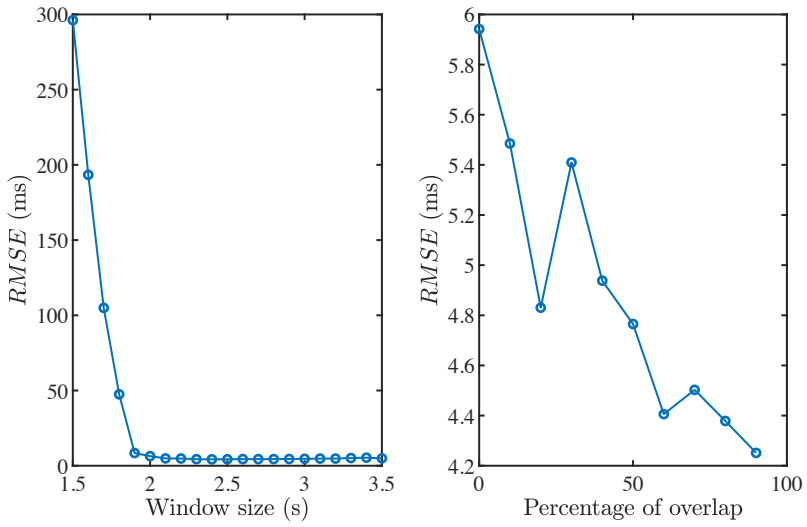

Figure 4.15: (a) Minimum root mean square error per window (b) Effect of the window overlap amount for $l_w = 2.4\ s$ (slice through the vertical axis of Figure 4.14 at $l_w = 2.4\ s$).

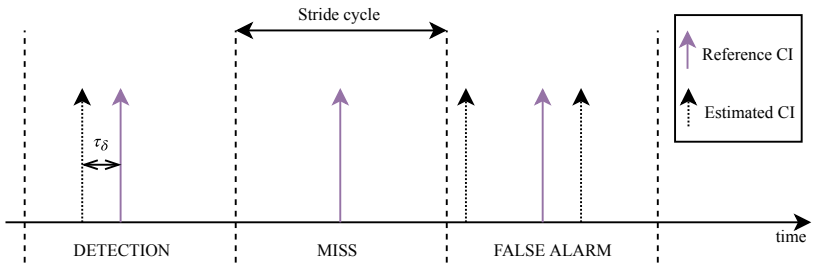

Figure 4.16: Metrics used for performance evaluation.

4.5.3 Evaluation metrics

The algorithm's performance is assessed based on five metrics, illustrated in Figure 4.16. These are the 1) Detection Rate (DR) - percentage of strides where exactly one ICI is detected, 2) Miss Rate (MR) - percentage of strides where no ICI is detected, 3) False Alarm Rate (FAR) - percentage of strides where more than one ICI is detected, 4) detection time error $\hat{\tau}_\delta$ computed only in cycles for which exactly one ICI is detected where for stride cycle j it is given by

$$\hat{\tau}_\delta(j) = \frac{\tau_\delta(j)100\%}{lj} \tag{4.37}$$

where $\tau_\delta(j)$ is the difference in number of samples between the true and the estimated ICI in stride cycle j with length l_j samples, and 5) detection accuracy ϱ - standard deviation of the collection of $\hat{\tau}_\delta$ values in a single recording. Similar reasoning applies for LCIs. The 100% is used in (4.37) to bring $\hat{\tau}_\delta$ and ϱ at the same scale as DR, MR and FAR. Different scaling can also be used but reporting the time error as a percentage of the stride duration helps in understanding the size of the error.

DR measures the capability of the algorithm at rejecting unsuitable candidates and selecting exactly 1 ICI and 1 LCI from the large pool of candidates. The detection accuracy metric ϱ on the other hand measures the algorithm's capability of selecting candidates that have consistent time errors with respect to ground truth. Small values of ϱ indicate high accuracy of detection. The average value of τ_δ represents a constant time offset that can be easily taken into account and subsequently corrected. These metrics are used in the analysis and result interpretation but for comparison it is easier if it's based on a single

metric value. Therefore, by combining DR and ϱ, a new evaluation score is defined as

$$D_{ss} = \frac{D_{ss}^{ici} + D_{ss}^{lci}}{2} \tag{4.38}$$

with

$$D_{ss}^{ici} = \Big(\frac{\mathrm{DR}}{100}\Big)^{\beta}\Big(1 - \frac{\varrho}{25}\Big)^{1-\beta} \tag{4.39}$$

$$\varrho = \begin{cases} \varrho & \text{if } \varrho \leq 25 \\ 25 & \text{if } \varrho > 25\ . \end{cases} \tag{4.40}$$

D_{ss}^{ici} denotes the evaluation score of the ICI estimates and $D_{ss}^{lci} \equiv D_{ss}^{ici}$ where the DR and ϱ values used in this case are obtained from the set of LCI estimates. The parameter $\beta \in [0, 1]$, controls the weighting of each metric in the final score, with larger values favouring DR and smaller values favouring ϱ. Depending on the application, β can be set appropriately. However, it is recommended that the two extreme values 0 and 1 are avoided as the results will be harder to interpret. Equation (4.39) is constructed in such a way so as to make the evaluation score dependable on both metrics. These metrics were chosen because they incorporate all the important performance aspects (detection and timing). The metric ϱ is chosen to be limited to a maximum of 25 and given that it is always positive, the range of possible values for $D_{ss}^{ici}, D_{ss}^{lci}$ and equivalently D_{ss}, is thus bounded to $[0, 1]$ where the higher the value is, the better the sets of CI estimates are, with $D_{ss} = 1$ denoting the highest achievable score obtained when DR $= 100\%$ and $\varrho = 0$. This can be seen in the examples of Figure 4.17 where the β value is varied and the D_{ss} is computed. In all the experiments that follow β was set to 0.5.

4.5.4 Hyperparameter optimization

In this Section the optimization procedures for choosing the hyperparameter values of the segmentation system are discussed and the results are presented. Appropriate tuning of the hyperparameters will have a significant positive impact on the performance of the overall system. The parameters that have a greater effect on the performance are the LPC order p and the GD window size N_w. These affect the set of generated candidates from

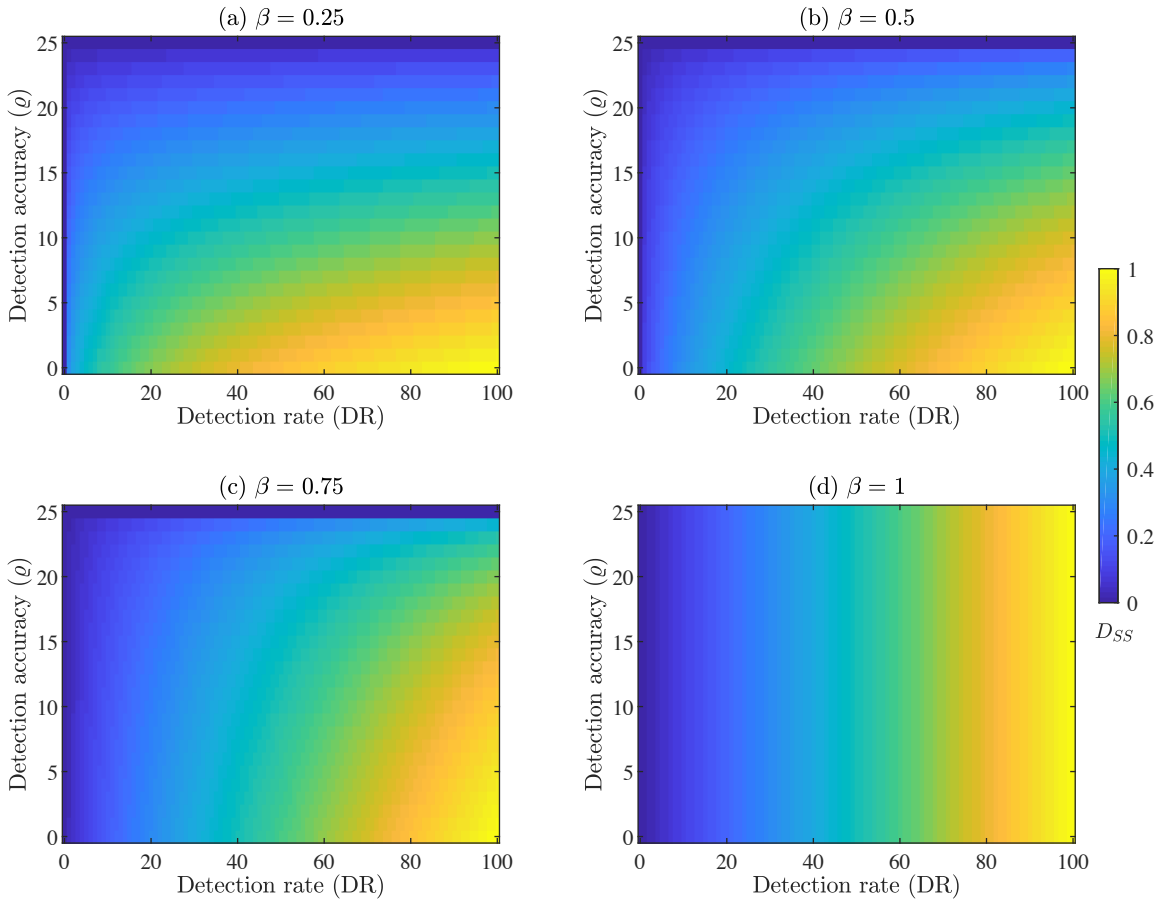

Figure 4.17: Distribution of D_{ss} scores given the detection rate and accuracy for $\beta = 0.25, 0.5, 0.75, 1$.

which the final choice is made. In addition to these, the weight vector $\boldsymbol{\Theta}$ is also tuned and the training procedure is discussed in a separate subsection. The algorithm parameter values used during training and evaluation were discussed in Section 4.4.2 and are now summarised in Table 4.2.

Parameter	Description	Value
F_s	Sampling frequency	16000 Hz
l_s	Slope window length	$0.003F_s$
α	Period deviation control parameter	0.001
ϵ	Stride duration variability tolerance factor	0.995
l_e	Energy window length	$0.2F_s$

Table 4.2: ASDSA parameter values used in the evaluation experiments.

Linear Predictive Coding order and group delay window size selection

As described in Section 4.4, candidate generation is an important building block of the stride detection and segmentation system. Given a set of candidates, it is possible for the algorithm to select a subset that would produce low ϱ and high DR. In this context, the original candidate set is considered to be 'good'. On the other hand, if a 'bad' set is given then no matter how efficient the algorithm is, ϱ will be high. In other words, the best value of ϱ and DR from a set of candidates is dictated by the set itself. Therefore, it is important to evaluate the suitability of the candidate set prior to evaluating the performance of ASDSA. Adding to this, the parameter values affecting the generation of the candidate set $\mathbf{\Omega}$ as discussed in Section 4.4.1 also need to be tuned. These are the LPC order p and the GD window size N_w. In order to objectively choose their values a cost function is defined as

$$C_I = \frac{1}{J}\sum_{j=1}^{J} 4\left(\frac{\tau_{ref_j} - \tau_{cl_j} + N_b}{l_j}\right)^2 \tag{4.41}$$

where j is the stride cycle number in a recording of J cycles long, l_j is the length in samples of stride cycle j extracted from ground truth data as described in Section 4.5.2, N_b is a bias parameter in number of samples, τ_{ref_j} and τ_{cl_j} are the sample instances of the reference CIs, from the set $\mathbf{\Gamma}_I$, and its closest candidate, identified from the set $\mathbf{\Omega}$, in stride j respectively. The bias parameter is used to find out whether there exists an offset associated with the difference between the reference CI and its closest candidate. The fraction inside the parenthesis is usually small so a normal quadratic would produce small cost values that would in return make the parameter choice more difficult. Therefore, multiplying by 4 rapidly increases the cost for small changes in $\tau_{ref_j} - \tau_{cl_j}$ (for fixed N_b).

The cost C_I is used to predict the efficacy of each pair $\{p, N_w\}$ for generating suitable ICI candidates as well as the suitability of the set itself. Similar expression to (4.41), denoted as C_L, can be obtained for LCI candidates by using $\mathbf{\Gamma}_L$ as the set of reference instants. The overall cost for a single $\{p, N_w\}$ combination is then obtained as

$$C_{IL} = \frac{C_I + C_L}{2} \ . \tag{4.42}$$

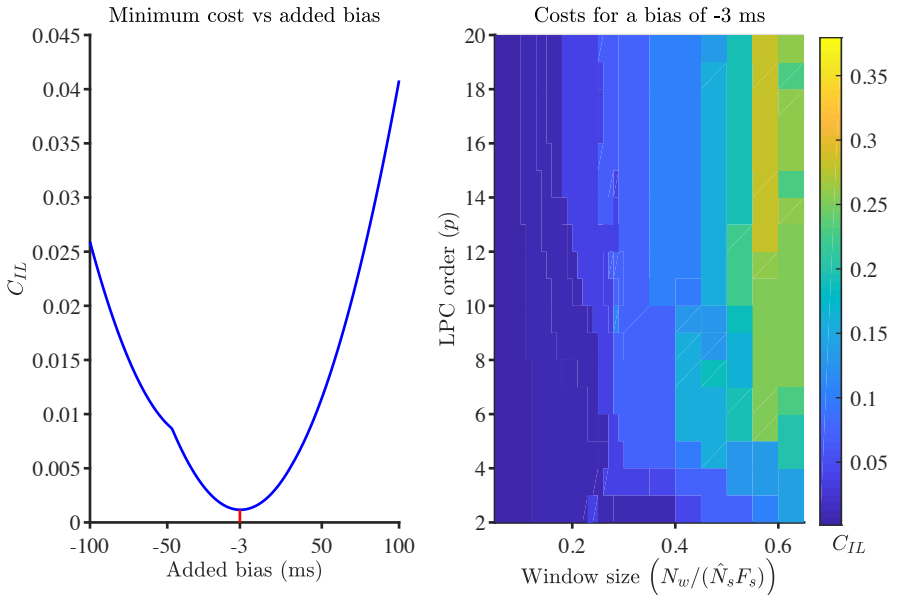

Figure 4.18: (a) Minimum cost per N_b (b) Detailed cost variations of $N_b = -3$ ms for all orders and window sizes.

The experiment was conducted for $N_b = [-100, 100]$ ms with 1 ms steps at $F_s = 16$ kHz sampling frequency for orders $p = 2, 3, \ldots, 20$ and group delay window sizes expressed as fractions of the stride rate, $N_w = [0.05, 0.06, \ldots, 0.3, 0.35, \ldots, 1]\hat{N}_s F_s$. For practical reasons the 0.05 limit was chosen to limit the number of generated candidates since a large number will increase the possibility of finding periodicity between spurious candidates and this would incorrectly result to $C_{pd} \approx 0$ and $C_{ss} \approx 0$. For the corner cases of obtaining a single candidate within the interval defined by (4.34) or a miss, i.e. failure of the energy-weighted group delay method generating a candidate within that interval due to a large window value, a higher cost than (4.41) is applied to that stride. The former is undesirable because the simultaneous detection of both CIs is of interest. The experiment was executed on the training database described in Section 4.5.1 and the results are shown in Figure 4.18. The minimum is found at $N_b = -3$ ms which corresponds to $p = 3$ and window size of 0.05 i.e. $N_w = {}^{\hat{N}_s F_s}/_{20}$. The results for sizes larger than 0.65 are not shown because they generate candidates that are not suitable for detecting both CIs.

Weight vector training

The elements of the weight vector $\mathbf{\Theta}$ represent the contribution to the final cost of each cost function component in (4.23). They are trainable parameters and when defined appropriately they can lead the DP algorithm to select the ICIs and LCIs correctly.

A training subset of the database, obtained as discussed in Section 4.5.1, consists of 19 knee sound signals and was used to determine the vector $\mathbf{\Theta}$ by employing an optimization procedure that searched each weight over the range $(0, 1]$ using a three stage grid search. In the first stage, the search space for each weight was set to $[0.1, 0.3, 0.5, 0.7, 0.9]$ and the algorithm was executed with all possible combinations, resulting to 5^4 executions for each recording. The results were then evaluated using the metrics in Section 4.5.3 and the weight values $\hat{\theta}$ of the combination that gave the highest D_{ss} were retained, forming the vector $\hat{\mathbf{\Theta}}$ which is an initial estimate of $\mathbf{\Theta}$. In each subsequent grid search stage, $\hat{\mathbf{\Theta}}$ is updated with the new values from the weight combination that gave the highest average D_{ss}. The difference in each stage is the search space which is incrementally made narrower and is centred at the previously selected $\hat{\theta}$ values. More specifically, in the second stage the range of values is $[\hat{\theta} - 0.1, \hat{\theta}, \hat{\theta} + 0.1]$ and in the third stage, the range $[\hat{\hat{\theta}} - 0.05, \hat{\hat{\theta}}, \hat{\hat{\theta}} + 0.05]$ is tested from which the final decision is made. $\hat{\hat{\theta}}$ is the second estimate of $\hat{\theta}$.

Since the weights can take any value, a huge computing power would be required to examine all possible weight combinations. The benefit of using this much power to improve the training results would be counteracted by the cost of producing the power as the improvement will not be significant given that the results are already very good as will be shown later. Hence, for practical considerations the search space is divided in these three stages which evidently make the problem tractable. With the training procedure outlined above, $\mathbf{\Theta}$ was determined as

$$\mathbf{\Theta} = [\theta_{sd}, \theta_{pd}, \theta_{er}, \theta_{ss}] = [0.4, 0.2, 0.6, 0.4] \tag{4.43}$$

with average DR and ϱ equal to 98.79% and 11.29 respectively, giving a D_{ss} score of 0.74 obtained using (4.38). For these weight vector values, the variation of the time errors for

ICI and LCI estimates is shown separately in Figure 4.19. The majority of the time errors for both types is concentrated around 0 indicating the high accuracy of detection but for ICIs the variation is slightly larger. Figure 4.20 compares the DR and ϱ values obtained for each weight combination in stage 3 and indicates the values that yielded the highest overall D_{ss} score (red cross). As can be seen from the figure, other values for $\mathbf{\Theta}$ gave better DR or ϱ. Depending on the application, the trade-off parameter β can be set to favour higher ϱ and lower DR or vice versa, and hence select a different set of weights.

4.5.5 Results and discussion

ASDSA was evaluated on a test database with a total of 90 recordings from which 52 are healthy and are obtained from 10 healthy knees, 28 recordings are OA (from 6 OA knees) and 10 are OAH (from 2 OAH knees). As stated in Section 4.5.1, the probability densities in the energy ratio (4.31) and swing-to-stance ratio cost components (4.32) are obtained using the recordings of the training database.

The results, presented in Table 4.3, show that ASDSA achieved an average DR of 96.88% for ICIs and 96.12% for LCIs which means a combined misdetection error of only 101 instants out of the 2888 in total. The slightly lower percentage of detection for LCIs is mainly due to the higher FAR which is 1.52% compared to less than 1% for the ICI detection. MR is the error type that has a stronger impact on the detection percentage as it is always larger than FAR, reported at 2.15% and 2.36% for ICIs and LCIs respectively. Similar performances with respect to ϱ of 15.64% (ICIs) and 15.68% (LCIs). The detection accuracy is expressed as a percentage of the stride duration (normalised to each ground truth stride). The time error bias, on the other hand, obtained as the average of all $\hat{\tau}_\delta$ values for the estimated LCIs is 2.25% (26.5 ms) which is about half that of the ICIs (63.6 ms). In both cases the bias is positive which means that the estimated instants are observed earlier (in time) relative to the reference instants. The small positive bias can also be seen in the histogram plots of the time errors in Figure 4.21. These plots show that the majority of $\hat{\tau}_\delta$, for both CI types, is concentrated around zero. This indicates the consistency of the selections made by ASDSA and their closeness to the ground truth.

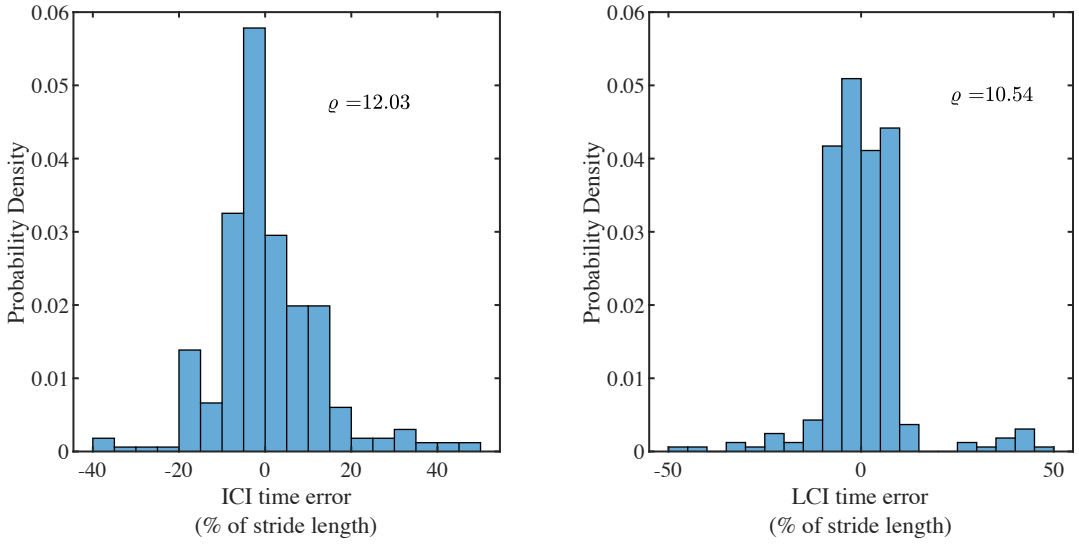

Figure 4.19: Histograms of (a) ICI and (b) LCI time errors, $\hat{\tau}_\delta$, for ASDSA on the training database with the selected weights in (4.43).

