

Title	Digital signal processing and artificial intelligence for the automated classification of food allergy
Authors	Twomey, Niall Joseph
Publication date	2013
Original Citation	Twomey, N.J. 2013. Digital signal processing and artificial intelligence for the automated classification of food allergy. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2013, Niall J. Twomey - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-10-03 03:51:25
Item downloaded from	https://hdl.handle.net/10468/1236

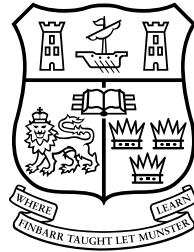


UCC

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Digital Signal Processing and Artificial Intelligence for the Automated Classification of Food Allergy

Niall Twomey



A thesis submitted to the National
University of Ireland in fulfillment
of the requirements for the Degree of
Doctor of Philosophy

Supervisor: Dr. William P. Marnane
Head of Department: Prof. Nabeel Riza

Department of Electrical and Electronic Engineering,
National University of Ireland, Cork.

Abstract

As a by-product of the ‘information revolution’ which is currently unfolding, life-times of man (and indeed computer) hours are being allocated for the automated and intelligent interpretation of data. This is particularly true in medical and clinical settings, where research into machine-assisted diagnosis of physiological conditions gains momentum daily. Of the conditions which have been addressed, however, automated classification of allergy has not been investigated, even though the numbers of allergic persons are rising, and undiagnosed allergies are most likely to elicit fatal consequences.

On the basis of the observations of allergists who conduct oral food challenges (OFCs), activity-based analyses of allergy tests were performed. Algorithms were investigated and validated by a pilot study which verified that accelerometer-based inquiry of human movements is particularly well-suited for objective appraisal of activity. However, when these analyses were applied to OFCs, accelerometer-based investigations were found to provide very poor separation between allergic and non-allergic persons, and it was concluded that the avenues explored in this thesis are inadequate for the classification of allergy.

Heart rate variability (HRV) analysis is known to provide very significant diagnostic information for many conditions. Owing to this, electrocardiograms (ECGs) were recorded during OFCs for the purpose of assessing the effect that allergy induces on HRV features. It was found that with appropriate analysis, excellent separation between allergic and non-allergic subjects can be obtained. These results were, however, obtained with manual QRS annotations, and these are not a viable methodology for real-time diagnostic applications. Even so, this was the first work which has categorically correlated changes in HRV features to the onset of allergic events, and manual annotations yield undeniable affirmation of this.

Fostered by the successful results which were obtained with manual classifications, automatic QRS detection algorithms were investigated to facilitate the fully automated classification of allergy. The results which were obtained by this process are very promising. Most importantly, the work that is presented in this thesis did not obtain any false positive classifications. This is a most desirable result for OFC classification, as it allows complete confidence to be attributed to classifications of allergy. Furthermore, these results could be particularly advantageous in clinical settings, as machine-based classification can detect the onset of allergy which can allow for early termination of OFCs. Consequently, machine-based monitoring of OFCs has in this work been shown to possess the capacity to significantly and safely advance the current state of clinical art of allergy diagnosis.

ACKNOWLEDGEMENTS

I must first and foremost thank and acknowledge my supervisor, Dr. Liam Marnane, for taking me on in this project and for his advice during my incarceration in UCC. I would also like to thank the Irish Research Council, the Tril Centre and Intel for funding this research.

Next, I would like to thank my internal and external examiners, Dr. Bill Wright and Dr. Fernando Schlindwein, for a surprisingly enjoyable viva.

I would like to thank Prof. Jonathan O'B Hourihane, Deirdre Daly and Claire Cullinane from the Department of Paediatrics and Child Health for their patience and support during my time recording the data which forms the basis of this thesis. Your feedback, patient answers and advice was very much appreciated!

I would also like to extend my thanks to Stephen Faul and Andrey Temko for their guidance in all aspects of signal processing and machine learning, and for hitting me over the ear when I deserved it. I would like to also thank all of the postgraduate students that I came to know during the course of my research; thank you all for your company, friendship and especially for the caffeine and card game indulgences during the years; they kept me relatively sane.

Finally, I must also thank my friends and family for your constant support and throughout this time, specifically Jen and Chengde. Sincerely, I appreciate your consistent encouragement, sporadic berating and occasional willingness to seem interested in my research more than I can say.

STATEMENT OF ORIGINALITY

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text.

Niall Twomey

September, 2013

CONTENTS

1 Allergy and Allergic Reactions	1
1.1 Introduction	1
1.2 Allergy	2
1.2.1 Varieties and symptoms of allergy	2
1.2.1.1 Variety	2
1.2.1.2 Symptoms	3
1.2.2 Management and treatment of allergic reactions	4
1.2.2.1 Mild reactions	5
1.2.2.2 Severe reactions	5
1.2.3 Risk factors, and quality of life	6
1.2.3.1 Risk and protection factors	6
1.2.3.2 Quality of life	7

1.3	Requirement for clinical diagnosis of allergy	8
1.4	Diagnosis of allergy	9
1.4.1	Blood testing	9
1.4.2	Skin testing	10
1.4.3	Challenge testing	10
1.4.3.1	Preliminary tests	11
1.4.3.2	Checkup	13
1.4.3.3	Observation	13
1.4.3.4	Failure protocol	14
1.4.3.5	Supplementary stages	14
1.5	Challenge-testing clinical experience	15
1.6	Machine-assisted classification of allergy	16
1.7	Layout of this thesis	17
2	Algorithms, methods and data collection	19
2.1	Introduction	19
2.2	Activity analysis	20
2.2.1	Introduction to inertial measurement	20
2.2.2	Applications of inertial sensing	22

2.2.2.1	Activity recognition	23
2.2.3	Activity-based analysis of OFC	25
2.3	Heart rate analysis	26
2.3.1	History and introduction	26
2.3.2	Applications of HRV analysis	28
2.4	Machine learning and classification	30
2.4.1	Introduction to classification	30
2.4.2	Background to machine learning	30
2.4.3	Novelty detection	33
2.4.3.1	One class SVMs and GMMs	34
2.4.3.2	Example of novelty detection	35
2.4.4	Applications of novelty detection	37
2.5	Food challenge data collection	38
2.5.1	Recording platform	38
2.5.2	Integration with oral food challenge	39
2.5.3	Other data	41
2.5.4	Data recording	41
2.6	Conclusion	43

3 Accelerometer-based analysis of oral food challenges	44
3.1 Introduction	44
3.2 Accelerometer-based activity analysis	47
3.2.1 Activity metrics	47
3.2.2 Energy expenditure estimation algorithms	49
3.2.2.1 A note on the energy expenditure estimation algorithms . .	50
3.2.2.2 Bouten et al	50
3.2.2.3 Chen et al	52
3.2.2.4 Crouter et al	55
3.3 Energy expenditure validation	57
3.3.1 Experimental setup	58
3.3.2 Pre-processing	59
3.3.2.1 Codeword conversion	59
3.3.2.2 Breath and acceleration synchronisation	62
3.3.2.3 Normalisation	63
3.3.3 Performance evaluation	63
3.3.4 Results	64
3.3.5 Discussion on energy expenditure estimation algorithms	66

3.3.6	Conclusion on energy expenditure estimation	67
3.4	Accelerometer-based analysis during OFCs	67
3.5	Probability density functions	68
3.6	Results	69
3.7	Conclusion	71
4	ECG-based analysis of OFCs	73
4.1	Introduction	73
4.2	ECG and HRV	74
4.2.1	ECG recording	74
4.3	HRV feature extraction	75
4.3.1	Overview	75
4.3.2	Epochs	75
4.3.3	Epoch overlap	76
4.4	Feature normalisation/calibration	77
4.5	HRV feature categories	79
4.5.1	Feature categories	79
4.5.2	Frequency domain feature analysis	80
4.5.3	Resampling + FFT	80

4.5.4	Direct PSD estimation of HRV	82
4.5.5	Comparison of HRV frequency analysis methods	85
4.6	Features	86
4.6.1	Notation	87
4.6.2	Time domain features	87
4.6.2.1	Mean heart rate	87
4.6.2.2	Standard deviation	88
4.6.2.3	Coefficient of variation	88
4.6.2.4	RMSSD	90
4.6.2.5	NN/PNN	91
4.6.2.6	Histogram	93
4.6.3	Sequential domain features	94
4.6.4	Poincaré features	95
4.6.5	Frequency domain features	98
4.7	Discussion	100
5	Machine learning for allergy classification	102
5.1	Introduction	102
5.2	Novelty detection for OFC	103

5.2.1	Choice of classification routine	103
5.2.2	Feature transformation	103
5.2.3	Gaussian mixture models	106
5.2.3.1	k -means clustering	108
5.2.3.2	Expectation maximisation	109
5.2.4	Postprocessing	112
5.3	Classification procedures	114
5.4	Classifier model selection	116
5.4.1	Performance evaluation	116
5.4.2	Parameter selection	117
5.4.2.1	Search space	117
5.4.2.2	Cost function	118
5.5	Classification metrics	119
5.5.1	Sensitivity/specificity	119
5.5.2	Time gain parameters	120
5.5.2.1	Time gain	120
5.5.2.2	Doses saved	121
5.5.2.3	Activation percentage	121

5.6	Results	122
5.6.1	A brief note on the structure of these results	122
5.6.2	Results obtained at epoch length of 60 seconds	123
5.6.3	Overall results	125
5.6.4	Inconsistent classification at different epoch lengths	125
5.6.4.1	Short-duration signatures of allergy	125
5.6.4.2	Longer signatures of allergy	127
5.6.4.3	Tolerance to non-allergic variances	127
5.6.5	Boosted allergy classification	131
5.6.5.1	Sensitivity/specificity	131
5.6.5.2	Time gain parameters	133
5.7	Discussion	135
5.7.1	Specificity of OFC classification	135
5.7.2	Robust classification	136
5.7.3	Parameter selection	138
5.7.3.1	Importance of correct parameter selection	138
5.7.3.2	Alternative parameter selection	139
5.7.4	Role of classification in OFCs	140

6 Automatic QRS detection	142
6.1 Introduction	142
6.2 QRS detection	142
6.2.1 QRS validation	143
6.2.2 Validation databases	144
6.2.3 Sensitivity and positive predictivity	145
6.2.4 Good detection window	146
6.2.5 Feature accuracy	146
6.2.6 Box-plots	147
6.3 Choice of QRS detector	149
6.4 Hilbert transform based QRS detection	150
6.4.1 Theory of Hilbert transform	150
6.4.2 Method of QRS detection with Hilbert transform	152
6.4.3 Beat identification	155
6.5 Filter-banks based QRS detection	156
6.5.1 Theory of filter banks	156
6.5.2 QRS detection with filter banks	157
6.5.2.1 Pre-processing	157

6.5.2.2	Beat-classification logic	158
6.5.2.3	Overall	163
6.6	Results obtained on MIT-BIH database	164
6.6.1	Sensitivity and positive predictivity	164
6.6.2	Percentage RMS difference	166
6.6.3	Conclusions on QRS detection on MIT-BIH database	167
6.7	Requirement for artefact detection	167
6.7.1	Introduction	167
6.7.2	Artefact detection algorithm	168
6.7.3	Demonstration of artefact detection	169
6.8	Results obtained on allergy database	171
6.8.1	Artefact detection	171
6.8.2	Sensitivity and positive predictivity	173
6.8.3	Percentage RMS difference	175
6.9	Overall discussion	176
7	Fully automated allergy detection	178
7.1	Introduction	178
7.2	Methods	179

7.3	Unmatched classification results	181
7.3.1	Results of unmatched classification	181
7.3.2	Discussion on unmatched classification results	182
7.4	Matched classification results	183
7.4.1	Sensitivity and specificity	183
7.4.2	Artefact detection	185
7.4.3	Time gain	187
7.5	Discussion	188
7.6	Conclusion	189
8	Overall summary, final conclusions and future work	191
8.1	Summary of this thesis	191
8.2	Primary contribution of this thesis	194
8.3	Possible avenues of future work	195
8.3.1	Data collection	195
8.3.2	Alternative novelty detectors	196
8.3.3	Real-time and portable implementation	197
8.3.4	Feature and epoch analysis	197
8.4	Publications resulting from this work	198

Appendices	201
A Alternative parameter selection routines	201
A.1 Introduction and Methods	201
A.2 Results	204
A.3 Discussion	206
A.4 Conclusion	206
B Investigation into the importance of features	208
B.1 Introduction and methods	208
B.2 Results	210
B.3 Discussion	212
B.4 Conclusion	214
References	215

LIST OF TABLES

2.1	Tabulation of the characteristics of the subjects who were recorded for this study.	42
3.1	MET activity and corrected MET activity values.	56
3.2	Tabulation of the physical characteristics of the subjects who participated in the accelerometer-based energy expenditure validation test.	58
3.3	Table of PRD values computed between true energy expenditure and the estimated energy expenditure values obtained from the algorithms investigated.	64
4.1	Table of HRV diagnostic frequency ranges for children.	98
5.1	Classification results obtained with the novelty detection routine at epoch length of 60 seconds.	124
5.2	Tabulation of the classification results of the allergic subjects where ‘1’ represents an allergic classification (TP) whereas ‘0’ represents a non-allergic classification (FN).	132

5.3	Classification result, time gain, doses saved and activation percentages obtained by the classification routine. The results in this table were obtained by fusing the results obtained for the individual epoch lengths together.	133
6.1	Differences in reported and calculated sensitivity and positive predictivity. .	164
6.2	Sensitivity and positive predictivity of QRS detectors on allergy database. . .	173
6.3	Distribution of mean and standard deviation of the PRD values calculated from automatically extracted QRS points.	176
7.1	Sensitivity and specificity of classification results obtained with the manual classification models on the automatically extracted HRV features (i.e. crossover classification results).	181
7.2	Sensitivity and specificity of classification results obtained by Afonso's and the Hilbert transform QRS detectors.	183
7.3	Specific time gain metrics obtained from fully automatic allergy classification based on Afonso and Hilbert transform QRS detectors.	188
A.1	Tabulation of sensitivity, specificity, and the time gain metrics which were obtained by selecting the mean, median and mode of the set of post-processing parameters from the training data. In the case of the mean method, imperfect specificity was obtained.	205
B.1	Classification metrics which were obtained with the time-, frequency-, Poincaré- and sequential-domain classification models ranked by order of importance of the feature category in question.	210

LIST OF ACRONYMS

ADC	analogue to digital converter.....	59
ANS	autonomic nervous system.....	80
BP	band-pass.....	168
BPM	beats per minute.....	76
BW	bandwidth.....	157
CPET	cardio pulmonary exercise testing.....	58
CE	European conformity.....	38
CSI	cardiac sympathetic index.....	97
CVI	cardiac vagal index.....	97
DFT	discrete Fourier transform.....	81
DoF	degrees of freedom.....	20
DSP	digital signal processing.....	29
ECG	electrocardiogram.....	17
EE	energy expenditure.....	49
$\widehat{EE}$	energy expenditure estimation.....	49
EE_{act}	energy expenditure due to activity.....	51
$\widehat{EE}_{act}$	energy expenditure estimation due to activity.....	51

EE_{true}	true energy expenditure	49
EEG	electroencephalogram	29
EM	expectation maximisation	109
Epi-pen	epinephrine pen	5
FFT	fast Fourier transform	82
FIR	finite impulse response	152
FN	false negative	120
FP	false positive	120
FT	Fourier transform	80
GMM	Gaussian mixture model	33
HF	high frequency	98
HR	heart rate	76
HRV	heart rate variability	28
IAA	integral of absolute acceleration	47
IAA_t	total integral of absolute acceleration	48
IAA_x	integral of absolute acceleration in the x -axis	47
IAA_y	integral of absolute acceleration in the y -axis	47
IAA_z	integral of absolute acceleration in the z -axis	47
IIR	infinite impulse response	156
IQR	inter-quartile range	148
KDE	kernel density estimation	68
km/h	kilometers per hour	52
kNN	k -nearest neighbours	31
LF	low frequency	98
LOO	leave-one-out	116
MEMS	micro-electro-mechanical systems	20

MET	metabolic equivalents.....	55
MIT-BIH	Massachusetts Institute of Technology Beth Israel Hospital	145
nesC	network embedded systems C.....	38
OFC	oral food challenge	11
PCA	principal component analysis.....	103
PDF	probability density function	68
+P	positive predictivity	146
PRD	percentage root mean square difference	63
PSD	power spectral density.....	82
REE	resting energy expenditure	51
RMR	resting metabolic rate.....	55
RMS	root mean square	90
RMSE	root mean square error.....	63
RMSSD	root mean square of successive difference	90
SHIMMER device	sensing health with intelligence, modularity, mobility and experimental reusability.....	38
SNR	signal to noise ratio.....	157
SVM	support vector machine.....	31
TGS	specific time gain.....	120
TGT	total time gain.....	120
TN	true negative	120
TP	true positive.....	120
ULF	ultra-low frequency.....	98
VLF	very low frequency	98

LIST OF FIGURES

1.1	Photograph of an adrenaline pen (theonlineallergist.com, 2013).	6
1.2	Oral food challenge flowchart which presents the means by which allergy is diagnosed in a clinical environment.	12
2.1	Illustration of the approximate growth of microelectromechanical systems accelerometers, gyroscopes and magnetometers in the research market over the past decade. (Google Inc., 2013).	21
2.2	Functional diagram of a MEMS accelerometer, identifying the mass and spring.	21
2.3	The original galvanometer developed by Willem Einthoven to record the ECG in the early 20 th century.	26
2.4	Labelled ECG waveform (Dublin Institute of Technology, 2013).	28
2.5	Example of how classification is obtained with two-dimensional data (top) with SVM based classification (bottom left) and GMM based classification (bottom right).	32

2.6	Example of how novelty detection algorithms can be employed to determine novel and normal data points. The upper image shows the distribution of the labels, and the lower-right figure shows one-class SVM while the lower-left figure shows the example of GMM-based novelty detection on the same data.	36
2.7	Diagram of the SHIMMER device, with the various components annotated (reproduced with permission from SHIMMER-research (2010)).	39
2.8	Modified oral food challenge flowchart which is employed to accommodate the introduction of the SHIMMER monitoring device for data collection during OFCs.	40
3.1	The x (anterio-posterior), y (medio-lateral), and z (vertical) acceleration directions in relation to a body.	47
3.2	Demonstration of the conversion process from the raw digital codewords obtained from the accelerometer (a) to a measurement of absolute gravity (b).	60
3.3	The $\widehat{EE}$ and EE_{true} values obtained for participant 2. The values obtained are overlaid.	65
3.4	Example histogram and probability density function of a feature.	68
3.5	Illustration of the differences between PDFs that describe two separate classes of data.	69
3.6	Histograms plotting the normalised IAA values of the allergic and non-allergic subjects who were investigated.	70
4.1	Einthoven triangle configuration for ECG electrode placement (University of Nottingham, 2013).	74

4.2 Illustration of relationship between the ECG and the epoch length for ECG recorded in OFC.	75
4.3 Illustration of relationship between the ECG, the epoch length and the epoch overlap.	77
4.4 Illustration of the raw HR ($\square$) which is not periodically sampled, and the HR re-sampled to 10 Hz via cubic spline interpolation.	81
4.5 PDF of mean heart rate, generated from allergic and non-allergic subjects . .	88
4.6 PDF of standard deviation of the heart rate, generated from allergic and non-allergic subjects.	89
4.7 PDF of coefficient of variation of the heart rate, generated from allergic and non-allergic subjects.	89
4.8 PDF of RMSSD of the heart rate, generated from allergic and non-allergic subjects.	90
4.9 PDF of PNN ₅₀ of the heart rate, generated from allergic and non-allergic subjects.	92
4.10 Histogram of the relative times between successive QRS complexes.	92
4.11 PDF of histogram of the heart rate, generated from allergic and non-allergic subjects.	93
4.12 Chart of the change between successive QRS complexes.	94
4.13 PDFs derived for the sequential domain features.	95
4.14 Original and rotated points plotted in a Poincaré Chart.	96
4.15 CSI and CVI PDF from Poincaré features.	97

4.16 PDF of the frequency domain features.	99
5.1 An illustration of PCA in two-dimensional feature space (subplot a) and two-dimensional component-space (subplot b).	105
5.2 A mixture of three equally-weighted Gaussians (dashed lines) which combine to represent a multi-modal non-normal distribution (solid black line). .	107
5.3 A demonstration of the difference in clustering which is obtained by <i>k</i> -means clustering and the expectation maximisation algorithm. With subfigures B and C, a line is drawn from each point to its associated cluster. .	110
5.4 Sample likelihood (subplot a) and histogram of the background likelihood (subplot b) of Subject 23.	112
5.5 Flowchart of classification procedure involving the recording of ECG, annotation of QRS complexes, feature extraction and the classification procedure of OFC.	114
5.6 Illustration of the data segmentation and testing routines employed in the allergy classification procedure.	115
5.7 The confusion matrix showing how sensitivity and specificity are obtained with regard to the ground truth (diagnosis) and predicted (classification) results.	119
5.8 A demonstration of early detection of allergy (with Subject 11). The segments at 45, 60 and 80 minutes which fall beneath the threshold were classified as allergy.	123
5.9 Example demonstrating how the generated likelihood for Subject 2 satisfies the allergy criteria at an epoch length of 60 seconds (subplot a) while failing to do so for epoch lengths of 120, 180 and 300 seconds (subplots b — d). . .	126

5.10 Example demonstrating how the generated likelihood for Subject 13 does not satisfy the allergy criteria at an epoch length of 60, 120 and 180 seconds (subplot a — c) but the criteria are then met for the epoch length of and 300 seconds (subplots d).	128
5.11 Example demonstrating how the generated likelihood for Subject 16 surpasses the threshold, but does not satisfy the allergy criteria due to the inclusion of the duration parameter and for all epoch lengths the subject is correctly classified as non-allergic.	130
5.12 Example of arrhythmia on the ECG trace (a) and the effect this has on the heart rate (b) on Subject 20.	137
5.13 The likelihood series chosen for Subject 2 which does not diverge from the background level significantly enough to classify allergy. PCA preserved 80% of the feature variance which was modelled with a GMM order of 32 at an epoch length of 60 seconds.	139
6.1 Relationship between a box-plot, and quartile ranges with a normal distribution. The locations marked Q_l and Q_u are the lower and upper quartiles respectively, and the median is marked as m .	148
6.2 The real and imaginary components resulting from the Hilbert Transform of the ECG.	152
6.3 The flowchart for QRS detection from the Hilbert Transform.	153
6.4 The stages employed by the Hilbert Transform QRS detection algorithm (the ECG data was obtained from Patient 113 in the MIT-BIH Database).	154
6.5 The generic filter banks flow chart incorporating both bandpass and synthesis filters. The $\downarrow$ and $\uparrow$ symbols represent down- and up-sampling respectively.	156

6.6	The idealised filter response of the filter banks, with M equally-wide sub-bands.	156
6.7	Overall simplified flowchart of Afonso's QRS detection method.	159
6.8	The effect of the various QRS validation levels from Afonso's QRS detection algorithm. In these charts, the $\circ$ symbols represent the candidate QRS points. Charts b — d have been down-sampled.	161
6.9	Incorrect QRS complex localisation (Patient 8 of MIT-BIH arrhythmia database). Manual QRS annotations are marked with $\times$ and automatic detections are marked with $\square$ (Hilbert transform algorithm) and $\circ$ (Afonso's algorithm).	165
6.10	Sensitivity and positive predictivity box-plots of QRS detection on the MIT-BIH arrhythmia database.	166
6.11	PRD box-plots of the mean (μ) and standard deviation (σ) of the heart rate over all subjects in the MIT-BIH arrhythmia database.	167
6.12	Normalised output of high-frequency (a,b) and low-frequency (c,d) energy estimators for artefact detection (data from Subject 8 of allergy database). . .	170
6.13	The breakdown of the number of artefact events which were detected by the artefact detection algorithm for each subject of the allergy database.	172
6.14	Sensitivity and positive predictivity box-plots of QRS detection on the allergy database.	174
6.15	Example of poor quality of the ECG signal after the application of denoising filters which contributed to poor sensitivity and positive predictivity values for Subject 3.	175

6.16	Boxplots of the PRD of the mean and standard deviation of the heart rate between the manual and automatic QRS points extracted.	176
7.1	The flow of how the matched (right) and unmatched (left) classification results are obtained.	179
7.2	Likelihood plots of Subject 1 for manual and automatic models at epoch lengths of 60 seconds. Subplots (a) — (c) show the likelihoods which were obtained with manual, Afonso and Hilbert models respectively. In all cases the threshold for allergy classification is off the scope of the figures.	184
7.3	Likelihood plots of Subject 7. Subplot (a) shows the threshold which was computed without the aid of artefact detection and how allergy classification does not classify allergy. Subplot (b) shows the threshold which was computed when artefact detection was incorporated and how allergy classification is successful with artefact-aware classification.	186
A.1	The estimated distribution of duration and multiplicative post-processing parameters which achieve 100% specificity and maximum sensitivity on the training dataset. The image is limited to d and n parameters of 75. The darker regions indicate a higher density of suitable parameters.	203

CHAPTER 1

ALLERGY AND ALLERGIC REACTIONS

1.1 INTRODUCTION

ALLERGY is defined as an abnormally high acquired sensitivity towards certain substances (Dorland, 1901). To the vast majority of the population allergens are harmless and do not interfere with their day-to-day living. However, for those who suffer from allergy, allergens can provoke acquired, predictable and rapid conditions which are called allergic reactions, and in some cases, these can be fatal.