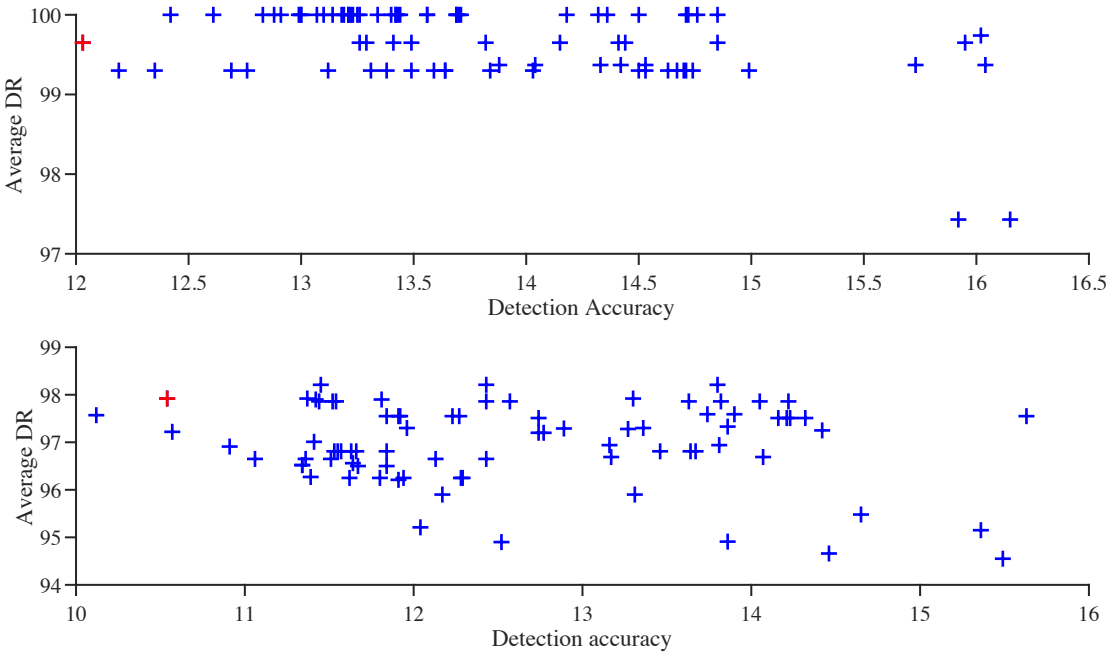

Figure 4.20: Average DR and detection accuracy, ϱ, for each weight combination in stage 3 of the training procedure for ICI (top) and LCI (bottom), for ASDSA on the training database. The red cross indicates the metric values that yielded the highest overall D_{ss} score.

Condition	CI	DR (%)	MR (%)	FAR (%)	ϱ (%)	bias (%)	$D_{ss}^{ici}/D_{ss}^{lci}$	D_{ss}
Healthy	initial	96.26	2.57	1.17	15.93	8.46	0.59	0.60
	last	96.15	2.33	1.52	15.49	3.95	0.61	
OA	initial	97.41	1.65	0.94	15.64	2.37	0.60	0.59
	last	97.18	1.88	0.94	16.71	-0.44	0.57	
OAH	initial	98.77	1.23	0	10.71	1.52	0.75	0.72
	last	93.25	3.68	3.07	12.33	0.36	0.69	
All	initial	96.88	2.15	0.97	15.64	5.86	0.60	0.60
	last	96.12	2.36	1.52	15.68	2.25	0.60	

Table 4.3: Results per knee condition, for ASDSA on the test database.

However, there is a small number of estimates that are as far as half a stride away from the ground truth instants and in consequence increase the bias and ϱ, decreasing therefore the D_{ss} metric score. An overall $D_{ss} = 0.6$ is achieved, with $D_{ss}^{ici} = D_{ss}^{lci} = 0.6$. The test D_{ss} is lower than the corresponding training score which is expected and is mainly due to the larger test ϱ for both ICI and LCI detection. Adding to this, the weight factors of each cost component are optimised for the training signals. The weights can be tuned further by using a larger database which is expected to improve the results as is the case for any algorithm that involves some sort of training.

For the signals that ASDSA generates estimates with large time errors ($> |30|\%$) it was observed that these bad estimates most commonly occur at the beginning of the signal. At this stage, the algorithm is operating with only the slope deviation cost, C_{sd}, out of the four cost function components and therefore, an uncertainty exists in the final selection made by only C_{sd}. This uncertainty arises when there are candidates that, other than those which are in the vicinity of the ground truth instants, generate a slope at the NZC close to the ideal value of -1 giving $C_{sd} \approx 0$. As described at the development of the C_{sd} cost in equation (4.25), these NZC are found from the linear prediction residual $u(r)$ and those that have a slope of -1 are clear impulses. Hence, the existence of more than 2 strong impulses in $u(r)$ in a single stride (relating to ICI and LCI) suggests that there are events which cannot be predicted by LPC. Events such as double heel strike, loud knee clicking or the impact when the foot becomes flat on the ground can cause strong

impulses. The impact of these events occurring at a later stage however, is not as strong as when they occur at the beginning of the signal, since the error is mitigated by the other cost components, making ASDSA therefore a robust algorithm.

The algorithm performs slightly better on healthy knees over OA, achieving D_{ss} scores of 0.60 and 0.59 respectively. Even though DR and D_{ss}^{ici} are higher in the latter case, the difference in the final average score is due to the higher ϱ for the detection of LCIs for OA knees. For OAH recordings, the detection performance is better for ICIs where ϱ is even better than what was obtained during training of $\mathbf{\Theta}$. When combined with the high DR percentage of 98.77%, it gives the highest D_{ss}^{ici} score of 0.75. On the contrary, the LCI's DR is the lowest amongst all knee conditions but still very high (over 93%) and due to the low ϱ, the D_{ss}^{lci} score achieved is the highest.

The difference in the performance of ICI and LCI detection for OAH signals could relate to the fact that OA patients modify their gait pattern so as to protect their OA knee joint from loading, and compensate for the resultant pain, as deduced from several studies [173,179,180]. As a result, more force is exerted on the opposing non-OA knee joint, correlated to the pain felt on the OA knee in each stride which is subjective. Therefore, the emitted sound signal, as picked up by the patella microphone, could at times be more similar to a normal knee, other times more similar to a pathological knee and at other times neither. Hence, the cost function of ASDSA likely needs an additional component to capture the attributes of this type of signals and the similarity in these cases can be measured by cross correlation, frequency spectra, or any other suitable method. However, this is a speculation based on previous clinical findings and observations derived from this study and is not explored within the scope of this current work. Attempting to question or validate this speculation here gives the risk of producing results that do not necessarily hold for other such cases of OAH knees due to the limited amount of OAH recordings with temporal ground truth data in the database of this study. However, OAH knees can safely be used for the case of evaluating the CI estimation performance of ASDSA as was done in these experiments and it is safe to conclude that even with these signals, the algorithm achieves a very high DR (96% on average) with good time accuracy and low bias.

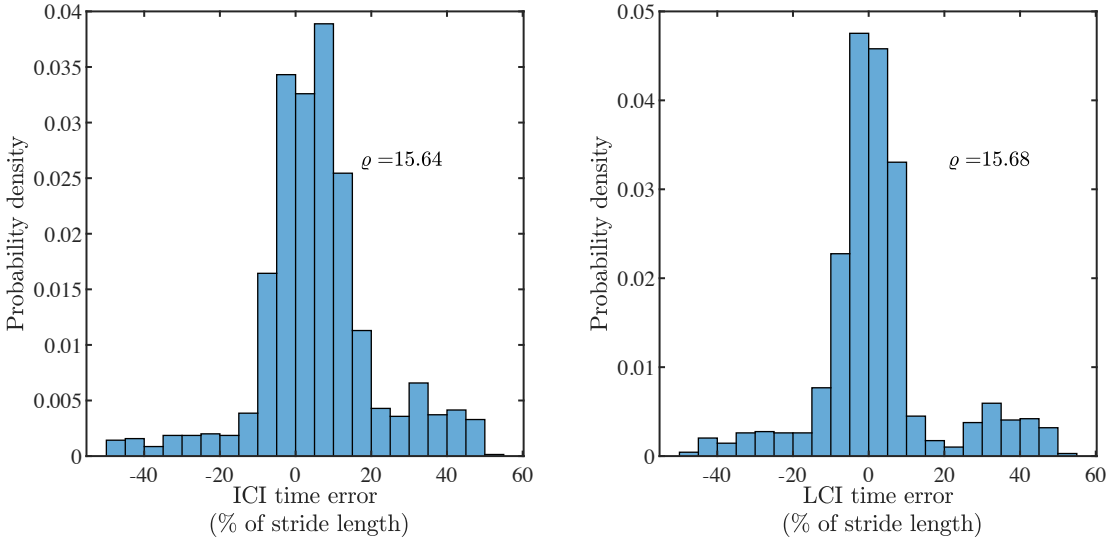

Figure 4.21: Histograms of (a) ICI and (b) LCI time errors, $\hat{\tau}_\delta$, for ASDSA on the test database.

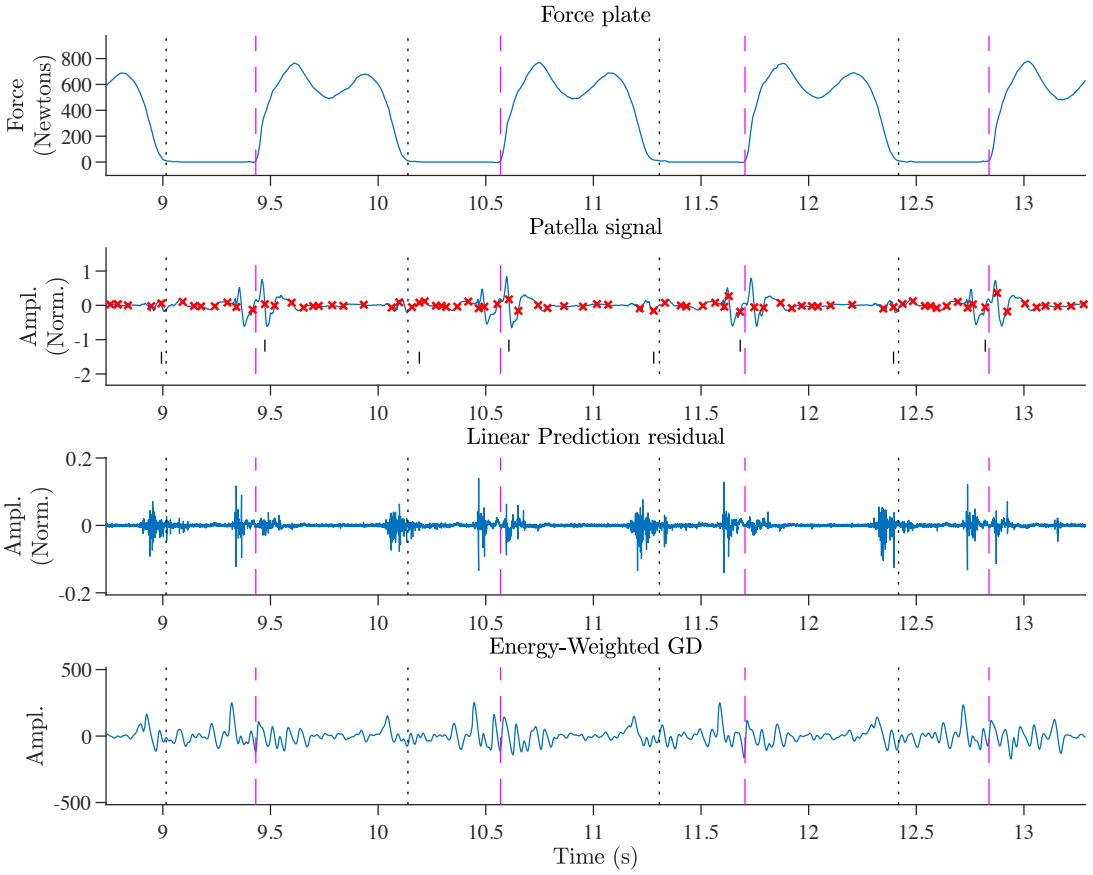

Figure 4.22: (a) Force plate signal with identified ICIs (dashed lines) and LCIs (dotted lines) (b) OA knee patella recording $s(n)$ (normalised amplitude) with candidates (crosses) and reference CIs (c) LP residual signal $u(r)$ with reference CIs (d) Energy-Weighted GD signal with reference CIs. Ticks in (b) correspond to selections made by the DP (ICI estimates on the first line and LCI estimates on the second line).

Figure 4.22 shows an example of the complete operation of ASDSA where the extracted ground truth instants are also shown for visual comparison. It can be clearly seen in this example that the algorithm correctly selects 1 ICI and 1 LCI in each stride and rejects all other candidates based on (4.20). A different patella signal of length 20 s is used to illustrate the ASDSA cost function components in Figure 4.23. The knee sound signal segment is shown in plot (c) in which the upper set of ticks represent the candidate CIs and the lower set of ticks indicate the selections made by DP. Figure 4.23 (a) illustrates the individual value of the 7 cost function components for each candidate and in plot (b) their weighted sum $\boldsymbol{\Theta}^T \boldsymbol{C}(z, r)$ is shown, using the weight values of (4.43). For a given ICI candidate z and an LCI candidate r, the cost function components $C_{sd}(z)$ and $C_{sd}(r)$ can be determined independently. On the contrary, the other cost components are dependent on the particular selection of CIs made by the DP. Therefore, in this illustrative example, the cost values for a single candidate are found using DP as those that gave the minimum weighted cost across all viable selections which include a range of possible previous and next candidates as discussed in the algorithm development. It might be odd however to notice that all 7 cost components are shown for a single candidate since it does not make sense to consider one as both an ICI and an LCI at the same time. This 'paradox' can be explained by the inherent nature of the algorithm that seeks to optimise the selection of an ICI-LCI pair at each DP step as can be clearly deduced from the cost function equation (4.20). This means that at each step a decision is made on the best ICI to add in the path which also indicates its paired LCI that was found from a pool of possible candidates. Therefore, in Figure 4.23 (a) the costs that depend on r are for the already selected LCI for the z^{th} ICI candidate.

It can be seen from Figure 4.23 that, in most cases, the swing-to-stance ratio cost C_{ss} discriminates well between candidates close to the reference ICIs and those that are further away. This is consistent with the high weighting (0.4) of this cost component. Near 19 s, however, its cost of approximately zero indicates an error and the successful rejection of the candidate is achieved by the other cost function components. This error arises when there are two other candidates at a certain distance (in time) away that make the ratio in (4.32) nearly identical to the ground truth. The cause of this error is the large number

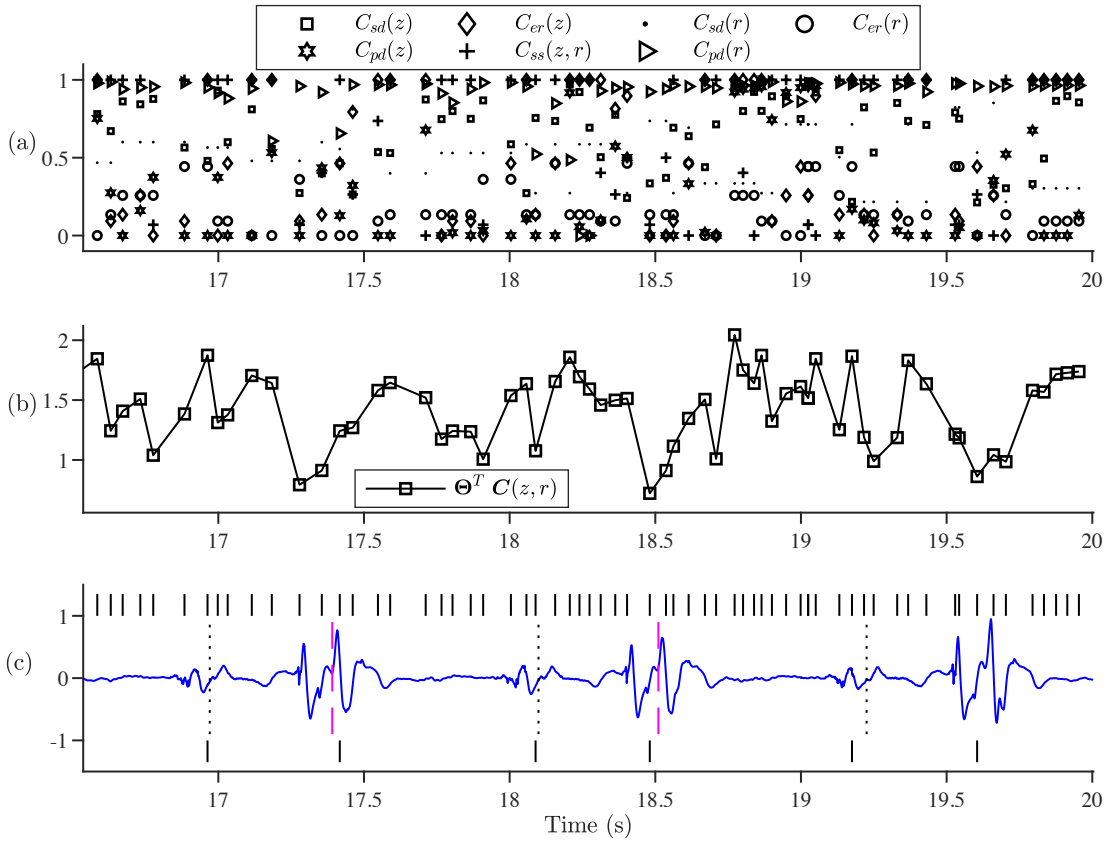

Figure 4.23: ASDSA cost function components. (a) Value of each component (b) Total weighted cost per candidate (c) Knee sound signal segment (normalised amplitude) with reference ICIs (dashed lines) and LCIs (dotted lines), CI candidates (upper ticks) and CIs selected by ASDSA (lower ticks).

of candidates. The component with the highest weighting is the energy ratio cost, C_{er}, and it can be seen that it can correctly discriminate the ICIs from spurious candidates. However, it sometimes falsely penalises the candidates that are near the LCIs, with a cost value similar to the ICIs, as for the candidate at approximately 19.1 s. Nevertheless, the contributions of the other cost function components are sufficient to lead the DP to select the ICI correctly, as in the case discussed above.

At first glance, in Figure 4.23 (c) and around 17.4 s there seems to be a more 'suitable' candidate to select. Suitable in the sense that it is more consistent in time with respect to the other two ICI selections. Indeed, the weighted sum of the time dependent cost components for that candidate $(C_{pd}(z), C_{ss}(z,r))$ is 0.248, which is less than that of the selected one (0.426) and the overall cost of selecting that candidate, when adding the

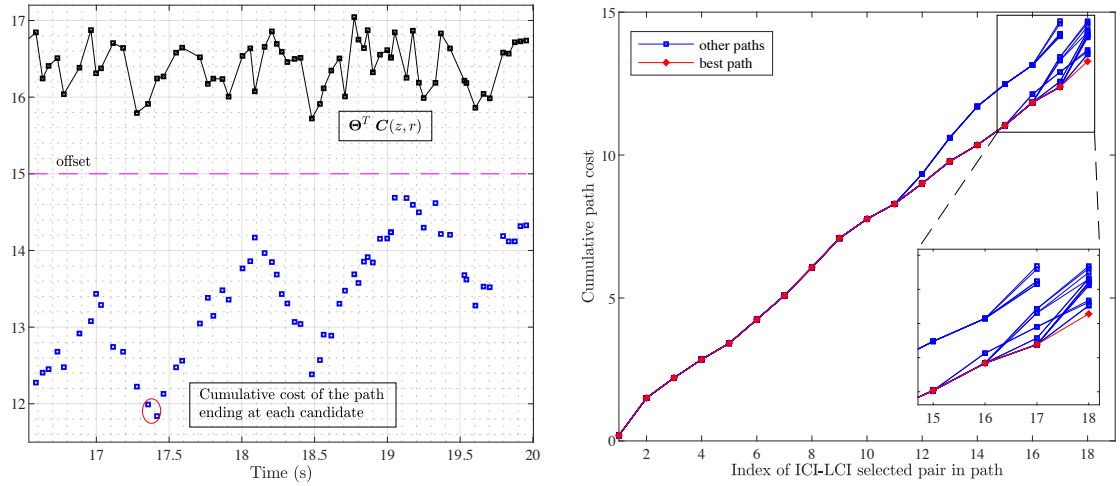

Figure 4.24: Left: Total weighted cost per candidate (top line) at an offset indicated by the dashed magenta line. Cumulative cost of the best path as determined by ASDSA that leads to each candidate (scatter plot). The figure is at the same time index as Figure 4.23. **Right:** All viable paths determined by ASDSA that lead to a selection in the last stride. Smaller plot is a zoomed in version at indices 15–18.

other costs, is still smaller (0.912 compared to 1.243). So a question arises as to why the algorithm did not select that particular candidate.

The left plot of Figure 4.24 answers the question by showing, for each candidate, the cumulative cost of the best path, as determined by DP, up to and including the candidate in question. The two encircled squares indicate the two candidates described in the previous paragraph. Since the plot has the same time index as Figure 4.23 it becomes easy to see that the cost of the path leading to the selected candidate is less than that of the other candidate even though, as seen before, the additional cost due to the candidate itself is smaller. This is another example of the issue discussed in Section 4.3.3 where it was shown that by choosing the best (cheapest in this case) nodes in a path does not necessarily lead to the optimal solution. The right plot of Figure 4.24 shows a comparison of the cumulative costs of all the viable paths that lead to a selection in the last stride from which the final choice is made. A number of paths are considered before selecting the cheapest one due to the fact that the termination point is not known in advance. It is also seen from the same plot that, initially, all the paths have equal cumulative costs which might suggest that the same ICI and LCI candidates are chosen.

4.6 Temporal accuracy improvement with post-processing

ASDSA performs extremely well in selecting one ICI and one LCI candidate per stride as indicated by DR (96.88% and 96.12% respectively). The small time error bias ($< 6\%$) also shows that the estimates are very close to the extracted ground truth instants. However, there is a number of estimates that are as far as half a stride away from the ground truth (see Figure 4.21). These estimates increase ϱ which in turn compromises the overall performance of the algorithm and therefore, refining their sample location will improve the performance (D_{ss}). For this reason, a post processing method was developed as an additional block to the existing ASDSA framework and as will be shown in the evaluation it significantly improves the overall performance, as measured by D_{ss}, by 18.33%.

The post processing method builds on the already good set of estimates generated by ASDSA. It uses the output of the dynamic programming block in Figure 4.7, that is, the $\boldsymbol{\Xi}_I$ and $\boldsymbol{\Xi}_L$ sets that hold the sample locations of the ICI and LCI estimates respectively. The method is developed based on the simple idea that, for normal walking and for a single recording, the waveform in the vicinity of a single CI is similar to the waveforms of all other instants of the same type. This similarity is captured, for ICIs, using the cross correlation between $s(n)$ and a template signal obtained using the elements of $\boldsymbol{\Xi}_I$ as

$$s_t(n) = \sum_{i=1}^{||\boldsymbol{\Xi}_I||} s(n + \zeta_i - {}^{l_r}/_2) \text{ for } n = 1, 2, \ldots, l_r \tag{4.44}$$

where l_r is the size of the template window in samples, centred at each estimate ζ_i and $||\boldsymbol{\Xi}_I||$ was defined in (4.20) as the number of ICI selections made by ASDSA. For an N-sample long $s(n)$, the cross correlation sequence is expressed as

$$r_s(n) = \sum_{m=-N}^{N-1} s(n+m)s_t(m) \ . \tag{4.45}$$

The locations of the $||\boldsymbol{\Xi}_I||$ highest peaks (in absolute value) of $r_s(n)$ are obtained as

$$\hat{\zeta}_i = \underset{n}{\text{argmax}}(\{|r_s(n)| : n = 1, 2, \ldots, 2N-1\}) + \frac{l_r}{2} + 1 \tag{4.46}$$

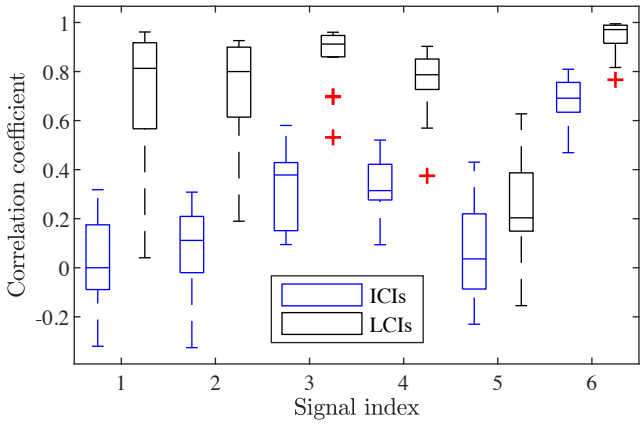

Figure 4.25: Box-plots of the correlation coefficients computed between the LCI template, formed with $l_r = 0.4F_s$, and each waveform obtained using a window of length 0.4 s centred at each ICI (in blue) and LCI (in black) for each signal. Outliers are indicated by red crosses.

for $i = 1, 2, \ldots, ||\boldsymbol{\Xi}_I||$ and subject to

$$|\hat{\zeta}_i - \hat{\zeta}_j| > \frac{\hat{N}_s}{2} \text{ for } j = 1, 2, \ldots, ||\boldsymbol{\Xi}_I|| \text{ and } j \neq i \tag{4.47}$$

since 1 ICI is desired per stride during normal walking. Given the variance in ASDSA estimates ζ, a time threshold larger than half a period was not used in order to reduce the impact of the first few highest peaks of $|r_s(n)|$ on other peak selections. The aim is to identify the strongest template matches in $s(n)$. If the separation is less than $\hat{N}_s/2$ for 2 or more $\hat{\zeta}_i$, then the one with the highest $|r_s(\hat{\zeta}_i)|$ value is retained and the rest are replaced by new peak locations. This is repeated until (4.47) is satisfied for all i and j. The set of refined ICI estimates $\hat{\zeta}_i$ is denoted as $\hat{\boldsymbol{\Xi}}_I$.

An LCI template is similarly found as in (4.44) by replacing $\boldsymbol{\Xi}_I$ with $\boldsymbol{\Xi}_L$ and using its associated ζ elements. The cross correlation sequence between $s(n)$ and the LCI template is computed as before and the $||\boldsymbol{\Xi}_L||$ highest peaks of the sequence are retained as the initial set of the new LCI estimates, subject to (4.47) as before. This set is subsequently refined by replacing the peaks that fall within a region centred at each element of $\hat{\boldsymbol{\Xi}}_I$. The lower limit of this region is defined at $\hat{\zeta}_i - 0.15\hat{N}_s$ and the upper limit at $\hat{\zeta}_i + 0.3\hat{N}_s$. These limit values were chosen on the basis of the bias values of Table 4.3 and that on average, based on the ground truth force plate data, the LCI instant occurs

at 60% of the stride cycle. The limits were set in order to avoid correlating the LCI template with the response waveform in $s(n)$ caused by the ICI, as it will produce strong correlation peaks. This statement is supported by the data in Figure 4.25 which shows, for 6 example signals from the test database, the box-plots (in blue) of the correlation coefficients computed between the LCI template, formed as in (4.44) with $l_r = 0.4F_s$, and each waveform obtained using a window of length 0.4 s centred at each ICI. The instants used in this test were obtained from the ground truth force plate signals. Similarly, the black box-plots are the correlation coefficients obtained using the ground truth LCIs with the same template as before. By comparing the boxes centred at each signal index, it becomes easy to see that even with ground truth data, overlapping in the correlation coefficient values occurs. Although, the overlap is not significant (except perhaps at signal index 5), it occurs in 16 out of the 90 test signals and it is therefore important to minimize this effect. This is achieved by defining the exclusion regions as discussed.

4.6.1 Evaluation

The same test database of 90 patella recordings was used for evaluating the post processing method. Parameter β in (4.38) was set to 0.5 which is the same value used for the tests in Section 4.5. The tuning parameter l_r was tested for the values $(0.03, 0.04, \ldots, 0.75)F_s$. The 30 ms window captures only part of the response waveform at each CI event whereas 750 ms is large enough to capture the full response waveform (the bias in ASDSA estimates is taken into account here). At the same time, 750 ms is a small enough window that avoids an overlap with the CI event (of the same type) from the neighbouring strides given that the maximum recorded speed is 9 km/h which gives an average stride duration of 0.7 s.

The results, in summary, reveal a significant improvement of 18.33% in D_{ss}, increasing from 0.6 for ASDSA to 0.71 for ASDSA with post processing. This score was obtained for ICI $l_r = 0.49F_s$ and LCI $l_r = 0.75F_s$ for which the values obtained for all the metrics are shown in Figure 4.26. To aid the visual comparison, each bar's length in the plot is normalised by the sum of the absolute value of the associated metric. For example, the D_{ss} blue bar's length is found as $|0.6|/|0.6|+|0.71|$ and similarly for the yellow bar it is

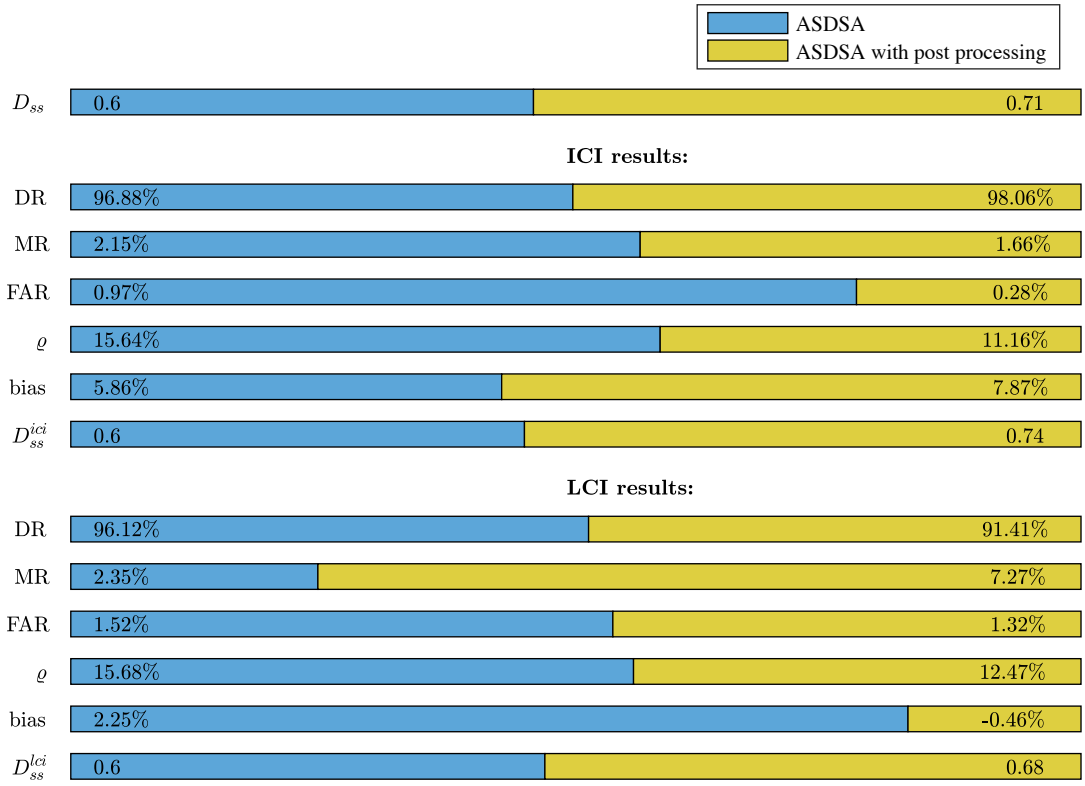

Figure 4.26: ASDSA vs ASDSA with post processing (ICI $l_r = 0.49F_s$, LCI $l_r = 0.75F_s$) results on the test database.

$|0.71|/|0.6|+|0.71|$. Hence, the changes in metric values due to the post processing method become immediately apparent.

It's clearly seen that by employing the post processing method, the estimation performance is significantly improved for each CI type. Both the D_{ss}^{lci} and D_{ss}^{ici} have increased, from 0.6 to 0.68 and 0.74 respectively. The improvement is due to the significantly reduced ϱ which, for ICIs, is even lower than the corresponding training ϱ. This however, came with an additional bias of 2% (or ≈20 ms) for ICIs. On the contrary, the LCI bias dropped down to only -0.46%, that is, an average time error of just 5.6 ms. As shown in Figure 4.27, even lower biases are achieved for different l_r combinations. However, neither of these are chosen as the best combination because the overall performance is measured by D_{ss}^{lci} and at those locations it is smaller than the maximum of 0.68. The lowest absolute bias obtained for LCIs is 1.46 ms (4.1×10^{-5}%; as a percentage of the stride cycle)

occurring at ICI $l_r = 0.34F_s$ and LCI $l_r = 0.58F_s$ (indicated by red cross in Figure 4.27). The corresponding ϱ is 13.54% and the resulting D_{ss}^{lci} is 0.65.

Figure 4.28 shows the histograms of $\hat{\tau}_\delta$ for both ICIs and ICIs. In comparison to the plots of Figure 4.21, the smaller ϱ values obtained with post processing are reflected by the narrower histograms. The heavier tail towards the positive $\hat{\tau}_\delta$ values in the ICI plot indicates the increased bias. In the LCI plot, a second peak in the distribution can be observed at approximately $\hat{\tau}_\delta = 30\%$. This peak arises due to the large bias in the estimates of 3 signals, as indicated in Figure 4.29 which also shows that the average $\hat{\tau}_\delta$ error value is within $\pm 10\%$ for the majority of the signals and only 8 signals give an error of over 20%.

With post processing, an increase in the ICI DR is observed (Figure 4.26) whereas a decrease of 4.7 percentage points is observed for LCIs. Given that the associated bias and ϱ are improved, the reduction in DR is attributed to the repositioning of estimates that were previously within the ground truth limits to new locations in $s(n)$ that are outside of these limits. Therefore, the number of misses increases, since the number of estimates does not change after post processing, and as the results show, the difference in DR is associated with an increase in MR. Figure 4.31 presents the variation in the detection metrics (DR, MR, FAR) with changing correlation window sizes l_r (in seconds) for each CI type. ICI results in plots (a), (b) and (c) show an improvement in all metrics for any window size used, with DR ranging from 97.1% to 98.2%. The opposite is true for the LCI detection where by employing post processing, the rate of detection is reduced. In general, higher DR values in plot (d) are obtained for large ICI l_r and small LCI l_r.

Figure 4.30 shows D_{ss} with varying l_r (in seconds). From plot (a) it's observed that for $\frac{l_r}{F_s} \geq 22$ ms the use of the post processing method improves ASDSA results. A clear maximum is obtained at $l_r = 0.49F_s$ with $D_{ss}^{ici} = 0.74$. This is a 23.3% improvement than without post processing. On the surface plot of Figure 4.30 (b), the contour lines are shown at the levels 0.37, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65 and 0.67. The 0.6 and 0.65 lines are indicated with black and the 0.67 line is highlighted in red. For most of the cases when ICI window size $l_r \geq 0.37F_s$ and depending on the LCI l_r value, the D_{ss}^{lci} is over 0.6

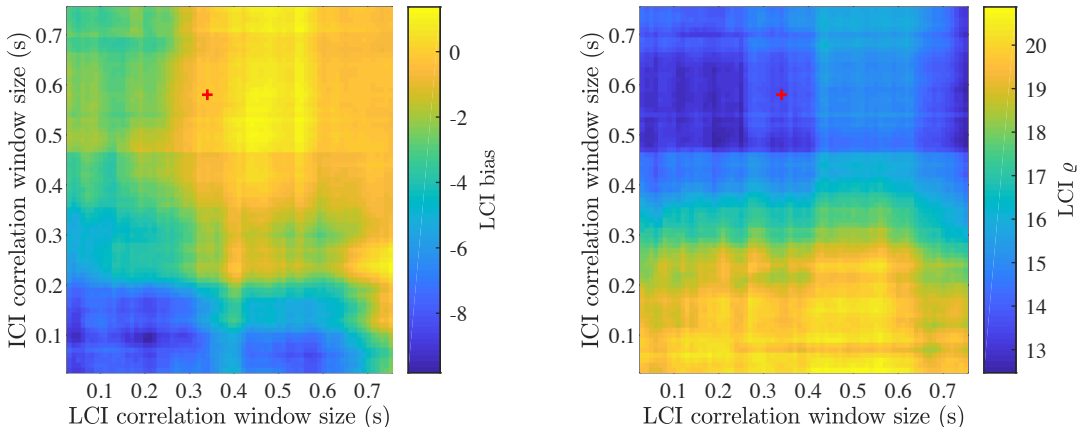

Figure 4.27: LCI time error (a) bias and (b) ϱ for a range of correlation window sizes l_r (in seconds). Red crosses indicate the location of the minimum absolute bias (1.46 ms or 4.1×10^{-5}%).

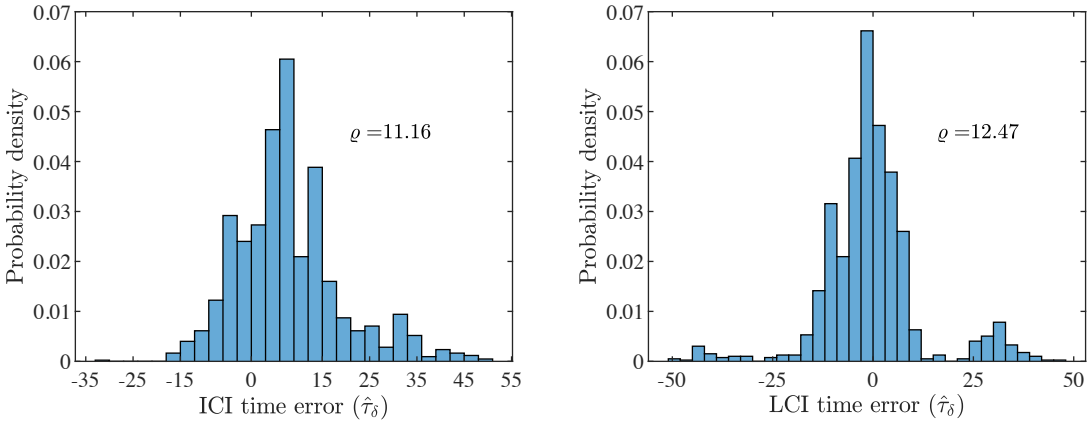

Figure 4.28: Time errors, $\hat{\tau}_\delta$, for (a) ICIs and (b) LCIs, for ASDSA with post processing on the test database and using $l_r = 0.49F_s$ for the ICI template and $l_r = 0.75F_s$ for the LCI template.

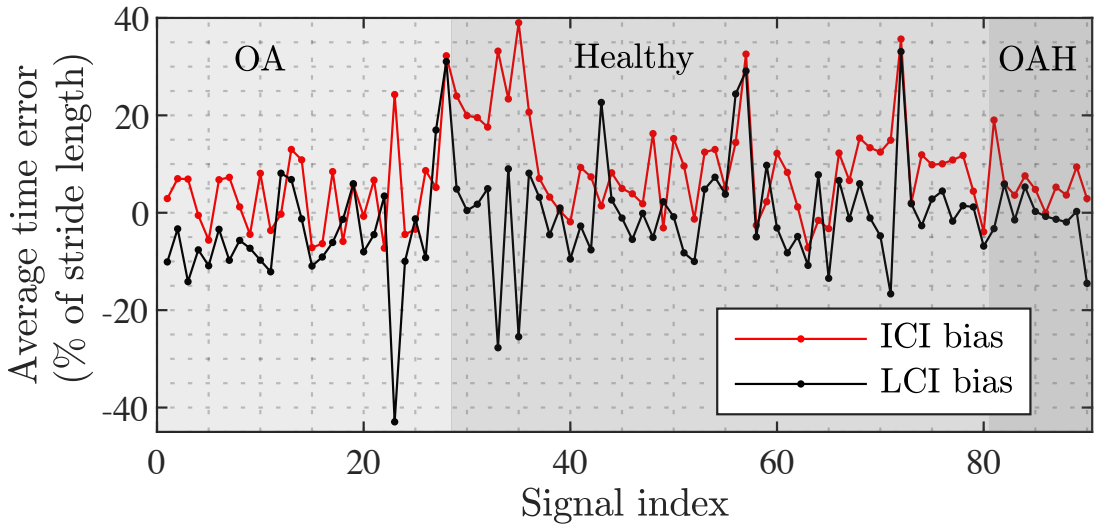

Figure 4.29: Time error bias per signal in the test database for ASDSA with post processing and using $l_r = 0.49F_s$ for the ICI template and $l_r = 0.75F_s$ for the LCI template. Shaded areas indicate the condition class of each signal.

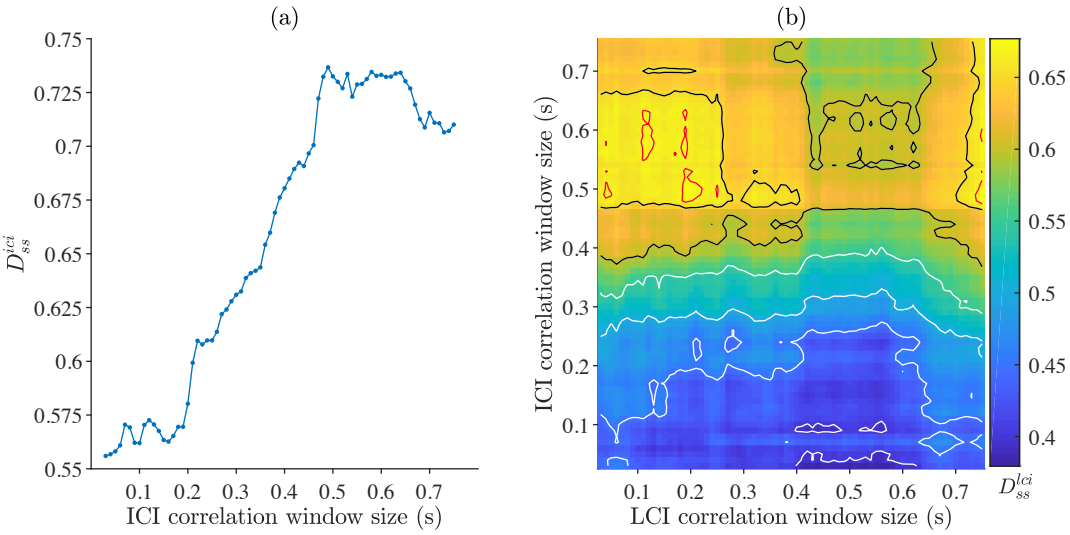

Figure 4.30: (a) D_{ss}^{ici} scores per ICI correlation window size in seconds and (b) D_{ss}^{lci} scores per ICI and LCI correlation window size combination, for ASDSA with post processing evaluated on the test database. Contour lines are shown at these levels: 0.37, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.67. The 0.6 and 0.65 lines are indicated with black and the 0.67 line is highlighted in red.

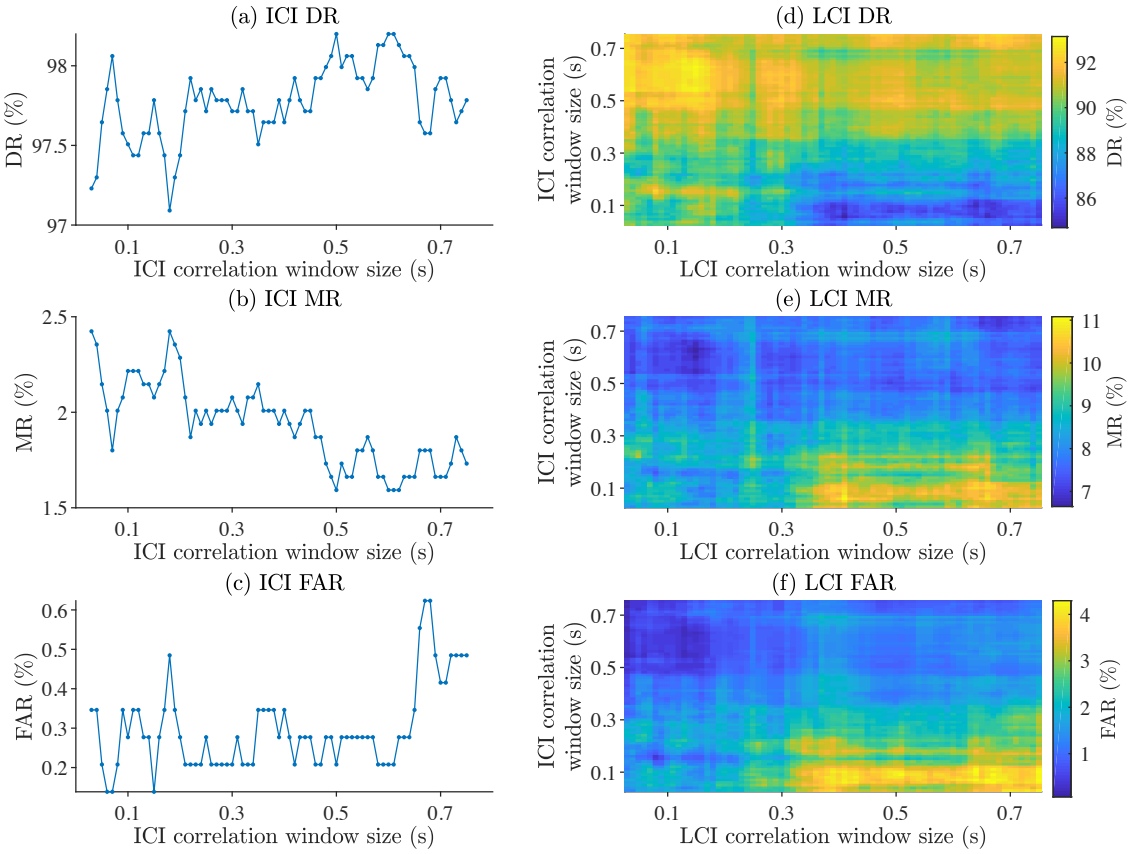

Figure 4.31: Detection metrics for ASDSA with post processing on the test database for a range of correlation window sizes l_r (in seconds), for ICIs in (a), (b), (c) and LCIs in (d), (e), (f).

and hence, better than the ASDSA results. The areas enclosed by the red contour lines indicate similar scores ($0.67 \leq D_{ss}^{lci} \leq 0.68$) but the highest score of 0.68 is obtained with an ICI window of size $l_r = 0.49F_s$ and a corresponding $l_r = 0.75F_s$ for LCIs. In terms of using ASDSA without post processing, the $D_{ss}^{lci} = 0.68$ is an improvement of 13.3%.