This chapter sets the background to the clinical knowledge of allergy and allergic reactions and it also introduces the scope of this thesis. The chapter focuses on the varieties, symptoms, diagnoses and treatment of food-based allergy. The possibility of automated allergy detection for real-time clinical allergy testing environments is then explored and the benefits which this type of analysis could provide are also introduced.

1.2 ALLERGY

Approximately 25 — 30% of people believe that they suffer from allergy (Miles et al., 2005), but it is estimated that only 6% of schoolchildren (Bock and Sampson, 1994) and 2.3% of adolescents (Pereira et al., 2005) suffer from food allergy. The reason for this disparity between perceived and true numbers is due to a lack of public understanding of allergy, and also because in most instances, suspicions of allergy are not confirmed diagnostically.

1.2.1 VARIETIES AND SYMPTOMS OF ALLERGY

While allergies can come in many varieties, it is only diagnosed if the symptoms present in a predictable manner.

1.2.1.1 VARIETY

A range of different types of allergy exist, and the primary varieties are listed here.

1. *Food allergy:*

Food allergy is caused by an adverse immune response against a food-type. In some cases, when a sufferer is very sensitive to the allergen, physical contact with the food-type is sufficient to provoke the symptoms of an allergic reaction. The majority of reactions occur after ingestion of the food. The quantity of food required to provoke a reaction differs from person-to-person, and it can be thought of as a specific threshold.

Typical food allergens include:

- Milk
- Egg
- Wheat
- Peanut
- Soy
- Cooked egg

2. *Environmental allergy:*

Allergic reactions can be provoked by pets, poor hygiene and insects. Reactions to environmental allergens can manifest due to the presence of the allergen in the air (dust mites, mold spores) or by physical contact (rubbing cat/dog, insect stings etc). The length of exposure required to provoke a reaction differs from person-to-person.

Environmental allergens include:

- Cats
- Dust mites
- Mould
- Dogs
- Pollen
- Insect sting/bites

3. *Drug/medication allergy:*

Allergic reactions can also be provoked by the administration of medication. The reaction manifests when the drug enters the sufferer's blood stream. These allergies can be very dangerous if the sufferer or medical staff are unaware of them. Often, allergies to medical drugs are due to dyes and stabilising agents found in the drug packaging rather than the drug itself. However, due to international regulation, colour coding is required to reduce counterfeiting.

Drug allergens include:

- Penicillin
- Anaesthetics

1.2.1.2 SYMPTOMS

The severity of the symptoms that the allergy sufferers encounter depends on the individual subject's susceptibility to the allergen and on the amount of the allergen that they have come into contact with. For subjects with severe sensitivities, an allergic reaction may be provoked simply by contact between the subject's skin and the allergen. For less

sensitive subjects, oral consumption of the food type is required in order to provoke a reaction.

The most dangerous of the symptoms of allergy involve restrictions of the airway (wheezing, shortness of breath and asthma attacks) and anaphylaxis. Anaphylaxis is an acute allergic reaction towards an allergen, and 3 — 15% of those who suffer from allergies will be afflicted with an anaphylactic reaction at least once (Matasar and Neugut, 2003; Yanishevsky and Hourihane, 2010; Järvinen et al., 2009), indicating that there is a very real risk of a severe reaction amongst sufferers of allergy. Interestingly, there is no universally accepted definition of anaphylaxis, and there is also disagreement about the criteria for the diagnosis of anaphylactic events (Sampson et al., 2006). In order to protect against reactions, sufferers of allergy must always be constantly vigilant and be equipped to treat and manage allergic reactions.

Symptoms of allergy can include:

- | | | |
|--------------------|------------------|-----------------------|
| • Hives | • Red eyes | • Vomiting |
| • Rashes | • Sneezing | • Diarrhoea |
| • Swelling | • Bloating | • Wheezing |
| • Running nose | • Mood change | • Shortness of breath |
| • Sinus congestion | • Abdominal pain | • Asthma attacks |
| • Itchy eyes | • Ear pain | • Anaphylaxis |

1.2.2 MANAGEMENT AND TREATMENT OF ALLERGIC REACTIONS

The best means of prevention of allergy is complete avoidance (Pereira et al., 2005; Sicherer and Sampson, 2006). For the cases of animal allergy, subjects are strongly encouraged not to have a pet in the home and to avoid them elsewhere. With some environmental

allergens, such as dust mites, a high degree of hygiene and cleanliness can assist in limiting the exposure to a subject, lowering their overall vulnerability to the allergen.

However, it is not possible to manage exposure to all environmental allergens, in the case with pollen for example, and if a subject is allergic to these unavoidable environmental allergens it is possible to self-administer precautionary antihistamines which will increase tolerance towards these symptoms for a short period of time.

This precautionary treatment is only viable in the case of environmental allergens. With food and chemical allergies, it is not appropriate to take medication on the chance of coming into contact with the substances. In the case when a person reacts to a food type, insect bite, etc, the means of countering the reaction depends of the extent of the reaction.

1.2.2.1 MILD REACTIONS

For mild reactions (hives, rashes, sneezing, abdominal pain, vomiting, etc) off-the-shelf antihistamines can be sufficient to stop the progression of a reaction. Zyrtec is commonly used in home and hospital environments to halt mild allergic reactions. This is a second-generation antihistamine which does not pass the blood/brain barrier, so drowsiness is not induced by consumption. Zyrtec can be taken in tablet or liquid form, and can begin to take effect in 10 — 20 minutes. In 2008 Zyrtec was the highest-grossing non-food product in the United States of America (Elliot, 2010), indicating the prevalence of allergy internationally, and the chronic nature of the condition.

1.2.2.2 SEVERE REACTIONS

For severe reactions (wheezing, shortness of breath, anaphylaxis) stronger rescue medications are required. Inhalers and epinephrine pens (Epi-pens) are used in these cases to stop the reaction.



Figure 1.1: Photograph of an adrenaline pen (theonlineallergist.com, 2013).

Epi-pens contain a dose of adrenaline* in a sealed syringe-like container, see Figure 1.1. If a subject requires adrenaline when in a state of anaphylaxis the top of the Epi-pen is removed, revealing a needle. The needle is then inserted in the thigh of the subject for ten seconds while the auto-injection mechanism releases the adrenaline into the subject's bloodstream. If the effects of the allergic reaction do not cease within ten minutes of administration, a second dose should be administered. After use of the Epi-pen the subject should go to the nearest hospital for a checkup and renewal of the Epi-pen prescriptions.

1.2.3 RISK FACTORS, AND QUALITY OF LIFE

1.2.3.1 RISK AND PROTECTION FACTORS

Some researchers have demonstrated that a number of factors can increase or decrease the likelihood of a person having allergic reactions. There are conflicting reports about the effect of rural and urban localisation on extensivity of allergy. Some studies (Braun-Fahrlander et al., 1997; Kilpelinen et al., 2000; Radon et al., 2004; Almqvist et al., 2003) report that children who have grown up on a farm have a lower frequency of allergy in comparison with urban and less-rural children. In contrast with this, other researchers were not able to find significant differences between rates of allergy based on rural and urban localisation (Viinanen et al., 2007; Omenaas et al., 1994; Azpiri et al., 1999).

*The terms adrenaline and epinephrine are synonyms for the same compound and are associated with British and American English respectively. The term Epi-pen employs the American naming convention and is adopted in this thesis because it is the most common term for the medical device.

It has also been reported that breastfeeding (Rautava et al., 2002) and exposure to household pets in a childhood home (Ahlbom et al., 1998) have been shown in some studies to be protection factors against allergy, but other studies demonstrate otherwise (Bergmann et al., 2002).

Allergy is therefore difficult to predict as a number of unknown dynamics determine a person's sensitivity to the allergens. It had been consistently reported, however, that suffering from allergy is a risk factor towards other other diseases. Allergy is heavily linked with asthma in childhood (Roberts et al., 2003; Black et al., 2000; Call et al., 1992), eczema (Bryld et al., 2003), wheezing and bronchial hyperresponsiveness (Arshad et al., 2005).

1.2.3.2 QUALITY OF LIFE

Allergy has a significant impact on the quality of life of its sufferers. In a quality of life survey, young sufferers of allergy reported much greater fear and anxiety of their condition in comparison to persons who suffer from other chronic diseases, such as insulin dependant diabetes mellitus (Avery et al., 2003). Owing to the relative ease of contamination, sufferers must live in a constant state of awareness towards their current state of health, and towards everything they consume. They must also live at constant risk of potential reactions and anaphylaxis, which increases stress and worry for the rest of their lives. Indeed, sufferers of severe allergies are reminded of their affliction daily because they must always carry Epi-pens in case of reactions.

Allergy affects the quality of life of more than just the sufferer and the quality of life of their family members are also impaired. Limitations of family activities are reported among suffering families (Sicherer et al., 2001), and in social situations the parents of suffers report significantly more disruption than non-suffering families (Oude Elberink et al., 2002). This is due to the persistent fear of the sudden death that follows sufferers of allergy (Primeau et al., 2000).

Adolescents are at the highest risk of death from anaphylaxis (Bock et al., 2007, 2001; Sampson et al., 1992), but young children and adults are also in danger, with up to 1% of these sufferers at mortal risk (Matasar and Neugut, 2003; Yanishevsky and Hourihane, 2010; Järvinen et al., 2009).

Without a confirmed diagnosis, a significant burden of anxiety is carried by suffering families (Primeau et al., 2000), and remains until a diagnosis is verified. Positive and negative diagnoses of allergy improve the quality of life of the family and individual involved, as, even with diagnosis of allergy, the burden of uncertainty is removed, and the family can adapt to living with the situation (DunnGalvin et al., 2010).

1.3 REQUIREMENT FOR CLINICAL DIAGNOSIS OF ALLERGY

It is very important for a subject to know if they are allergic to a substance. Precautionary, but unsubstantiated, avoidance of the substance may be considered a safe option, but it can also leave a subject poorly guarded against allergic events. In Ireland, for example, school-going children suspected as being allergic to a food-type are provided with a prescription for four Epi-pens, even before the diagnosis of allergy has been confirmed. One pair of Epi-pens is kept by the guardians of the subject, and the second pair is kept in the subject's school.

While great efforts may be taken to avoid an allergen in the home environment, it is in the unregulated environments that the subject is in greatest danger. For example at school lunchtime, cross-contamination (introduction of allergen from secondary source through contact) of food is common. If a subject has not been diagnosed (or is not waiting to be diagnosed), he/she will not be able to obtain the Epi-pens required to treat severe allergic reactions. They are, therefore, at the highest risk of sudden death as they are without appropriate preventative medicine should anaphylaxis arise.

For this reason, it is very important that subjects who might be at risk of allergy have an early clinical diagnosis. Yet, a very high percentage of adults who believe themselves to be allergic are not, and may therefore be needlessly avoiding food substances and impairing their own quality of life and that of their family.

1.4 DIAGNOSIS OF ALLERGY

Proof of a subject's susceptibility towards an allergen can be obtained clinically in controlled environments. Three types of test exist to assess the vulnerability and these are discussed in the subsequent sections.

1.4.1 BLOOD TESTING

Bio-chemically, symptoms of allergic reactions occur as a result of antibodies stimulating cells in such a way that allergic reactions manifest in a physical manner. Studies have shown that subjects with higher levels of these antibodies have a higher probability of suffering from allergy. However, high levels of these antibodies are not sufficient for a diagnosis of allergy, and low concentrations are not sufficient to rule out allergy.

Blood tests can be considered as an overall 'likelihood of allergy' test. With a blood sample, the plasma can be stimulated with a serum which contains a solution of a specific allergen. By analysing the concentrations of antibodies which are present before and after the serum has been introduced to the blood plasma, the likelihood of existing allergy towards the allergen can be assessed. These tests require specific laboratory equipment and trained personnel.

1.4.2 SKIN TESTING

Skin testing — also referred to as ‘puncture’, ‘scratch’ or ‘prick’ testing in relevant literature — is a quick means of assessing a subject’s allergic susceptibility towards an allergen. A list of potential allergens is drawn up based on a subject’s history. Samples of each are then obtained, and these are mixed with deionised water, producing a water-based solution of the allergen. Needles are then placed in the solution and these are used to scratch the skin of the subject. As well as scratching the skin with potential allergens, the skin is also scratched with a needle that was exposed only to deionised water, providing a true-negative scratch.

If a subject is allergic to one of the substances, a reaction can occur around the area of the scratch mark. The symptoms of the reaction can range from reddening, to inflammation, to an outbreak of hives on the skin. The extent of the reaction is then measured. The larger the response to the scratches with regard to the true negative, the more likely it is that the subject is allergic to the allergen. While allergic reactions will not present on the true negative scratch, the scratching process alone can introduce reddening (and sometimes inflammation) that must be accounted for in the measurement of the other test scratches.

Reaction of the skin to suspected allergens is an indication of the subject’s susceptibility towards the allergen, but, in the same manner as the blood tests, skin tests are not conclusive for a diagnosis of allergy, even in light of a strong reaction to the scratches (Sampson, 1999).

1.4.3 CHALLENGE TESTING

The only clinically validated means of diagnosing food allergy is the oral food challenge (OFC) (Yanishevsky and Hourihane, 2010; Järvinen et al., 2009). During the challenge the subject is required to consume one age-appropriate portion of a food-type. This has the potential to act as a medical poison for some subjects. The portion (e.g. one egg, a

glass of milk, eight peanuts etc) is divided up into five sub-portions doubling in size from approximately $\frac{1}{32}$ of a portion to $\frac{1}{2}$ of a portion. The smallest dose is always administered first. In the case of peanuts, a supplementary sub-portion is introduced where a peanut is rubbed on the subject's lower lip. Depending on the sensitivity of the subject towards peanuts, the lip may swell after this contact, and this is sufficient for diagnosis of allergy. The flowchart for the oral food challenge (OFC) is shown in Figure 1.2, and every stage of this figure will be discussed in this section.

If a subject is allergic to a food-type they might be able to consume a small quantity of the food-type without reacting. The amount they must consume to provoke allergic symptoms can be thought of as a 'reaction threshold,' so to induce a reaction requires this threshold to be surpassed. Reactions will be more severe with the consumption of a greater amount of the food. For the comfort and safety of the subjects who react during the food challenge, it is desirable for the subject to consume the smallest amount possible.

If the subject consumes the full portion of the food-type they are being tested against they are said to have 'passed' the test and may introduce the food-type into their diet. If a reaction to the food-type occurs during the challenge the subject is said to have 'failed' the test and the subject must avoid the allergen.

1.4.3.1 PRELIMINARY TESTS

The subject arrives with their parent/guardian and they are admitted to the day ward. Labelled skin tests are performed before the OFC begins. The subjects are left for fifteen minutes and if the skin has reacted to these allergens in this time, the size of the reaction is measured and recorded. Depending on the extent of the reactions to the solutions, supplementary food challenges may be scheduled. The food-type the subject is being tested against will always be included in the skin test before the OFC, but oral consumption is still required, regardless of the extent of reactions which are obtained from the skin tests.

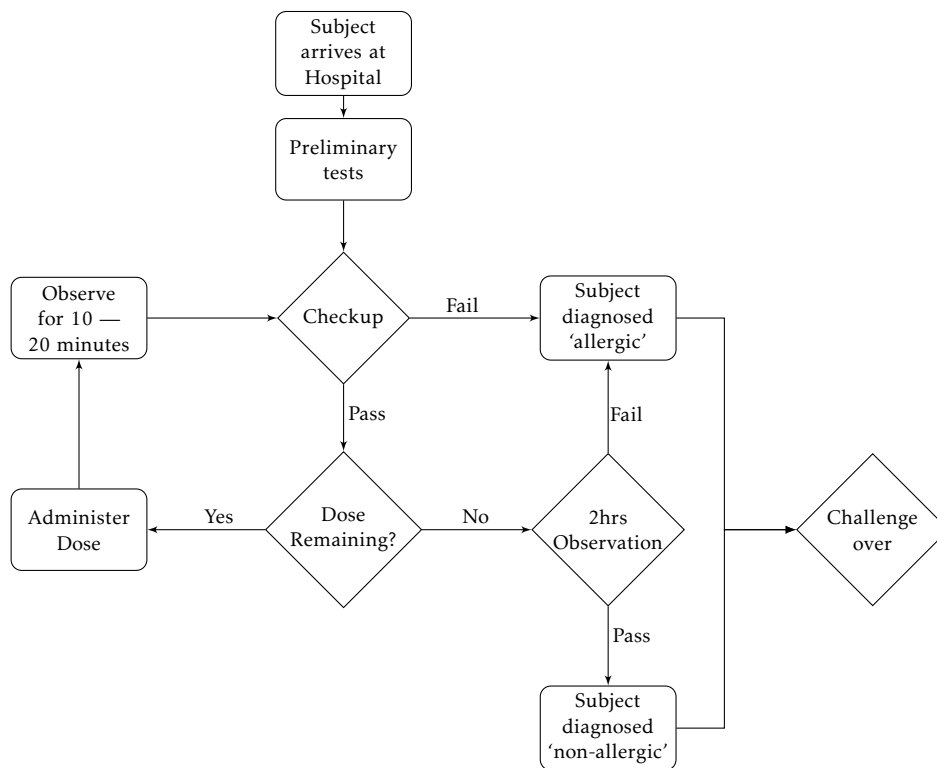


Figure 1.2: Oral food challenge flowchart which presents the means by which allergy is diagnosed in a clinical environment.

If the subject has been sick over the preceding two weeks, or if the subject suffered from an allergic reaction over the same time-period, the challenge will not proceed as the subject's immune system might be compromised, and the results of the OFC may be inconclusive.

1.4.3.2 CHECKUP

The subject is then given his/her first checkup by the allergists. The symptoms of allergic reactions can manifest in an outbreak of hives and rashes on the skin, so it is important to know of any existing skin blemishes *a priori* so they are not mistaken as the physical manifestations of an allergic reaction in later checkups. Indeed, as there is a definite link between allergy and dermatological conditions such as eczema (Bryld et al., 2003), the subjects in question can present with many non-allergy related skin conditions which might be mistaken for allergic rashes later. Therefore, a survey of these is required for accurate diagnosis.

The subject's heart rate, blood pressure, blood oxygen saturation and respiration rate are measured and logged by the allergist. The first sub-portion of the suspect allergen is then administered to the subject.

1.4.3.3 OBSERVATION

After the dose is administered the subject is observed from a nearby station by the nursing staff for 10 — 20 minutes.

If the subject is thought to be reacting to the allergen during this waiting period, the subject is given a checkup by the nursing staff. If the subject fails the checkup the subject is not required to consume any more food, and the failure protocol (see next subsection) will be followed. If the subject passes this intermediate checkup, the remainder of the observation time is allowed to pass.

When the observation period has completed, the subject is given another checkup. If the physiological signals recorded from the subject during the checkup are within the normal range for the subject's age group, and if no manifestations of an allergic reaction have been observed, the next largest dose is administered, and the subject is observed from the observation station for another 10 — 20 minutes. The 'checkup', 'sub-portion administration' and 'observation' sequence repeats until all sub-portions have been fully consumed, or until the subject reacts to the food type.

1.4.3.4 FAILURE PROTOCOL

If, at any stage during the OFCs, a subject reacts negatively to the food-type, rescue medications (i.e. antihistamines) can be administered. At this stage the subject will be diagnosed as being allergic to the food-type, and the guardians of the subject are informed about how to avoid the allergen and shown how to use Epi-pens in case of emergencies. Typically the reactions are not severe, and symptoms such as sneezing, itchy eyes and hives will present. In these situations the administration of Zyrtec is sufficient to halt the effects of the reaction. After this time period subjects are monitored for two supplementary hours as further allergic reactions can lay dormant for this length of time after the final dose of allergen has been consumed. Further antihistamines are administered to the subject during this time period if necessary. Approximately 1 — 3% of OFCs will require the administration of adrenaline even under the close supervision of the allergists.

1.4.3.5 SUPPLEMENTARY STAGES

Many of the symptoms the allergists look for are subjective and differ from subject to subject. For a definitive diagnosis, the allergist must continue to administer the allergen until a reaction occurs. If the allergist is unsure about the cause of a symptom (if the subject seems distressed it may be a result of restlessness rather than the onset of a reaction) they will wait for another 10 minutes or repeat the size of the previous sub-portion as the specific cases require.

If the subject consumes all sub-portions of food without reacting, they are deemed to have passed the food challenge. However, these subjects are also monitored for a further two hours as delayed reactions can still manifest after the challenge has finished. If the subject reacts during this time the failure protocol is followed and the subject is diagnosed allergic. If a subject does not react during the waiting period, they are deemed to have passed the food challenge.

1.5 CHALLENGE-TESTING CLINICAL EXPERIENCE

Over 600 OFCs have been performed by the Department of Paediatrics and Child Health in Cork University Hospital. These controlled OFCs are monitored by staff who have been trained to recognise changes in behaviour and changes in physiological signals which are typical of a subject who has reacted to the food type they are being tested against. However, even with the close supervision provided by the staff there have been seven cases where administration of adrenaline was required.

The allergists who conducted the OFCs identified two events which are suggestive of oncoming allergic reactions (Bindslev-Jensen et al., 2002). These are:

Change in activity: There is a tendency for the subject to become quiet, introverted and still before the onset of a reaction (Bindslev-Jensen et al., 2002). The desire to play disappears. Biologically this is due to inflammatory mediators of the cardiovascular system compensating for the effects of histamine and other mediates of allergic response. At this stage, the magnitude of the reaction is not yet sufficient to present perceivable symptoms. This quieting state does not always transpire for subjects who fail a food challenge as reactions can occur before the subject begins to feel unwell depending on the subject's susceptibilities to the allergen in question.

Change in heart rate: There is a general tendency for the heart rate of subjects to increase as a result of a reaction (Bindslev-Jensen et al., 2002). This was observed during the

checkups that the allergists performed on the subjects during the OFCs. A change in heart rate during a checkup is one of the factors that is used to determine if a subject is allergic to the food-type. However, a change alone is not sufficient to diagnose a subject as being allergic as it has not been definitively linked to the presence of allergy. In the presence of a likely allergic reaction, the absence of a change in heart rate is not sufficient to discount allergy as this is a subjective measure which is only recorded at twenty minute intervals.

Blood pressure and blood oxygen saturation are also measured during checkups. These signals will be the last to change as a result of a reaction, and other symptoms of allergic reactions will be observable before they change. However, if these signals exceed the normal range for the subject's demographics the test will immediately be halted and the failure protocol will be followed.

1.6 MACHINE-ASSISTED CLASSIFICATION OF ALLERGY

Based on the observations which were discussed in the previous section, it was therefore proposed to investigate the effectiveness of machine-assisted diagnosis of food challenge monitoring. In this, the activity and heart rate of the subjects are monitored in a non-invasive and remote manner, and these are later interrogated in order to enquire into the existence of signatures of allergy in the signals.