The post processing method was also employed with the LP residual signal instead of the raw knee sound signal by replacing $s(n)$ in (4.44) and (4.45) with $u(r)$. Evaluating on the same database for the same range of l_r values showed that the performance was not improved, achieving a maximum D_{ss} of 0.59. This is not surprising however, given that the post processing algorithm is based on waveform similarity in the vicinity of the CIs. The baseline signal of $s(n)$ contains important temporal information regarding the timing of the foot impacts and by performing LP it is removed.

4.7 Conclusion

This Chapter presented a dynamic programming based algorithm for estimating the instants of first and last contact of the foot to the floor during treadmill walking, requiring only the sound signal emitted from the knee. The algorithm is designed as an offline solution to the stride detection and segmentation problem but can easily be adapted to run in real-time. This can be achieved by using the pre-previous candidate instead of the next possible candidate in path, for the period deviation cost in (4.28). The evaluation results have shown that the algorithm performs with significant robustness, being able to select, from a large set of candidates, 1 ICI and 1 LCI per stride. In a database of 90 signals from which 2888 ground truth CIs can be extracted, ASDSA detected and estimated the location of 2787 instants, a detection error of only 3.5%. The average time error, obtained as a percentage of the stride duration, was 5.86% for ICIs and 2.25% for LCIs and the corresponding detection accuracy ϱ was 15.64% and 15.68% respectively. ASDSA performed similarly on OA and healthy knee AE signals but significantly better on OAH. The OAH database however, is too small in order to explore the reasons for this difference in performance. Training of the cost function's weight components in (4.20) was performed using a database that was subsequently excluded from testing the algorithm's

performance. An analysis was also performed on the choice of suitable LPC order and GD window size for the candidate generation step. It was later shown that the algorithm's performance in terms of temporal accuracy can be improved by post processing the estimates. The method developed is based on matched filtering with a template formed by the ASDSA contact instant estimates. The subsequent evaluation results showed an overall improvement, as measured by D_{ss}, of 18.33%. Even though the average detection error (MR+FAR) increased from 3.45% to 5.3%, the detection accuracies, ϱ, for both ICIs and LCIs improved by 4.48 and 3.21 percentage points respectively. Furthermore, an average time error of only -0.46% (5.6 ms) was obtained for LCIs. Using different window values, an even smaller bias of 1.46 ms can be achieved, however, the trade-off is that the associated D_{ss}^{lci} value is smaller than the maximum achieved for the former case.

Chapter 5

Stride-based Analysis and Classification of Knee Sounds

THE acoustic emission pulse events that were first introduced in Chapter 3 are further explored in this Chapter. In particular, the discriminant capabilities of these acoustic events are investigated on a per stride and stride phase basis for the task of normal vs abnormal knee joint sound signal classification. Section 5.1 describes the motivation for this work and the hypothesis that will be examined in the rest of the Chapter. Section 5.3 establishes the ground work by exploring various characteristics that can be obtained from the pulse waveforms. Section 5.4 describes the methodology developed for obtaining stride and stride phase based representations, which also includes feature importance analysis. Additionally, experiments are also presented in sections 5.5 and 5.6 to show the use of the Automatic Stride Detection and Segmentation Algorithm (ASDSA) that was developed in the previous Chapter and compare the results obtained with ground truth segmentation using force plate signals. Important outcomes and conclusions are summarized at the end of the Chapter.

5.1 Introduction

The work presented throughout this thesis focuses on the study of the acquisition and analysis of sounds generated by the knee during walking with particular focus on the effects due to OA. Chapter 3 investigated the information that can be extracted from the knee sound signals using a fixed frame-size segmentation approach while Chapter 4 presented an algorithm that enables the adaptive segmentation of the knee signals into strides and stride phases. In this context, Chapter 5 investigates the relationship between knee sounds with functional (i.e. gait biomechanical variables) and clinical outcomes, which is currently unknown.

The novel contributions of this work are the development of stride-synchronous and stride-phase-synchronous representations of the knee sound signal using the Acoustic Emission (AE) pulse events and their waveform characteristics and furthermore, the use of these representations to classify a knee as normal or abnormal. In particular, it was found that the number of threshold crossings, the threshold crossing rate and the energy, together with the number of pulse events are the most important features for classification, with the peak-to-peak amplitude being the least relevant. Adding to these, the classification accuracy was found to improve when the stride-synchronous based representations were used compared to features obtained over a longer time frame. In general, the work presented in this chapter establishes the foundations for any future work that uses AE pulses emitted during walking for characterizing knee health and relating to the main gait biomechanical variables, namely, the stride and its two fundamental phases, the stance and the swing.

It was found in Section 3.2.5 that OA knees emit 50% more AE pulses which have an energy that is on average 142% larger than the pulses emitted by healthy knees. These results were obtained by considering the pulse feature values as distributions formed by the collection of such pulses in a 20 second signal segment. As was observed in the experiments of sections 3.2.3 and 3.2.4, the discriminant power of the AE pulse features was, individually, lower than other features obtained from the STFT spectrum of the 20 s signal segments. However, given that a significant difference was found in the number of

emitted AE pulses between healthy and OA knees, these acoustic events are further explored here on a per stride and stride phase interval. The hypothesis is that the potentially discriminant information of these events is enhanced when analysed on shorter time scales relevant to the walking pattern of each subject rather than on large time scales. Using the automatic Contact Instant (CI) estimation algorithm that was developed in Chapter 4, the segmentation of a knee sound signal in strides and stride phases is now possible. The hypothesis is therefore examined in the following sections.

In [57] acoustic events were similarly observed in the knee sound signal that were of short duration, high frequency and high amplitude and were defined by the authors as clicks. In the aforementioned study a bandpass filter of 7 kHz to 16 kHz was applied on the raw knee signal and the sum of the logarithmic amplitude values was subsequently computed across frequency. The clicks were then identified as the peaks that satisfied an amplitude threshold found by employing a moving average window of 1000 samples duration at a 44.1 kHz sampling frequency. The detection of such acoustic events using the approach in [57] is susceptible to the choice of the STFT window size and type, as well as to the level of background noise. The inappropriate window choice or the high level of noise could produce frequency bins with large logarithmic amplitudes and lead therefore, to the detection of peaks that are irrelevant to the joint sounds. In addition, the authors were only interested in the location of these clicks relative to the knee angle measured by two inertial measurement units during flexion/extension [57, 62]. In this thesis, the high frequency acoustic events are referred to as AE pulses since the waveforms and the features that can be extracted from these waveforms are of interest and not just the time instant of occurrence.

Töreyin et al. [60], explored the consistency of knee acoustic events (clicks) during complex motions including walking. The sound signals in their study were captured by a MEMS-based microphone attached at the lateral side of the patella. The authors proposed a method that searches for high frequency and short duration clicks which also have consistent waveform shapes. In their approach the authors suggested to condition the raw knee sound signals by bandpass filtering with a bandwidth of 1 kHz to 20 kHz and

subsequently detecting all samples with amplitude value greater than the RMS value of the corresponding signal. This resulted in a large amount of candidate clicks to be selected. Next, time-windows of duration 5 ms and centred at each of these samples were used to extract potential AE pulse waveforms from which the ensemble average was computed. This ensemble average was used to identify similar waveforms in the knee sound signal by cross correlating with all the potential waveforms previously extracted and assigning first order Gaussian fits to the cross correlation signals. The clicks were then identified as those that have a Gaussian fit variance of less than 1.5 ms.

In applying the method of [60] to the signals obtained as described in Chapter 2 it was found that the threshold value of 1.5 ms was too low for the majority of the signals as it resulted in zero detected acoustic events. Therefore, to compare this method with the algorithm proposed in Section 3.2.1 II, the threshold value of 3 ms was chosen as it was found to reject the majority of the initially large number of candidate clicks (suspected false alarms) and to select a reasonable amount, relative to the proposed algorithm. However, further tuning of this parameter is outside the scope of this thesis and the authors did not provide a method for choosing the most appropriate threshold value. Furthermore, searching only for acoustic events with consistent waveforms discards other potentially informative pulses. As was shown in Section 3.2.1 II, there are many pulses with different waveform characteristics.

5.2 Technical background

The Support Vector Machine (SVM) classification approach is employed for the binary classification task considered in this Chapter. SVM was used in Chapter 3 as well but the classifier optimization was outside the scope of that work. On the contrary, SVM hyperparameter optimization is employed in this chapter and therefore, a technical review is presented in order to enable the understanding of the parameters that will be optimized.

SVM is a simple yet powerful classifier that is widely used for binary classification tasks [181, 182], as well as for multi-class problems [183] and for sequence classifica-

tion [184], with the use of suitable kernels. The concept of SVM for binary classification is to seek for a hyperplane that separates the classes such that the margin between them is maximized [104]. Assuming that the data is not perfectly separable, which is common with real-world data, SVM can use a soft margin approach in which the training error minimization is traded off against the number of training observations that fall within the margin [103]. There are two standard formulations of soft margins, the l_1-norm and the l_2-norm minimization problems, both of which employ a penalty term and slack variables, raised to the power of 1 and 2 respectively [96]. In this work the l_1-norm formulation is used due to its advantages over the l_2-norm in high dimensional feature spaces and in the presence of redundant features [185]. For classes i, j with Q and M number of observations respectively and the class vector $\boldsymbol{y} \in \{i, j\}$, the SVM l_1-norm formulation is defined, in its dual form, by the minimization problem [103]

$$\min_{\alpha} \left[\frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{B} \boldsymbol{\alpha} - \boldsymbol{e}^T \boldsymbol{\alpha} \right] \tag{5.1}$$

subject to

$$\boldsymbol{y}^T \boldsymbol{\alpha} = 0 \tag{5.2}$$

with

$$0 \leq \alpha_p \leq C, \quad p = 1, 2, \ldots, P \tag{5.3}$$

where $P = Q + M$ and $\boldsymbol{\alpha}$ is a vector of α_p which are the Lagrange multipliers. Consider the training vectors $\boldsymbol{x}_q$ and $\boldsymbol{x}_m$ with corresponding labels y_q and y_m where $y_q, y_m \in \boldsymbol{y}$ and the indices $q \in 1, 2, \ldots, Q$ and $m \in 1, 2, \ldots, M$. In (5.1) $\boldsymbol{e}$ is a vector of ones and each element in $\boldsymbol{B}$ is obtained by $B_{qm} = y_q y_m K(\boldsymbol{x}_q, \boldsymbol{x}_m)$ where $K(\cdot, \cdot)$ is a kernel. The minimization problem in (5.1) is solved using the sequential minimal optimization [186, 187], which is the standard algorithm for this problem. Other algorithms can also be employed such as the iterative single data algorithm which is appropriate for large data sets [188]. Finally, a new observation vector $\boldsymbol{x}$ is classified based on the sign of the decision function given by

$$f(\boldsymbol{x}) = \sum_{p=1}^{P} y_p \alpha_p K(\boldsymbol{x}, \boldsymbol{x}_p) + b \tag{5.4}$$

where b is the hyperplane intercept point. In this formulation only the $\boldsymbol{x}_p$ of $\alpha_p > 0$ are used which are the support vectors.

The parameter C in (5.3) is the penalty term that trades off misclassification of training observations against simplicity of the decision surface. A low C makes the decision surface smooth (i.e. misclassification becomes less important), while a high C attempts to classify all training examples correctly. Therefore, C is a hyperparameter that needs to be optimised. This is explored in Section 5.6 where the effect of different kernel types $K(\cdot, \cdot)$ is also investigated.

5.3 Analysis of acoustic events per stride phase

5.3.1 Stride segmentation methods

Chapter 4 presented the development of ASDSA, a dynamic programming based approach that identifies the Initial Contact Instants (ICIs) and Last Contact Instants (LCIs) using only the knee sound signal $s(n)$. These instants can be used to segment $s(n)$ into consecutive strides and stride phases, namely, the stance phase (foot in contact with the floor) and the swing phase (foot suspended over the floor). From the experiments carried out in the previous Chapter it was found that the ASDSA estimates have a time bias error, relative to the ground truth obtained from the force plate signals. Therefore, at this stage of development, the ground truth segmentation points are used, unless otherwise stated, and are obtained as described in Section 4.5.2. ASDSA is employed later in Section 5.6.

5.3.2 Acoustic emission pulse waveform features

The work carried out in Chapter 3 identified certain frequency bands that contain significantly discriminant information for the task of normal vs abnormal knee signal classification. In particular, the lowest band was identified at frequencies 220 Hz to 420 Hz. Given these observations, the knee sound signal $s(n)$ is high-pass filtered using the Parks-McClellan optimal FIR filter design, [80], with a cut-off frequency of 200 Hz. The filtered signal $s_0(n)$ is then standardised with zero mean in order to remove the trend (baseline

signal) caused by the response of the foot impact and focus the analysis on the fluctuations about the trend. This is achieved by splitting $s_0(n)$ into 1.5 ms frames with 50% overlap, finding the mean amplitude value per frame and subsequently subtracting, from each signal's sample amplitude, the value found by shape-preserving piecewise cubic interpolation, [81], using the per frame mean amplitude values as query points. The AE pulses are then extracted using the proposed algorithm as described in Section 3.2.1 II, for which the control variable $\eta = 10^{-14}$, as explained in the algorithm's description. The pulse definition parameters t_0 and r_0 are set equal to the number of samples equivalent to 0.5 ms in order to allow the detection of very short pulses. The practical limit for t_0 is set by the design of the algorithm at 3 samples which is equivalent to approximately 0.19 ms since $F_s = 16$ kHz is used. For such short duration acoustic events it is practically impossible to derive a meaningful waveform and its characteristics (peak-to-peak amplitude, number of threshold crossings etc.). Following this, an AE pulse is considered to be identified within a stride phase if it starts and ends in the same phase, otherwise it is discarded.

Additionally to the number of pulses detected, 5 features that aim to characterise the pulse waveforms are also extracted and investigated. These features are the (a) peak-to-peak amplitude expressed in dB, (b) duration in ms (excluding the rest time r_0 since it is constant in all of the pulses; see Section 3.2.1 II), (c) pulse energy in dB, (d) number of threshold crossings (TC) and (e) threshold crossing rate (TCR) which is obtained by dividing the number of threshold crossings by the pulse length in samples.

5.3.3 Experimental setup

The average value of each feature is obtained per stride and stride phase in order to examine whether a single feature can capture differences between healthy and OA knee acoustic signals segmented at the CIs. In this experiment a database of 11 healthy and 7 OA knees is used. When the corresponding sound signals are segmented using the ground truth CIs they result in 981 strides, 1050 stance phase and 981 swing phase segments for the healthy knees and 517, 564 and 517 segments respectively for the OA knees. The difference in number of segments is due to the discontinuities that were observed in the

collected force plate signals as detailed in Section 4.5.2.

5.3.4 Results and discussion

The results are shown in Figure 5.1 and Table 5.1. Figure 5.1 displays violin plots of the extracted features per stride phase. The left, blue side depicts the feature histogram for healthy knee signals whereas the right, red side displays the corresponding for OA knee signals. The histograms are obtained by the kernel distribution method using a normal kernel [189]. The average and median values of each histogram are also displayed in order to aid the visual comparison. It can be clearly seen from Figure 5.1 (a) that, on average, OA knees generate more AE pulses in each stride phase than healthy knees. During a single stride, 71.3% more pulses are emitted by OA knees and 69.8% and 70.5% more during the stance and swing phases respectively. Interestingly, AE pulses from both knee classes are emitted, on average, more frequently during the stance phase rather than the swing, during which the knee performs wider range of motion (for an OA subject: $54.8^\circ \pm 5.5^\circ$ during the swing compared to $10.3^\circ \pm 4.0^\circ$ during stance) [190]. The visual comparison along with basic statistics (Table 5.1) suggests that this feature has strong capabilities in discriminating between healthy and OA knees. In addition, the results highlight the benefit of using this feature on a per stride phase basis rather than on longer time scales in which a 50% difference was observed as described in Section 3.2.5. The peak-to-peak amplitude, the number of TC and the duration also indicate large differences between the distribution means in each of the three stride phases as shown in plots (c), (d) and (e) respectively. The percentage difference between the distribution means for these features ranges from 17.2% to 33.3% as summarised in Table 5.1. Moreover, a larger mean difference is observed between the distributions of the stance phase compared to those of the whole stride. The difference is further increased in the swing phase for the distributions formed by the number of TC and the duration but is decreased for the peak-to-peak amplitude feature.

Multi-modal distributions with short tails are observed for the features in Figure 5.1 plots (a) and (b) and with long tails in (c). On the other hand, skewed uni-modal dis-

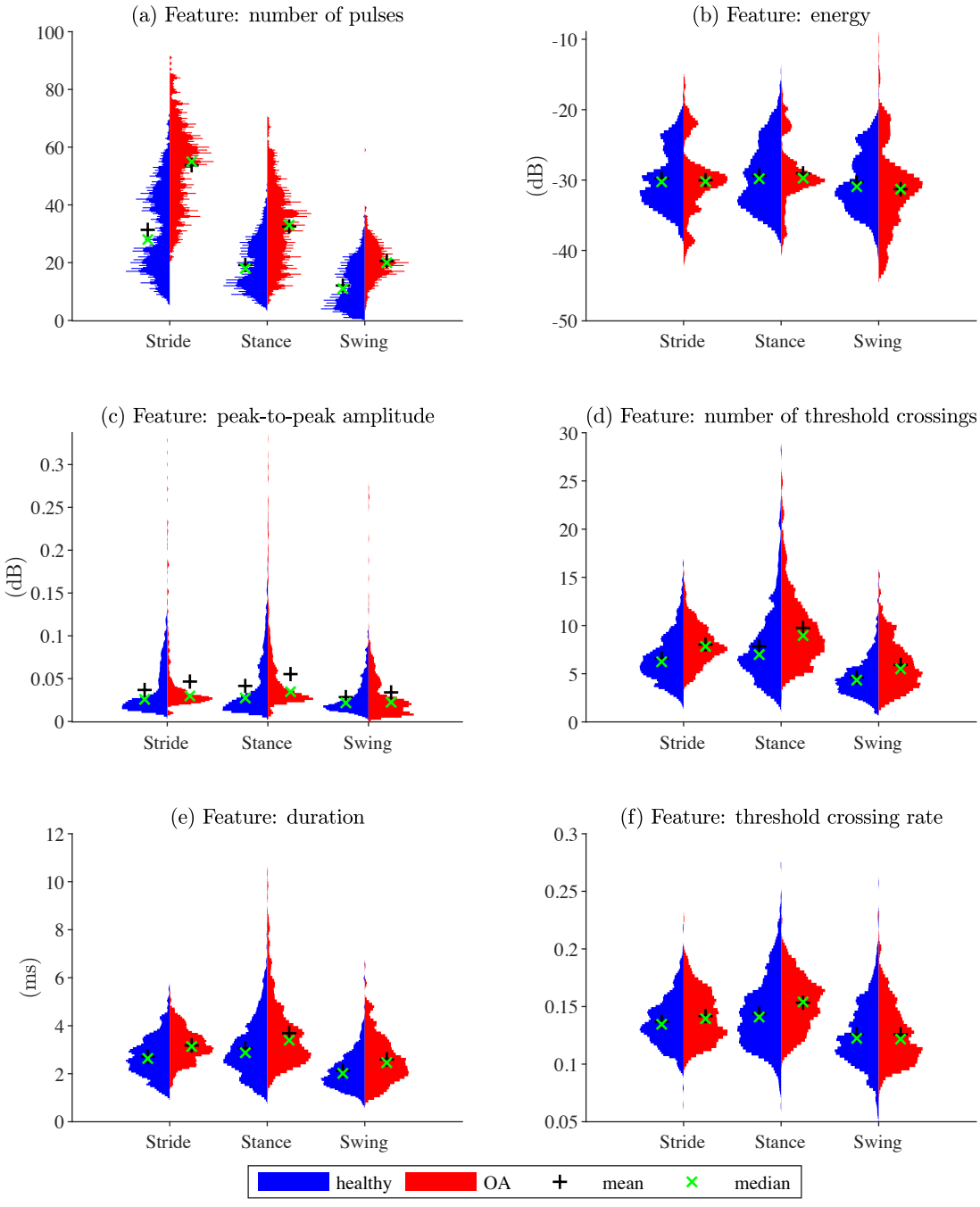

Figure 5.1: Histograms of acoustic emission pulse features obtained per stride phase from healthy (blue) and OA signals (red). The average and median values of each histogram are also shown. Plots (b) to (f) display the average feature values per stride phase.

tributions are observed in plots (d), (e) and (f). Differences of less than 7% are observed between the means of the healthy and OA distributions formed by the energy and TCR features. Significant distribution overlap is observed in these two cases for all three stride phases.

Feature	Phase	Healthy		OA		Percentage difference of means
		Mean	Median	Mean	Median	
Number of pulses	Stride	31.36	28	53.72	55	71.3%
	Stance	19.17	18	32.55	33	69.8%
	Swing	12.03	11	20.58	20	70.5%
Peak-to-peak amplitude (dB)	Stride	0.037	0.026	0.047	0.030	26.7%
	Stance	0.042	0.028	0.055	0.035	33.3%
	Swing	0.029	0.022	0.034	0.023	19.3%
Duration (ms)	Stride	2.71	2.64	3.18	3.14	17.2%
	Stance	3.06	2.88	3.69	3.40	20.5%
	Swing	2.12	2.02	2.60	2.46	22.4%
Energy (dB)	Stride	−29.74	−30.25	−30.02	−30.21	1.0%
	Stance	−29.36	−29.81	−29.05	−29.73	−1.1%
	Swing	−30.28	−30.94	−31.30	−31.30	3.3%
Threshold crossings	Stride	6.59	6.24	8.02	7.83	21.7%
	Stance	7.85	6.99	9.76	8.96	24.4%
	Swing	4.63	4.33	5.91	5.50	27.7%
Threshold crossing rate	Stride	0.137	0.135	0.142	0.140	3.5%
	Stance	0.144	0.141	0.153	0.154	6.6%
	Swing	0.126	0.122	0.126	0.122	−0.02%

Table 5.1: Average and median values of the acoustic emission pulse feature distributions shown in Figure 5.1.

In summary, shorter pulses with smaller peak-to-peak amplitude and lower frequency (TC) are generally observed during the swing phase compared to the stance phase for both knee conditions. OA knees in general, emit more and longer pulses with higher peak-to-peak amplitude and frequency in all three stride phases. This is also seen in the tails of the distributions where in all cases except TCR, the pulses from OA knees exhibit larger extreme values.

To investigate the potential of designing a classifier with low error rate, the sta-

Feature	Phase	H_0	p-value	Feature	Phase	H_0	p-value
Number of pulses	Stride	1	$\ll 0.01$	Energy	Stride	1	0.0025
	Stance	1	$\ll 0.01$		Stance	1	$\ll 0.01$
	Swing	1	$\ll 0.01$		Swing	1	$\ll 0.01$
Peak-to-peak amplitude	Stride	1	$\ll 0.01$	Threshold crossings	Stride	1	$\ll 0.01$
	Stance	1	$\ll 0.01$		Stance	1	$\ll 0.01$
	Swing	1	$\ll 0.01$		Swing	1	$\ll 0.01$
Duration	Stride	1	$\ll 0.01$	Threshold crossing rate	Stride	1	$\ll 0.01$
	Stance	1	$\ll 0.01$		Stance	1	$\ll 0.01$
	Swing	1	$\ll 0.01$		Swing	0	0.314

Table 5.2: Two-sample Kolmogorov-Smirnov test results at 1% significance level, for the comparison between each pair of distributions formed by each acoustic emission pulse feature from healthy and OA signals respectively, as shown in Figure 5.1.

tistical significance of the dissimilarity between the feature distributions of Figure 5.1 is quantified. The two-sample Kolmogorov-Smirnov Test (KST) is one of the standard tests most commonly used for such tasks and is therefore employed in the following. Let the null hypothesis H_0 state that the distribution formed by any of the 6 features obtained from the AE pulses detected in healthy knee sound signals originates from the same continuous distribution as the distribution formed by the same feature obtained from the AE pulses detected in OA knee sound signals. To simplify, each of the left blue sides of the violin plots in Figure 5.1 is tested whether it is similar to the corresponding right, red side. The two-sample KST examines H_0 by calculating the maximum absolute difference between the empirical CDFs [92]. The results are shown in Table 5.2. At 1% significance level H_0 is rejected for all features except for the TCR in the swing phase. The test on the energy feature distribution on a per stride basis, returns an asymptotic p-value of 0.0025 whereas for all other features the p-value is well below the statistical significance level of 1%. These results further confirm the differences between the AE pulses emitted by healthy and OA knees.

Motivated by these statistical differences the following section proposes a stride-synchronous and stride-phase-synchronous representation of $s(n)$ using only the AE pulses and their characteristics. Based on these representations the classification performance for the task of normal (clinically healthy) vs abnormal knee joint sound signals (clinically OA)

will be quantified. Several experiments will be conducted to test the sensitivity of this approach relative to the segmentation points obtained from the ground truth force plates.

5.4 Classification using stride-phase representations

5.4.1 Proposed methodology

The j^{th} signal segment of the i^{th} knee in the data-set, denoted as $s_{i,j}(n)$ with duration τ_s, is obtained as described in Section 3.2.1. RMS normalisation, HPF with 200 Hz cut-off frequency and mean subtraction are used for signal conditioning prior to AE pulse identification. Subsequently, the segment is divided into strides and further into stride phases. The identified AE pulses are then categorized into each phase, depending on their time of occurrence in the signal, and the 5 waveform features described in the previous section are extracted from each pulse. A single stride is then modelled by a 1×6 feature vector $\boldsymbol{x}$ that consists of the average value of each waveform feature within the same stride along with the number of pulse events. The stride-synchronous representation of $s_{i,j}(n)$ with L strides can be constructed as

$$\boldsymbol{X_s} = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_L^T\right] \tag{5.5}$$

where for clarity of notation the explicit dependency of $\boldsymbol{X_s}$ on i and j is dropped. If the feature averaging is instead performed within each stance and swing phase separately, the matrices $\boldsymbol{X_{st}}$ and $\boldsymbol{X_{sw}}$ can then be constructed as in (5.5). These matrices can be considered as stance-phase-synchronous and swing-phase-synchronous respectively, both of which on their own are incomplete representations of $s_{i,j}(n)$ since they model only part of the signal but are complementary to each other and hence are taken together. Therefore, the sets $\boldsymbol{X_s}$ and $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$ are two different representations of $s_{i,j}(n)$ which will be used for classification.

As previously stated, the classification of $s_{i,j}(n)$ into normal and abnormal is of interest. The question is therefore, how to transform $\boldsymbol{X_s}$ and $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$ into rank 1 feature vectors so that standard supervised learning algorithms, such as SVM, can

be employed. One option is to concatenate the rows of the matrices. Although this ensures that all of the information is retained, the concatenation produces large feature vectors ($1 \times 6L$ and $1 \times 12L$, for L strides). As a result, a large number of observations is required to adequately fill up the feature space in order to reduce the effects of the 'curse of dimensionality' (causes the overfitting to the training data), as mentioned in, for example [79, 98, 191, 192]. Given the limited size of the database (see Chapter 2) and the unavailability of any other similar database that could be used, this approach is rejected. In addition, this approach does not effectively utilise the potentially relevant (to the classification task) information in the dynamics of the representations ($\boldsymbol{X_s}, \boldsymbol{X_{st}}, \boldsymbol{X_{sw}}$), as is the intent in this section. Therefore, alternative options must be explored.

Feature extraction with dimensionality reduction is a common alternative approach and can be applied to $\boldsymbol{X_s}$ and $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$. A multitude of techniques exist in the literature, such as LDA [98], Independent Component Analysis (ICA) [193, 194], manifold learning for non-linear dimensionality reduction (usually based on the assumption that the data manifold is locally linear) [195, 196], and more recently a family of neural networks called autoencoders [197], with the adversarial autoencoders [198], being the state of the art method for regularizing autoencoders. Perhaps the most popular joint feature extraction and dimensionality reduction method is Principal Components Analysis (PCA) [199], an unsupervised technique that seeks orthogonal projections of the original data such that the variance in each projection is maximised. A common pitfall of PCA however, is that the new features created, called principal components, are not directly interpretable [199]. This is undesirable for the work in this Chapter as the interest is to find out to which extent the AE pulses can be used as a biomarker for OA detection during dynamic functional activities. Exploring the information inherent in these AE pulses is therefore essential.

To this respect, it is hypothesised that for an $s_{i,j}(n)$, the information of each feature in $\boldsymbol{X_s}, \boldsymbol{X_{st}}$ and $\boldsymbol{X_{sw}}$ can be captured by the mean value and the variance across all strides and stance/swing phases respectively. This is motivated by the results displayed in Figure 5.1 and Table 5.1 in which large differences were observed between the means of the global feature distributions. Additionally, the variance is used to capture the variability

of these features from stride to stride (equally stance to stance and swing to swing). Essentially with the variance, the consistency in number of emissions of such acoustic pulse events will be examined along with the consistency in their waveform characteristics. Ultimately, the evaluation for the choice of features is made based on the classification performance as is commonly the case with such tasks.

Following the above, for an $s_{i,j}(n)$ with L strides, the vectors $\boldsymbol{\mu}_s$ and $\boldsymbol{\sigma}_s$ containing the means and variances respectively of each feature in $\boldsymbol{X_s}$ are obtained as

$$\boldsymbol{\mu}_s = \frac{1}{L}\boldsymbol{X_s}[1,1,\ldots,1]^T \tag{5.6}$$

$$\boldsymbol{\sigma}_s = \frac{1}{L-1}(\boldsymbol{X_s}-\boldsymbol{\mu}_s)(\boldsymbol{X_s}-\boldsymbol{\mu}_s)^T \ . \tag{5.7}$$

Hence, the vector $\boldsymbol{x_s} = [\boldsymbol{\mu}_s^T, \boldsymbol{\sigma}_s^T]$ is used to represent $s_{i,j}(n)$. Alternatively, $s_{i,j}(n)$ is represented by the 1×24 feature vector $\boldsymbol{x_p}$ that is constructed by the means and variances obtained for each feature in $\boldsymbol{X_{st}}$ and $\boldsymbol{X_{sw}}$ separately, as in (5.6) and (5.7). Subsequent experiments are conducted in order to measure the classification performance obtained with the two alternative feature representations. In order to make the discussion of the results more clear, let the indices $m = 1, 2, \ldots, 6$ correspond to the number of pulses, peak-to-peak amplitude, duration, energy, TC and TCR respectively. Additionally, let the notation $\boldsymbol{X_s}[m]$ to indicate the m-th feature from the feature set $\boldsymbol{X_s}$. Similar thinking applies for $\boldsymbol{X_{st}}$ and $\boldsymbol{X_{sw}}$.

5.4.2 Experimental methods

For the following experiments the segment length was fixed at $\tau_s = 20$ seconds, as in the experiments of Chapter 3. Furthermore, the stride segmentation is achieved by the force plate signals, as described in Section 4.5.2, in order to rule out the possibility of the classification performance being affected by inaccurate identification of the segmentation points. The former is investigated later in Section 5.5 and the results are compared to the performance obtained when segmenting with ASDSA in Section 5.6.

The knee database used in the experiments consists of 11 healthy knees from which

63 segments with available force plate data are obtained and 7 OA knees from which 36 segments can be used. For the evaluation, a 3-fold cross-validation procedure is employed in which the database is divided into 3 randomly constructed groups (folds) in a ratio of healthy to OA knees of 3:3, 4:2, 4:2 and each group is made up with the segments of their constituting knees. Two groups are used for training the classifier which is then tested on the group left out. This is repeated until all 3 groups are evaluated. Prior to this, the training data is scaled by subtracting the mean and normalising by the variance. The same normalization values are then applied to the test set. SVM with linear kernel is used as the classifier due to its capability of training on data with small number of observations to number of features ratio and its robustness to overfitting [200]. Further advantages of SVM have been previously described in the Analysis subsection of 3.3.2.

Cross-validation variability analysis

Given the database size (18 knees), 3 folds are a good choice for constructing representative groups for training and testing the SVM model. Moreover, the classifier results are expected to suffer from high variability due to the small database size. However, by repeating the cross-validation procedure and averaging the results of each execution by the number of executions will reduce this variability. Therefore, a variability error analysis is performed in order to identify a good number of cross-validation executions that will allow stability and repeatability of the evaluation results.

The results of the analysis obtained with the feature vector $\boldsymbol{x_s}$ are displayed in Figure 5.2. The evaluation metrics used have been previously introduced in Chapter 3 and are the error rate (E_r), $F_{0.5}$ score [101], Matthew's Correlation Coefficient (MCC) [102], S_c as defined in (3.30) and the Area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve [105]. It can be seen that for small number of executions (< 1000) all five metrics exhibit relatively large variability in their values. Variance in AUC becomes stable after approximately 2500 executions although, the mean value continues to exhibit small fluctuations (in the order of 10^{-3}) until approximately 5500. MCC, E_r, $F_{0.5}$ and hence S_c on the other hand, initially show larger changes in both the mean

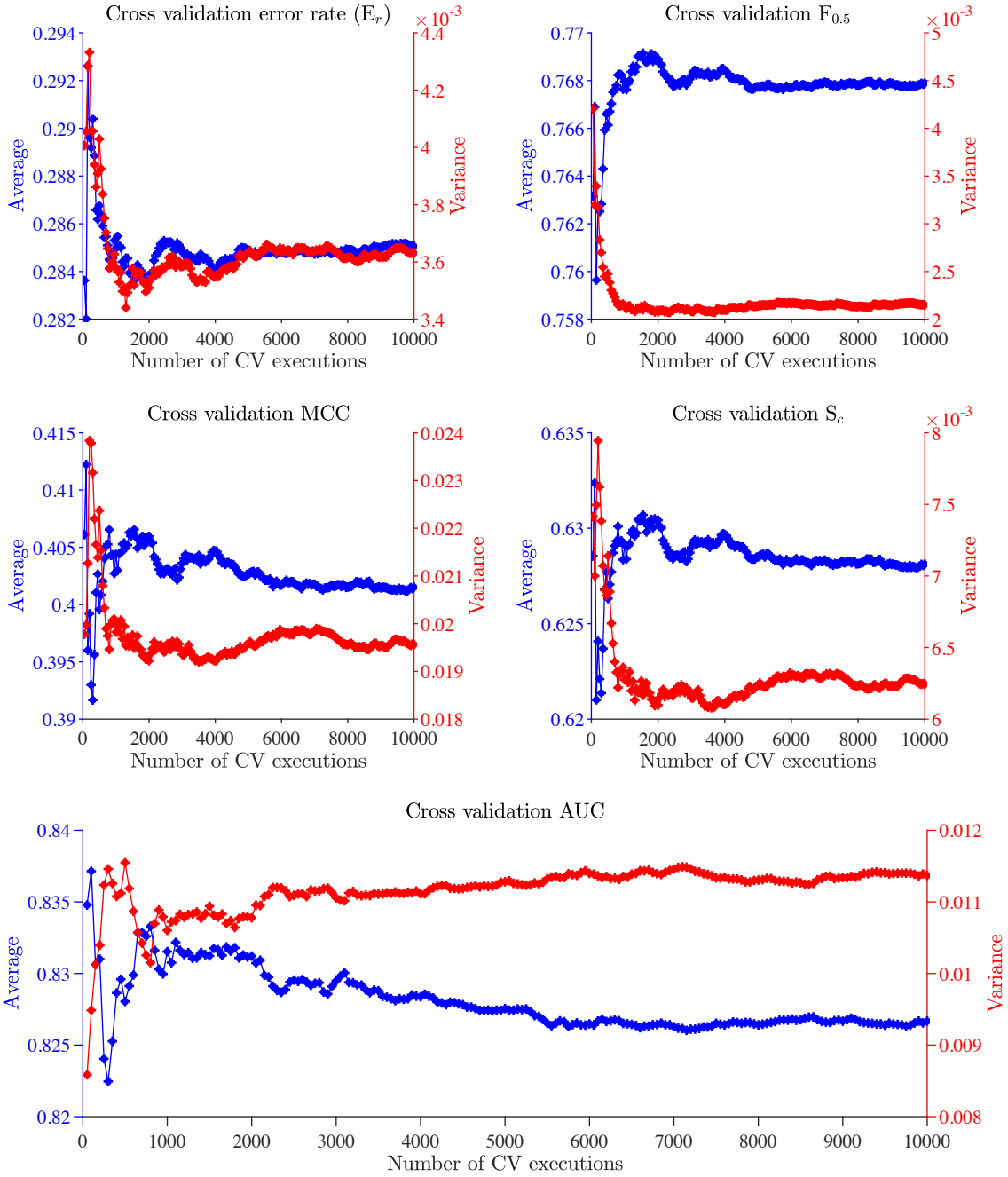

Figure 5.2: Variability performance analysis for the number of cross-validation executions. Results shown for each metric are averaged over the number of executions (in steps of 50) and obtained using all features from the stride-synchronous feature set representations $\boldsymbol{X_s}$ (i.e. the vectors $\boldsymbol{x_s}$), of the 20 second signal segments.

and the variance but become approximately stable after about 5000 cross-validation executions. Given these results it is reasonable to assume that repeating the cross-validation procedure 10,000 times and averaging the classification results at the end, produces stable and repeatable results for the evaluation of the SVM model.

Furthermore, given the results displayed in Figure 5.2, it is sufficient to assume that for this and the following experiments, the final value of each metric is accurate to 1 d.p. (decimal place). The third d.p. may vary slightly but it is difficult to accurately measure this variability. Therefore, the value in the second d.p. is accurate with an error of ± 1 unit in its value, due to the plausible inaccuracy in the third d.p (rounding up or down). Bearing in mind the above, the results in the following sections are reported to 3 d.p.

Feature importance analysis

The classification experiments to follow are conducted using: (a) $\boldsymbol{x_s}$, the vector obtained from the stride-synchronous representations $\boldsymbol{X_s}$ and (b) $\boldsymbol{x_p}$, the vector obtained from the stance-synchronous, $\boldsymbol{X_{st}}$, and swing-synchronous, $\boldsymbol{X_{sw}}$, representations. In addition, feature importance analysis is carried out to find which features are more relevant to the classification task. The approach employed is SVM-RFE which uses Recursive Feature Elimination (RFE) and is based on the coefficients that define a vector that is orthogonal to the separating hyperplane [201]. In the first iteration of RFE, a linear SVM model is trained with all the features. The feature with the smallest absolute weight is then discarded from subsequent iterations as it is the least informative [201]. The remaining features are used to train a new model in the next iteration. This process is repeated until only 1 feature remains in the set which is the most relevant to the classification task at hand. Therefore, this process generates a feature importance ranking based on the order of elimination. In the 3-fold cross-validation procedure employed in this section, the feature rankings obtained per fold are averaged and then are averaged again across the cross-validation repetitions to obtain the final rankings.

5.4.3 Baseline method

The method presented in [60] and described in the introduction is used as the baseline since it is the only method that was developed and applied on knee sound signals obtained during walking. However, this method can only identify 'interesting' peaks in the signal and cannot be used to extract the waveforms of the acoustic pulse events which were shown to have potential discriminant information (see Figure 5.1).

Integrating the baseline's detection method with the algorithm presented in Section 3.2.1 II, is not possible because the waveform extraction part of the algorithm relies on the amplitude threshold that is set on the time-domain signal and was used to identify the acoustic events in the first place. As stated in the introduction, the baseline method uses a threshold on the variance of the first order Gaussian fits to the correlation peaks in order to identify the AE events [60]. Therefore, to extract the waveforms, a threshold is chosen at 3 standard deviations from the correlation peak. In this way, the edges of the AE pulse can be identified. Hence, the peak-to-peak amplitude, duration and energy can be extracted whereas TC and TCR cannot, given the lack of an amplitude threshold. In the same way that $\boldsymbol{X_s}$ and $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$ are constructed, the baseline method is used to obtain the equivalent sets from which the feature vectors denoted as $\boldsymbol{b_s}$ and $\boldsymbol{b_p}$ respectively are obtained using (5.6) and (5.7). These are 8-dimensional and 16-dimensional feature vectors respectively (mean and variances of number of pulses, peak-to-peak amplitude, duration and energy).

The variability analysis of Figure 5.2 was performed for all feature sets (including the baseline) and showed that after 10,000 cross-validation executions the metric values obtained are stable and hence repeatable. It is also worth mentioning that for a single $s_{i,j}(n)$ of 20 seconds, the baseline method takes from a few minutes ($2 \sim 3$) to several minutes (> 10) to execute, depending on the number of candidate events that are initially detected. Compared to 0.37 seconds (on average) taken for the proposed algorithm to execute, this is a huge difference. Therefore, this shows the potential of the proposed method to be used in a real-time application. The execution times were measured under the same conditions on a machine with processor speed 2.6 GHz, RAM of 8 GB and

Feature vectors	$\overline{\mathrm{E}}_r$	$\overline{\mathrm{F}}_{0.5}$	$\overline{\mathrm{MCC}}$	$\overline{\mathrm{S}}_\mathrm{c}$	$\overline{\mathrm{AUC}}$
$\boldsymbol{x_s}$	0.285	0.768	0.402	0.628	0.827
$\boldsymbol{x_p}$	0.325	0.732	0.279	0.562	0.683
$\boldsymbol{b_s}$ (baseline)	0.419	0.670	0.121	0.457	0.593
$\boldsymbol{b_p}$ (baseline)	0.279	0.779	0.397	0.632	0.729

Table 5.3: Average 3-fold cross-validation results per feature set, including the baseline method. The over-line denotes the metric value averaged over 10,000 cross-validation executions.

using MATLAB [202]. The long execution time of the baseline method is due to the cross correlation computation and Gaussian fitting for each peak.

5.4.4 Evaluation and discussion

The classification results are presented in Table 5.3. It can be clearly seen that classifying the knee signals using the features obtained from the stride-synchronous representations $\boldsymbol{X_s}$ rather than the $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$ set achieves higher results in all of the metrics. In a database of 99 signals, an average AUC of 0.827 is achieved, a 21% improvement over the AUC obtained with the stride phase feature set $\boldsymbol{x_p}$. The corresponding average $\mathrm{S_c}$ is 0.628 with the respective individual components obtained as $[\overline{\mathrm{E}}_r, \overline{\mathrm{F}}_{0.5}, \overline{\mathrm{MCC}}] = [0.285, 0.768, 0.402]$. In comparison to the baseline, the feature set $\boldsymbol{x_s}$ clearly outperforms $\boldsymbol{b_s}$ and $\boldsymbol{b_p}$ by 39.5% and 13.4% respectively (in AUC). Marginal differences are observed in the other metrics (-1.3% to $+1.4\%$) between $\boldsymbol{b_p}$ and $\boldsymbol{x_s}$. On the other hand, $\boldsymbol{b_p}$ gives an AUC that is 6.73% higher than the score obtained with $\boldsymbol{x_p}$. However, given that $\boldsymbol{x_p}$ includes more features than $\boldsymbol{b_p}$, this performance difference may have resulted by the capability of the fewer features to better highlight the class differences and not by the pulse events themselves, as detected by the two algorithms. To further explore this, the importance of each feature for the classification task is examined in the following, using SVM-RFE as described in Section 5.4.2.

The feature importance analysis produces the rankings shown in Tables 5.4 and 5.5 for the features in $\boldsymbol{X_s}$ and $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$ sets respectively. The tables show the classification AUC values obtained when using the feature of each line in addition to all the previous more relevant features, starting with the most important feature at the first rank. The

Rank	AUC	Feature	Description
1	0.795	Mean of $\boldsymbol{X_s}[1]$	number of pulses
2	0.795	Variance of $\boldsymbol{X_s}[4]$	energy
3	0.852	Mean of $\boldsymbol{X_s}[6]$	TCR
4	**0.857**	Variance of $\boldsymbol{X_s}[1]$	number of pulses
5	0.853	Variance of $\boldsymbol{X_s}[6]$	TCR
6	0.838	Variance of $\boldsymbol{X_s}[3]$	duration
7	0.851	Mean of $\boldsymbol{X_s}[5]$	TC
8	0.848	Variance of $\boldsymbol{X_s}[5]$	TC
9	0.846	Mean of $\boldsymbol{X_s}[4]$	energy
10	0.836	Mean of $\boldsymbol{X_s}[3]$	duration
11	0.844	Mean of $\boldsymbol{X_s}[2]$	peak-to-peak amplitude
12	0.827	Variance of $\boldsymbol{X_s}[2]$	peak-to-peak amplitude

Table 5.4: Summary of the importance of the features in the set $\boldsymbol{X_s}$. They are ranked from the most important (1) to the least important (12). The number in bold indicates the highest AUC obtained for this feature set.