This monitoring should be performed in an unobtrusive manner without interfering with the subjects, staff or the usual progression of challenges. With the recorded data, characterisation of activity and heart rate will be performed with the goal of detecting, predicting and classifying a subject as being 'allergic' or 'non-allergic'. The classification is designed to complement the allergists diagnoses during OFCs as the allergists cannot be replaced. Therefore, classification must be 'tuned' in an objective manner to ensure that no false positive classifications occur as this would introduce an unnecessary violation to the subject's quality of life indefinitely.

Monitoring of OFCs carries the potential to introduce many advances to the current state of the art. If a correlation between the measurements and the onset of allergy is discovered, automatic classification introduces the possibility of early-detection of allergy. With this, challenges could be stopped earlier, antihistamines could be administered sooner, and the stress on the subject and their family during these tests could be significantly reduced by suppressing the extent of the reaction. Indeed, with machine-assisted monitoring, real-time and non-invasive visualisation of the heart rate throughout the challenges could be obtained. This would also introduce high-resolution clinical information with which the allergists can make their diagnoses.

1.7 LAYOUT OF THIS THESIS

For the purposes of this investigation, the remainder of this thesis takes the following form:

Chapter 2 discusses and reviews the in more detail the algorithms which are employed in later chapters and also provides an overview of the subjects whose data were recorded during the OFC.

Chapter 3 discusses the collection of accelerometer and energy-expenditure data for the purpose of validation of the algorithms which were employed to assess OFCs. This chapter also presents the results obtained when employing these algorithms for the purpose of classification of allergy.

Chapter 4 introduces the heart rate variability features which are extracted from the electrocardiogram (ECG) data which were recorded during OFCs. This chapter also plots probability density functions which demonstrate how effective the individual features were at discriminating between allergic and non-allergic subjects.

Chapter 5 discusses the machine learning algorithms which are employed for the classification of allergy. This chapter also presents the results which are obtained when classification of allergy is performed based on manual QRS annotations of the ECG.

Chapter 6 discusses the concepts of automatic QRS extraction. Two QRS detection algorithms are discussed and are validated against online ECG databases. The effectiveness of the algorithms is then assessed with the ECG recorded during OFCs.

Chapter 7 presents the results which are obtained by combining the capabilities of the topics discussed in Chapters 4, 5 and 6 for the fully automated assessment of allergy classification.

Chapter 8 first summarises the work presented in the following chapters. The main contributions of this thesis are then presented and a number of directions that should be investigated in future work are also listed.

CHAPTER 2

ALGORITHMS, METHODS AND DATA COLLECTION

2.1 INTRODUCTION

THIS chapter provides an overview of the analytic techniques which will be employed later in this thesis. In the previous chapter, it was stated that the allergists, who oversee oral food challenges, observed that subjects who react to the food-type have a tendency to become quiet and still, and it was also stated that there is a tendency for the heart rate of the subject to change before allergic reactions.

In later chapters, accelerometer-based activity, and heart rate variability analysis will be employed to objectively assess the extent to which these changes occur. However, as this is the first work to investigate these signals during allergic events it is not possible to cite review results from other researchers for allergy classification. Therefore, the state of the

art of each of these disciplines is discussed separately in this chapter. Furthermore, a core aspect of this thesis is also in regard to the utilisation of machine learning algorithms for automated decision making. Classification algorithms are also introduced and discussed in this chapter.

As this chapter discusses four separate areas (activity, heart rate, machine learning and data collection) which are — for the sake of clarity in this chapter — independent, these topics are discussed in isolation. Later, if chapters require cooperation between these individual methods, the means by which they engage will be discussed.

Finally, this chapter will discuss the hardware which was employed to record data during the OFCs, and the means by which this was integrated into the oral food challenge for data collection is outlined.

2.2 ACTIVITY ANALYSIS

2.2.1 INTRODUCTION TO INERTIAL MEASUREMENT

Inertial sensors sense acceleration. Recently inertial sensors have been miniaturised with the invention of micro-electro-mechanical systems (MEMS) fabrication technologies. Inertial sensors can accurately sense acceleration in a number of axes (or degrees of freedom (DoF)), and they can be employed to monitor complicated applications with good accuracy.

The most common types of inertial sensors are accelerometer-, gyroscope- and magnetometer-based. These measure acceleration, angular velocity and magnetic fields (i.e. cardinal orientation with regard to Earth's magnetic poles) respectively. Individually, these sensors will typically measure up to three DoFs. However, when combined together in similar planes on one board, a 9-DoF sensor is obtained and these are capable of assessing full kinematic mobility.

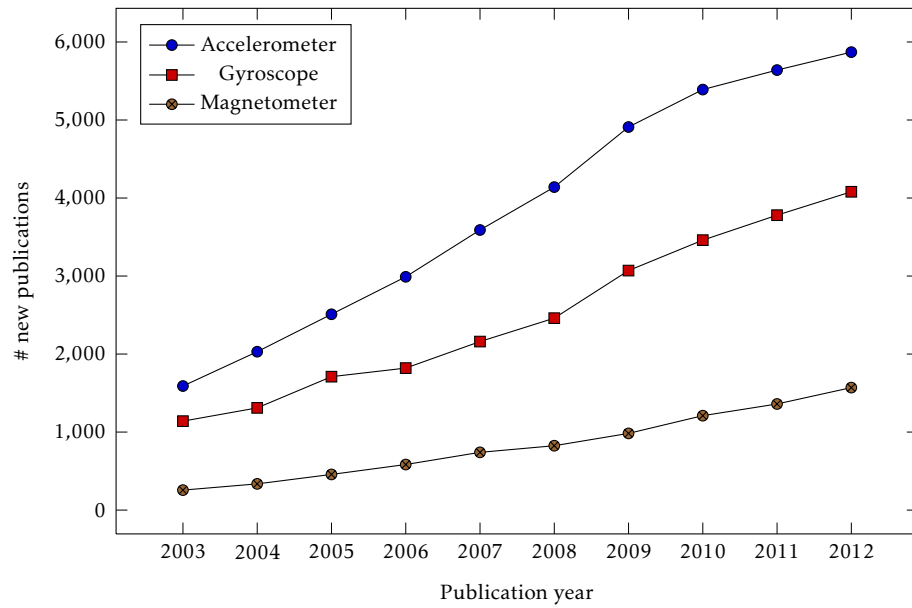


Figure 2.1: Illustration of the approximate growth of microelectromechanical systems accelerometers, gyroscopes and magnetometers in the research market over the past decade. (Google Inc., 2013).

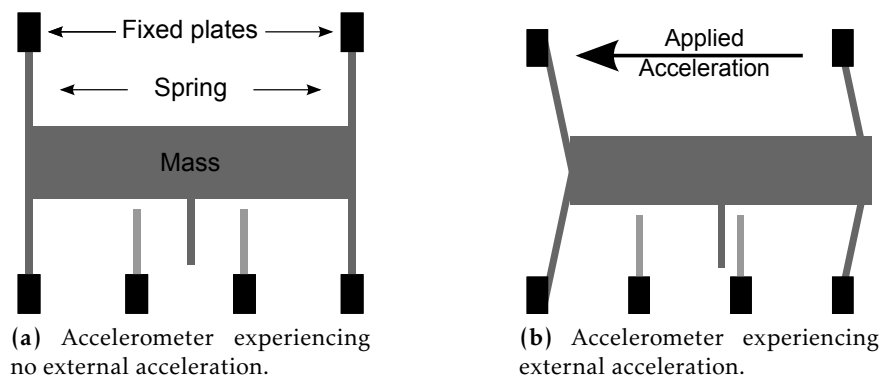


Figure 2.2: Functional diagram of a MEMS accelerometer, identifying the mass and spring.

Accelerometers are the most popular of these sensors both in production volumes and research markets. Figure 2.1 shows the trends of new research articles published over the past 10-years containing the phrases ‘MEMS accelerometers’, ‘MEMS gyroscopes’ and ‘MEMS magnetometers’ on Google Scholar (Google Inc., 2013). This chart shows steady growth of research for each inertial sensor, but, in particular, the popularity of accelerometers has outgrown the other inertial sensors.

Figure 2.2 shows a high-level functional diagram which demonstrates how MEMS accelerometers sense changes in acceleration. These devices consist of fixed plates which stay immobile relative to a device, and mobile plates that are held in position by ‘springs’. Between the springs is a mass which, under the influence of applied acceleration, will physically lag relative to the acceleration that is experienced. This causes the distance between the fixed outer plates and the plate on the moveable mass to change. When these plates are driven with a voltage, the change in distance generates a change in capacitance between the plates. The change in capacitance is then measured by supplementary circuitry, and it is the output of this circuitry which provides the measurement of acceleration.

Inertial sensors have a wide number of applications and have been applied to good effect in a large number of areas, including automotive (Galvin et al., 2000), aerospace (Hayton et al., 2001) and crash testing (Castro et al., 1997) to name a few. However, as the inertial sensors are to be employed to monitor children undergoing OFCs, this section focuses on the application of inertial sensors for human monitoring.

2.2.2 APPLICATIONS OF INERTIAL SENSING

The algorithms which are discussed in this section relate to monitoring of human bodies for the purpose of estimating activity, energy expenditure and other metrics. In all cases accelerometers are employed, but in some cases other inertial measurements are also utilised.

2.2.2.1 ACTIVITY RECOGNITION

Inertial sensing can be applied to recognition of postures, gait and activities (i.e. sitting, standing, walking running, etc). This form of sensing has many applications. For example, Minnen et al. (2005) inferred high-level behaviour from low-level gestures for the purpose of the detection of behavioural syndromes such as autism, Asperger's syndrome, etc. To perform this, three microphones and two accelerometers were worn, and an on-body computer logged the data which were recorded. Simple classification routines were utilised by the researchers, but to good effect. The suitability of the hardware for the application was not discussed even though with three microphones, two accelerometers and associated data logging hardware, this may not be appropriate for persons with behavioural syndromes.

Pober et al. (2006) discussed the application of quadratic discriminant analysis to assess activities. It was shown in this study that the classification of activities is non-trivial, because for some (for example with walking uphill and on flat surfaces) the accelerations which are obtained are very similar. It is possible that with measurement of more DoFs that the angle of ascent could have been inferred to allow discrimination between flat and tilted walks, but this is not discussed. However, with the classification routines which were utilised, good results were obtained. The sample size which was used was also quite small ($n = 6$). While this number of participants is suitable for the validation of pre-existing algorithms, it is small for the generation of new algorithms. The authors also concede this and indicate that there is also value in investigating alternative classification routines.

Staudenmayer et al. (2009) presented a routine where 48 volunteers performed light, moderate and intense exercise for 10 minutes a piece. The classification effectiveness that was achieved was very good. Indeed, Staudenmayer et al. (2009) also demonstrates how energy expenditure can be estimated through accelerometer analysis, and when compared to the ground truth, the results which were obtained were competitive with other researchers, such as Crouter et al. (2006); Freedson et al. (1998); Swartz et al. (2000). The results which were obtained involved activities requiring high degrees of energy, such

as racquetball and basketball. It is shown elsewhere, however, that tests of this nature should not exceed 20 minutes in length as the metabolic response of persons after this time is unreliable (Winter et al., 2006). This is because as the body is exerted to intense or lengthy physical exertion, different attributes will affect the metabolic response which is experienced. Therefore, while the results obtained by Staudenmayer et al. (2009) were accurate, there is the possibility that they may still have been corrupted by this factor.

Benbasat and Paradiso (2002) presented a generic motion recognition framework which employed six DoFs to track movements. This platform was designed so that it could be used by researchers with minimal requirement for knowledge of the underlying algorithms. However, while the platform performed very well and adapted to many situations appropriately, it occupied a volume of approximately 500 cm³, which is not practical for many applications, in particular in the medical setting.

Recently, with the integration of inertial sensors into mobile phones, activity analysis has become readily accessible to smartphone users. Interestingly, it was found that by analysis of a single accelerometer from within a smart phone in a user's pocket, it was possible to predict the identity of users (Kwapisz et al., 2010). Movement-based identification has been performed before in image analysis (Kale et al., 2003) and with multi-sensor gait analysis (Annadhorai et al., 2008) previously, but the novelty of Kwapisz et al. (2010) was the first which obtained this from single sensors in the pocket.

Another active area of research is in the estimation of energy expenditure through accelerometer analysis. In particular, with the popularity of mobile phones, and with their integration with accelerometers, overall assessment of daily energy expenditure can be achieved with accelerometer data alone. Indeed, this is a primary application of a number of youth-based anti-obesity initiatives where, integrated with social networking websites, daily activity can be posted as a means of encouraging exercise (Lanningham-Foster et al., 2009). Research into activity-based intervention monitoring to combat obesity is gaining momentum (Oude Luttikhuis et al., 2009) and these means of research are not only applicable for countering obesity in youth (Lobstein et al., 2004; Wang and Lobstein,

2006), but it also applicable to the adult generation which is also susceptible to this epidemic (WHO, 2000).

Accelerometers can also be employed for the analysis of rehabilitation exercises (after a fall, for example). This can be achieved because MEMS accelerometers can provide reliable and objective measures of mobility, gait and balance (Culhane et al., 2005). Accelerometer-based rehabilitation has also been employed to measure the mobility of sufferers of strokes by Uswatte et al. (2005, 2006). These studies have shown that accelerometers have been shown to objectively quantify mobility scores which are within 10% of human-recorded scores (Hester et al., 2006), which shows that accelerometry is an accurate measurement for fully mobile and constrained applications.

2.2.3 ACTIVITY-BASED ANALYSIS OF OFC

In the previous chapter it was stated that there is a tendency for subjects to become quiet and that there is a tendency for them to play less in light of allergic reactions. This effect can be objectively measured by the measurement of activity and by the estimation of energy expenditure, and accelerometry is an appropriate tool for this measurement. Indeed, it has been shown here that accelerometry has the capacity to analyse both activity and energy expenditure in free living and constrained environments alike.

For accelerometer-based analysis of OFCs, the primary means of assessment will focus on activity– (i.e. movement) and accelerometer-based energy expenditure. Postural and activity classifications are not performed on this data because the subjects are required to stay on a bed throughout the challenges, and will, therefore, typically be lying supine.

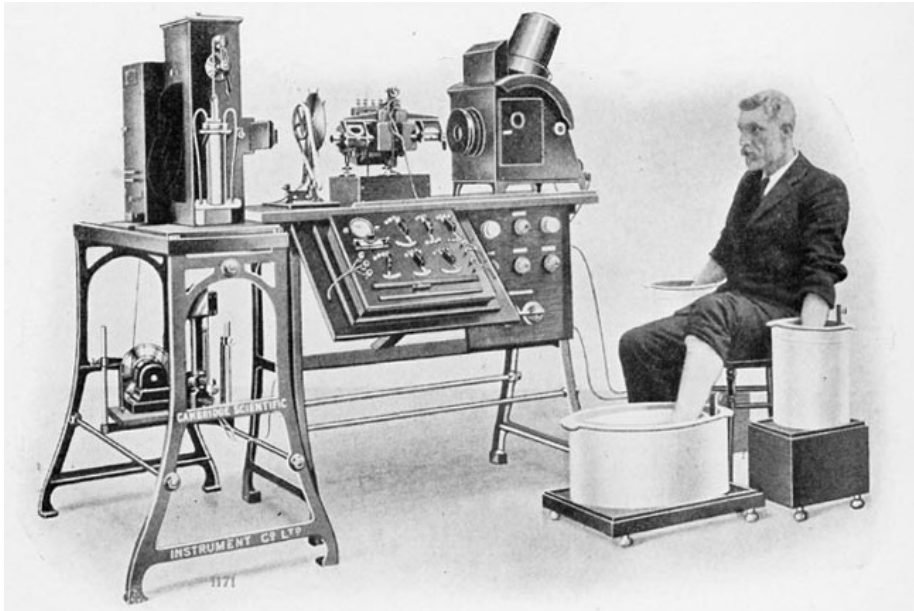


Figure 2.3: The original galvanometer developed by Willem Einthoven to record the ECG in the early 20th century.

2.3 HEART RATE ANALYSIS

2.3.1 HISTORY AND INTRODUCTION

Clinically, ECG analysis is one of the most important area of research for the assessment of cardiac health. The development of modern machines which can record the ECG was pioneered by Willem Einthoven with the invention of the string galvanometer (Barold, 2003). A representation of this machine is shown in Figure 2.3 where the subject has placed his hands and legs in buckets which contain a salt water solution. This is performed to increase the conductivity of the electrical signals of the heart to the galvanometer. The signals were then amplified by a large electromagnetic amplifier, and a string was deflected based on the signals which are obtained. This deflection was transferred to paper and provided a representation of the ECG.

The naming conventions of the ECG waveforms that are used today were initially coined by Einthoven. However, while the terminology is similar, the technology which is employed in the modern recording of the ECG has changed dramatically in size, power and proficiency. Indeed, with today's technology, the ECG can be recorded with circuitry no bigger than a coin for many hours (Burns et al., 2010), whereas with Einthoven's contraption, five assistants were required to operate the machine, and water was required to cool the active mechanisms (Churchill, 2008).

Even at its primitive beginnings, the ECG was shown to be diagnostically valuable, and doctors, such as Einthoven and later Thomas Lewis, pioneered the investigation of the effects of disease on the ECG. Since this era, the development of cardiology has evolved, and now 12-lead ECG is the state of the art for preliminary diagnosis of heart disease. This type of analysis can not only detect irregularities in the ECG, but with training, cardiologists can also localise the precise area of the heart which is subject to the condition without having to perform more invasive tests (Fuchs et al., 1982), such as contrast-enhanced ultrasound (Martegani et al., 2008; Furlow, 2009). However, upon such diagnoses, more invasive procedures may be scheduled in order to obtain more detailed information about the condition under investigation.

Due to the expertise required to apply and interpret 12-lead ECG, these analyses are typically only performed by trained cardiologists when clinical diagnosis of heart disease is required. The lengths of these recordings are typically quite short. For medium-to-long term ECG recordings, 3- and 5-lead ECG are used, which can result in ECG traces similar to that shown in Figure 2.4 (Dublin Institute of Technology, 2013) in which the segments of the ECG are annotated as described by Einthoven. While these 3-lead recordings can highlight cardiac arrhythmias and other heart conditions, their specificity is quite low, and these recordings are often discarded after recording.

ECGs recorded with 3- and 5-leads are more typically used in the assessment of the variability of the heart over time. Statistical measurements obtained between sets of R-R intervals of the heart rhythm are extracted, and these are known as features (see Figure

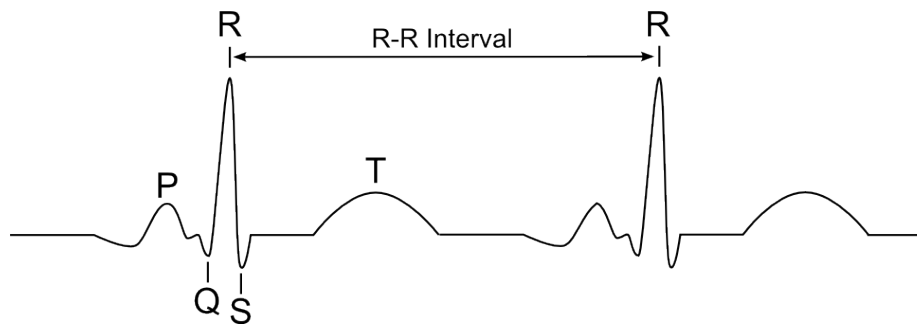


Figure 2.4: Labelled ECG waveform (Dublin Institute of Technology, 2013).

2.4). The variation of these is computed mathematically to quantify the state of the heart. This is known as heart rate variability (HRV) feature analysis.

2.3.2 APPLICATIONS OF HRV ANALYSIS

The ECG is the best way of measuring variation in the heart (Sabiston Jr, 1981). The ECG has previously been shown to change due respiration (Baldzer et al., 1989; Yamamoto et al., 1991; Jan et al., Nov; Kemp et al., 2010), exercise (Robinson et al., 1966; Tulppo et al., 1996; Cole et al., 2000; Sandercock et al., 2005), stress (Falkner et al., 1979; Kostis et al., 1982; Bernardi et al., 2000; Obrist et al., 2007; Riese et al., 2004; Bořil et al., 2012; Bailón et al., 2010), hypo-tension (Hernando et al., 2011), heart disease (Dekker et al., 2000; Tsuji et al., 1996; Antelmi et al., 2004; Nolan et al., 1998; Licht et al., 2008), anxiety (Licht et al., 2009), asphyxia (Boardman et al., 2002) and later in this thesis, investigations are performed into whether the heart rate variability due to allergic reactions.

The heart rate variability (HRV) features, and in particular the sympathetic and parasympathetic indices, are strong indications of the risk of sudden adult death (Bradley and Floras, 2003; Lombardi et al., 2001; Seccareccia et al., 2001; Dekker et al., 1997; Huikuri et al., 2001). This is a condition in which otherwise healthy adults die suddenly. Risk factors for this include a number of well-known arrhythmia and other heart conditions which can be measured with the ECG. Screening can be performed with ECG and persons

can be alerted to this which will then allow for medication and exercise regimes to be used by the patient so that the risk of death is reduced.

Much work has been performed into digital signal processing (DSP) related investigation and conditioning of the ECG (Sörnmo and Laguna, 2005, 2006; Schlindwein et al., 2006). For example, the effect of breathing can be ascertained from signal processing analysis of the ECG (Bailón et al., 2007). Indeed, HRV analysis has been used in order to assess seizure in neonatal (Doyle et al., 2010) and adult (Jeppesen et al., 2010) hospitalised patients. The results which have been obtained with these correlations have shown that HRV-based seizure detection is difficult to obtain in general, and that electroencephalogram (EEG) analysis provides the best results. However, it is the case that some patient's HRV features react strongly to seizure, and for these reasons their analysis can offer good subjective seizure appraisal when applied with appropriate persons (Doyle et al., 2010).

While the role of HRV has been stated as being capable of detecting abnormalities in the ECG, it can also be used to characterise normality in adults (Nunan et al., 2010) and children (Aziz et al., 2012) which can be employed to rule out certain diseases.

Heart monitoring can be used in many situations, even outside of medically diagnostic applications. For example, HRV analysis can be employed in order to assess the anxiety and stress of persons throughout their daily work commute and employ. With medical professionals, for example, stress can impact their work, decision making and vigilance. Therefore, with HRV analysis, times at which these personnel become stressed can be assessed automatically (Jovanov et al., 2011), which leads to the possibility of alerting the medic of their state of stress. If conditions allow, relaxation methods could be followed. With car driver analysis, the heart rates can be detected and drivers who might be suffering from road rage can be identified (Bořil et al., 2012; Healey and Picard, 2005). This can provide an objective feedback mechanism for drivers and can contribute to obtaining safer roads. In both of these examples, however, the participants must be compliant with analysis and towards the feedback for benefits to be obtained from this.

Therefore, the use of the HRV features for the detection of signatures of allergy will be investigated in order to discriminate between allergic and non-allergic subjects. HRV features have been shown to be of good clinical use for the assessment of many stressful environments and many acquired and chronic medical conditions, and it is believed that signatures of allergy will be uncovered by analysis of these features.

2.4 MACHINE LEARNING AND CLASSIFICATION

2.4.1 INTRODUCTION TO CLASSIFICATION

Applications in which activity and HRV analyses have been employed have been described in the previous sections. However, in order to use these analyses in an automated manner, it is generally necessary to employ classification algorithms. A number of machine learning algorithms which would be appropriate for classification of OFC are described here. It is important to note that the algorithms which are employed are generally independent of their applications and can be used to solve a very wide range of other applications when deployed appropriately. For example, classification can be used for many applications including email spam detection (Andersen et al., 2008), fraud detection (Phua et al., 2004), vibration analysis of bridges, cars, engines, etc Oh et al. (2009), to name a few. Machine learning has also applied to activity analyses (Ravi et al., 2005; Mannini and Sabatini, 2010) and HRV (Kononenko, 2001) . The use of classification for these will be discussed in a later section.

2.4.2 BACKGROUND TO MACHINE LEARNING

Classification is the process of automatically and intelligently assigning labels to data. The label which is assigned is subject to the application which is being investigated. For example, with accelerometer-based classification the task at hand might be to determine

the activity which is being performed, and possible labels might include walking, running, jumping, etc. Likewise, for ECG analysis, the task might be to determine if a specific event has occurred, and the labels might be normal rhythm, arrhythmia, ectopia, etc (the meanings of these conditions are discussed later).

Two main branches of classification are available: supervised and unsupervised. In supervised learning, all the data which is employed is labelled according to the application, whereas in unsupervised learning, supplementary algorithms are first required to automatically assign labels to the data, and then classification can be performed with these labels. In this section (and indeed for the remainder of this thesis) the classification routines which are discussed relate to supervised classification only.

Two independent stages are involved with supervised classification: training and testing. The training stage involves employing data with known labels and ‘feeding’ this data into a learning algorithm. Once this process has completed, other data (which are also labelled) are used to test the classifier. By this process the performance of the classifier can be assessed as selected classification labels will be compared to the ground truth. The process where the training data and testing data are separated is the preferred routine which can be employed as it can assess the generalisation error of the problem at hand (Vapnik and Kotz, 1982).

The most common type of classifier is the discriminative classifier. This type of classifier is trained on more than one data label. Then, by the algorithms within its framework, the classifier can learn to discriminate between the classes it was trained with. It is not guaranteed that the ground truth will be recovered by classifiers, however, and the accuracy of classification is subject to the application in question and the data which are available.

Many different types of classifier exist, including — but not limited to — support vector machines (SVMs), Gaussian mixture models (GMMs), k -nearest neighbourss (kNNs), naive Bayes, etc. When multiple learners cooperate together for tackle a machine learning problem procedures are termed ensemble or boosting methods. Eventually, all of these

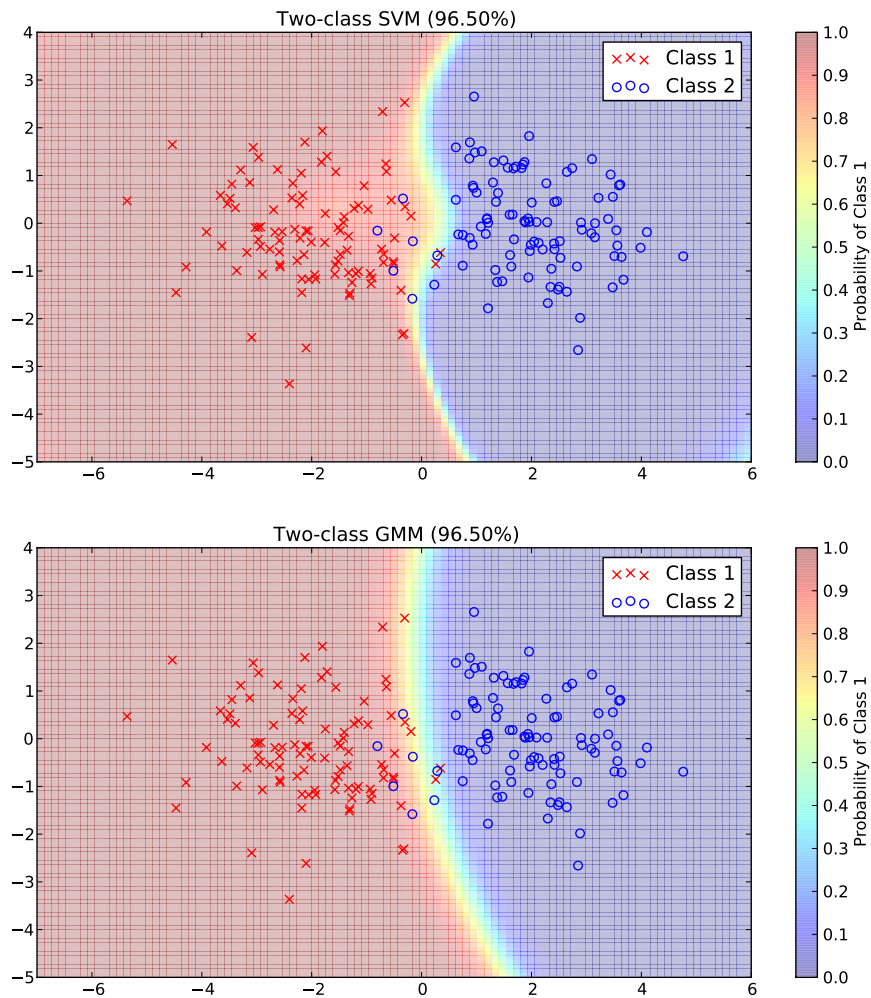


Figure 2.5: Example of how classification is obtained with two-dimensional data (top) with SVM based classification (bottom left) and GMM based classification (bottom right).

algorithms reduce to determining which of a number of labels is the most likely, having been trained on one dataset and being tested on new data. Clearly, for successful application of machine learning, the training dataset must contain a wide range of examples that are representative of the problem under investigation. An example of machine learning can be visualised in Figure 2.5, where SVM and Gaussian mixture model (GMM) classifiers are employed in order to classify two classes (i.e. labels) of data which are marked as $\times$ and $\circ$. Areas on the left hand side of the images (highlighted in red) are the regions which have been selected as being more indicative of the $\times$ class, while those highlighted on the right hand side (in blue) are more indicative of the $\circ$. It can be seen that for this simple example, the classification routine attempts to split the feature space in two regions, one for each class of data. An obstacle which is common to all classifiers is the means in which data should be best modelled in order to minimise (and ideally eliminate) uncertainty between the classes. This example shows that complicated decision boundaries can be obtained for relatively well-behaved data.

With the example in Figure 2.5, a case where two classes have been analysed is presented. It can sometimes occur that more than two classes require classification, for example previously it was stated that activity classification might assign walking, running, and standing labels. However, it can also be the case that knowledge of only one class is available. In this case, the classification routine is termed ‘novelty’ or ‘abnormality’ detection.

2.4.3 NOVELTY DETECTION

The task of novelty detection algorithms is to detect data which are not from the distribution of the training data. These are termed ‘novel’ data points. This type of classification is the means by which allergy must be classified because only one class of labelled data is available, and this is the data which were recorded before the administration of the first dose of the allergen. This is guaranteed to be non-allergic because the allergen has not yet been administered. Classification of allergy, therefore,

involves training classification routines on this non-allergic data and determining the boundaries of novelty from this. Data found to be outside this boundary would then be considered novel (i.e. allergic).

With the allergy data, one could hypothesise about the temporal labels during the OFC and thereby obtain a discriminative problem (e.g. one could assume that features 10 minutes prior to a reaction belongs to the allergic class, for example). However, these hypotheses will only enable the classifier to learn this assumption rather than facilitate detection of the true signatures of allergy, and this may obtain sub-optimal performance. Consequently, novelty detection is the preferred method for allergy detection.

Novelty detection applications are not as popular as multi-class problems, but they have been employed for many applications as expert labels are often difficult — and sometimes impossible — to obtain. These algorithms are trained on the data which are labelled and obtain ‘boundaries of novelty’ and the data within these boundaries are ‘normal’ while those from without are novel. The two most popular types of novelty detection routines involve support estimation by a one-class SVMs, and density estimation by GMMs.

2.4.3.1 ONE CLASS SVMs AND GMMs

The discriminative SVM was mentioned previously and an example of how it partitions the data was shown in Figure 2.5. SVMs compute a maximum-margin hyperplane between the distributions of the classes which are investigated (Flach, 2012). However, with one class SVMs, only one class is available for training, and a maximum margin hyperplane cannot be obtained. One-class SVMs therefore learn a hypersphere which ‘surrounds’ a specified percentage of the training data, and the radius of this sphere is employed to estimate the support of a distribution (Schölkopf et al., 2000, 2001). Then, data which lies within this sphere are considered normal, and data which is outside the sphere are novel. The extent of the novelty of this data can be calculated by computing the perpendicular distance from that point to the hypersphere. In effect, the radius of this sphere is a threshold which is selected during the learning process.

The GMM classifier sets out to model a distribution with a mixture of Gaussians. With this, the ‘probability’ that new data are normal can be assessed by ‘reading’ the value off the learnt distribution. It is, however, necessary to apply thresholding for GMM-based novelty detection after the distribution has been modelled. This is acceptable as this is also performed by the one-class SVM, but this process is ‘hidden’ by the learning algorithm.

2.4.3.2 EXAMPLE OF NOVELTY DETECTION

Figure 2.6 shows an example of novelty detection algorithms being employed. Here, the same data that was used in Figure 2.5 were employed, and the novelty detection routines were trained on data from Class 1 only. In the images, the red areas are more indicative of normal data, and blue areas are more novel. The upper image shows the decision surface obtained from the one-class SVM. While this algorithm learns a hypersphere, the shape of the separation is not circular because of the ‘kernel trick’ (Aizerman et al., 1964) that was employed. This projects the original data into high-dimensional feature spaces, and it is in these feature spaces where the sphere is obtained. While the examples in Figure 2.5 effectively separated the feature plane, novelty detection is more likely to isolate a small region of the plane and define this as being ‘normal’.

With both examples here, the normal data are well modelled and the highest probabilities of ‘normal’ data are found in the centres of the distribution. The further away from this that the data points are found indicate more novel features. It is difficult to say which routine is superior due to the fact that both have been employed very successfully in many applications. In the example in Figure 2.6 the one class SVM learns the boundary of normal data which provides a quick roll-off between ‘normal’ and ‘abnormal’ data. GMM-based modelling models the distribution, so it has a slower roll-off than the SVM-based procedure.

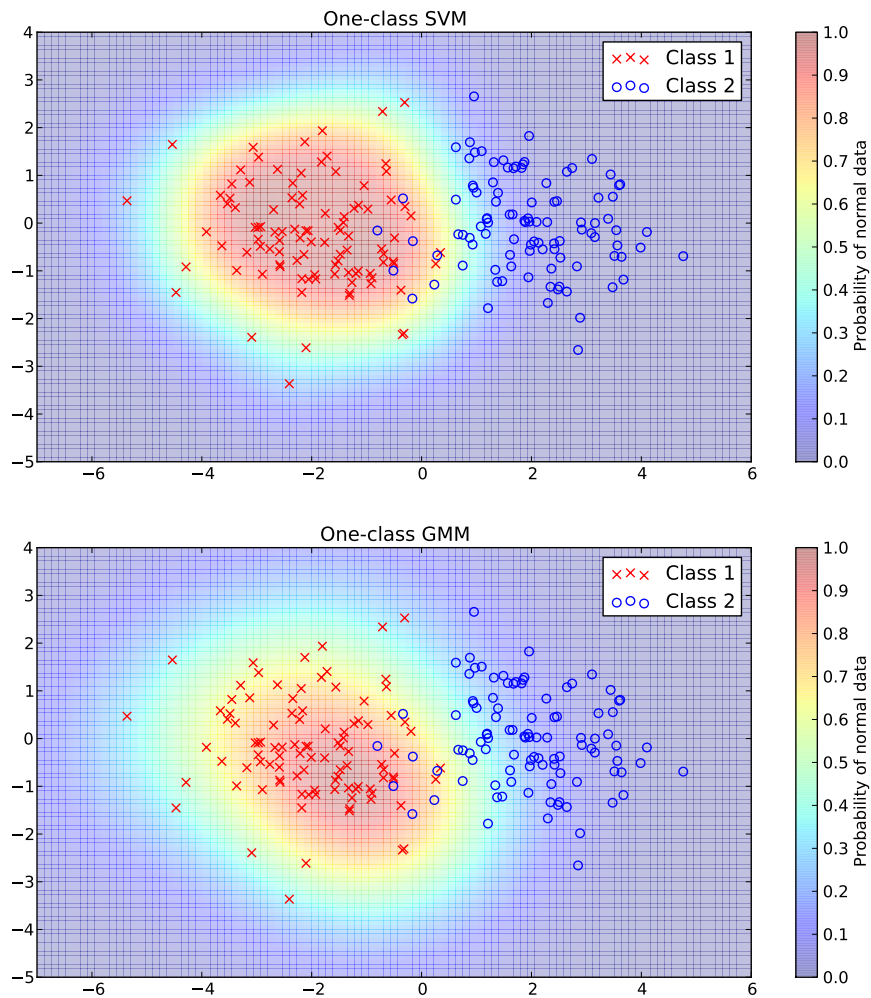


Figure 2.6: Example of how novelty detection algorithms can be employed to determine novel and normal data points. The upper image shows the distribution of the labels, and the lower-right figure shows one-class SVM while the lower-left figure shows the example of GMM-based novelty detection on the same data.

2.4.4 APPLICATIONS OF NOVELTY DETECTION

GMM- and SVM-based novelty detection algorithms have been employed in a number of machine learning applications to very good effect. This section will focus on physiological-based classification as this is the application which is under investigation in this thesis.

Novelty detection has been used for the automatic and online classification of epileptic seizure with intracranial EEG (Gardner et al., 2006; Gardner, 2004). This classification routine involved the use of one-class SVMs and obtained excellent sensitivity of classification. However, while excellent sensitivity was obtained, the authors did not report the specificity of the routine. The methodology investigated also only used three energy-based features to describe the EEG, which is a very low number of features in comparison to the state of the art in seizure detection (Temko et al., 2009, 2011b; Thomas et al., 2009).

Roberts (1999, 2000), investigated the use of extreme value theory (the branch of statistics which deals with abnormally high or low values in distributions) in conjunction with GMMs in order to assess a number of different applications in the medical and image processing fields. It was shown that GMM-based novelty analysis of hand tremors, epileptic seizure, vigilance (during repetitive, boring or long-term tasks) and anaesthesia are applications with which novelty detection can be employed to classify medical conditions. These papers do not provide any metrics of the overall performance of classification, but rather demonstrated the applicability of novelty detection for these applications. However, they do show that GMM-based analysis is useful for movement (i.e. accelerometer-based) applications, and that it is also appropriate in medical studies.

Speech recognition has also employed novelty detection (Markov and Nakamura, 2008) for improved speaker identity, and up until very recently GMMs are the state of the art for speaker identification (Fine et al., 2001; Faundez-Zanuy and Monte-Moreno, 2005).

2.5 FOOD CHALLENGE DATA COLLECTION

2.5.1 RECORDING PLATFORM

The requirements for the recording platform are that the acceleration and ECG must be recorded concurrently during OFCs. Then, DSP is employed to identify the applicability of the acceleration and ECG data for classification of allergy.

The sensing health with intelligence, modularity, mobility and experimental reusability (SHIMMER device) (SHIMMER research, 2010; Burns et al., 2010) is the wireless sensing platform that was chosen for data collection during the OFCs, and is illustrated in Figure 2.7. This is a sensing device which features Bluetooth and 802.15.4 radios, tri-axial accelerometer (Freescale MMA7260Q, with programmable sensitivity*) with programmable axial sensitivity (selectable from 1.5 g, 2 g, and 4 g), a micro SD card slot (up to 2 GB) and a ultra-low power microprocessor (MSP430). The microprocessor is programmed with the network embedded systems C (nesC) event-driven programming language, and is supported by the TinyOS component-based operating system. The SHIMMER device also supports ‘daughterboards’ which extend the functionality of the platform, and the ECG daughterboard allows for ECG recordings. The SHIMMER device offers a small form factor (53 mm × 32 mm × 25 mm), which is approximately 10 times smaller than the platform that was previously discussed in this chapter by Benbasat and Paradiso (2002). This is a very important consideration for the monitoring of children during OFC. The SHIMMER device has been awarded European conformity (CE) certification mark (SHIMMER-research, 2010), signifying it fulfills European Union health and safety requirements for clinical-based use, and it is therefore appropriate for data collection during OFCs. The acceleration signals and the ECG traces were sampled at 256 Hz.

*These accelerometers are sufficiently sensitive for gait- and energy expenditure-based analysis and have been used for many such application domains.

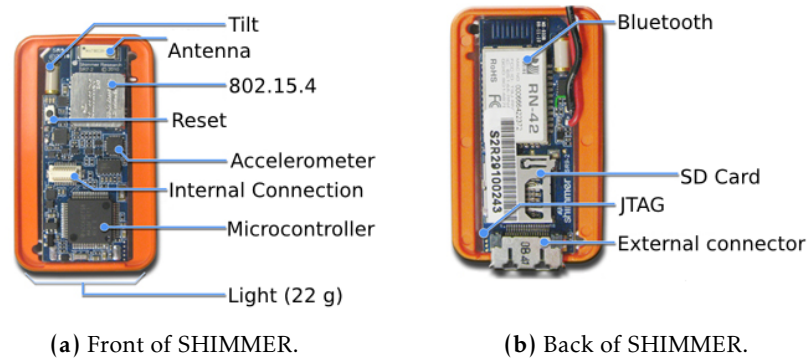


Figure 2.7: Diagram of the SHIMMER device, with the various components annotated (reproduced with permission from SHIMMER-research (2010)).

2.5.2 INTEGRATION WITH ORAL FOOD CHALLENGE

The Department of Paediatrics and Child Health in Cork University Hospital conduct OFCs on a regular basis. A collaboration between the Department of Electrical and Electronic Engineering of University College Cork and the Department of Paediatrics and Child Health of Cork University Hospital was organised, and ethical approval was sought and obtained from the ethics board to collect accelerometer and ECG data from subjects undergoing OFC with the SHIMMER device. Participation in data collection was voluntary, and if the parents or the child did not wish to partake, the OFC progressed as normal.

With the requirements for data collection defined, the procedure of the OFC was modified slightly in order to accommodate data collection. Figure 2.8 presents a modified flowchart in which two supplementary stages are introduced to facilitate the use of the SHIMMER device for data collection.

The first supplementary step required is to obtain informed consent from the subject's guardians for monitoring the subject's physiological signals during the food challenge in accordance with the ethical approval. Allergists discuss with the guardian the data-collection procedure and provide them with literature they must read before signing the consent forms. If consent was provided, a chest strap holding the SHIMMER device

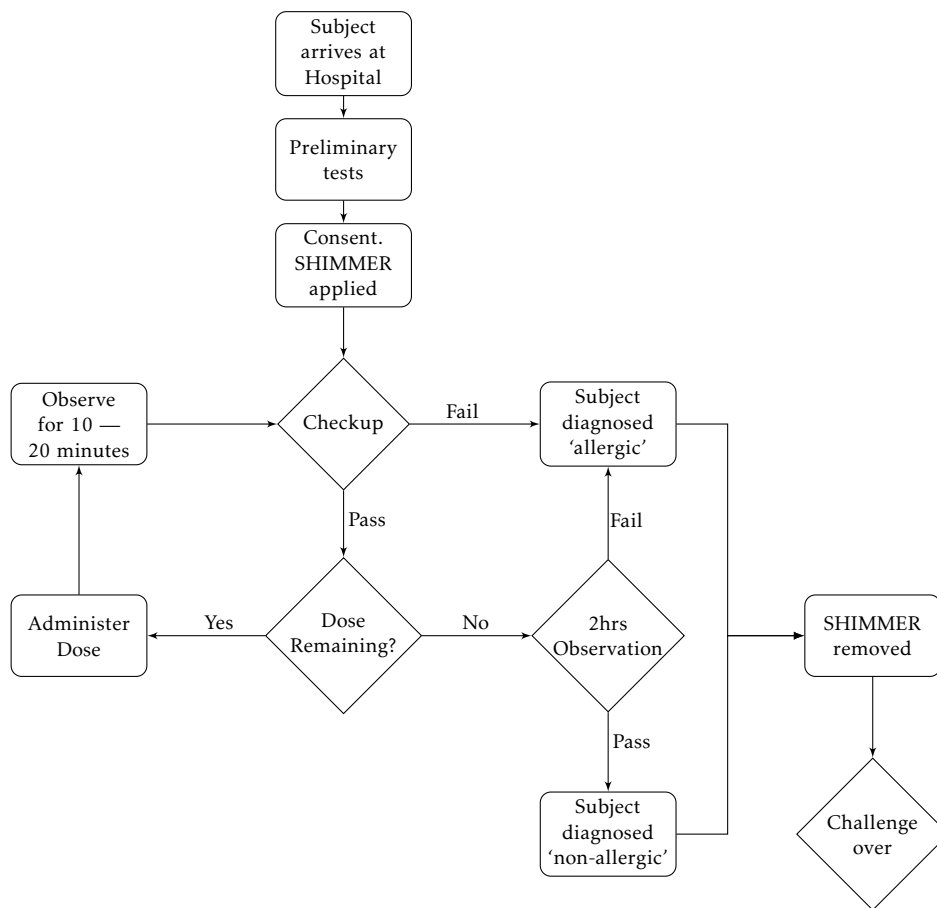


Figure 2.8: Modified oral food challenge flowchart which is employed to accommodate the introduction of the SHIMMER monitoring device for data collection during OFCs.

was fastened around the trunk of the subject and the ECG electrodes are configured to record data. The SHIMMER device then connects to a host computer over which data is transmitted. The second supplementary stage involves removing the SHIMMER device from the subjects after the test has concluded. The OFC process is unchanged at this stage, and diagnoses are unaffected.

2.5.3 OTHER DATA

While the blood pressure and blood oxygen saturation levels are recorded periodically by the allergists who conduct the OFCs, these were not recorded during the data collection stage. The primary reason for this is because the eventual allergy classification framework is envisioned to require minimal interaction by the allergists, and to also record the physiological data in a non-invasive manner which will introduce as little discomfort as possible to the subject.

It would not be in line with this methodology, therefore, to require the allergist to log with blood pressure and blood oxygen saturation levels in a computer after every checkup. Moreover, it is also not feasible to delegate equipment to record these (in particular the blood pressure) as they are invasive sensors that require compliance and subject-participation for meaningful readings.

It is also the case that blood pressure and oxygen saturation signals will be the final physiological conditions which change under the influence of allergy, and these will only change after perceivable symptoms have manifested.

2.5.4 DATA RECORDING

In total, the acceleration and ECG data of 24 subjects were recorded during OFCs. Of these subjects 15 reacted to the food type during OFC and were diagnosed allergic. These subjects are tabulated in Table 2.1. All subjects are ten years of age and under. The

Table 2.1: Tabulation of the characteristics of the subjects who were recorded for this study.

Index	Gender	Age	Recording length (minutes)	Allergen	Diagnosis
1	male	1.5 years	14	wheat	Allergic
2	male	6 years	95	peanut	
3	male	9 years	90	egg	
4	male	12 months	100	milk	
5	male	8 years	120	peanut	
6	female	9 years	33	peanut	
7	male	6 years	57	soy	
8	male	5 years	100	peanut	
9	female	8 years	50	egg (cake)	
10	male	3 years	82	milk	
11	female	6 years	85	peanut	
12	female	5 years	40	milk	
13	female	3 years	105	milk	
14	male	8 years	125	soy	
15	female	9 months	96	wheat	
16	male	6 years	69	egg	Non-allergic
17	male	10 years	91	egg (cake)	
18	female	4 years	125	soy	
19	male	6 years	33	peanut	
20	female	1.5 years	110	milk	
21	female	7 months	57	milk	
22	male	12 months	81	milk	
23	female	4 years	58	wheat	
24	male	2 years	90	peanut	

shortest test was obtained with Subject 1 whose OFC lasted for 14 minutes, and the longest challenge lasted 130 minutes (Subject 19). While the lengths of the challenges should be consistent, in particular for non-allergic subjects, the OFC is a dynamic test in which delays can occur, and this explains the variances in challenge length.

2.6 CONCLUSION

It has been shown in this chapter that inertial measurement and ECG analysis can each be employed for the classification of many physiological and clinical applications. In particular, changes in HRV have been associated with many chronic and acquired medical conditions. It has, therefore, been reasoned here that, while these have not been used for allergy detection, acceleration and ECG analyses are capable of assessing the observations of the allergists in an objective and quantifiable manner. Later chapters will now investigate the applicability of these algorithms and for machine learning based classification of allergy. The work presented in the remainder of this thesis is original and has not been investigated by other researchers before.

CHAPTER 3

ACCELEROMETER-BASED ANALYSIS OF ORAL FOOD CHALLENGES

3.1 INTRODUCTION

THIS chapter introduces the concept of accelerometer-based activity and energy expenditure estimation. This is investigated because allergists have reported that there is a tendency for the activity and energy levels of subjects to change before an allergic reaction presents, and these can be measured with accelerometers.