Rank	AUC	Feature	Rank	AUC	Feature
1	0.750	Mean of $\boldsymbol{X_{st}}[1]$	13	0.690	Variance of $\boldsymbol{X_{st}}[2]$
2	0.756	Mean of $\boldsymbol{X_{sw}}[1]$	14	0.736	Mean of $\boldsymbol{X_{sw}}[4]$
3	0.769	Variance of $\boldsymbol{X_{sw}}[1]$	15	0.707	Mean of $\boldsymbol{X_{sw}}[6]$
4	0.756	Variance of $\boldsymbol{X_{st}}[6]$	16	0.703	Mean of $\boldsymbol{X_{st}}[2]$
5	0.751	Mean of $\boldsymbol{X_{sw}}[5]$	17	0.705	Variance of $\boldsymbol{X_{st}}[1]$
6	**0.831**	Mean of $\boldsymbol{X_{st}}[6]$	18	0.700	Variance of $\boldsymbol{X_{st}}[5]$
7	0.805	Mean of $\boldsymbol{X_{sw}}[3]$	19	0.692	Variance of $\boldsymbol{X_{st}}[3]$
8	0.772	Mean of $\boldsymbol{X_{st}}[5]$	20	0.691	Variance of $\boldsymbol{X_{sw}}[3]$
9	0.767	Variance of $\boldsymbol{X_{sw}}[6]$	21	0.679	Variance of $\boldsymbol{X_{sw}}[5]$
10	0.766	Variance of $\boldsymbol{X_{sw}}[4]$	22	0.679	Mean of $\boldsymbol{X_{st}}[4]$
11	0.757	Mean of $\boldsymbol{X_{st}}[3]$	23	0.686	Mean of $\boldsymbol{X_{sw}}[2]$
12	0.751	Variance of $\boldsymbol{X_{st}}[4]$	24	0.683	Variance of $\boldsymbol{X_{sw}}[2]$

Table 5.5: Summary of the importance of the features in the set $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$. They are ranked from the most important (1) to the least important (24). The number in bold indicates the highest AUC obtained for this feature set. The index in the brackets refers to each individual feature as described in Table 5.4.

numbers in square brackets indicate the feature index. A brief description of each feature is also shown in Table 5.4. Both tables indicate that the most important feature for the classification task is the average number of pulses. Using only this feature gives a classification AUC score of 0.795 with the $\boldsymbol{X_s}$ set and 0.756 with the $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$ set. Most importantly the tables show that using all features for classification degrades the performance. In particular, using the means of the features with indices 1 and 6 and the

variances of 1 and 4 from the $\boldsymbol{X_s}$ set achieves an AUC that is 3.6% higher than when using all features (0.857 over 0.827). Table 5.5 shows that using only 6 out of the 24 features of $\boldsymbol{x_p}$ achieves a 21.7% higher AUC of 0.831, compared to 0.683 obtained when using all 24. The results also indicate that the peak-to-peak amplitude feature is the least relevant for this classification task. Given these observations, the top 4 features from the Table 5.4 and the top 6 features from the Table 5.5 are used to construct the vectors $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ and $\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$ respectively. These vectors are used in the experiments to follow.

In comparison to this Chapter's work, the methodology followed in Chapter 3 looked at longer time scales (20 seconds was used in the experiments) and used an 11-dimensional statistical feature vector parameterisation for the classification of $s_{i,j}(n)$. This vector was extracted from the feature distributions formed by considering all the pulses detected in $s_{i,j}(n)$. Using LDA and QDA the highest S_c was 0.476 and was obtained with the duration feature (see Figure 3.8). In comparison, the highest S_c scores obtained for the results in Tables 5.4 and 5.5 were 0.646 and 0.645 respectively. It is clear therefore, that using the pulse features obtained from either the stride-synchronous or the stance and swing-phase-synchronous representations produces better classification performances. However, in the work of Chapter 3 the aim was to investigate various feature representations of the knee sound signals and quantify their discriminant capabilities individually. In the current Chapter, the focus is to incorporate the spatio-temporal information obtained by gait detection and investigate whether it helps or deteriorates the knee sound analysis and detection of OA. The significant improvement in the performance, approximately 36% in S_c, obtained by using either stride-synchronous features $\boldsymbol{X_s}$ or stride-phase-synchronous features $\{\boldsymbol{X_{st}}, \boldsymbol{X_{sw}}\}$, highlights the benefit of analysing the knee sound signals on the basis of acoustic pulse detection in relation to the gait biomechanical variables (gait phases).

5.5 Contact instant time error analysis

The signal segmentation method used in this Chapter so far is based on the CIs derived from the force plate signals. However, the expensive equipment needed for the acquisition of these signals is only available within the controlled settings of a laboratory (for further

information on the topic read Section 4.2.1). As described in Chapter 1, one of the aims of this thesis work is to investigate the possibility of OA detection during dynamic functional activities outside the clinical laboratory. For this reason, ASDSA was developed in Chapter 4 as an alternative method for CI estimation. The algorithm requires only the knee sound signal picked up by a single microphone attached at the patella. In the algorithm's evaluation (see sections 4.5 and 4.6.1) it was shown that a small bias of 7.87% and -0.46% (as a percentage of the stride duration) exists in the Initial Contact Instant (ICI) and Last Contact Instant (LCI) estimates respectively (Figure 4.26). Therefore, prior to applying ASDSA for stride detection and segmentation with subsequent feature modelling, an analysis must be conducted in order to investigate the effect of the time bias on the classification performance.

To this respect, an experiment is designed in which a time bias is added to both ground truth ICIs and LCIs and the classification performance is then computed using $\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$. The experiment investigates whether precise stride phase segmentation is necessary to obtain high classification results or whether dividing the signal into segments of length equivalent to stance/swing and at points irrelevant to the gait phases, suffices or even generates better results.

The bias values that are tested, range from -50% to 50% in steps of 1%, computed as a percentage of the average stride duration that is found from the force plate signals. This range ensures that all possible segmentation points for stance/swing within a single stride are covered. The added bias is constant to all of the CIs which ensures that the detection accuracy, ϱ (defined in Section 4.5.3), is 0. The same 3-fold cross-validation procedure with 10,000 repetitions is employed as before and the results are displayed in Figure 5.3 which shows the AUC and S_c vs added bias.

It can be clearly seen that within the range -29% to -10% of bias value added, the metric scores are higher than at the 0% bias case. An improvement of 1.17% to 11.17% is observed in AUC, where the highest value of 0.924 is obtained with a bias of -13%. This is similar to the best AUC of 0.921 obtained with the MFCC based feature set $\boldsymbol{\Phi}_{\boldsymbol{M}}$ and a linear kernel SVM, as was shown in Section 3.3.3 IV of Chapter 3. The

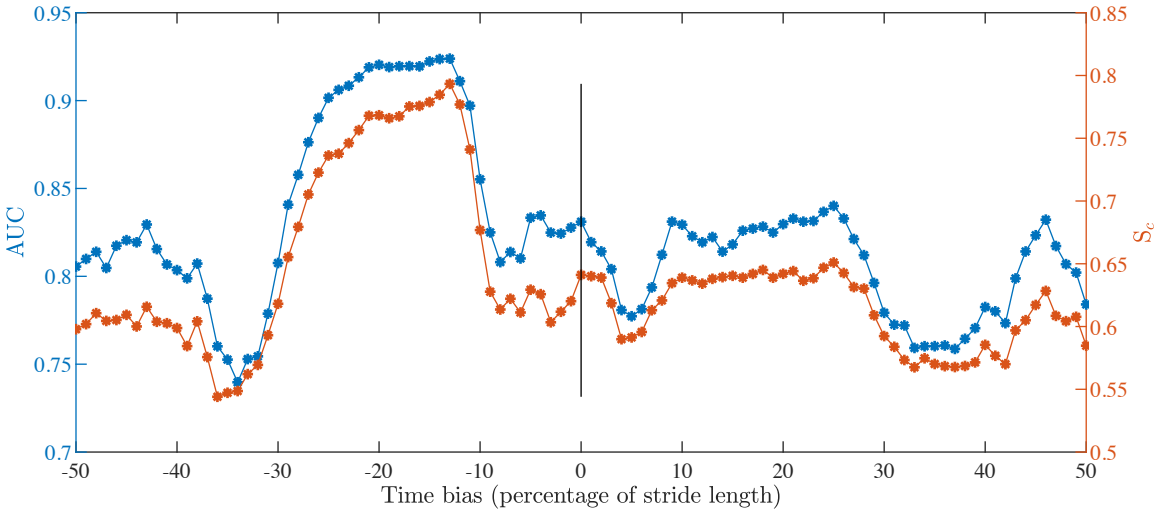

Figure 5.3: Contact instant time error analysis results vs time bias added to each ICI and LCI. The bias is presented as a percentage of stride length. The vertical (solid) line indicates 0% bias.

S_c performance line displayed in the same plot of Figure 5.3, shows an improvement compared to the no bias case that ranges from 2.21% to 23.74% and within the same bias range as for AUC. A peak in the performance is observed at a bias of -13%, at which point a maximum of 0.793 is achieved, with the respective individual components being $[\overline{\mathrm{E}}_r, \overline{\mathrm{F}}_{0.5}, \overline{\mathrm{MCC}}] = [0.164, 0.883, 0.660]$. The S_c of 0.793 is a significant improvement of 66.7% over the S_c of 0.476 that was obtained using only the pulse duration feature, represented by the 11-dimensional statistical vector $\boldsymbol{f}$, in Chapter 3.

The most significant observation that can be derived from Figure 5.3 is that not all pulses are equally important for the classification task. This suggests that there are individual pulses which are either more descriptive of OA (i.e. OA pulse signatures) or are indicative of the absence of OA, that is, if a pulse event, or events, with particular characteristics is emitted by a knee joint, then the joint is definitely not OA. However, the first possibility does not preclude the second. The above are motivated by the significant differences observed in the metric values (Figure 5.3) that are obtained as the ICI and LCI locations vary within the stride. The significantly large range (0.544 to 0.793 in S_c and 0.740 to 0.924 in AUC) can only be attributed to the particular pool of pulses that fall within the analysis segments in each bias case. This is because the features, the classifier and the repetitive cross-validation procedure, that as was shown in Figure 5.2 also ensures

the stability and repeatability of the results, are all fixed parameters in the experiment. Therefore, the only change is the location of the segmentation points that define the stance and swing phases. The metric values obtained within the range -29% to -10% of added bias, are better than any other bias case. This leads to the conclusion that the pulse events most important for the classification task, occur during both the stance and the swing phase and that their discriminant power is better utilised when these events fall in the same analysis segment.

In summary, the results obtained in this section further support the use of the AE pulses and their waveform characteristics as a means to discriminate between healthy and OA knees. It is further highlighted here, from a classification point of view, the benefit of modelling the stride phases of the knee sound signal instead of analysing it over longer time scales.

5.6 Evaluation using ASDSA

In this section, ASDSA with the post processing algorithm will be employed in order to compare the classification performance against the results obtained when ground truth segmentation is used. Furthermore, the larger database that was used in Chapter 3 will also be employed in order to examine the generalisability of the methodology developed in Section 5.4.1 and compare to the methodology proposed in Chapter 3. For the aforementioned experiments, SVM hyperparameter optimization will be conducted and the results will be reported based on the set of parameters that generated the best classification performance for each method.

5.6.1 Experimental setup

The following classification experiment will be conducted on two databases, $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ where $\boldsymbol{B_1} \subset \boldsymbol{B_2}$. Database $\boldsymbol{B_1}$ consists of 11 healthy knees and 7 OA knees from which 63 and 36 $s_{i,j}(n)$ 20 second signal segments can be obtained respectively. This is the database that was used in this Chapter's experiments so far. Database $\boldsymbol{B_2}$ (used in Chapter 3)

consists of 19 healthy knees and 21 OA knees from which 249 and 297 $s_{i,j}(n)$ are obtained respectively. Force plate data is only available for the signals in $\boldsymbol{B_1}$ and therefore, for stride detection and segmentation of the signals in $\boldsymbol{B_2}$ only ASDSA can be used. The number of observations in each database is governed by the segment duration parameter τ_s which was set equal to 20 seconds as in Chapter 3, see Section 3.2.1 for the reasoning behind this value.

5.6.2 Methods

For the evaluation, a k-fold cross-validation procedure is employed [96], with $k = 3$ for database $\boldsymbol{B_1}$ and $k = 5$ for $\boldsymbol{B_2}$. Appropriate scaling of the features in the training and test folds is conducted because SVM is not scale invariant [203]. For further details and in particular to see how the data is split, see Section 5.4.2 for the $k = 3$ case and Section 3.2.3 for $k = 5$. The cross-validation procedure is repeated as described in Section 5.4.2. In the experiment of this section several segmentation methods are compared along with SVM hyperparameter tuning for each method. Therefore, repeating the cross-validation procedure 10,000 times as suggested in Section 5.4.2 requires significant computational resources and time. To avoid this, the number of repetitions is reduced to 1000 at the expense of marginally increasing the variability in the final result (see Figure 5.2). This will affect the average metric values in their third d.p. but the second d.p. value can still be considered as accurate with an error of ± 1 unit.

The linear kernel SVM with $C = 1$ was successfully used in this Chapter for classifying normal and abnormal knee signals and showed significantly high results, see for example Figure 5.3. However, there is no guarantee that it is the optimal kernel type to use for this task and for this data. There are several other kernels that can be employed, see [204] for a list of available kernels and an analysis of their data transformation properties. With appropriate tuning of the associated kernel parameters it could possibly lead to better results. Therefore, SVM hyperparameter optimization is examined in the following.

For the kernel type $K(\cdot, \cdot)$, three different functions are considered [204]:

Parameter	Description	Values
$K(\cdot,\cdot)$	Kernel type	$K_1,\ K_2,\ K_3$
C	Penalty parameter	{0.01, 0.1, 0.5, 1, 2, ..., 20, 30, 40, ..., 100, 1000}
γ	Kernel coefficient	$\{5,\ 2,\ 1/L,\ 10^n\}$ for $n = 1, 0, \ldots, -6$
d	Polynomial order	{2, 3, 4, 5}

Table 5.6: SVM hyperparameters to be optimized. L corresponds to the number of features used in the classification.

1. Linear (K_1):

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle \tag{5.8}$$

2. Gaussian RBF (K_2):

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(-\gamma \left\| \langle \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \right\|^2) \tag{5.9}$$

3. Polynomial (K_3):

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle + 1)^d \tag{5.10}$$

where $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$ denotes the inner product between the two training feature vectors. To this date, there is no scientifically proven optimization technique for finding the best kernel as it depends on the particular data and application. In this work, the cross-validation procedure is conducted in conjunction with grid search for choosing the most suitable kernel along with the best values for the set of hyperparameters $\Lambda = \{C, \gamma, d\}$. The penalty term C is common to all kernels and acts as a regularization parameter as described in Section 5.2. Parameter γ is the RBF kernel coefficient. Technically, it is the inverse of the standard deviation of the RBF kernel and defines how much influence a single training example has on other examples. With increasing γ, the variance of the Gaussian function becomes smaller and therefore, examples must be close together (i.e. a dense region) in order to be considered similar. Parameter d is the polynomial order. With increasing d values the SVM decision boundary becomes more complex and requires more data to avoid overfitting and more time to train. The d values chosen for the experiment are 2 to 5. Table 5.6 summarises the SVM hyperparameters and the range of values to be tested. Each appropriate combination is constructed for grid search.

Three different segmentation methods are considered: (a) Ground truth segmentation using the force plate signals, (b) ASDSA with post processing and (c) Fixed frame-size with duration $\tau_s = 20$ seconds. Note that the fixed frame-size approach divides $s_i(n)$ into $s_{i,j}(n)$ of duration τ_s and methods (a) and (b) further divide $s_{i,j}(n)$ intro strides and stride phases as explained in Section 5.4.1. Method (a) is tested on database $\boldsymbol{B_1}$ with the feature vectors $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ and $\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$. Method (b) is tested on both databases ($\boldsymbol{B_1}, \boldsymbol{B_2}$) using both $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ and $\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$. For method (c), the framework developed in Chapter 3 is employed. In this context, let matrix $\boldsymbol{A}$ denote the collection of all the pulse features obtained from a single $s_{i,j}(n)$. Each matrix row holds the values of the 5 features, for a single AE pulse. Following the notation of Chapter 3, let $\boldsymbol{\phi_A} = \mathcal{D}(\boldsymbol{A}) = [\boldsymbol{f}_1^A, \boldsymbol{f}_2^A, \ldots, \boldsymbol{f}_5^A]$ be the feature matrix obtained for $s_{i,j}(n)$ where $\mathcal{D}(\cdot)$ is the matrix operator defined in Section 3.2.2. Hence $\widetilde{\boldsymbol{\phi}}_{\boldsymbol{A}} = \left[N_p, \mathrm{vec}(\boldsymbol{\phi_A})^T\right]^T$ is the vector that represents $s_{i,j}(n)$, where N_p is the number of rows of matrix $\boldsymbol{A}$, i.e. number of pulses.

5.6.3 Results and discussion

The average cross-validation metric scores obtained for each method along with the set of SVM hyperparameters that gave these scores are shown in Table 5.7. The displayed kernel type and C, γ and d where appropriate, comprise the hyperparameter set of the SVM model that gave the best results for each alternative method. The choice was made based on the ranking in both the AUC and $\mathrm{S_c}$ that is, the average of the two rankings is taken. This is because it is preferred to choose the model that performs well in all metrics rather than a model which scores well in one.

It can be clearly seen from Table 5.7 that classifying using the stride-based representation $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ achieves higher AUC compared to using the AE pulse based features $\widetilde{\boldsymbol{\phi}}_{\boldsymbol{A}}$ obtained from a longer time scale 20 seconds. This is supported by the results with both databases. For the smaller database, $\boldsymbol{B_1}$, ASDSA with post processing and using $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ achieves AUC= 0.84 over 0.77 obtained with $\widetilde{\boldsymbol{\phi}}_{\boldsymbol{A}}$, a 9.1% difference. Compared to the ground truth segmentation, the proposed CI detection algorithm generates comparable results for $\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$ but for $\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$ it generates an AUC that is 12% lower. The results obtained

Segmentation method	Database	Feature vector	$\overline{\mathrm{E}}_r$	$\overline{\mathrm{F}}_{0.5}$	$\overline{\mathrm{MCC}}$	$\overline{\mathrm{S}}_\mathrm{c}$	$\overline{\mathrm{AUC}}$	Kernel	C	d	γ
Ground truth	$\boldsymbol{B_1}$	$\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$	0.277	0.776	0.441	0.647	0.865	K_1	90	-	-
Ground truth	$\boldsymbol{B_1}$	$\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$	0.266	0.783	0.452	0.656	0.855	K_1	0.1	-	-
ASDSA with post processing	$\boldsymbol{B_1}$	$\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$	0.277	0.779	0.433	0.645	0.840	K_1	20	-	-
ASDSA with post processing	$\boldsymbol{B_1}$	$\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$	0.324	0.741	0.302	0.573	0.761	K_1	100	-	-
ASDSA with post processing	$\boldsymbol{B_2}$	$\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$	0.218	0.725	0.529	0.679	0.864	K_3	18	2	-
ASDSA with post processing	$\boldsymbol{B_2}$	$\widetilde{\boldsymbol{x}}_{\boldsymbol{p}}$	0.269	0.650	0.437	0.606	0.804	K_3	0.1	2	-
Fixed (20 seconds)	$\boldsymbol{B_1}$	$\widetilde{\boldsymbol{\phi}}_{\boldsymbol{A}}$	0.248	0.795	0.466	0.671	0.770	K_1	2	-	-
Fixed (20 seconds)	$\boldsymbol{B_2}$	$\widetilde{\boldsymbol{\phi}}_{\boldsymbol{A}}$	0.240	0.702	0.525	0.662	0.848	K_2	11	-	5

Table 5.7: Average cross-validation results per feature set. The over-line denotes the metric value averaged over the number of cross-validation executions.

in both databases with ASDSA suggest that classifying using the stride phase representation is sensitive to the bias of the segmentation instants. Nevertheless, the results obtained with ground truth segmentation show that stride phase representations can also achieve good classification performance (AUC = 0.855) as with the stride-based approach (AUC = 0.865).

Having more data improves the results as can be seen for both the ASDSA segmentation method (using either feature vector) and the fixed segmentation approach using Chapter's 3 proposed statistical feature parameterisation. The performance improvement is expected for algorithms that are consistent in their function and for class distributions that are effectively represented in both the smaller and the larger databases. Therefore, the results show that ASDSA with the post processing method have consistent performance in ICI and LCI detection.

Regarding the chosen model hyperparameters, it can be clearly seen in Table 5.7 that K_1 always gives the best results for the smaller database ($\boldsymbol{B_1}$) which supports the kernel choice for the experiments carried out in the previous sections. On the contrary, for the larger database ($\boldsymbol{B_2}$) the K_2 and K_3 kernels generate hyperplanes that better separate the data. These are more complex kernels and in general require more training data than the linear kernel in order to avoid overfitting. However, if K_2 and K_3 are properly tuned they can produce better classification performance, as shown in Table 5.7. In this respect, the results for $\boldsymbol{B_2}$ showed that the best K_1 model, using any segmentation method, is ranked 31^{st} in terms of the performance, with an S_c that is 5.5% lower than the best non-linear kernel based model, of the same method.

5.7 Conclusion

This Chapter presented an analytical study of the AE pulse events emitted by the knee joint during walking and investigated their discriminant properties for the classification task of normal (healthy) vs abnormal (OA) knee joint signals in relation to the main gait events (stride, stance phase and swing phase). For this purpose, stride-synchronous

and stride-phase-synchronous representations were proposed and several experiments were carried out to test the sensitivity of this approach to the stride segmentation points.

Among the several AE pulse waveform features that were investigated based on statistical distributions in Section 5.3, the peak-to-peak amplitude, duration and number of TC were found as the most discriminant factors between healthy and OA knees in any stride phase, in terms of the statistical differences between the distribution means. However, from a classification perspective, it was found using SVM-RFE that the number of TC, TCR and energy, together with the number of pulse events are the most important features, with the peak-to-peak amplitude being the least relevant (Tables 5.4 and 5.5).

The experiment in Section 5.5 showed that the AE pulse events that are most important for the classification task occur in both the stance and swing phases and was found that their discriminant power is better utilised when these events fall in the same analysis segment (Figure 5.3). More importantly, the results indicated that not all pulses are equally important for the classification task. Given that audible joint sounds during walking have not been previously explored for OA detection, these are important findings.

The significance of the experimental results presented in Section 5.6 is that they show the benefit of segmenting the knee signals into strides with subsequent AE pulse feature extraction compared to using larger analysis segments (20 seconds). In particular, by employing ASDSA with post processing for segmentation and using only 4 features ($\widetilde{\boldsymbol{x}}_{\boldsymbol{s}}$), compared to 56 ($\widetilde{\boldsymbol{\phi}}_{\boldsymbol{A}}$) for the latter method, achieved an average AUC of 0.864 that is 1.9% higher and an average S_c of 0.679 (2.6% higher) with corresponding components $[E_r, F_{0.5}, MCC] = [0.218, 0.725, 0.529]$. The 0.218 is the lowest error rate achieved by any of the methods of Table 5.7. In addition, this experiment also showed the benefit of using ASDSA in practice and when the force plate data is not available.

Extra care was taken in all the experiments to reduce the risk of overfitting the training data. For this reason, k-fold cross-validation procedure was employed. In addition, by repeating this procedure ensures that the variance of the SVM model is reduced, as can bee seen in Figure 5.2. Also, hyperparameter tuning (includes a regularization term) using grid search was conducted to construct suitable SVM models for the data.