Energy can be measured in a number of ways. Physically, the gold-standard of energy expenditure measurement is calorimetry. This measures the heat that radiates from a body and, in combination with metrics such as the size and mass of that body, calorimetry calculates true energy expenditure. This process is feasible for the measurement of energy

from small immobile objects, but it is not feasible to use this process to measure the true energy expenditure of people in a non-invasive and comfortable manner.

Therefore, indirect calorimetry was devised. This exploits the fact that in order to expend heat, humans (and indeed all animals) must breathe in oxygen and expel carbon dioxide and other gases. By measuring the volumes of O_2 and CO_2 entering and leaving a subject as they breathe, and by combining this with standard physiological characteristics (such as mass, height, etc), the true energy expenditure of a subject can be inferred. This process is called indirect calorimetry as energy is not directly measured. Two indirect calorimetry processes exist for the measurement of energy expenditure in humans. Both of these methods measure the same data with similar accuracy.

The first method utilises a mask that must be worn by a subject who breaths through this into a tube. The oxygen and carbon dioxide content of the subject's breath is analysed by specialised and calibrated hardware (Cosmet, 2013). This mechanism is carefully designed to ensure that that minimal difficulties are introduced to a subject's respiratory process by the applied hardware. The system, based on the gas levels recorded, then produces a report of energy expended every time a subject has completed a respiratory cycle, i.e. the subject must both inspire and expire before energy expenditure readings are obtained as both processes provide metrics that are required for the analysis. The regression process was described in detail by Ferrannini (1988).

The second method that can be used is called the doubly labelled water method (Schoeller et al., 1988). This process employs isotopes of hydrogen (deuterium) and oxygen (oxygen-18). These isotopes are not radioactive or dangerous, and can be found naturally in the atmosphere. Here, they are used as chemical markers. The doubly labelled water method is conducted within a sealed room where the air and temperature conditions are monitored and controlled by trained technicians. The isotopes are introduced into the atmosphere of the room in known and controlled volumes. By monitoring the gases expressed by the respiration process which are marked by these isotopes, and by monitoring the change

in deuterium and oxygen-18 in the room, true energy expenditure can be computed with similar equations that were employed with the first method.

Both of these methods measure true energy expenditure, but the doubly labelled water method is the lesser used of the two as it is more expensive to operate and requires a dedicated room with sophisticated controls and experienced personnel to correctly operate the protocol. It is not practical to employ indirect calorimetry in every situation, however, and it is not suited for recording data during oral food challenges where subjects are young and often become sick and require medical assistance and the comfort of their parents.

Therefore, a recent area of research has been in the estimation of true energy expenditure based on the monitoring of signals which can be obtained in a non-invasive, low-power and low-cost manner. Signals obtained from uni-axial (single-dimensional) and tri-axial (three-dimensional) accelerometers have been shown to be highly correlated with true energy expenditure. Figure 3.1 shows a diagram of a body and the three axes of acceleration that are monitored by tri-axial accelerometers. Here, antero-posterior is the x -axis, mediolateral is the y -axis and vertical is the z -axis. Indeed, accelerometry can provide a number of advantages over indirect calorimetry and Chapter 2 demonstrated how acceleration signals can be employed for activity analysis too. In this chapter, a distinction is drawn between activity and energy expenditure analysis — activity is the measure of movement and mobility, whereas energy expenditure employs activity analysis to estimate energy.

This chapter compares a number of accelerometer-based activity and energy expenditure estimation algorithms. These algorithms are validated with independent data recorded during new experiments in order to ensure that the values which result from these algorithms are accurate. The accelerations of subjects undergoing oral food challenges which were recorded are then assessed yielding activity and energy expenditure estimation readings. Based on the results of these, the applicability of accelerometer-based allergy detection is discussed.

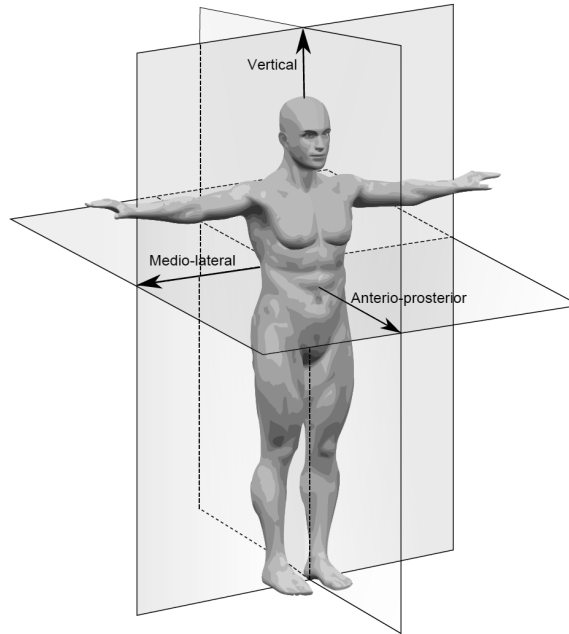


Figure 3.1: The x (anterio-posterior), y (medio-lateral), and z (vertical) acceleration directions in relation to a body.

3.2 ACCELEROMETER-BASED ACTIVITY ANALYSIS

3.2.1 ACTIVITY METRICS

A number of metrics can be calculated on raw acceleration values over a time-window of length T_e seconds. The most common activity analysis metrics which form the basis of all of the algorithms discussed in this chapter are described here.

- The integral of absolute acceleration (IAA) computes the absolute value of acceleration. The integral of absolute acceleration in the x -axis (IAA_x), integral of absolute acceleration in the y -axis (IAA_y) and integral of absolute acceleration in the z -axis

(IAA_z) can be computed for tri-axial accelerometers and these are defined by

$$IAA_x(t) = \int_{t=\tau}^{t=\tau+T_e} |a_x(t)| dt, \quad (3.1)$$

$$IAA_y(t) = \int_{t=\tau}^{t=\tau+T_e} |a_y(t)| dt, \quad (3.2)$$

$$IAA_z(t) = \int_{t=\tau}^{t=\tau+T_e} |a_z(t)| dt, \quad (3.3)$$

where a_x , a_y and a_z are the acceleration values in the x , y and z directions. These are in g-units (i.e. multiples of 9.81 ms^{-2}). In subsequent sections, when a number is followed by the letter g (e.g. 1 g), this identifies multiples of the unit.

- The total integral of absolute acceleration (IAA_t) combines the three IAA values together. This value is sometimes referred to as activity and acceleration counts (Chen and Bassett Jr, 2005) and is defined by

$$IAA_t(t) = \int_{t=\tau}^{t=\tau+T_e} (|a_x(t)| + |a_y(t)| + |a_z(t)|) dt \quad (3.4)$$

$$= IAA_x(t) + IAA_y(t) + IAA_z(t). \quad (3.5)$$

- The magnitude of acceleration is the square root of the sum of the squares of the raw accelerations. This metric provides an indication of overall acceleration but the directionality of the signal is removed by the squaring and summation. This is computed by

$$IAV(t) = \int_{t=\tau}^{t=\tau+T_e} \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2} dt. \quad (3.6)$$

- Instantaneous velocities in x , y and z directions can be computed by

$$v_x(t) = \int_{t=0}^{t=\tau} a_x(t) dt + v_x(t = 0), \quad (3.7)$$

$$v_y(t) = \int_{t=0}^{t=\tau} a_y(t) dt + v_y(t = 0), \quad (3.8)$$

$$v_z(t) = \int_{t=0}^{t=\tau} a_z(t) dt + v_z(t = 0). \quad (3.9)$$

However, these values must be integrated from $t = 0$ as acceleration is the continuous derivative of velocity.

- The mean kinetic energy can be computed by

$$KE_{\text{tot}}(t) = \frac{m}{2T_e} \int_{t=\tau}^{t=\tau+T_e} (v_x(t)^2 + v_y(t)^2 + v_z(t)^2) dt, \quad (3.10)$$

where m is the mass of the body on which acceleration is being measured, and v_x , v_y and v_z are the velocities in the x , y and z directions respectively.

- The mean power can be computed by

$$P = \frac{m}{2T_e} \int_{t=\tau}^{t=\tau+T_e} \left| \frac{d}{dt} (v_x(t)^2 + v_y(t)^2 + v_z(t)^2) \right| dt. \quad (3.11)$$

Equations 3.1 — 3.11 are fundamental equations which are employed by researchers for activity analysis and energy expenditure estimation algorithms (Bouten et al., 1994, 1997a,b; Crouter et al., 2006; Chen and Sun, 1997). These equations are presented in the continuous domain, but are employed in the discrete domain for the work here. Replacing the time variable, t , with the sample index, k , yields the discrete-domain equivalent of these equations. The continuous equations were presented in this section as they are more intuitively understood than the discrete equivalents.

3.2.2 ENERGY EXPENDITURE ESTIMATION ALGORITHMS

The goal of accelerometer-based energy expenditure (EE) is to obtain an energy expenditure estimation ($\widehat{EE}$) based on activity analysis which is as close as possible to true energy expenditure (EE_{true}). Four different algorithms are described here which have been designed to accomplish this.

The algorithms here are selected for investigation because they encompass a number of different aspects of data modelling. Firstly, Bouten et al. (1994) provided algorithms which

performed an overall assessment of energy expenditure based on one regression equation. Chen and Sun (1997) developed regression equations which were linear and non-linear in nature, and these equations are designed to adapt to the physiological features of the subjects being investigated. Finally, Crouter et al. (2006) employed a 'decision tree'-like algorithm to estimate energy expenditure based on the extent of activity which was detected in the preceding minute.

3.2.2.1 A NOTE ON THE ENERGY EXPENDITURE ESTIMATION ALGORITHMS

In the next three subsections, a number of algorithms are described which can be employed for the estimation of expended energy by persons wearing accelerometers. This thesis does not cover the background on how to obtain such models, and for information on these processes, and for justification of the regression units, the original publications that are cited in the subsequent subsections should be reviewed.

For every algorithm presented below, the models are parameterised with values that were found to minimise a specific cost function (e.g. the squared difference between the true and estimated energy expenditure). Each model is therefore non-trivially parameterised by seemingly arbitrary constants, but it should be noted that these were obtained by employing regression techniques to learn the relationship between acceleration signals and true energy expenditure. These parameters are therefore, by definition of the regression techniques, of appropriate units to convert the input values to the units of the target variables (Bishop et al., 2006; Flach, 2012). Therefore, the specific parameters that were obtained from the training process will not be discussed, but are presented here to facilitate speedy reproduction of these algorithms.

3.2.2.2 BOUTEN ET AL

Bouten et al. (1994), provided regression equations for human-based energy expenditure estimation with a tri-axial accelerometer. Analysis of 30 seconds of accelerometer

data is required for this estimation. Based on the data which were available, linear relationships were discovered between acceleration counts and energy expenditure values. Two regression equations were generated to estimated energy expenditure estimation due to activity ($\widehat{EE}_{act}$), and are defined by

$$\widehat{EE}_{act,x} = -0.176 + 0.0851 \times IAA_x, \quad (3.12)$$

$$\widehat{EE}_{act,t} = 0.104 + 0.023 \times IAA_t, \quad (3.13)$$

where $\widehat{EE}_{act,x}$ and $\widehat{EE}_{act,t}$ are the energy expenditure due to activity (EE_{act}) estimates based on IAA_x and IAA_t respectively. As these equations model EE_{act} , the resting energy expenditure (REE) must be added to the results obtained to achieve total energy expenditure. REE is calculated by Equations (3.14) and (3.15) for male and female subjects respectively (Hemokinetics, 1993).

$$REE_m = \frac{\begin{pmatrix} 215 & \times \text{weight (kg)} \\ + 12 & \times \text{height (m)} \\ - 513 & \times \text{age (years)} \\ + 4687 \end{pmatrix}}{100,000} \quad (3.14)$$

$$REE_f = \frac{\begin{pmatrix} 150 & \times \text{weight (kg)} \\ + 9 & \times \text{height (m)} \\ - 353 & \times \text{age (years)} \\ + 49854 \end{pmatrix}}{100,000} \quad (3.15)$$

Therefore, the total energy expenditure estimation is given by

$$\widehat{EE} = \widehat{EE}_{act} + REE. \quad (3.16)$$

Bouten et al. (1994) generated Equations (3.12) and (3.13) by recruiting subjects who walked on a treadmill between 3 and 7 kilometers per hour (km/h). An indirect calorimeter was worn by the participants and the acceleration was recorded at the same time. The parameters of these equations were obtained by a least squares analysis of the parameters which minimised the squared difference between EE_{true} and $\widehat{EE}$.

Equation (3.12) estimates the energy expenditure with recorded accelerations from the IAA_x direction only. This is because the majority of the effort required for movement is stated in this publication as being expended in this direction.

However, Equation (3.13) estimates energy expenditure based on IAA_t . This allows the regression to produce estimates which are representative of whole-body movement. In this work, Equation (3.13) is used over Equation (3.12) because it was stated as providing better results, and because it is intolerant to orientation errors as the overall acceleration is measured (see later).

3.2.2.3 CHEN ET AL

While Bouten et al. produced two generalised equations estimating energy expenditure, Chen and Sun (1997), allowed for specialisation of their models based on the age, gender, mass and height of the subjects whose acceleration was recorded.

The doubly labelled water method was employed here to obtain the reference energy levels. The participants who were recruited were within the controlled room for two 24 hour periods, and in total over 6000 hours of data was recorded. The range of activities

which the participants performed during their stay included sedentary, light, moderate and vigorous activities.

The horizontal and vertical acceleration vectors were separated, as it was reasoned by these researchers that different effort would be expended in vertical and horizontal directions (Chen and Sun, 1997). The horizontal and vertical accelerations are computed by Equations (3.17) and (3.19) respectively.

$$H(k) = \sqrt{a_x(k)^2 + a_y(k)^2} \quad (3.17)$$

$$V(k) = \sqrt{a_z(k)^2} \quad (3.18)$$

$$= |a_z(k)| \quad (3.19)$$

Linear and non-linear algorithms were computed by Chen and Sun (1997), and these both allow targeted specialisation of the algorithm towards specific physical characteristics (i.e. age, height, weight, etc) in order to obtain the minimum error of estimation.

- *Linear algorithm*

The linear algorithm assumes the form of

$$\widehat{EE}_{act}(k) = a_L \times H(k) + b_L \times V(k), \quad (3.20)$$

where $H(k)$ and $V(k)$ are calculated by Equations (3.17) and (3.19) respectively.

The parameters a_L and b_L were computed by minimising the difference between $\widehat{EE}$ and EE_{true} , and the parameters which were found to achieve this on the data available are calculated by

$$a_L = \frac{\begin{pmatrix} 5.76 & \times \text{weight (kg)} \\ + 11.95 & \times \text{height (cm)} \\ + 6.89 & \times \text{age (years)} \\ - 2,001 \end{pmatrix}}{1000}, \quad (3.21)$$

$$b_L = \frac{5.96 \times \text{mass(kg)} + 349.5}{1000}. \quad (3.22)$$

- *Non-linear algorithm*

The relationship between acceleration and energy expenditure measurements is not necessarily linear (Chen and Sun, 1997). In a similar manner to the linear algorithm, the horizontal and vertical accelerations were also separated for non-linear equations. The non-linear regression equations take the following form

$$\widehat{EE}_{\text{act}}(k) = a_N \times H(k)^{p1} + b_N \times V(k)^{p2}. \quad (3.23)$$

The optimal scaling and power parameters (i.e. a_N , $p1$, b_N and $p2$) were computed and were given by

$$p1 = \frac{2.66 \times \text{mass(kg)} + 146.72}{1000}, \quad (3.24)$$

$$p_2 = \frac{-3.85 \times \text{mass}(\text{kg}) + 968.28}{1000}, \quad (3.25)$$

$$a_N = \frac{12.81 \times \text{mass}(\text{kg}) + 843.22}{1000}, \quad (3.26)$$

$$b_N = \frac{\begin{pmatrix} 38.90 & \times \text{weight}(\text{kg}) \\ -682.44 & \times \text{gender} \\ +692.44 \end{pmatrix}}{1000}, \quad (3.27)$$

where the ‘gender’ parameter of Equation (3.27) is 1 for male subjects, and 2 for female subjects.

3.2.2.4 CROUTER ET AL

Crouter et al. (2006), regressed acceleration readings to a measurement of energy expenditure termed metabolic equivalents (MET). 1 MET is the metabolic rate and energy expended by person in rest (Ainsworth et al., 1993). The ratio of the energy expended performing a task (walking, reading, etc) to that expended in a resting state is the metabolic equivalent of the task, and thus the measurement is dimensionless.

Ainsworth et al. defined the rates of metabolic exercise and some of these are reproduced in Table 3.1. Recently, a number of corrections have been made to these figures, (Ainsworth et al., 2000, 2011), and the corrected MET values for normal- and over-weight persons are also shown in Table 3.1. The correction factor between the original and revised MET values is calculable and it is given by Equation (3.28). It is a function of the Harris-Benedict resting metabolic rate (RMR) values which are calculated with Equations (3.29) and (3.30) for female and male subjects respectively.

Table 3.1: MET activity and corrected MET activity values.

Activity	Original	Corrected _(female)		Corrected _(male)		Rating
		<77kg	≥77kg	<91kg	≥91kg	
Rope jumping	12.3	13.5	16.5	12.9	15.4	Vigorous
Running	9.8	10.7	13.1	10.3	12.3	
Bicycling	7.5	8.2	10.0	7.9	9.4	Moderate
Pushing stroller	4.0	4.4	5.4	4.2	5.0	
Calisthenics	3.5	3.8	4.7	3.7	4.4	
Shopping	2.3	2.5	3.1	2.4	2.9	Light
Watching TV	1.3	1.4	1.7	1.4	1.6	

$$\text{Corrected MET value} = \text{MET} \times \frac{3.5}{\text{RMR}} \quad (3.28)$$

$$\begin{aligned} \text{RMR}_f = & 5.0 \times \text{Height(cm)} + \\ & 13.7 \times \text{Weight(kg)} - \\ & 4.7 \times \text{Age(years)} + 66.5 \end{aligned} \quad (3.29)$$

$$\begin{aligned} \text{RMR}_m = & 1.8 \times \text{Height(cm)} + \\ & 9.6 \times \text{Weight(kg)} - \\ & 4.7 \times \text{Age(years)} + 655.1 \end{aligned} \quad (3.30)$$

The algorithm of Crouter et al. (2006) is a function of the acceleration counts that were calculated over the preceding minute and the coefficient of variation of the counts over the preceding ten seconds. Models were generated to predict the energy expenditures under these conditions. Listing 3.1 details the process of the algorithm (Crouter et al., 2006). The algorithm selects different regression equations depending on activity that

Listing 3.1: Crouter et al.'s energy expenditure estimation algorithm.

```

1
2 if ( AccelerationCountsmin ≤ 50 ) {
3   EEMET = 1.0;
4
5 } else {
6
7   if ( CV(AccelerationCounts10s) ≤ 10 ) {
8     EEMET = 2.379833 × exp{ 0.00013529 × AccelerationCountsmin };
9
10  } else {
11    EEMET = 2.330519 + 0.001646 × (AccelerationCountsmin)
12             - (1.2017×10-7 × (AccelerationCountsmin)2)
13             + (3.3779×10-12 × (AccelerationCountsmin)3);
14  }
15 }
16

```

was recorded over previous time windows. By this process the algorithm is unique to the energy expenditure field as its formulation concedes that it is beneficial to generate multiple regression equations for different operating points. It was tested over a large set of light to vigorous activities (including those tabulated in Table 3.1) and the authors stated that it was the most accurate algorithm in comparison to algorithms by Freedson et al., 1998; Swartz et al., 2000; Hendelman et al., 2000.

3.3 ENERGY EXPENDITURE VALIDATION

The algorithms which have been discussed have been stated to perform well by the authors. However, these algorithms must be replicated and validated as, when these algorithms are employed to assess OFCs, no true energy expenditure data are available. Therefore the accelerometer-based estimates are the only energy expenditure metrics which will be available from these recordings, and validation of the algorithms and validation of the implementation of the algorithms is required.

In this section, the validation experiment is discussed. Acceleration and EE_{true} metrics were recorded on a small set of subjects ($n = 5$), and the $\widehat{EE}$ estimation algorithms are

Table 3.2: Tabulation of the physical characteristics of the subjects who participated in the accelerometer-based energy expenditure validation test.

ID	Age (years)	Height (m)	Weight (kg)
1	23	1.75	77.5
2	25	1.83	81.1
3	22	1.87	86.7
4	24	1.79	87
5	30	1.70	88
$\mu \pm \sigma$	24.8 ± 3.1	1.79 ± 0.07	84.2 ± 4.45

employed on the acceleration data, and these results are compared to the EE_{true} data. The physical characteristics of each subject were recorded and are presented in Table 3.2.

3.3.1 EXPERIMENTAL SETUP

A treadmill (Powerjog GX100) was set up to operate from 3 — 7 km/h at 1 km/h increments. The treadmill ran at each speed for four minutes at a gradient of 1%, which has previously been stated as emulating walking on a flat surface (Jones and Doust, 1996). The recruited participants walked on this while EE_{true} was recorded by a cardio pulmonary exercise testing (CPET) indirect calorimeter (Cosmet, 2013). The CPET was calibrated before each recording. A gas mask was worn by the subject and it was secured around the subject's nose and mouth. Care was taken to ensure that the volunteers were comfortable with the mask and that no difficulty in breathing was introduced by the equipment. A hose leaving the mask is attached directly to the gas analysers in the CPET calorimeter that was used. To ensure that no gas entered or exited the seal about the subject's face, the subject was asked to momentarily block the outlet of their gas mask. If they were unable to express air during this brief time the mask was deemed securely fastened and the test continued.

The time for which the subject walked at each speed was chosen at four minutes due to works published by Chatagnon and Busso (2006), and Winter et al. (2006). Chatagnon and

Busso (2006) provides the upper bound of the time that an exercise must be performed for a body to have reached a steady metabolic state as being between 2 and 3 minutes, subject to the fitness of the individual. Winter et al. (2006) states that metabolic tests of this nature should not exceed 20 minutes in length as reliable metabolic response after this time is unreliable because as the body is exerted to more physical exertion, different attributes of the body will affect the metabolic rates of subjects. Employing the treadmill speeds with consideration to Chatagnon and Busso (2006), and Winter et al. (2006) dictates that each speed should be recorded for four minutes.

3.3.2 PRE-PROCESSING

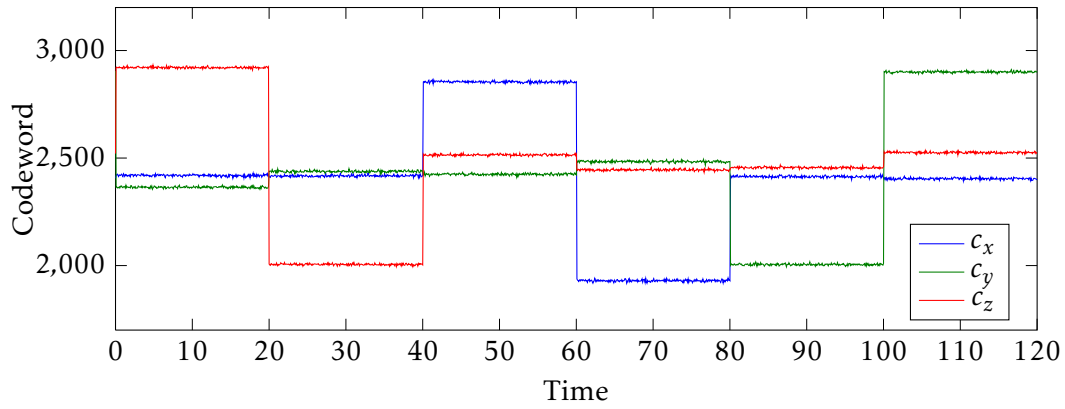
3.3.2.1 CODEWORD CONVERSION

The accelerometer on the SHIMMER device produces analogue voltages which the analogue to digital converter (ADC) samples at 12-bits of resolution (allowing for 4096 possible values). Therefore, the data which is obtained by the microprocessor is in the form of 12-bit digital codewords, and these codewords must be converted to measurements of acceleration which are measured in g's.

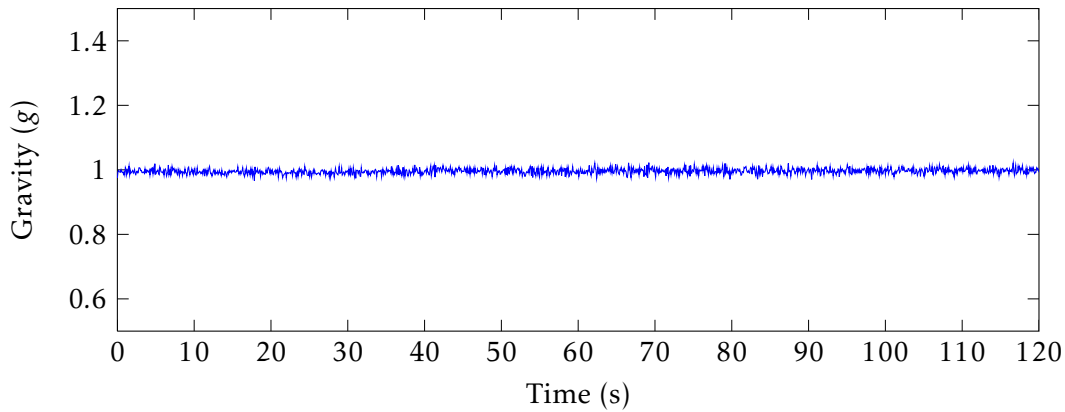
At rest, the absolute value of the three axes of a tri-axial accelerometer will sum to 1 g, due to the orthogonal orientation of the axes of the sensor. The absolute value of acceleration is obtained in the discrete domain by

$$a_{\text{abs}}(k) = \sqrt{a_x(k)^2 + a_y(k)^2 + a_z(k)^2}. \quad (3.31)$$

Figure 3.2a shows the raw codewords (c_x , c_y , and c_z) which were recorded by the microprocessor when the SHIMMER device was orientated on its six faces for equal periods of time. It should be noted that this data was spliced so that each face of the SHIMMER device provides equal portions of data which ensures that no axis dominates



(a) Accelerometer codewords on X, Y and Z axes



(b) Combination of raw accelerometer data obtaining the absolute acceleration

Figure 3.2: Demonstration of the conversion process from the raw digital codewords obtained from the accelerometer (a) to a measurement of absolute gravity (b).

the calibration process. The signals associated with movements between faces were removed.

On a perfectly flat and still surface, when the SHIMMER device rests on a given face, the axis experiencing gravity will read either 1 g or -1 g (depending on the orientation of the sensor). The remaining axes should read 0 g. In order to convert from codewords to gravitational units, each codeword must have the 0 g offset removed ($c_{x,0g}$, $c_{y,0g}$, and $c_{z,0g}$) and must then be scaled by the 1 g value ($c_{x,1g}$, $c_{y,1g}$, and $c_{z,1g}$). The absolute acceleration in g's is then computed with reference to these codewords by

$$a_{\text{abs}}(k) = \sqrt{\left(\frac{c_x(k) - c_{x,0g}}{c_{x,1g}}\right)^2 + \left(\frac{c_y(k) - c_{y,0g}}{c_{y,1g}}\right)^2 + \left(\frac{c_z(k) - c_{z,0g}}{c_{z,1g}}\right)^2}. \quad (3.32)$$

The optimal values of the converting codewords which can transform raw readings to gravitational units can be discovered by searching for c_{0g} and c_{1g} in the range $[0, 4095]$ for each acceleration axis. The values which would be selected are those which minimise the standard deviation of the resulting vector of a_{abs} . This process results in a search space of 4096^3 bits per transforming metric, and a brute force search is not feasible to make this discovery. Therefore, a binary search algorithm was implemented to select these optimal values in an efficient manner with guaranteed convergence obtained in 12 iterations.

The result of the search yields the acceleration vector shown in Figure 3.2b which has obtained average acceleration of 1 ± 0.0082 g ($\mu \pm \sigma$) with the values shown in Figure 3.2a as the input. This process provides codeword conversion values which are tolerant to small orientation errors because the searching routine is performed on the absolute value of acceleration. The robustness of this algorithm can be seen in Figure 3.2b where the orientations of the SHIMMER device were not perfectly orthogonal but the original signal variance is absent from Figure 3.2b.

3.3.2.2 BREATH AND ACCELERATION SYNCHRONISATION

The EE_{true} levels captured by the CPET were recorded on a breath-by-breath basis, and the time at which each energy expenditure value is logged is rounded to the nearest second by the hardware. The breathing rate of a human is non-periodic and non-stationary, in particular when the subject is exercising (Chon et al., 2009). Therefore, the EE_{true} recordings and the acceleration values computed are not sampled at a common rate, and the two datasets must be synchronised in time together to allow for comparisons between EE_{true} and $\widehat{EE}$.

The selected synchronisation method involved averaging the EE_{true} values which occurred within the previous 10 seconds to the current EE_{true} value. The 10 second window was chosen as the minimum respiratory rate for healthy adults typically requires 10 seconds per breath (Lindh et al., 2009), and this rate will increase with exercise. Therefore, a 10 second time window allows for minimum latency and phase difference between EE_{true} values obtained while allowing for the maximum range of respiratory rates.

Consider the (unlikely) situation where a subject does not breathe for 20 seconds, for example. With the time-window averaging algorithm, no EE_{true} signals will be logged over this time. This is an intuitive consequence to the absence of breathing as without respiratory effort, the CPET will not log any new energy expenditure data and the averaging method follows this trend. As a result of holding their breath, when the subject next breathes, a higher concentration of CO_2 will be expressed and a higher volume of O_2 will be inspired which is recorded by the CPET and will yield higher EE_{true} . By using the time-window averaging method for synchronisation, the dynamics of EE_{true} values will be accurately reflected in the synchronised array, while with others, for example cubic spline interpolation, the true dynamics will not be followed.

3.3.2.3 NORMALISATION

Crouter's algorithm regresses to METs, which was previously described. As neither Chen nor Bouten regressed to this unit of energy it was necessary to scale the reference and the regressed energy values to a common standard so that all algorithms can be compared on the same grounding. This was accomplished by

$$\widehat{f} = \frac{f - \min(f)}{\max(f - \min(f))}, \quad (3.33)$$

where f is the signal which is normalised to $\widehat{f}$, and the min and max functions select the minimum and maximum values of the data series respectively. Equation (3.33) is employed on both the EE_{true} and $\widehat{EE}$ signals and scales the range to between 0 and 1.

3.3.3 PERFORMANCE EVALUATION

The root mean square error (RMSE) can be used to quantify the difference between EE_{true} and $\widehat{EE}$ mathematically, and it is defined by

$$\text{RMSE} = \sqrt{\text{mean} \left(\left(\frac{\widehat{EE} - EE_{\text{true}}}{EE_{\text{true}}} \right)^2 \right)}, \quad (3.34)$$

where mean computes the average of the data array. The RMSE can be converted into the percentage root mean square difference (PRD) which yields a percentage difference representation of the two signals by

$$\text{PRD} = \text{RMSE} \times 100\%. \quad (3.35)$$

Table 3.3: Table of PRD values computed between true energy expenditure and the estimated energy expenditure values obtained from the algorithms investigated.

Algorithm	PRD (%)	
	μ	σ
Bouten	7.54	1.93
Chen linear	8.07	1.32
Chen non-linear	6.57	0.33
Crouter	13.50	1.41

PRD values of 0 indicate that the arrays are identical, and increasing PRD values indicate that two arrays are becoming more dissimilar.

3.3.4 RESULTS

Table 3.3 tabulates the PRD results which were obtained for each energy expenditure estimation algorithm in the order in which the algorithms were presented earlier. Chen's non-linear algorithm performed best of all of the algorithms investigated over all the subjects yielding an average PRD of approximately 6.6% while also obtaining the lowest standard deviation of PRD. Figure 3.3c shows an example where EE_{true} and $\widehat{EE}$ are overlaid. It can be seen that the values which were obtained follow the trend in the changes of EE_{true} .

It is not entirely surprising that Chen's linear and nonlinear algorithms performed well because the algorithms were designed to automatically specialise for persons of a known weight, gender, height etc, and because the algorithms were also generated with reference to the largest amount of data of all the algorithms considered. The fact that the nonlinear algorithm performance is best suggests accelerometer based energy expenditure regression is not a linear process averaged over all participants. However, the differences are small.

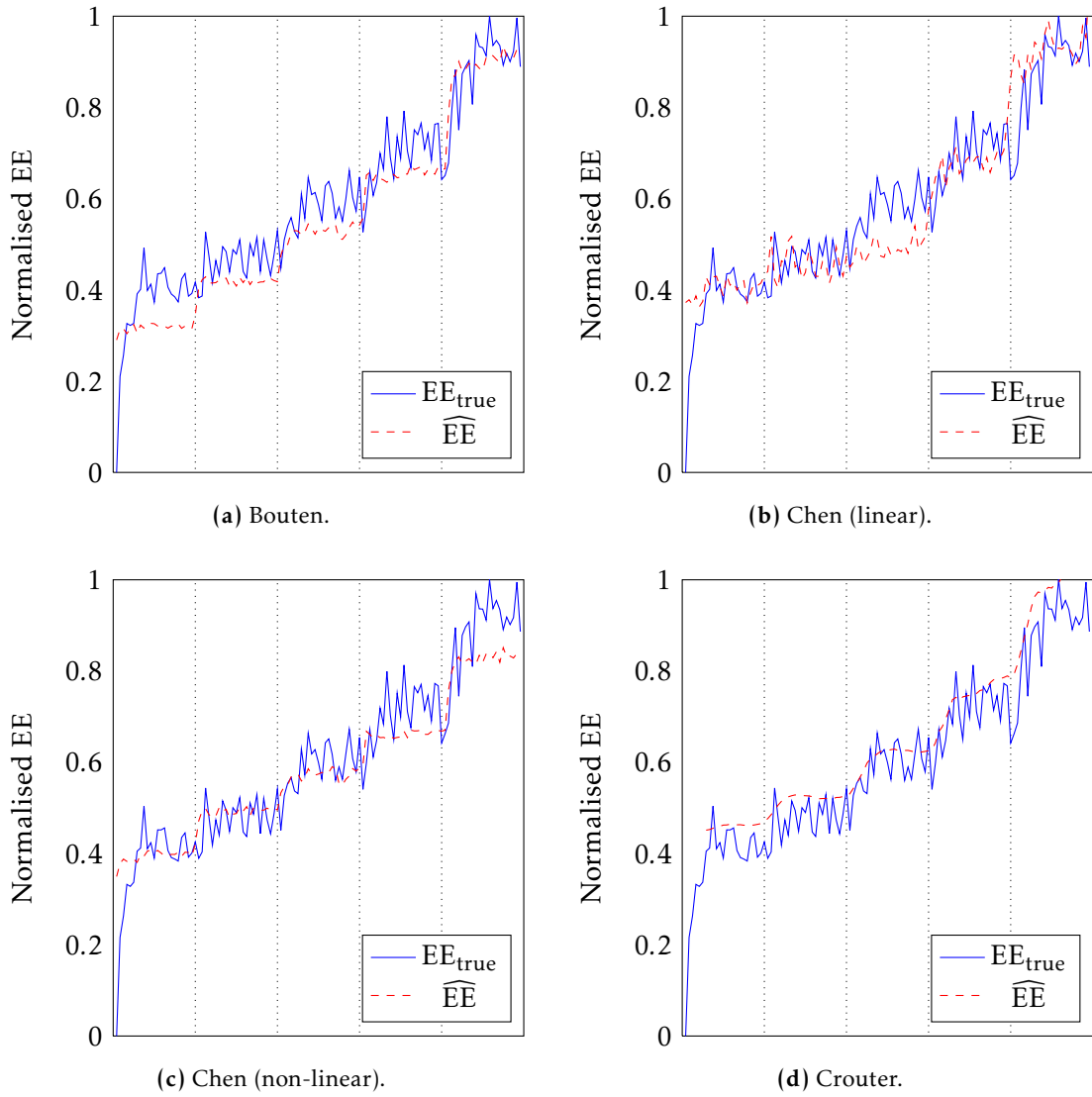


Figure 3.3: The $\widehat{EE}$ and EE_{true} values obtained for participant 2. The values obtained are overlaid.

Crouter's algorithm (Figure 3.3d) consistently overestimated the energy expenditure which contributed heavily to it performing poorest out of the algorithms investigated, yielding a mean PRD of 13.50%. However, the energy trace was smooth and did follow the trend of the metabolic response, see Figure 3.3d. The smooth nature of this signal trace is due to how the algorithm considers not only recent data, but also data from the preceding minute. This introduces low-pass filtering behaviour to the algorithm.

3.3.5 DISCUSSION ON ENERGY EXPENDITURE ESTIMATION ALGORITHMS

It is very clear from the PRD values and figures presented that accelerometers are a well suited instrument for estimating energy expenditure in the situations investigated here.

That Chen's algorithms provide excellent estimations of energy expenditure is a very strong indication that subject tailored regression was one of the strongest factors for accurate regression to energy expenditure values. Chen's non-linear algorithm resulted in the lowest overall PRD. This result supports the argument that accelerometry and energy expenditure are non-linearly related.

The algorithms of Bouten and Crouter do not target the individual but rather were developed to accommodate the standard user. This carries a convenient 'plug and play' feature to the algorithms (i.e. no setup or customisation is required for their use) but in counterpoint the accuracy of energy expenditure estimation is traded, which can be seen by high PRD values for Crouter's algorithm in Table 3.2.

The reported accuracy of all of the algorithms is subject to the reference that their results are compared against. As the reference energy expenditure levels are not periodic and the time stamps are rounded to the nearest second, data synchronisation was performed on the original reference data to coordinate its values to those which were estimated by the $\widehat{EE}$ algorithms. If the synchronisation algorithm is not chosen carefully, this process can introduce undesirable artefacts to the reference data. The method employed here

involved averaging the breath data over time window, which achieves a low pass filtering effect which improves the reliability of the synchronised EE_{true} .

Indeed, the reference energy levels can themselves be corrupted before any analysis is performed. This can happen on the treadmill if the subject yawns, coughs or talks. Participants were requested to refrain from doing this, but it is neither possible or reasonable to stop a subject from coughing during these recordings if the subject must. These artefacts will accentuate, exaggerate or otherwise affect the accuracy of the reference values and the times they are measured. As low-pass filtering is achieved by synchronisation, these artefacts will be reduced in the EE_{true} reference data.

3.3.6 CONCLUSION ON ENERGY EXPENDITURE ESTIMATION

This study shows that accelerometer-based energy expenditure estimation algorithms can achieve very accurate estimates of true energy expenditure. The SHIMMER device was also found to be an ideal platform with which to measure the acceleration data. As the results obtained here were very good and followed the dynamics of the changes of EE_{true} , the analysis of accelerometer-based metrics for classification of allergy during oral food challenges will be discussed in the next section.

3.4 ACCELEROMETER-BASED ANALYSIS DURING OFCs

For the assessment of activity and energy expenditure based analysis of OFCs, the SHIMMER device was applied to the torso and dominant wrist before the test began, and then streamed tri-axial accelerometer data to a nearby computer. At this stage, activity analysis and energy expenditure estimation algorithms which were previously validated were applied to this new data.

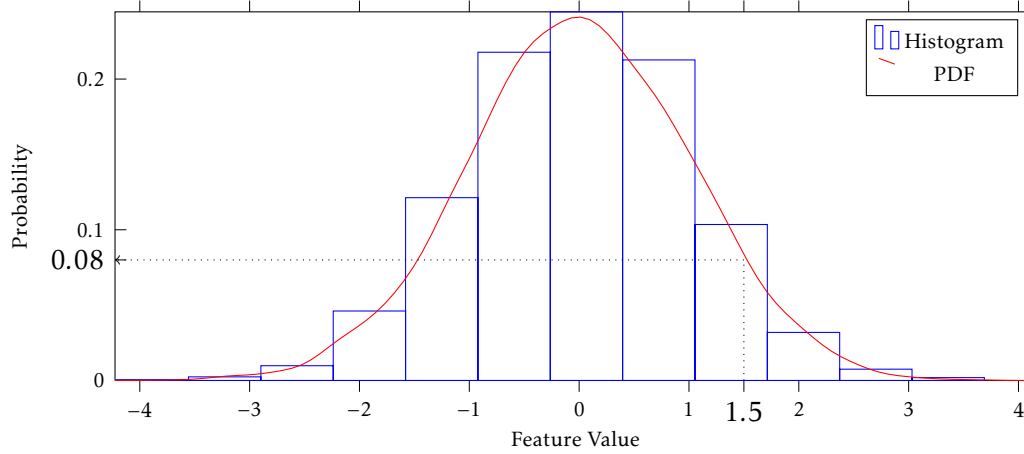


Figure 3.4: Example histogram and probability density function of a feature.

3.5 PROBABILITY DENSITY FUNCTIONS

Figure 3.4 shows a sample histogram (blocks) and associated probability density function (PDF) (solid curve) for a randomly generated normal distribution with a zero-mean and a unity standard deviation. In general the histograms are normalised so that the area under the curve sums to 100%. This diagram is indicative of what the distribution of accelerometer metrics might look like. The probability of achieving an arbitrary x -value (within a finite range) can be determined visually by projecting vertically upwards from the selected point until the PDF curve is intersected. The probability of a feature of this value occurring is then determined from the y -value at the intersection. This is illustrated by the dotted arrow in Figure 3.4 for a feature value of $x = 1.5$. Here, kernel density estimation (KDE) (Sheather and Jones, 1991) was employed in all cases in this thesis to compute the histograms. This is a procedure which will produce ‘smooth’ histograms that can be more representative of the true distribution of the process than centred histogramming (Sheather and Jones, 1991).

This example shows the overall distribution of a feature. By computing PDFs on two subsets of the data (first the allergic and second the non-allergic subjects, for example) the characteristic differences between the two classes can be visualised. Figure 3.5 shows

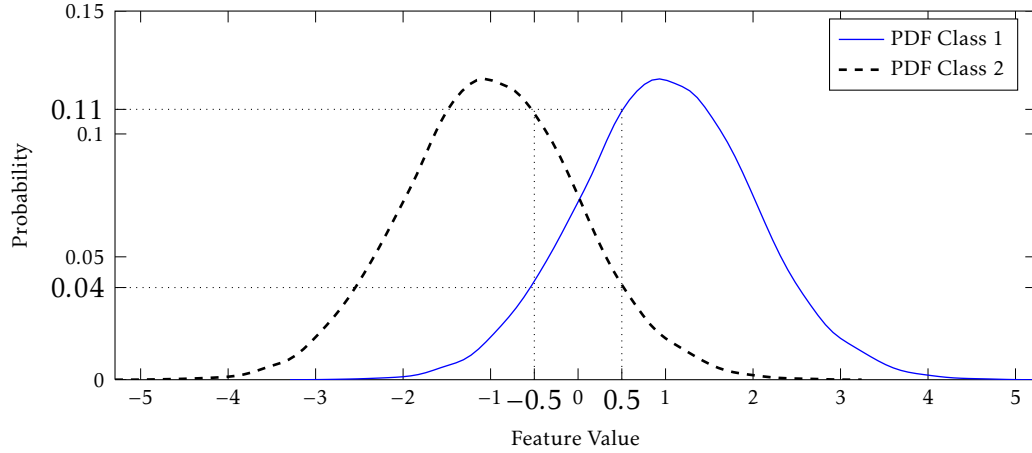


Figure 3.5: Illustration of the differences between PDFs that describe two separate classes of data.

how PDFs of the two classes might differ. Two PDFs are shown in this image, representing the allergic and non-allergic classes. The means of these PDFs are centred at +1 and -1. Large differences between the means of the two curves indicate that the feature might be a suitable candidate for separation between allergic and non-allergic subjects. This is because with a smaller overlap there is a smaller probabilistic uncertainty about the class of the data (i.e. does the data originate from an allergic or non-allergic subject?). In contrast to this, if a large overlap exists between the PDFs the metric might not (by itself) facilitate good class separation. In Figure 3.5 the probability densities for the two classes are shown for an $x = -0.5$ where it can be seen that a higher probability of PDF Class 2 is obtained in comparison to Class 1. However, at $x = 0.5$ the probabilities obtained for each class are reversed.

3.6 RESULTS

Section 3.2.2 presented the regression equations which were employed when computing activity and $\widehat{EE}$ for the algorithms being considered, and IAA was used by all of these. It can be seen that all of these equations consider IAA_x , IAA_y and IAA_z , and Bouten and Crouter's algorithms also consider IAA_t . PDFs were generated to show the expected values

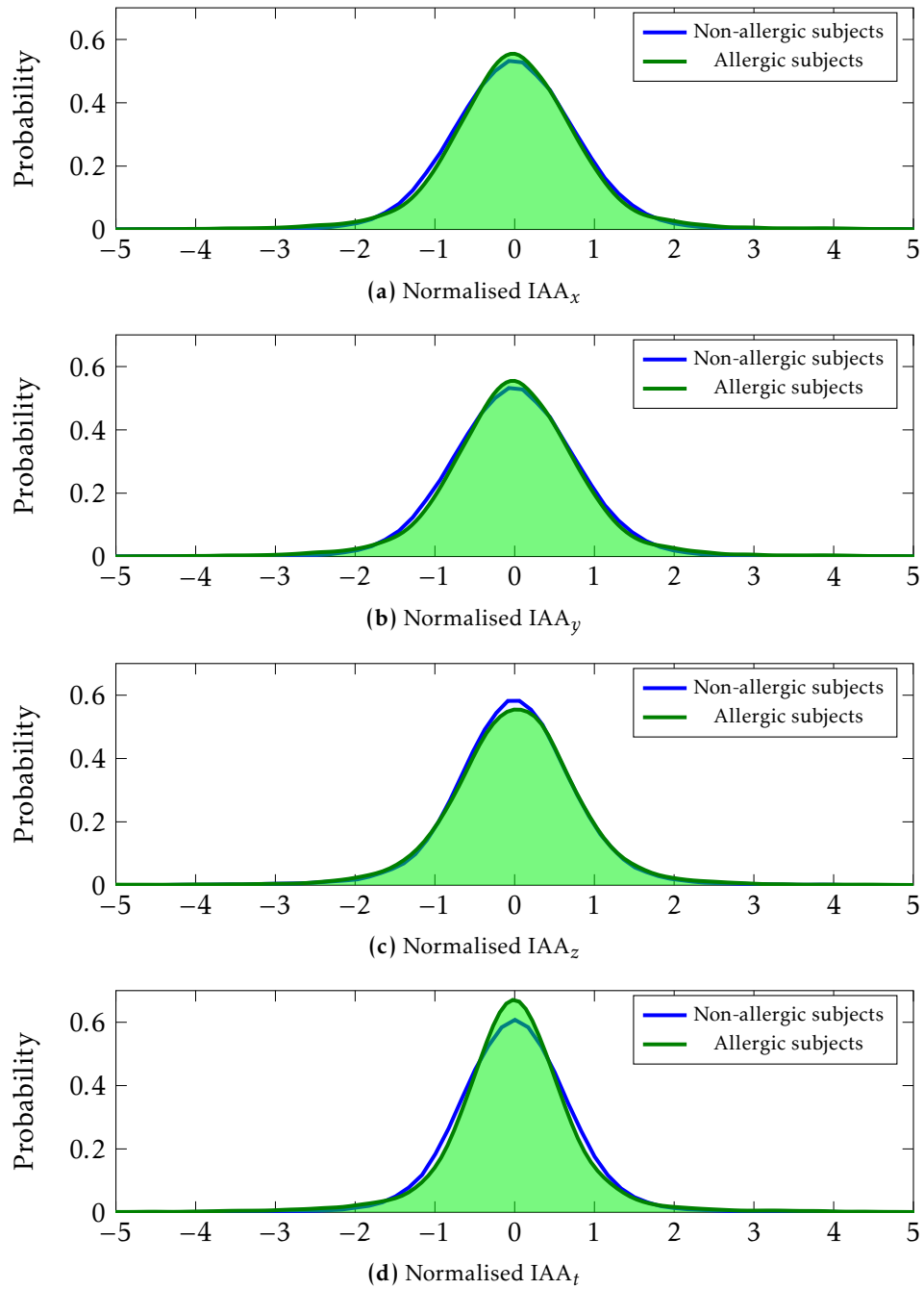


Figure 3.6: Histograms plotting the normalised IAA values of the allergic and non-allergic subjects who were investigated.

of these during OFCs, and Figure 3.6 shows the set of histograms of the IAA metrics that was obtained.