Chapter 6

Conclusion

THIS thesis addressed the audible sounds emitted by the knee joint during walking, captured by a single microphone attached at the patella, and investigated whether these sounds can be used for the detection of OA and if discriminant information can be effectively extracted from such signals. Several methods have been presented for this purpose which focused on investigating features from the time, frequency, cepstral and modulation domains. These methods showed that healthy and OA knees can be discriminated with high accuracy (accuracy= 86.1%, AUC= 0.92) and indicated that they could potentially be used to enable knee joint acoustic assessment outside the lab using a quick and cost effective approach. Furthermore, these methods could potentially facilitate effective intervention toward knee OA management. This can reduce the costs of clinical testing by improving the selection of patients that require further tests (e.g. with MRI).

Firstly, a framework was developed to test the hypothesis that the sounds emitted by knees with OA are statistically different from those emitted by clinically healthy knees (Section 3.2). Among various features investigated, the magnitude of the STFT spectrum ($\boldsymbol{\Psi_F}$) and the modulation magnitude spectrum ($\boldsymbol{\Psi_{FF}}$) were shown to have the most statistically different coefficients between the two classes (Figure 3.6). Tests were conducted using the two-sample Kolmogorov-Smirnov Test and the Bhattacharyya distance which was also extended to cover univariate gamma distributions in (3.22).

Next, an 11-dimensional statistical parameterisation was proposed for the feature

distributions formed by the two knee condition classes (Section 3.2.2). Subsequent classification experiments highlighted the effectiveness of this parameterisation for capturing differences between OA and healthy knees (Section 3.2.3 and Section 3.3.3). In particular, it was found that the most discriminant features are the low order MFCCs (2^{nd} and 7^{th}) and the STFT features that correspond to the frequency bands 220 Hz to 420 Hz and 1 kHz to 3.4 kHz (Table 3.2 and Section 3.3.3). From the methods that have been studied, the highest accuracy achieved was 86.1% (Table 3.3) using a combination of MFCC and STFT based features represented by the proposed 11-dimensional statistical parameterisation.

Various experiments were carried out to investigate the effect on the classification performance of the analysis frame-size and the number of filters in the filter-bank for the extraction of spectrum and cepstrum features (Section 3.3.3). The results showed that the frame-size is a classification performance trade-off and that it is better, from a classification point of view, to use mel-frequency spaced filters over linear-spaced ones for extracting the cepstral coefficients and for reducing the dimensionality of the STFT spectrum. In addition, the results suggested that a combination of linear and mel scaled filters in the filter-bank could capture even more discriminant information than either two separately.

The work was then focused at developing stride detection methods that use the knee sound signal alone. The investigation lead to the development of ASDSA (Section 4.4), a dynamic programming based algorithm that can efficiently detect the ICIs and LCIs with a detection rate of 96.88% and 96.12% respectively and a corresponding time bias error of only 5.86% and 2.25%, as a percentage of the stride duration (Figure 4.26). Various experiments were conducted to study the behaviour of the ASDSA cost function and its components (Section 4.5). It was found that as the algorithm traverses through the signal, the candidate selection can sometimes deviate from the true CIs, defined by the vertical component of theGRF force plate signal (Section 4.5.2). This is due to the similar penalisation of candidates that are close to each other which is in turn a direct effect of the large number of generated candidates from which ASDSA must choose the set that minimizes (4.20). Over time (within the 20 second analysis segment) this deviation can result to the selection of ICIs that fall midway between the true ICI and LCI or are

close to the true LCI location which directly affects the LCI estimate (Figure 4.22). This effect degrades the overall time accuracy metric of the estimates in the entire database giving detection accuracies of $\varrho = 15.64\%$ and $\varrho = 15.68\%$ for ICI and LCI detection respectively (Figure 4.26). For this reason, a post processing method based on cross correlating the estimates with a global template was then proposed (Section 4.6). This template is constructed from the ensemble average of the ASDSA estimates (ICI and LCI separately). It was demonstrated that the ICI time accuracy is improved by 28.6% and the respective for LCIs is improved by 20.5% (Figure 4.26). It is concluded that for best results, ASDSA together with the post processing algorithm should be used.

The ASDSA algorithm (with the post processing method) enabled the study of the knee sounds in relation to spatiotemporal information, namely, the gait pattern (Chapter 5). In this context, stride-synchronous and stride-phase-synchronous representations of the knee signals were proposed using the AE pulse events (Section 5.4). Among the several pulse waveform characteristics that were analysed, the energy, the number of TC and TCR together with the number of pulse events per stride phase were found to be the most important features for classification (Tables 5.4 and 5.5). By employing ASDSA as the segmentation method, it was demonstrated in Section 5.6 that using the stride-synchronous approach improves the classification accuracy by 9.2% compared to extracting the pulse features over a longer time scale (20 seconds) and using the 11-dimensional statistical parameterisation (Table 5.7). In addition, the results in Section 5.5 showed that there are specific AE pulse events which are more important for classification and that these occur in both the stance and the swing phase during walking.

6.1 Future research

Several fruitful directions for future research can be identified from the conclusions described above. For example, the combination of features from different domains was not extensively explored for classification as the research interest was focused at investigating the feature domains independently in order to identify which ones contained more descriptive information of OA. Therefore, feature domain combination can be an interesting

topic for future research.

Given the results and the insights derived from the work in Section 3.3.3 it follows that it will be interesting to examine whether an adaptive filter-bank can be beneficial for the analysis. This can be a topic for future research.

Furthermore, the spread of values in the number of AE pulses emitted by a knee, as shown in Figure 5.1 provides an insight that a trend may exist with changing speed. This is because sound recordings from various walking speeds were included in the experiment. However, this relationship was not investigated in this work simply because there aren't enough samples to sufficiently represent the spread of walking speeds in the data. Nevertheless, it is an interesting question that is worth exploring in the future when more data becomes available.

Moreover, given the results in Section 5.5 an interesting future direction could be to find the specific AE pulses, using the waveform characteristics introduced in this thesis, that relate to OA in the knee (i.e. OA pulse signatures) or the absence of OA as suggested in the same section.

Finally, the ASDSA algorithm opens up many potential applications for future research, for example, a possible direction is to extend the algorithm for multi-microphone data. More than one microphones can be attached at different positions on the knee and an interesting work is to explore whether fusing this data with the ASDSA algorithm improves the detection performance. In addition, the algorithm permits gait analysis with low time bias error without the need for expensive and bulky equipment as is the case with the state of the art force plate based systems. Therefore, it will be interesting to examine the consistency of the estimations for this application in a larger database. Adding to the above, given that ASDSA with the post processing method generate the best results, it will be interesting to investigate the fusion of these two methods by, for example, making the post processing method into an additional cost component in ASDSA.

The above are only a few interesting suggestions for further research primarily on the thesis topic but also in other fields as well. The results and the findings described in all of the thesis Chapters can spark many more exciting research directions.

Bibliography

[1] R. Bitton, "The economic burden of osteoarthritis," *American Journal of Managed Care*, vol. 15, pp. 230—235, 2009.

[2] I. J. Wallace, S. Worthington, D. T. Felson, R. D. Jurmain, K. T. Wren, H. Maijanen, R. J. Woods, and D. E. Lieberman, "Knee osteoarthritis has doubled in prevalence since the mid-20th century," *Proceedings of the National Academy of Sciences*, vol. 114, no. 35, pp. 9332–9336, 2017.

[3] "Osteoarthritis in general practice: Data and perpectives," Arthritis Research UK, Tech. Rep., 2013.

[4] A. O. Amoako and G. G. A. Pujalte, "Osteoarthritis in young, active, and athletic individuals," *Clinical Medicine Insights: Arthritis and Musculoskeletal Disorders*, vol. 7, pp. 27–32, 2014.

[5] D. Bhatia, T. Bejarano, and M. Novo, "Current interventions in the management of knee osteoarthritis," *Journal of pharmacy & bioallied sciences*, vol. 5, pp. 30—38, 2013.

[6] A. Chen, C. Gupte, K. Akhtar, P. Smith, and J. Cobb, "The global economic cost of osteoarthritis: How the uk compares," *Arthritis*, vol. 2012, 2012.

[7] M. Hiligsmann and J. Reginster, "The economic weight of osteoarthritis in europe," *Medicographia*, vol. 35, pp. 197–202, 2013.

[8] A. Litwic, M. H. Edwards, E. M. Dennison, and C. Cooper, "Epidemiology and burden of osteoarthritis," *British Medical Bulletin*, vol. 105, no. 1, pp. 185–199, Jan. 2013.

[9] C. Palazzo, C. Nguyen, M.-M. Lefevre-Colau, F. Rannou, and S. Poiraudeau, "Risk factors and burden of osteoarthritis," *Annals of Physical and Rehabilitation Medicine*, vol. 59, no. 3, pp. 134–138, 2016.

[10] U. S. Bone and J. Initiative, *The Burden of Musculoskeletal Diseases in the United States*, 3rd ed., Rosemont, IL., 2014.

[11] N. Lane and D. J. Walace, *All about osteoathritis: the definitive resource for arthritis patients and their families*. Oxford University Press, 2002.

[12] M. Mahesh, "Fluoroscopy: Patient radiation exposure issues," *RadioGraphics*, vol. 21, no. 4, pp. 1033–1045, 2001.

[13] D. Hunter, A. Guermazi, F. Roemer, Y. Zhang, and T. Neogi, "Structural correlates of pain in joints with osteoarthritis," *Osteoarthritis and Cartilage*, vol. 21, no. 9, pp. 1170—1178, Sep. 2013.

[14] G. E. Gold, "Dynamic and functional imaging of the musculoskeletal system," *Semin Musculoskelet Radiol*, vol. 7, pp. 245–248, 2003.

[15] D. Schiphof, M. van Middelkoop, B. de Klerk, E. Oei, A. Hofman, B. Koes, H. Weinans, and S. Bierma-Zeinstra, "Crepitus is a first indication of patellofemoral osteoarthritis (and not of tibiofemoral osteoarthritis)," *Osteoarthritis and Cartilage*, vol. 22, no. 5, pp. 631–638, 2014.

[16] R. A. Kapur, "Acoustic emission in orthopaedics: A state of the art review," *Journal of Biomechanics*, vol. 49, no. 16, pp. 4065–4072, 2016.

[17] F. Flandry and G. Hommel, "Normal anatomy and biomechanics of the knee," *Sports Medicine and Arthroscopy Review*, vol. 19, pp. 82–92, Jun. 2011.

[18] R. P. Welsh, "Knee joint structure and function," *Clinical orthopaedics and related research*, vol. 147, pp. 7–14, 1980.

[19] J. B. Morrison, "The mechanics of the knee joint in relation to normal walking," *J. Biomechan.*, vol. 3, pp. 51–61, 1970.

[20] J. M. Jordan, C. G. Helmick, J. B. Renner, G. Luta, A. D. Dragomir, J. Woodard, F. Fang, T. A. Schwartz, L. M. Abbate, L. F. Callahan, W. D. Kalsbeek, and M. C. Hochberg, "Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in african americans and caucasians: the johnston county osteoarthritis project," *The Journal of Rheumatology*, vol. 34, no. 1, pp. 172–180, 2007.

[21] C. E. Chaisson, D. R. Gale, E. Gale, L. Kazis, and D. T. Skinner, K. Felson, "Detecting radiographic knee osteoarthritis: what combination of views is optimal?" *Rheumatology*, vol. 39, pp. 1218—1221, 2000.

[22] J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the Rheumatic Diseases*, vol. 16, no. 4, pp. 494–502, 1957.

[23] M. D. Kohn, A. A. Sassoon, and N. D. Fernando, "Classifications in brief: Kellgren-Lawrence classification of osteoarthritis," *Clinical Orthopaedics and Related Research*, vol. 474, no. 8, pp. 1886–1893, Aug. 2016.

[24] L. Moreira, N. G. aes, and N. Azevedo, "8 - imaging strategies for bioinspired materials," in *Bioinspired Materials for Medical Applications*. Woodhead Publishing, 2017, pp. 215–239.

[25] P. M. M. Cashman, R. I. Kitney, M. A. Gariba, and M. E. Carter, "Automated techniques for visualization and mapping of articular cartilage in mr images of the osteoarthritic knee: a base technique for the assessment of microdamage and submicro damage," *IEEE Transactions on NanoBioscience*, vol. 99, no. 1, pp. 42–51, Mar. 2002.

[26] A. Guermazi, F. Roemer, and H. Genant, "Role of imaging in osteoarthritis: diagnosis, prognosis and follow-up," *Medicographia*, vol. 35, pp. 164—171, 2013.

[27] F. W. Roemer, F. Eckstein, D. Hayashi, and A. Guermazi, "The role of imaging in osteoarthritis," *Best Practice & Research Clinical Rheumatology*, vol. 28, no. 1, pp. 31–60, 2014.

[28] W. E. Blodgett, "Auscultation of the knee joint," *Boston Med. Surg.*, vol. 146, no. 3, pp. 63–66, 1902.

[29] E. Bircher, "Zur diagnose der meniscusluxation und des meniscusabrisses," *Zentralbl. Chir.*, vol. 40, pp. 1852–1857, 1913.

[30] S. Tavathia, R. M. Rangayyan, C. B. Frank, G. D. Bell, K. O. Ladly, and Y. Zhang, "Analysis of knee vibration signals using linear prediction," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 9, pp. 959–970, 1992.

[31] A. Steindler, "Auscultation of joints," *J. Bone Joint Surg.*, vol. 19, pp. 121–124, 1937.

[32] H. Fischer and E. W. Johnson, "Analysis of sounds from normal and pathologic knee joints," in *3rd Int. Congr. Phys. Med.*, 1960, pp. 50–57.

[33] M. L. Chu, I. A. Gradisar, L. D. Zavodney, and G. F. Bowling, "Detection of knee joint diseases using acoustical pattern recognition technique," *J. Biomechan.*, vol. 9, pp. 111–114, 1976.

[34] M. L. Chu, I. A. Gradisar, M. R. Railey, and G. F. Bowling, "An electroacoustical technique for the detection of knee joint noise," *Med. Res. Eng.*, vol. 12, no. 1, pp. 18–20, 1976.

[35] M. L. Chu, I. A. Gradisar, L. Zavodney, and G. F. Bowling, "Computer aided acoustical correlation of papathologic cartilage generated noise," in *30th Ann. Conf. Eng. Med. Biol.*, 1977, p. 175.

[36] M. L. Chu, I. A. Gradisar, and R. Mostardi, "A noninvasive electroacoustical evaluation technique of cartilage damage in pathological knee joints," *Med. Biol. Eng. Comput.*, vol. 16, pp. 437–442, 1978.

[37] S. C. Abbott and M. D. Cole, "Vibration arthrometry: A critical review," *Critical reviews in biomedical engineering*, vol. 41, pp. 223–242, Jan. 2013.

[38] R. M. Rangayyan, S. Krishnan, , G. D. Bell, C. B. Frank, and K. O. Ladly, "Analysis of vibroarthrographic signals with features related to signal variability and radial-basis functions," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 11, pp. 1068—–1074, Nov. 1997.

[39] S. Krishnan, R. M. Rangayyan, G. D. Bell, C. B. Frank, and K. O. Ladly, "Adaptive filtering, modelling and classification of knee joint vibroarthrographic signals for non-invasive diagnosis of articular cartilage pathology," *Med. Biol. Eng. Comput.*, vol. 35, no. 6, pp. 677—–684, Nov. 1997.

[40] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank, "Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 6, pp. 773–783, 2000.

[41] K. S. Kim, C. G. Song, and J. H. Seo, "Time-frequency analysis of vibroarthrographic signals for non-invasive diagnosis of articular pathology," *The Transactions of the Korean Institute of Electrical Engineers*, vol. 57, no. 4, pp. 729–734, 2008.

[42] K. S. Kim, J. H. Seo, J. U. Kang, and C. G. Song, "An enhanced algorithm for knee joint sound classification using feature extraction based on time-frequency analysis," *Comput Methods Programs Biomed*, vol. 94, pp. 198–206, 2009.

[43] Y. Wu and S. Krishnan, "An adaptive classifier fusion method for analysis of knee-joint vibroarthrographic signals," in *International Conference on Computational Intelligence for Measurement Systems and Applications*, 2009, pp. 190–193.

[44] S. Cai, S. Yang, F. Zheng, M. Lu, Y. Wu, and S. Krishnan, "Knee joint vibration signal analysis with matching pursuit decomposition and dynamic weighted classifier fusion," *Computational and Mathematical Methods in Medicine*, vol. 2013, 2013.

[45] R. M. Rangayyan and Y. F. Wu, "Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions," *Med Biol Eng Comput*, vol. 46, no. 3, pp. 223–32, Mar. 2008.

[46] K. S. Kim, S. O. Lee, J. H. Seo, and C. G. Song, "Classification of joint pathology using an acoustical analysis of knee joint sound," in *Biomedical Circuits and Systems Conference (BioCAS)*, Nov. 2006.

[47] K. S. Kim, J. H. Seo, and C. G. Song, "An acoustical evaluation of knee sound for non-invasive screening and early detection of articular pathology," *Journal of Medical Systems*, vol. 36, pp. 715–722, 2012.

[48] Y. Wu, S. Cai, S. Yang, F. Zheng, and N. Xiang, "Classification of knee joint vibration signals using bivariate feature distribution estimation and maximal posterior probability decision criterion," *Entropy*, vol. 15, pp. 1375–1387, 2013.

[49] Y. Wu, *Knee Joint Vibroarthrographic Signal Processing and Analysis*, ser. SpringerBriefs in Bioengineering. Springer, 2015.

[50] S. M. A. Al-Obaidi, M. S. Leong, R. R. Hamzah, and A. M. Abdelrhman, "A review of acoustic emission technique for machinery condition monitoring: Defects detection & diagnostic," in *Mechanical and Electrical Technology IV*, ser. Applied Mechanics and Materials, vol. 229. Trans Tech Publications, 11 2012, pp. 1476–1480.

[51] B. Mascaro, J. Prior, L.-K. Shark, J. Selfe, P. Cole, and J. Goodacre, "Exploratory study of a non-invasive method based on acoustic emission for assessing the dynamic integrity of knee joints," *Med. Eng. Phys.*, vol. 31, pp. 1013–1022, 2009.

[52] L.-K. Shark, H. Chen, and J. Goodacre, "Discovering differences in acoustic emission between healthy and osteoarthritic knees using a four-phase model of sit-stand-sit movements," *Open Med. Inform.*, vol. 4, pp. 116–125, 2010.

[53] L.-K. Shark, J. Goodacre, and H. Chen, "Knee acoustic emission: A potential biomarker for quantitative assessment of joint ageing and degeneration," *Med. Eng. Phys.*, vol. 33, pp. 534–545, 2011.

[54] H. Chen, "Discovery of acoustic emission based biomarker for quantitative assessment of knee joint ageing and degeneration," Ph.D. dissertation, University of Central Lancashire, 2011.

[55] S. Hersek, M. Baran Pouyan, C. N. Teague, M. N. Sawka, M. L. Millard-Stafford, G. F. Kogler, P. Wolkoff, and O. T. Inan, "Acoustical emission analysis by unsupervised graph mining: A novel biomarker of knee health status," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1291–1300, Jun. 2018.

[56] S. Krishnan, R. Rangayyan, G. Bell, and C. Frank, "Auditory display of knee-joint vibration signals," *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3292–3304, 2001.

[57] C. N. Teague, S. Hersek, H. Töreyin, M. L. Millard-Stafford, M. L. Jones, G. F. Kogler, M. N. Sawka, and O. T. Inan, "Novel methods for sensing acoustical emissions from the knee for wearable joint health assessment," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1581–1590, Aug. 2016.

[58] H. Töreyin, S. Hersek, C. N. Teague, and O. T. Inan, "A proof-of-concept system to analyze joint sounds in real time for knee health assessment in uncontrolled settings," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 2892–2893, May 2016.

[59] O. T. Inan, D. C. Whittingslow, C. N. Teague, S. Hersek, M. B. Pouyan, M. Millard-Stafford, G. F. Kogler, and M. N. Sawka, "Wearable knee health system employing novel physiological biomarkers," *Journal of Applied Physiology*, vol. 124, no. 3, pp. 537–547, 2018.

[60] H. Töreyin, H. K. Jeong, S. Hersek, C. N. Teague, and O. T. Inan, "Quantifying the consistency of wearable knee acoustical emission measurements during complex motions," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1265–1272, Sep. 2016.

[61] H. Jeong, M. B. Pouyan, D. C. Whittingslow, V. Ganti, and O. T. Inan, "Quantifying the effects of increasing mechanical stress on knee acoustical emissions using unsupervised graph mining," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 3, pp. 594–601, Mar. 2018.

[62] C. Teague, S. Hersek, H. Töreyin, M. L. Millard-Stafford, M. L. Jones, G. F. Kogler, M. N. Sawka, and O. T. Inan, "Novel approaches to measure acoustic emissions as biomarkers for joint health assessment," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Jun. 2015, pp. 1–6.

[63] C. Yiallourides and P. A. Naylor, "Time-frequency analysis and parameterisation of knee sounds for non-invasive detection of osteoarthritis," *IEEE Trans. Biomed. Eng.*, Jul. 2019, Submitted.

[64] C. Yiallourides, A. H. Moore, E. Auvinet, C. Van Der Straeten, and P. A. Naylor, "Acoustic analysis and assessment of the knee in osteoarthritis during walking," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 281–285.

[65] C. Yiallourides, V. Manning-Eid, A. H. Moore, and P. A. Naylor, "A dynamic programming approach for automatic stride detection and segmentation in acoustic emission from the knee," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 401–405.

[66] A. Oppenheim and R. W. Schafer, *Digital Signal Processing*, 2nd ed. Prentice-Hall Inc., Englewood Cliff, NJ, 1993.

[67] G. Giannakis, *Cyclostationary Signal Analysis*, Nov. 2009, pp. 1–32.

[68] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[69] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "Measuring time-frequency information content using the renyi entropies," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.

[70] J. W. William, L. B. Mark, and O. H. Alfred, "Uncertainty, information, and time-frequency distributions," *Proc. SPIE 1566, Advanced Signal Processing Algorithms, Architectures, and Implementations II*, vol. 1566, pp. 1566–1579, 1991.

[71] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. University of California Press, 1961, pp. 547–561.

[72] P. Flandrin, R. G. Baraniuk, and O. Michel, "Time-frequency complexity and information," in *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. iii, April 1994, pp. III/329–III/332 vol.3.

[73] L. Spain, B. Rajoub, D. Schlüter, J. Waterton, M. Bowes, L. Shark, P. Diggle, and J. Goodacre, "Biomarkers for knee osteoarthritis: New technologies, new paradigms," *Future Rheumatology*, vol. 10, no. 4, pp. 287–297, 8 2015.