In every case, a very strong overlap can be seen between the allergic and non-allergic PDFs. In order for these metrics to be employed for classification, separation between the allergic and non-allergic curves must be apparent, but these PDFs show very similar distributions.

The set of PDFs show a very high correlation between the allergic and non-allergic classes, and this is also reflected in the $\widehat{EE}$ values which were computed. The $\widehat{EE}$ values which were obtained also offered no distinguishable difference in metrics in a similar manner to the IAA metrics which are shown in Figure 3.6. Therefore, as the PDFs which are shown in Figures 3.6a — 3.6d present as overlapped normal distributions, and because functions of normal distributions are themselves normal distributions (Leon-Garcia and Leon-Garcia, 2009), the $\widehat{EE}$ estimates offered no better separability than the IAA metrics.

It should be stated that there will be differences between acceleration obtained from adults and children. The algorithms employed were tested on adults and worked well, but when tested on children they worked poorly. However, PDFs that were displayed in Figure 3.6 show activity-based metrics, which are the basis of the activity and energy expenditure estimation equations. With these data no separability is obtained, and this also occurs with the energy-expenditure based algorithms. Therefore, it is believed that the poor performance is not due to age-based discrepancies, but due to the inadequacy of the accelerometer-based approach for allergy classification.

3.7 CONCLUSION

The PDFs did not, in any case, present with a mean shift or any exploitable anomaly between the allergic and non-allergic classes. This indicates that the subjects cannot be separated by activity analysis by the methods applied here, and therefore classification of allergy cannot be resolved by these means. Therefore, accelerometer-based analysis is not

an appropriate means of classifying allergy, and will not be employed in the remainder of this thesis.

CHAPTER 4

ECG-BASED ANALYSIS OF OFCs

4.1 INTRODUCTION

CHAPTER 3 discussed the use of accelerometer-based activity and energy expenditure metrics for use in the classification of oral food challenges. It was shown previously that the analysis of these achieved very poor separability between the allergic and non-allergic classes, and as such are insufficient measures for the classification of allergy.

During the oral food challenges the SHIMMER device recorded the ECG of the subjects who underwent these challenges. While it has been observed by the allergists who conduct the OFCs that there is a tendency for the heart rate of subjects to change before the onset of an allergic reaction, the effect of allergy on the heart has not been definitively quantified and this chapter investigates this.

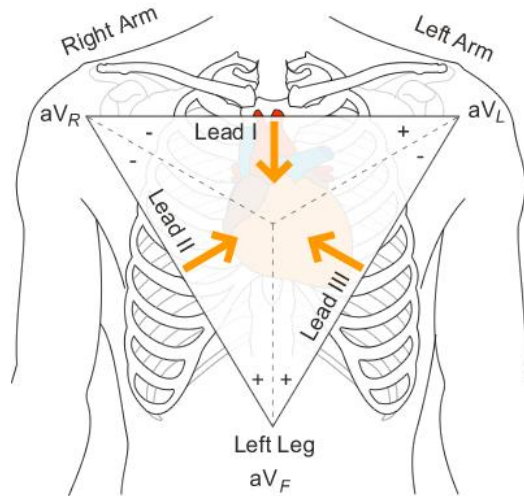


Figure 4.1: Einthoven triangle configuration for ECG electrode placement (University of Nottingham, 2013).

4.2 ECG AND HRV

4.2.1 ECG RECORDING

As 12-lead ECG is generally only performed for diagnosis of cardiac disease and stress tests, for general heart monitoring (i.e. in a hospital ward) 3-lead ECG is recorded. This employs the limb electrodes which are arranged in the Einthoven Triangle configuration (Wilson et al., 1946) shown in Figure 4.1 (University of Nottingham, 2013).

Figure 2.4 shows the P-, Q-, R-, S- and T-waves which characterise the ECG (Dublin Institute of Technology, 2013). While all of the waves can yield diagnostically relevant information, the QRS complex is the principal feature of the ECG which is utilised for the identification of heart beats. The intervals between sets of R-R intervals (as shown in Figure 2.4) are employed to describe the heart rate variability mathematically.

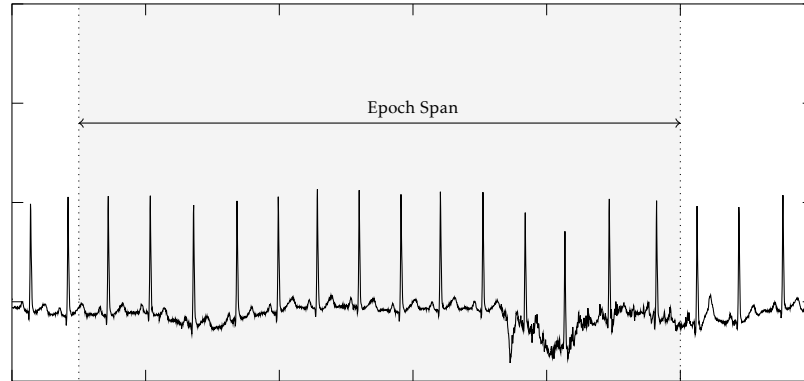


Figure 4.2: Illustration of relationship between the ECG and the epoch length for ECG recorded in OFC.

4.3 HRV FEATURE EXTRACTION

4.3.1 OVERVIEW

By considering HRV features which are extracted from allergic and non-allergic subjects independently, the characteristic differences between the two classes can be assessed. This can be employed to provide meaningful descriptors between allergic and non-allergic data. In this chapter these differences are analysed and quantified to determine if HRV-based classification of allergy is worthy of investigation.

4.3.2 EPOCHS

HRV feature extraction is performed based on the times of QRS complexes which were found within time-windows (known as epochs) of ECG data. This is illustrated by Figure 4.2, where the QRS points found within the shaded region are employed by the feature extraction for this epoch.

Longer epochs will naturally consider a greater number of QRS points. These will measure longer term variation of the HRV, whereas shorter epochs will obtain information about shorter term characteristics of the heart. Long- and short-term epochs are diagnostically interesting measurements and are both considered in this work.

Various lengths of epoch were investigated for this research. The European Society of Cardiology and the North American Society of Pacing and Electrophysiology stated that epochs between 120 and 300 seconds should be considered when extracting HRV features on adults (Rawenwaaij-Arts et al., 1993). The subjects recorded during OFCs are children, some of which have resting heart rates (HRs) exceeding 160 beats per minute (BPM) (Giddens and Kitney, 1985) which is approximately twice the heart rate of the average resting adult. Therefore, 60 second epochs are also considered for this work. The full set of epoch lengths which are investigated here are {60, 120, 180, 300} seconds. This set of epoch lengths was chosen because the effect of allergy on the HRV features has not been characterised. Therefore, analysis of all of these sets of epochs will qualify whether signatures of allergy are better obtained with longer- or shorter-duration epoch lengths.

4.3.3 EPOCH OVERLAP

Figure 4.3 illustrates the relationship between the ECG, the epoch length and the epoch overlap. The epochs in this illustration are 9 seconds in length, with one second increments for illustrative purposes only as this epoch length is insufficient for meaningful feature extraction (van Ravenswaaij-Arts et al., 1993). Feature extraction performed on the QRS points found within the bounds of Epoch 1 quantify the behaviour of the heart between 1 and 10 seconds, while features extracted from QRS points found within the boundaries of Epoch 2 characterise the heart rhythms between 2 and 11 seconds.

One second increments in time were utilised between epochs, which represents approximately 98% overlap with epoch lengths of 60 seconds. This was selected in order to increase the number of data points which are available from the OFC. Subject 1 presented

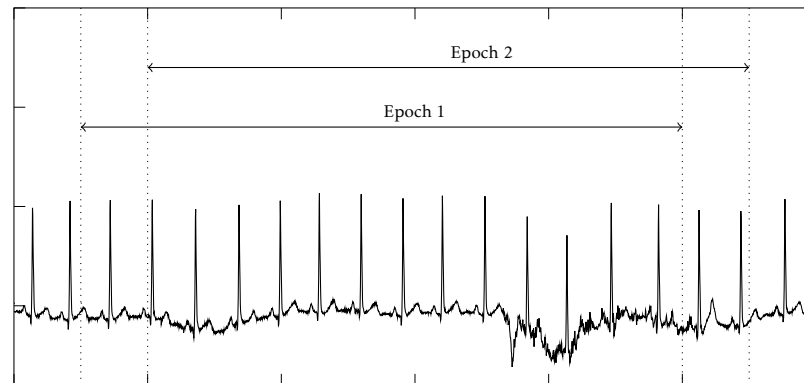


Figure 4.3: Illustration of relationship between the ECG, the epoch length and the epoch overlap.

with the shortest challenge which lasted approximately 3 minutes after the first dose of the allergen was administered. Without any epoch overlap, only three features would be extracted for this subject after administration of the first dose. However, with 1 second epoch increments, 180 points describe this period. Subject 19 presented with the longest OFC recording. Without epoch overlaps, this entire challenge would be described by approximately 130 data points, but with the overlap, approximately 9000 points will be used.

Indeed, for classification purposes, due to the short and varied length of the OFCs, classification algorithms would not have sufficient quantities of data with which to perform classification without epoch overlap (Duda et al., 1995; Bishop et al., 2006; Cherkassky and Mulier, 2007; Catal and Diri, 2009), and therefore, as well as employing one second increments these for distribution analysis in this chapter, small epoch increments are also requirement for later chapters.

4.4 FEATURE NORMALISATION/CALIBRATION

Chapter 2 presented Table 2.1 which tabulated the subjects investigated in this study. From this table, it can be seen that the ages of the subjects varied from 7 months to 10

years. Infants under one year of age will typically have a HR of ≈ 120 BPM (with a range of 80 — 160 BPM) while that of a ten year-old child will typically have a HR of ≈ 90 BPM (with a range of 70 — 110 BPM) (O'Brien et al., 1986; Tanaka et al., 2001; Kliegman et al., 2007; Aziz et al., 2012). Because of the age-related differences in the baseline resting heart rate it is not appropriate to directly compare the features extracted from different subjects. Therefore, a calibration procedure is performed in order to allow indirect and valid comparisons between the features extracted from the different subjects.

Typically feature data is normalised by subtracting the mean of the data and dividing by its standard deviation. In the case of OFCs this is not appropriate because this process will force the features obtained from the allergic and non-allergic subjects to have similar statistical properties and could reduce and possibly eliminate — any characteristic signatures of allergy in the HRV.

Therefore, the calibration process which is employed here computes the mean and standard deviation of the features before the problem foods were administered to the subjects. The entirety of the recording is then normalised by these values. This process guarantees that the features are normalised by non-allergic HRV data as no allergen would have been consumed by subjects. Therefore, if features deviate from this distribution as a result of allergy, this will be observed in the PDFs which were generated for the allergic subjects (see Chapter 3). Likewise, if, for non-allergic subjects, the features obtained for the remainder of the challenge should not change. The time before the first dose of the allergen is administered to the subjects is termed the 'background' or 'baseline' region henceforth, and it is guaranteed to present non-allergic HRV features. To define this period concretely, it is the time after the skin tests have concluded, but before the administration of the first dose of the allergen.

Normalisation is then achieved by

$$\widehat{\mathbf{f}} = \frac{\mathbf{f} - \mu_b}{\sigma_b}, \quad (4.1)$$

where $\mathbf{f}$ is the non-normalised feature vector, $\widehat{\mathbf{f}}$ is the normalised feature, μ_b and σ_b compute the mean and standard deviation of the background segment of feature $\mathbf{f}$. All background lengths recorded are approximately ten minutes in length, and all PDFs presented in the following sections are computed from normalised features.

While the normalised background data is characterised with zero mean and unity variance, non-background data will not preserve these traits. OFCs will induce stresses to the recruited subjects and consequently the HRV features will tend to deviate from this background baseline. As a result, the PDFs that are presented later in this chapter will demonstrate significant probability mass approximately three times beyond the normal ranges for a normal distribution (i.e. far beyond 5σ). This is explained because the units of the x -axis are of the units of $\sigma_{\text{background}}$ and not the overall standard deviation, and larger multiples of this indicate features that are less similar to this background.

4.5 HRV FEATURE CATEGORIES

4.5.1 FEATURE CATEGORIES

It should be stated here that HRV features extracted over an epoch are employed in order to measure the characteristics of the distribution. Therefore, much of the information that is extracted from the data relates to the statistical properties of such distributions, and the extractions of time domain features such as the mean, standard deviation, etc, is performed.

However, other feature categories are also extracted. For example, sequential domain features measure the relative ‘acceleration’ and ‘deceleration’ that the heart rate has experienced over an epoch length, whereas Poincaré features are employed to assess non-linear dynamics of the heart rate, and this has been shown to have a close relationship to the sympathetic indices of the autonomic nervous system (ANS). These features are also used for many other applications in different fields (Cogdell and Piatetski-Shapiro, 1990).

Other feature types can be extracted which are a measure of the frequency-domain characteristics of the heart, which have been correlated to the ANS too. Frequency domain analyses are popular methodologies employed in many engineering and scientific fields.

4.5.2 FREQUENCY DOMAIN FEATURE ANALYSIS

When computing the frequency spectrum of the heart rate, care must be taken because the heart rate does not beat periodically (Moody, 1993; Badilini and Blanche, 1996). Therefore, direct application of the Fourier transform (FT) cannot be employed on the heart rate data (Moody, 1993; Ebdem, 2002; Clifford and Tarassenko, 2005), and the spectrum must be estimated. Two popular means of extracting the frequency power spectrum from the irregular data series exist in HRV literature and are discussed below.

4.5.3 RESAMPLING + FFT

With this method, the heart rate series is re-sampled at a periodic frequency to provide a uniformly sampled data series (Clifford and Tarassenko, 2005, Laguna et al., 1998). A variety of re-sampling methods exist in literature including nearest-neighbor, linear, cubic spline and piecewise cubic Hermite techniques (Srikanth et al., 1998). Figure 4.4 shows the raw HR signal and the interpolated signal re-sampled after performing cubic spline interpolation. Once the data series has been re-sampled, the FT can be used to compute the power spectrum from this data series.

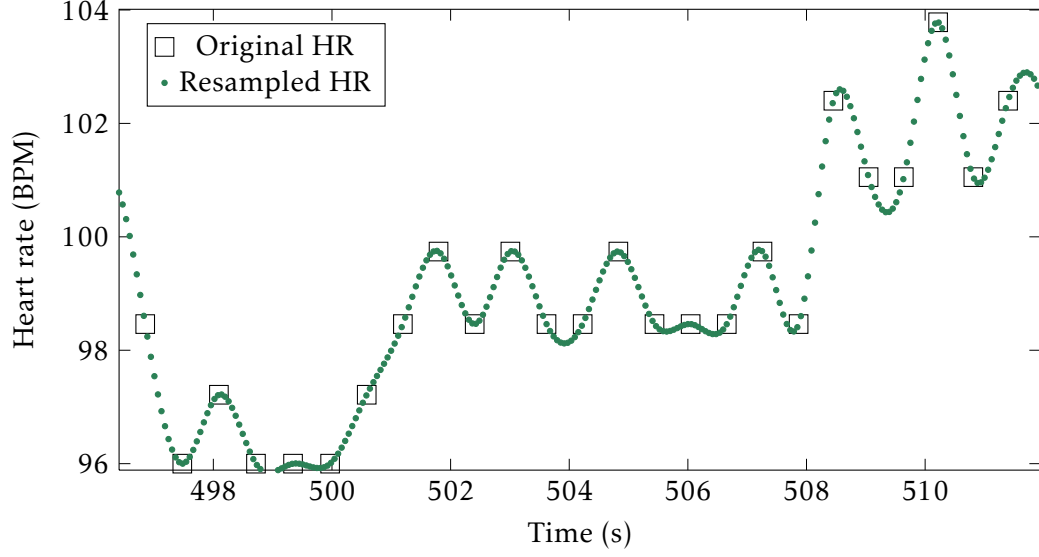


Figure 4.4: Illustration of the raw HR ($\square$) which is not periodically sampled, and the HR re-sampled to 10 Hz via cubic spline interpolation.

The FT is defined by Equation (4.2) and allows translation of a real, continuous time-domain signal, $x(t)$, to a frequency-domain representation, $X(\omega)$.

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad (4.2)$$

The discrete Fourier transform (DFT) is used to compute the FT of data segments of finite length N in the discrete domain. The data segments which are analysed must be periodically sampled at a frequency f_s , and the DFT is described for functions of multiples of the sampling frequency, and yields powers at frequencies which are multiples of this.

$$X(\omega_n) = \sum_{k=1}^N x(t_k)e^{-j\omega_n t_k} \quad (4.3)$$

The periodogram of the DFT is the estimate of the power spectral density (PSD) of a signal which is defined by

$$P_x(\omega_n) = \frac{1}{N} \left| \sum_{k=1}^N x(t_k) e^{-j\omega_n t_k} \right|^2 \quad (4.4)$$

$$= \frac{1}{N} \left[\left(\sum_{k=1}^N x(t_k) \cos(\omega_n t_k) \right)^2 + \left(\sum_{k=1}^N x(t_k) \sin(\omega_n t_k) \right)^2 \right]. \quad (4.5)$$

Using complexity analysis (Lewis and Papadimitriou, 1997) computation of the DFT can be shown to scale exponentially with regard to N (i.e. $O(N^2)$). Today, the DFT is rarely computed from first principles, and complexity optimisation methods (such as butterflying, memoisation and look-up tables) can be employed to reduce the computational cost of the algorithm dramatically to $O(N \log N)$ iterations as is the case with the fast Fourier transform (FFT) (Cooley and Tukey, 1965; Frigo and Johnson, 1998).

4.5.4 DIRECT PSD ESTIMATION OF HRV

The Lomb periodogram (Lomb, 1976; Flannery et al., 1992) is a least-squares optimisation technique which minimises the squared error between a signal, $x(t_k)$, and a reference signal, $s(t_k; \omega)$ directly without resampling. Whereas the FT estimation weighs the results based on the time interval, the Lomb periodogram weighs the data on a per-point basis (Biala et al., 2010). This is achieved by computing the square error between a signal and the least squares estimation by

$$\epsilon = \sum_{k=1}^N \left(x(t_k) - s(t_k; \omega) \right)^2, \quad (4.6)$$

where ϵ is the square error value, and $s(t_k; \omega)$ is a set of reference sinusoids and is defined by

$$s(t_k; \omega) = a_1 \cos(\omega t_k) + a_2 \sin(\omega t_k), \quad (4.7)$$

where a_1 and a_2 are the amplitudes of the constituent components of the sinusoid. The full error equation is written as follows.

$$\epsilon(a_1, a_2) = \sum_{k=1}^N \left(x(t_k) - a_1 \cos(\omega t_k) - a_2 \sin(\omega t_k) \right)^2 \quad (4.8)$$

The least squares algorithm optimises the amplitudes, a_1 and a_2 , to achieve the minimum square error against the reference. This is achieved by computing two partial derivatives of Equation (4.8) with regard to the amplitudes of the reference sinusoids, and equating the results to 0, i.e.

$$\frac{\delta \epsilon}{\delta a_1} = \sum_{k=1}^N -2 \cos(\omega t_k) \left(x(t_k) - a_1 \cos(\omega t_k) - a_2 \sin(\omega t_k) \right) = 0, \quad (4.9)$$

$$\frac{\delta \epsilon}{\delta a_2} = \sum_{k=1}^N -2 \sin(\omega t_k) \left(x(t_k) - a_1 \cos(\omega t_k) - a_2 \sin(\omega t_k) \right) = 0. \quad (4.10)$$

It is desirable to have

$$\sum_{k=1}^N \cos(\omega t_k) \sin(\omega t_k) = 0 \quad (4.11)$$

for computational complexity and orthogonality reasons, as Equations (4.9) and (4.10) can be rewritten in matrix form by

$$\begin{pmatrix} \sum_{k=1}^N x(t_k) \cos(\omega t_k) \\ \sum_{k=1}^N x(t_k) \sin(\omega t_k) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^N \cos^2(\omega t_k) & \sum_{k=1}^N \cos(\omega t_k) \sin(\omega t_k) \\ \sum_{k=1}^N \cos(\omega t_k) \sin(\omega t_k) & \sum_{k=1}^N \sin^2(\omega t_k) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \quad (4.12)$$

Therefore, when the condition of Equation (4.11) is true, Equation (4.12) becomes orthogonal. To obtain this condition, a time delay factor, τ , is introduced, and τ must satisfy the criteria that

$$\sum_{k=1}^N \cos(\omega(t_k - \tau)) \sin(\omega(t_k - \tau)) = 0. \quad (4.13)$$

Solving Equation (4.13) for τ yields

$$\tau = \frac{1}{2\omega} \arctan \left(\frac{\sum_{k=1}^N \sin(\omega t_k)}{\sum_{k=1}^N \cos(\omega t_k)} \right). \quad (4.14)$$

Now, all sinusoidal components are subject to the τ correction factor, and applying this to Equation (4.12) provides

$$\begin{pmatrix} \sum_{k=1}^N x(t_k) \cos(\omega(t_k - \tau)) \\ \sum_{k=1}^N x(t_k) \sin(\omega(t_k - \tau)) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^N \cos^2(\omega(t_k - \tau)) & 0 \\ 0 & \sum_{k=1}^N \sin^2(\omega(t_k - \tau)) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \quad (4.15)$$

The optimal amplitudes of a_1 and a_2 can be computed and the power at a specific frequency can be estimated by

$$P(\omega) = \frac{1}{2} \left(\sum_{k=1}^N x^2(t_k) - \epsilon \right). \quad (4.16)$$

The full Lomb periodogram is then written by

$$P(\omega) = \frac{1}{2} \left(\frac{\left(\sum_{k=1}^N x(t_k) \cos(\omega(t_k - \tau)) \right)^2}{\sum_{k=1}^N \cos^2(\omega(t_k - \tau))} + \frac{\left(\sum_{k=1}^N x(t_k) \sin(\omega(t_k - \tau)) \right)^2}{\sum_{k=1}^N \sin^2(\omega(t_k - \tau))} \right). \quad (4.17)$$

4.5.5 COMPARISON OF HRV FREQUENCY ANALYSIS METHODS

As discussed in Sections 4.5.3 and 4.5.4, there are multiple means by which the PSD of the heart rate can be computed. A number of open-source HRV tools are available for free from software repositories which allow users to choose the PSD estimation method

(de Carvalho et al., 2002; Hamilton, 2002; Parvin et al., 2002; McSharry and Clifford, 2004). Investigations into the benefits of the direct and indirect PSD estimates indicates that the Lomb periodogram is preferable over the re-sampling methods (Moody, 1993, Laguna et al., 1998, Chang et al., 2001, Clifford and Tarassenko, 2005). Indeed, McSharry and Clifford, 2004, designed an open-source ECG generation model where the true spectrum of the heart signal is set by the user before generating the ECG signals. Comparing the results of Lomb and resampling techniques with the knowledge of the true underlying dynamics of the PSD, it was stated that the Lomb periodogram performed in a superior fashion.

For these reasons, the frequency domain features which are described later were extracted with the Lomb periodogram spectral estimation method.

4.6 FEATURES

The QRS points which were employed in the generation of the PDFs were manually annotated. This was performed so that the PDFs which were generated were representative of the true dynamics of the cardiovascular system for the allergic and non-allergic subjects.

Two PDFs are shown in each chart in this section. The PDF which relates to the allergic subjects can contain a significant portion of non-allergic HRV features. This is because it is not known when the allergy begins to affect the ECG. Indeed, obtaining this is the principal focus of this thesis.

In all subsequent PDF charts, the more densely filled (green) curves represent the allergic PDFs and those less-densely filled (blue) are the PDFs which were generated for non-allergic subjects.

4.6.1 NOTATION

For the description of features, all data is considered to be in the discrete domain. A subject's OFC is represented by M feature vectors, with one vector per epoch. A specific epoch is identified by the subscript j , and the time difference between the QRS points within the j^{th} epoch are defined as the vector $\mathbf{RR}_j$. The i^{th} element of $\mathbf{RR}_j$ is accessed by $\mathbf{RR}_j(i)$. Each epoch contains N QRS points, and the value of N is not necessarily similar for all epochs. The elements of $\mathbf{RR}$ are in the units of seconds.