[74] C. L. Hill, D. G. Gale, C. E. Chaisson, K. Skinner, L. Kazis, M. E. Gale, and D. T. Felson, "Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis," *The Journal of Rheumatology*, vol. 28, no. 6, pp. 1330–1337, 2001.

[75] E. M. Roos and L. S. Lohmander, "The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis," *Health Qual Life Outcomes*, vol. 1, 2003.

[76] T. F. Quatieri, *Discrete-Time Speech Signal Processing*. Prentice Hall, 2002.

[77] M. Kendall, *The Advanced Theory of Statistics*, 4th ed. Macmillan, 1979.

[78] W.-Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[79] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier, 2009.

[80] L. Rabiner and O. Herrmann, "The predictability of certain optimum finite-impulse-response digital filters," *IEEE Transactions on Circuit Theory*, vol. 20, no. 4, pp. 401–408, July 1973.

[81] F. Fritsch and R. Carlson, "Monotone piecewise cubic interpolation," *SIAM Journal on Numerical Analysis*, vol. 17, no. 2, pp. 238–246, 1980.

[82] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, Mar. 2003.

[83] J. A. Baglivo, *Mathematica Laboratories for Mathematical Statistics*. SIAM-ASA, 2004, ch. Linear Least Squares Analysis, pp. 205–226.

[84] K. Paliwal, K. Wojcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Communication*, vol. 52, no. 5, pp. 450–475, 2010.

[85] G. H. Seng and T. T. Swee, "Spectral coefficients system for osteoarthritis detection," *International journal of circuits, systems and signal processing*, vol. 7, pp. 151–159, 2013.

[86] G. H. Seng and T. T. Swee, "Osteoarthritis detection system using optimal dynamic feature configuration," *International journal of circuits, systems and signal processing*, vol. 7, pp. 231–239, 2013.

[87] J. Makhoul and L. Cosell, "LPCW: An LPC vocoder with linear predictive spectral warping," *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, pp. 466–469, 1976.

[88] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, Jan 1974.

[89] D. Cousineau, S. Brown, and A. Heathcote, "Fitting distributions using maximum likelihood: Methods and packages," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 4, pp. 742–756, Nov. 2004.

[90] C. Yang, Z. He, and W. Yu, "Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis," *BMC Bioinformatics*, vol. 10, no. 1, Jan. 2009.

[91] F. Scholkmann, J. Boss, and M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, Nov. 2012.

[92] F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68—78, 1951.

[93] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, p. 99–109, 1943.

[94] S. Nadarajah, "On the use of the generalised gamma distribution," *International Journal of Electronics*, vol. 95, no. 10, pp. 1029–1032, 2008.

[95] A. R. Webb and K. D. Copsey, *Statistical pattern recognition*. Wiley, 2011.

[96] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., Feb. 2009.

[97] T. J. O'Neil, "Error rates of non-bayes classification rules and the robustness of fisher's linear discriminant function," *Biometrika*, vol. 79, no. 1, pp. 177–184, 1992.

[98] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[99] N. Henze and B. Zirkler, "A class of invariant consistent tests for multivariate normality," *Communications in Statistics - Theory and Methods*, vol. 19, no. 10, pp. 3595–3617, 1990.

[100] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley and Sons, 2001.

[101] D. M. Powers, "Evaluation: From precision, recall and f-measure to ROC, informedness, markedness & correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, Feb. 2011.

[102] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000.

[103] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. New York, NY, USA: Cambridge University Press, 2000.

[104] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[105] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers," *Pattern Recognition Lett.*, vol. 31, pp. 1–38, Jan 2004.

[106] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Lett.*, vol. 27, pp. 861–874, 2006.

[107] A. Godfrey, S. D. Din, G. Barry, J. Mathers, and L. Rochestera, "Instrumenting gait with an accelerometer: A system and algorithm examination," *J. of Medical Engineering & Physics*, vol. 37, no. 4, pp. 407–407, Apr 2015.

[108] L. Chaitow and J. DeLany, "Chapter 3 - gait analysis," in *Clinical Application of Neuromuscular Techniques, Volume 2*, 2nd ed. Oxford: Churchill Livingstone, 2011, pp. 61–84.

[109] J. B. Webster and B. J. Darter, "Chapter 4 - Principles of normal and pathologic gait," in *Atlas of Orthoses and Assistive Devices*, 5th ed. Elsevier Science Publishers B. V., 2018, pp. 49–62.

[110] S. Ounpuu, "The biomechanics of walking and running," *Clinics in sports medicine*, vol. 13, pp. 843–863, 11 1994.

[111] T. L. Switaj and F. G. O'Connor, "Chapter 43 - gait analysis," in *The Sports Medicine Resource Manual*. W.B. Saunders, 2008, pp. 536–542.

[112] J. D. O'Sullivan, C. M. Said, L. C. Dillon, M. Hoffman, and A. J. Hughes, "Gait analysis in patients with parkinson's disease and motor fluctuations: Influence of levodopa and comparison with other measures of motor function," *Movement Disorders*, vol. 13, no. 6, pp. 900–906, 1998.

[113] S.-B. Koh, K.-W. Park, D.-H. Lee, S. J. Kim, and J.-S. Yoon, "Gait analysis in patients with parkinson's disease: Relationship to clinical features and freezing," *JMD*, vol. 1, no. 2, pp. 59–64, 2008.

[114] M. Sveshlik, E. B. Zwick, G. Steinwender, W. E. Linhart, P. Schwingenschuh, P. Katschnig, E. Ott, and C. Enzinger, "Gait analysis in patients with parkinson's disease off dopaminergic therapy," *Archives of Physical Medicine and Rehabilitation*, vol. 90, no. 11, pp. 1880–1886, 2009.

[115] S. Okuda, S. Takano, M. Ueno, Y. Hara, Y. Chida, T. Ikkaku, F. Kanda, and T. Toda, "Gait analysis of patients with parkinson's disease using a portable triaxial accelerometer," *Neurology and Clinical Neuroscience*, vol. 4, no. 3, pp. 93–97, 2016.

[116] P. Chung, Y. Hsu, C. Wang, C. Lin, J. Wang, and M. Pai, "Gait analysis for patients with alzheimer's disease using a triaxial accelerometer," in *2012 IEEE International Symposium on Circuits and Systems*, May 2012, pp. 1323–1326.

[117] Y. Hsu, P. Chung, W. Wang, M. Pai, C. Wang, C. Lin, H. Wu, and J. Wang, "Gait and balance analysis for patients with alzheimer's disease using an inertial-sensor-based wearable instrument," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1822–1830, Nov 2014.

[118] J. Perry and J. M. Burnfield, *Gait Analysis: Normal and Pathological Function*, 2nd ed. Slack Incorporated, 2010.

[119] J. Rueterbories, E. G. Spaich, B. Larsen, and O. K. Andersen, "Methods for gait event detection and analysis in ambulatory systems," *J. of Medical Engineering & Physics*, vol. 32, no. 6, pp. 545–552, 2010.

[120] S. Khandelwal and N. Wickstrom, "Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database," *Gait & Posture*, vol. 51, pp. 84–90, 2017.

[121] J. T. Viitasalo, P. Luhtanen, H. V. Mononen, K. Norvapalo, L. Paavolainen, and M. Salonen, "Photocell contact mat: A new instrument to measure contact and flight times in running," *Journal of Applied Biomechanics*, vol. 13, no. 2, pp. 254–266, 1997.

[122] A. H. Hansen, D. S. Childress, and M. R. Meier, "A simple method for determination of gait events," *Journal of Biomechanics*, vol. 35, no. 1, pp. 135–138, Jan 2002.

[123] N. Whelan, R. Healy, I. Kenny, and A. Harrison, "A comparison of foot strike events using the force plate and peak impact acceleration measures," in *International Conference of Biomechanics in Sports*, 06 2015, pp. 114–117.

[124] H. Rouhani, M. O. Abe, K. Nakazawa, M. R. Popovic, and K. Masani, "Heel strike detection using split force-plate treadmill," *Gait & Posture*, vol. 41, no. 3, pp. 863–866, 2015.

[125] J. Nilsson, V. P. Stokes, and A. Thorstensson, "A new method to measure foot contact," *Journal of Biomechanics*, vol. 18, no. 8, pp. 625–627, 1985.

[126] J. M. Hausdorff, Z. Ladin, and J. Y. Wei, "Footswitch system for measurement of the temporal parameters of gait," *Journal of Biomechanics*, vol. 28, no. 3, pp. 347–351, 1995.

[127] A. Willemsen, J. van Alste, and H. Boom, "Real-time gait assessment utilizing a new way of accelerometry," *Journal of Biomechanics*, vol. 23, no. 8, pp. 859–863, 1990.

[128] A. Hreljac and R. N. Marshall, "Algorithms to determine event timing during normal walking using kinematic data," *Journal of Biomechanics*, vol. 33, no. 6, pp. 783–786, 2000.

[129] S. Ghoussayni, C. Stevens, S. Durham, and D. Ewins, "Assessment and validation of a simple automated method for the detection of gait events and intervals," *Gait & Posture*, vol. 20, no. 3, pp. 266–272, December 2004.

[130] J. K. De Witt, "Determination of toe-off event time during treadmill locomotion using kinematic data," *Journal of Biomechanics*, vol. 43, no. 15, pp. 3067–3069, 2010.

[131] J. C. Handsaker, S. E. Forrester, J. P. Folland, M. I. Black, and S. J. Allen, "A kinematic algorithm to identify gait events during running at different speeds and with different footstrike types," *Journal of Biomechanics*, vol. 49, no. 16, pp. 4128–4133, 2016.

[132] J. A. Zeni, J. G. Richards, and J. S. Higginson, "Two simple methods for determining gait events during treadmill and overground walking using kinematic data," *Gait & Posture*, vol. 27, no. 4, pp. 710–714, May 2008.

[133] F. Alvim, L. Cerqueira, A. D. Netto, G. Leite, and A. Muniz, "Comparison of five kinematic-based identification methods of foot contact events during treadmill walking and running at different speeds." *Journal of Applied Biomechanics*, vol. 31, no. 5, pp. 383–388, 2015.

[134] B. D. Hendershot, C. E. Mahon, and A. L. Pruziner, "A comparison of kinematic-based gait event detection methods in a self-paced treadmill application," *Journal of Biomechanics*, vol. 49, no. 16, pp. 4146–4149, 2016.

[135] N. Eckardt and A. Kibele, "Automatic identification of gait events during walking on uneven surfaces," *Gait & Posture*, vol. 52, pp. 83–86, 2017.

[136] J. Leitch, J. Stebbins, G. Paolini, and A. B. Zavatsky, "Identifying gait events without a force plate during running: A comparison of methods," *Gait & Posture*, vol. 33, no. 1, pp. 130–132, 2011.

[137] R. E. Fellin, W. C. Rose, T. D. Royer, and I. S. Davis, "Comparison of methods for kinematic identification of footstrike and toe-off during overground and treadmill running," *Journal of Science and Medicine in Sport*, vol. 13, no. 6, pp. 646–650, 2010.

[138] E. Desailly, Y. Daniel, P. Sardain, and P. Lacouture, "Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait," *Gait & Posture*, vol. 29, no. 1, pp. 76–80, 2009.

[139] C. M. O'Connor, S. K. Thorpe, M. J. O'Malley, and C. L. Vaughan, "Automatic detection of gait events using kinematic data," *Gait & Posture*, vol. 25, no. 3, pp. 469–474, 2007.

[140] J. Morris, "Accelerometry - a technique for the measurement of human body movements," *Journal of Biomechanics*, vol. 6, no. 6, pp. 729–736, 1973.

[141] A. Mansfield and G. M. Lyons, "The use of accelerometry to detect heel contact events for use as a sensor in fes assisted walking," *J. of Medical Engineering & Physics*, vol. 25, no. 10, pp. 879–885, 2003.

[142] Y. Shimada, S. Ando, T. Matsunaga, A. Misawa, T. Aizawa, T. Shirahata, and E. Itoi, "Clinical application of acceleration sensor to detect the swing phase of stroke gait in functional electrical stimulation," *The Tohoku Journal of Experimental Medicine*, vol. 207, no. 3, pp. 197–202, 2005.

[143] J. M. Jasiewicz, J. H. Allum, J. W. Middleton, A. Barriskill, P. Condie, B. Purcell, and R. C. T. Li, "Gait event detection using linear accelerometers or angular velocity transducers in able-bodied and spinal-cord injured individuals," *Gait & Posture*, vol. 24, no. 4, pp. 502–509, 2006.

[144] J. J. Kavanagh and H. B. Menz, "Accelerometry: A technique for quantifying movement patterns during walking," *Gait & Posture*, vol. 28, no. 1, pp. 1–15, 2008.

[145] S. Khandelwal and N. Wickstrom, "Novel methodology for estimating initial contact events from accelerometers positioned at different body locations," *Gait & Posture*, vol. 59, pp. 278–285, 2018.

[146] M. Sekine, T. Tamura, T. Togawa, and Y. Fukui, "Classification of waist-acceleration signals in a continuous walking record," *J. of Medical Engineering & Physics*, vol. 22, no. 4, pp. 285–291, 2000.

[147] R. W. Selles, M. A. G. Formanoy, J. B. J. Bussmann, P. J. Janssens, and H. J. Stam, "Automated estimation of initial and terminal contact timing using accelerometers; development and validation in transtibial amputees and controls," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 81–88, March 2005.

[148] W. Zijlstra, "Assessment of spatio-temporal parameters during unconstrained walking," *European Journal of Applied Physiology*, vol. 92, no. 1-2, pp. 39–44, Jun 2004.

[149] A. Mesaros, T. Heittola, and T. Virtanen, "Tut database for acoustic scene classification and sound event detection," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug 2016, pp. 1128–1132.

[150] M. Tanaka and H. Inoue, "A study on walk-recognition by frequency analysis of footsteps," *IEEJ Transactions on Electronics, Information and Systems*, vol. 119, pp. 762–763, Jun 1999.

[151] D. T. Alpert and M. Allen, "Acoustic gait recognition on a staircase," in *2010 World Automation Congress*, Sep 2010, pp. 1–6.

[152] J. T. Geiger, M. Hofmann, B. Schuller, and G. Rigoll, "Gait-based person identification by spectral, cepstral and energy-related audio features," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 458–462.

[153] S. Hwang and J. Gim, "Listen to your footsteps: Wearable device for measuring walking quality," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '15, 2015, pp. 2055–2060.

[154] M. U. B. Altaf, T. Butko, and B. H. Juang, "Acoustic gaits: Gait analysis with footstep sounds," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 2001–2011, Aug 2015.

[155] C. Wang, X. Wang, Z. Long, J. Yuan, Y. Qian, and J. Li, "Estimation of temporal gait parameters using a wearable microphone-sensor-based system," *Sensors*, vol. 16, no. 12, p. 2167, Dec 2016.

[156] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[157] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., ser. signal processing. Upper Saddle River, New Jersey, USA: Prentice Hall, 1999.

[158] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice Hall, 1996.

[159] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 4th ed. Wiley, 2009.

[160] S. Mitra and S. Mitra, *Digital Signal Processing: A Computer-based Approach*, ser. Connect, learn, succeed. McGraw-Hill, 2011.

[161] M. Brookes, P. A. Naylor, and J. Gudnason, "A quantitative assessment of group delay methods for identifying glottal closures in voiced speech," *IEEE Trans. Speech Audio Process.*, vol. 14, 2006.

[162] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Trans. Speech Audio Process.*, vol. 15, no. 1, pp. 34–43, Jan. 2007.

[163] W. Hess, *Pitch Determination of Speech Signals*. Springer-Verlag, 1983.

[164] W. Hess, "Pitch and voicing determination," in *Advances in speech signal processing*. New York: Marcel Dekker, 1992, pp. 3–48.

[165] S. Upadhya and N. Wankhede, "Pitch estimation using autocorrelation method and AMDF," in *International Conference on Advances in Computing and Management*, 2012, pp. 249–253.

[166] M. Rashidul Hasan and T. Shimamura, "An efficient pitch estimation method using windowless and normalized autocorrelation functions in noisy environments," *International journal of circuits, systems and signal processing*, vol. 6, pp. 197–204, Jan 2012.

[167] R. Tanawongsuwan and A. Bobick, "A study of human gaits across different speeds," Georgia Institute of Technology, Report, 2003.

[168] Y. Hiranaka and H. Yamasaki, "Envelope representations of pinna impulse responses relating to three-dimensional localization of sound sources," *J. Acoust. Soc. Am.*, vol. 73, no. 1, pp. 291–296, 1983.

[169] J. Loudon, M. Swift, and S. Bell, *The Clinical Orthopedic Assessment Guide*, 2nd ed. Human Kinetics Europe Ltd, 2008.

[170] J. Nedoma, J. Stehlik, I. Hlavacek, J. Danek, T. Dostalova, and P. Preckova, *Mathematical and Computational Methods and Algorithms in Biomechanics: Human Skeletal Systems*, ser. Bioinformatics: Computational Techniques and Engineering. Wiley, 2011.

[171] R. Bellman and S. Dreyfus, *Applied Dynamic Programming*. Princeton, N.J.: Princeton University Press, 1962.

[172] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–260, Apr. 1967.

[173] K. R. Kaufman, C. Hughes, B. F. Morrey, M. Morrey, and K. An, "Gait characteristics of patients with knee osteoarthritis," *Journal of Biomechanics*, vol. 34, pp. 907–915, 2001.

[174] J. B. Cronin and M. C. Rumpf, "Effect of four different step detection thresholds on nonmotorized treadmill sprint measurement." *Journal of strength and conditioning research*, vol. 28 10, pp. 2996–3000, 2014.

[175] A. Belli, P. Bui, A. Berger, A. Geyssant, and J.-R. Lacour, "A treadmill ergometer for three-dimensional ground reaction forces measurement during walking," *Journal of Biomechanics*, vol. 34, no. 1, pp. 105–112, 2001.

[176] K. Jordan, J. H. Challis, and K. M. Newell, "Walking speed influences on gait cycle variability," *Gait & Posture*, vol. 26, no. 1, pp. 128–134, 2007.

[177] Y. Liu, K. Lu, S. Yan, M. Sun, D. K. Lester, and K. Zhang, "Gait phase varies over velocities," *Gait & Posture*, vol. 39, no. 2, pp. 756–760, 2014.

[178] F. Hebenstreit, A. Leibold, S. Krinner, G. Welsch, M. Lochmann, and B. M. Eskofier, "Effect of walking speed on gait sub phase durations," *Human Movement Science*, vol. 43, pp. 118–124, 2015.

[179] S. Farrokhi, S. Tashman, A. B. Gil, B. Klatt, and G. K. Fitzgerald, "Are the kinematics of the knee joint altered during the loading response phase of gait in individuals with concurrent knee osteoarthritis and complaints of joint instability? a dynamic stereo x-ray study," *Clinical biomechanics (Bristol, Avon)*, vol. 27, pp. 384–389, November 2011.

[180] B. J. Fregly, "Gait modification to treat knee osteoarthritis," *HSS journal: the musculoskeletal journal of Hospital for Special Surgery*, vol. 8, pp. 45–48, February 2012.

[181] G. Nalbantov, J. C. Bioch, and P. J. F. Groenen, "Solving and interpreting binary classification problems in marketing with svms," in *From Data and Information Analysis to Knowledge Engineering*, M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, and W. Gaul, Eds. Springer Berlin Heidelberg, 2006, pp. 566–573.

[182] N. Smith and M. J. F. Gales, "Speech recognition using SVMs," in *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press, 2002, nIPS-2002.

[183] Z. Chen, N. Pears, M. Freeman, and J. Austin, "Road vehicle classification using support vector machines," in *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*, vol. 4, Nov. 2009, pp. 214–218.

[184] V. Wan and S. Renals, "SVMSVM: Support vector machine speaker verification methodology," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, Apr. 2003, pp. 221–224.

[185] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani, "1-norm support vector machines," in *Proceedings of the 16th International Conference on Neural Information Processing Systems*, ser. NIPS'03. Cambridge, MA, USA: MIT Press, 2003, pp. 49–56. [Online]. Available: http://dl.acm.org/citation.cfm?id=2981345.2981352

[186] J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Advances in kernel methods - Support Vector Learning, Tech. Rep., 1998.

[187] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using second order information for training support vector machines," *The Journal of Machine Learning Research*, vol. 6, pp. 1889–1918, Dec. 2005.

[188] V. Kecman, T.-M. Huang, and M. Vogt, *Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 255–274.

[189] A. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-PLUS Illustrations*, 11 1997, vol. 94.

[190] I. McCarthy, D. Hodgins, A. Mor, A. Elbaz, and G. Segal, "Analysis of knee flexion characteristics and how they alter with the onset of knee osteoarthritis: a case control study," *BMC Musculoskeletal Disorders*, vol. 14, no. 1, p. 169, May 2013.

[191] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *AMS CONFERENCE ON MATH CHALLENGES OF THE 21ST CENTURY*, 2000.

[192] M. Köppen, "The curse of dimensionality," in *5th online world conference on soft computing in industrial applications (WSC5)*, 2000.

[193] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4-5, pp. 411–430, May 2000.

[194] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley-Interscience, 2001.

[195] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[196] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[197] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[198] A. Makhzani, J. Shlens, N. Jaitly, and I. Goodfellow, "Adversarial autoencoders," 2016. [Online]. Available: arXiv:1511.05644

[199] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. Springer-Verlag, 2002.

[200] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, Aug. 2010.

[201] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1, pp. 389–422, Jan. 2002.

[202] *MATLAB 2019a*, The Mathworks Inc., Natick, 2019.

[203] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," vol. 101, pp. 1396–1400, 01 2003.

[204] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

Appendix A

Equipment technical specifications

Basik Pro from Schertler

Output impedance	1000 ohm
Frequency response	20 Hz to 20 kHz
SNR	60 dB

Table A.1: Basik Pro contact microphone specifications obtained from the manufacturer's website.

RME Babyface

Input AD	2×XLR (2×Mic/Line), servo-balanced
Output DA (Main, 2 Channels)	2×XLR, servo-balanced
Output DA (Phones, 2 Channels)	2×1/4" TRS (Stereo), unbalanced
Input Digital	1×ADAT optical or SPDIF optical
Output Digital	1×ADAT optical or SPDIF optical
MIDI	1×MIDI I/O via breakout cable (2×5-pin DIN jacks)
Dynamic range AD	108 dB RMS unweighted, 111 dBA
THD AD	< -100 dB (< 0.001%)
THD+N AD	< -98 dB (< 0.0012%)
Crosstalk AD	> 110 dB
Dynamic range DA	112 dB RMS unweighted, 115 dBA
THD DA	< -104 dB (< 0.00063%)
THD+N DA	< -100 dB (< 0.001%)
Crosstalk DA	> 110 dB

Table A.2: RME Babyface USB audio interface specifications obtained from the manufacturer's website.