4.6.2 TIME DOMAIN FEATURES

4.6.2.1 MEAN HEART RATE

The mean heart rate (HR) measures the average heart rate over a given epoch. It is defined by

$$\text{HR} = \frac{60}{\mu}, \quad (4.18)$$

where μ is the average time between subsequent heart beats in a given epoch, and is defined by Equation (4.19)

$$\mu = \frac{1}{N} \sum_{i=1}^N \mathbf{RR}_j(i). \quad (4.19)$$

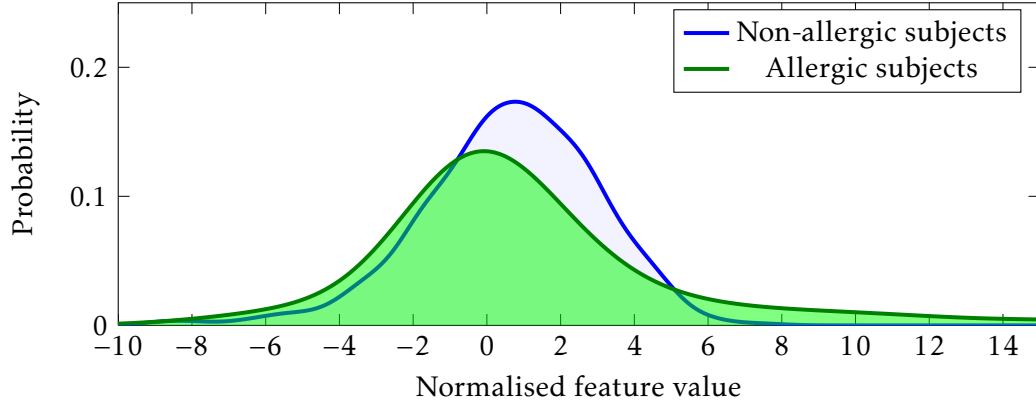


Figure 4.5: PDF of mean heart rate, generated from allergic and non-allergic subjects

4.6.2.2 STANDARD DEVIATION

The standard deviation of the R-R intervals measures the variability and diversity of the QRS complexes found within a given epoch. It is defined by

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{RR}_j(i) - \mu)^2}, \quad (4.20)$$

where μ is defined by Equation (4.19). A higher standard deviation of the heart rate indicates high variance of heart beats within an epoch, and conversely a low standard deviation indicates consistency in the heart rhythm. The PDF of the standard deviation is shown in Figure 4.6.

4.6.2.3 COEFFICIENT OF VARIATION

The coefficient of variation is a normalised measure of the variance of a series of data. It is calculated by dividing the standard deviation of the heart rate by the mean of the heart rate and it is defined by

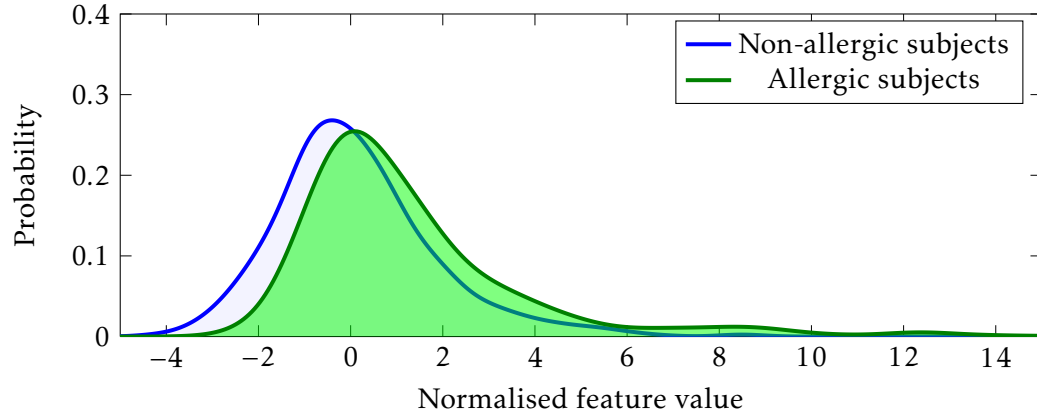


Figure 4.6: PDF of standard deviation of the heart rate, generated from allergic and non-allergic subjects.

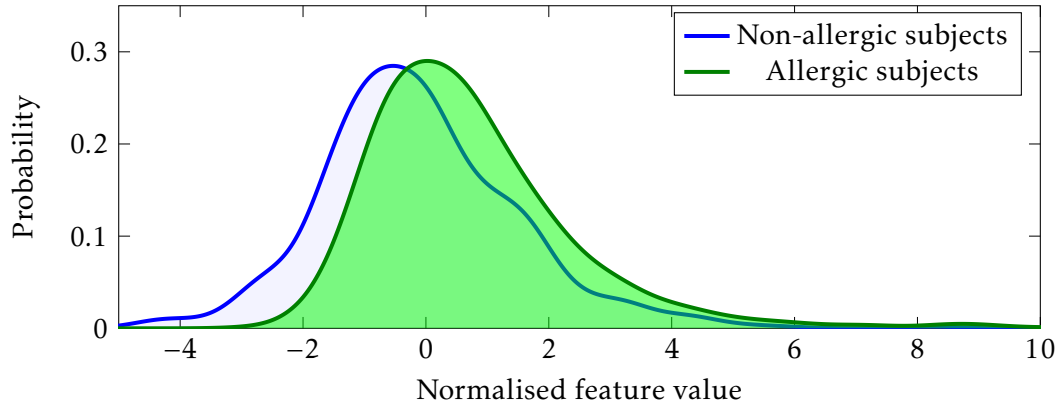


Figure 4.7: PDF of coefficient of variation of the heart rate, generated from allergic and non-allergic subjects.

$$\text{Coefficient of Variation} = \frac{\sigma}{\mu}. \quad (4.21)$$

The effect of this normalisation procedure on the PDFs of the mean and standard deviation features (Figures 4.5 and 4.6) is shown in Figure 4.7. The distributions of the standard deviation and coefficient of variation features are similar, but the coefficient of variation appears to provide better separation than μ and σ alone.

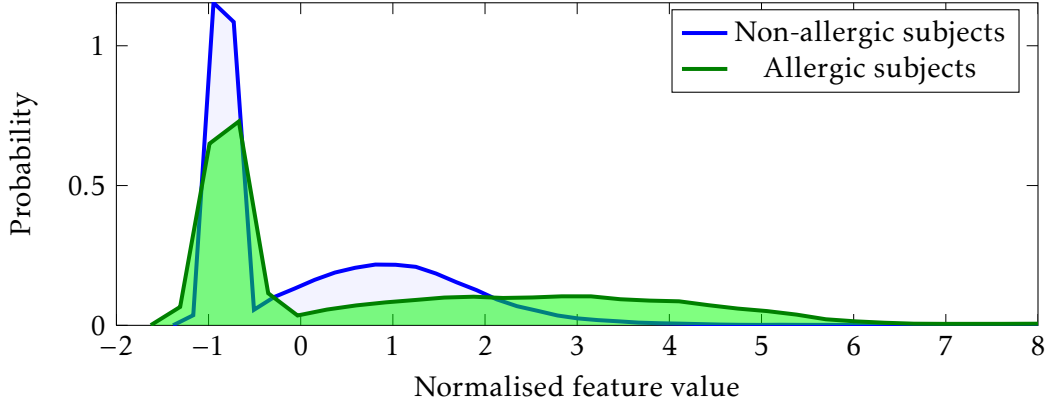


Figure 4.8: PDF of RMSSD of the heart rate, generated from allergic and non-allergic subjects.

4.6.2.4 RMSSD

The root mean square of successive difference (RMSSD) feature measures the root mean square (RMS) of the successive differences between the times of the QRS complex. It is defined by Equation (4.22).

$$\text{RMSSD} = \sqrt{\frac{1}{N-2} \sum_{i=2}^{N-1} (\text{RR}_j(i) - \text{RR}_j(i-1))^2} \quad (4.22)$$

Figure 4.8 shows the PDF of this feature. A high correlation between the allergic and non-allergic subject data can be seen in this figure. However, the probability of the allergic category is greater towards the more positive values of the normalised feature. The shape of this feature is not normal. It is uncertain why this shape occurred, but as the QRS points are manually annotated there is confidence that this distribution is accurate.

Listing 4.1: Calculation of NN_x .

```

1 // Input parameters
2 //   RR: pointer to the RR array
3 //   N:   Length of RR array
4 //   x:   Difference threshold
5
6 // Returns
7 //   nn:   The number of QRS points that differ
8 //         by over  $x_{ms}$  in the given epoch
9
10 int nn( double *RR, int N, float x ) {
11     int n = 0;
12
13     for ( int i=1; i<N; i++ )
14         if ( RR[i]-RR[i-1] ≥ x )
15             n++;
16
17     return n;
18 }

```

Listing 4.2: Calculation of pNN_x .

```

1 // Input parameters
2 //   RR: pointer to the RR array
3 //   N:   Length of RR array
4 //   x:   Difference threshold
5
6 // Returns
7 //   pnn:  The percentage of QRS points that differ
8 //         by over  $x_{ms}$  in the given epoch
9
10 double pnn( double *RR, int N, float x ) {
11     return (double)nn( RR, N, x ) / (double)( N - 2 );
12 }

```

4.6.2.5 NN/PNN

The NN_x and PNN_x features describe the number of successive QRS points within an epoch that differ by the time x . Their calculation is described in 0-indexed C code shown in Listings 4.1 and 4.2.

Typically values of 25 ms and 50 ms are used. Figure 4.9 shows the allergic and non-allergic distributions of PNN_{50} .

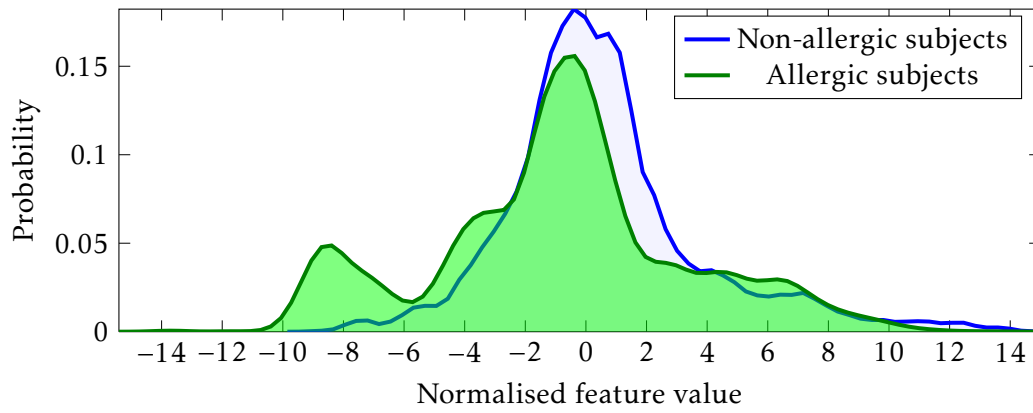


Figure 4.9: PDF of PNN_{50} of the heart rate, generated from allergic and non-allergic subjects.

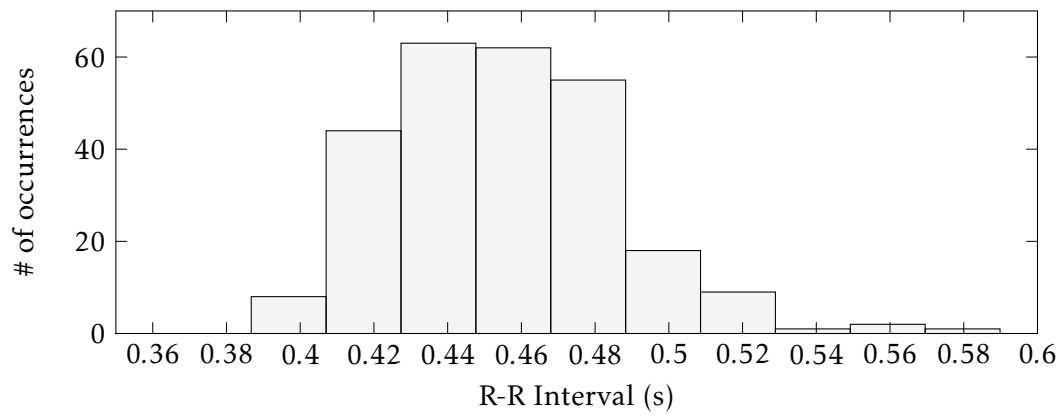


Figure 4.10: Histogram of the relative times between successive QRS complexes.

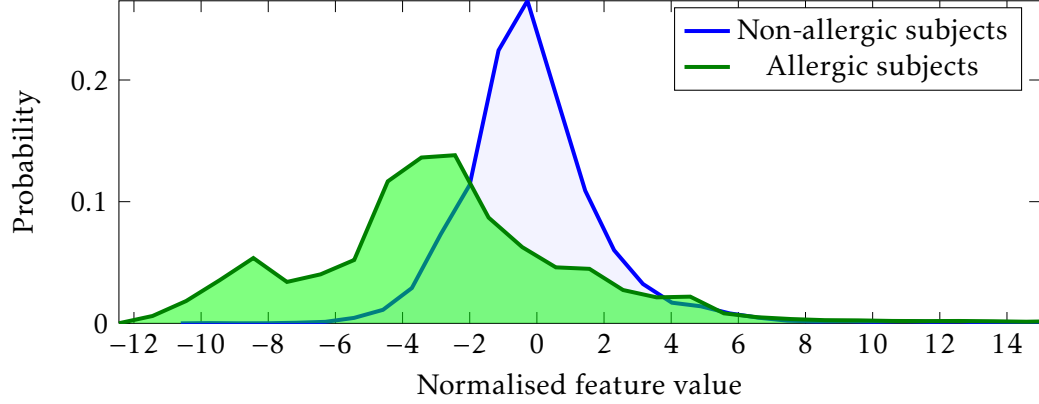


Figure 4.11: PDF of histogram of the heart rate, generated from allergic and non-allergic subjects.

4.6.2.6 HISTOGRAM

Features can also be calculated through generation of a histogram of the R-R intervals from within an epoch. Figure 4.10 shows a typical histogram of R-R intervals which was constructed with 10 bins. The feature value is then calculated by

$$\text{hist} = \frac{h}{w}, \quad (4.23)$$

where w is the width of the histogram, i.e. the difference between the maximum and minimum values on the x-axis, and the h is the height of the most frequently occurring bin.

Smaller values of this feature indicate a wide dispersal of R-R intervals over the epoch owing to the smaller h and larger w values, while larger values of this feature indicate consistency between R-R intervals over an epoch. This is reflected in the PDF of this feature shown in Figure 4.11 where a higher probability of allergy can be seen with lower values of this feature. Indeed, this feature shows the best difference between the allergic and non-allergic subjects of all the features which were discussed up until now.

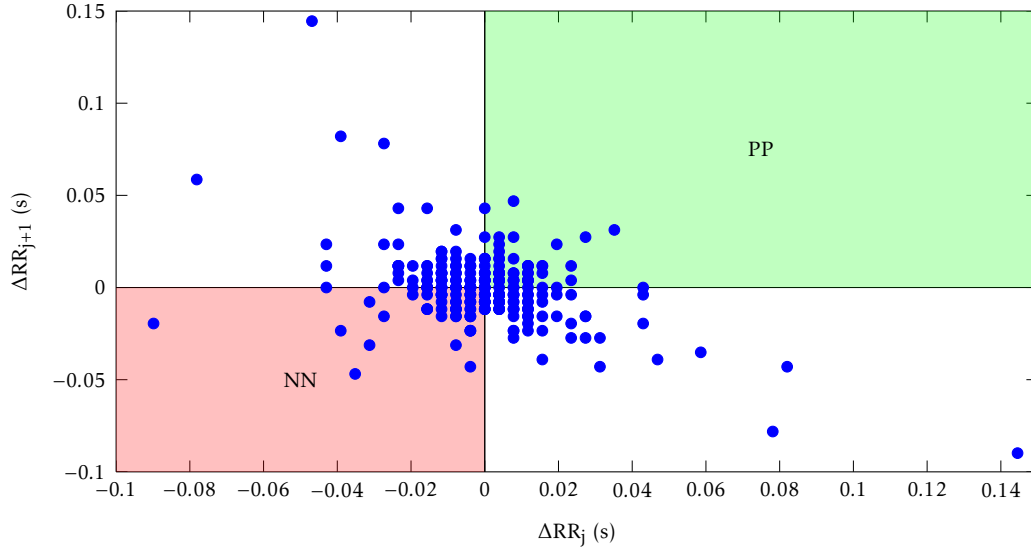
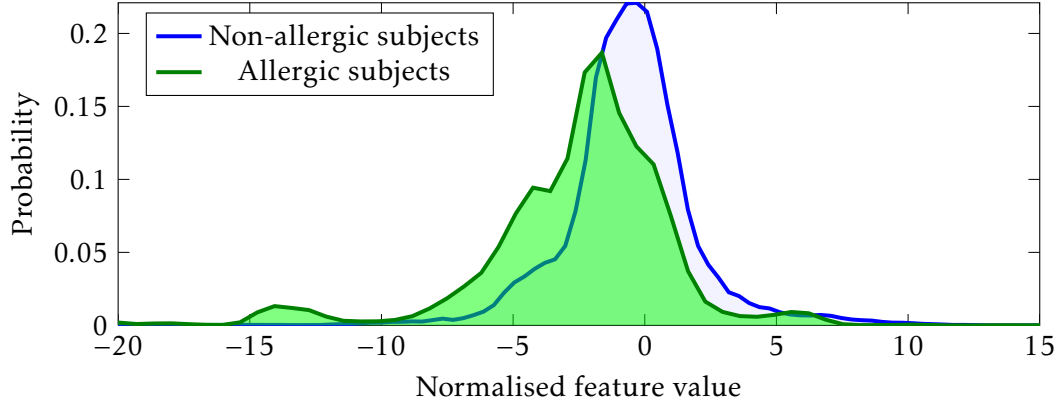


Figure 4.12: Chart of the change between successive QRS complexes.

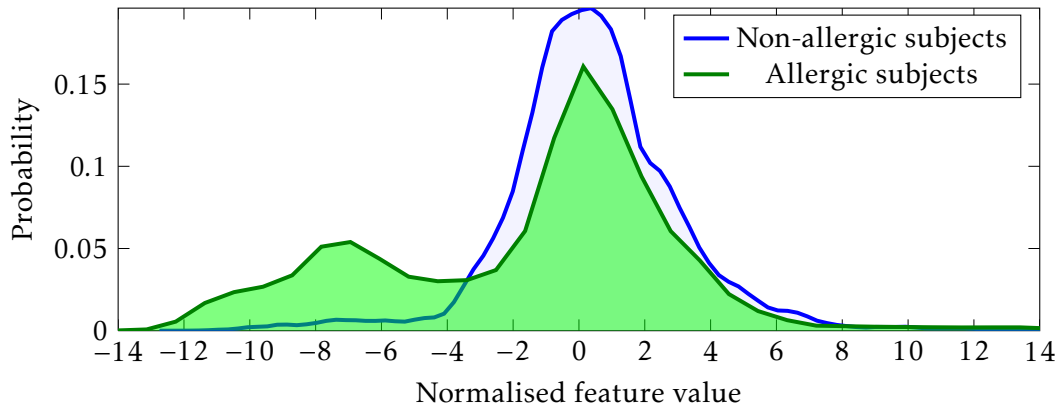
4.6.3 SEQUENTIAL DOMAIN FEATURES

Features extracted from the sequential domain compute the inter-beat variability of the QRS complexes found within an epoch. The relative increases and decreases of the heart rate are computed by differentiating the vector of R-R intervals between each successive QRS complex (effectively obtaining the relative proportions of ‘acceleration’ of and ‘deceleration’ of the heart rate over an epoch (Schechtman et al., 1992)). Quantification of this can be visualised by plotting the i^{th} result of the difference of the R-R intervals against the $(1 + i)^{\text{th}}$. Figure 4.12 shows this plot and the graph is segmented into four quadrants.

The shaded region in the upper right hand quadrant of Figure 4.12 indicates that two consecutive increments in time between successive QRS complexes occurred, which is representative of a slowing heart rate. This region is referred to as the PP quadrant. The shaded region in the lower left hand quadrant of Figure 4.12 indicates the presence of two consecutive R-R intervals with a decreasing interval, which indicates a speeding heart rate. This quadrant is referred to as NN quadrant.



(a) PDF of PP feature of the heart rate, generated from allergic and non-allergic subjects from the sequential domain.



(b) PDF of NN feature of the heart rate, generated from allergic and non-allergic subjects from the sequential domain.

Figure 4.13: PDFs derived for the sequential domain features.

The sequential trend features are calculated by counting the number of occurrences of points within the PP and NN quadrants. These are then divided by the total number of points in the chart. The PDF of the features are shown in Figures 4.13a and 4.13b. A significant amount of similarity exists between these two features. However, in both cases, lower values of the feature are more indicative of allergy.

4.6.4 POINCARÉ FEATURES

Poincaré features are another means of assessing the beat-to-beat variability and non-linear dynamics of R-R intervals. Poincaré features are obtained from a Poincaré chart,

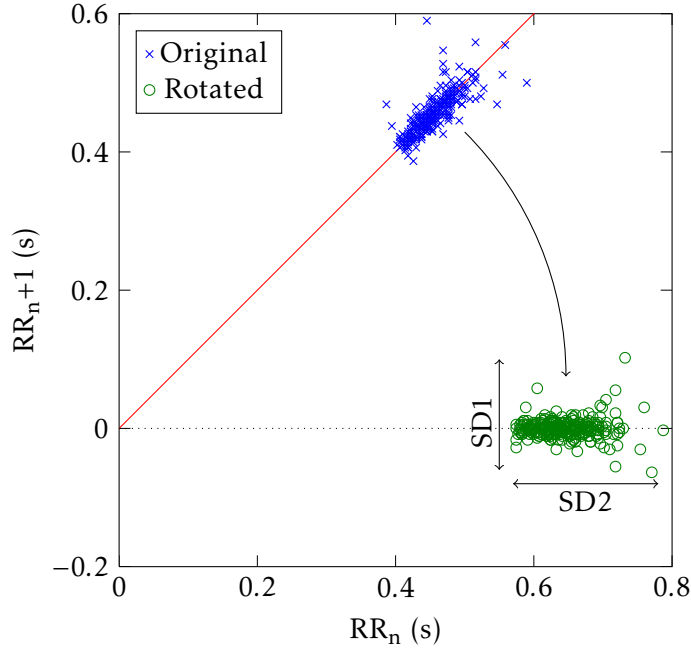
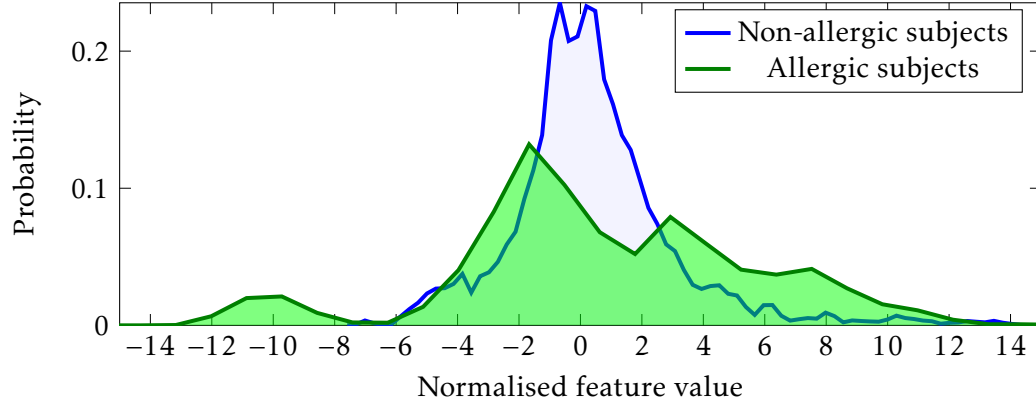


Figure 4.14: Original and rotated points plotted in a Poincaré Chart.

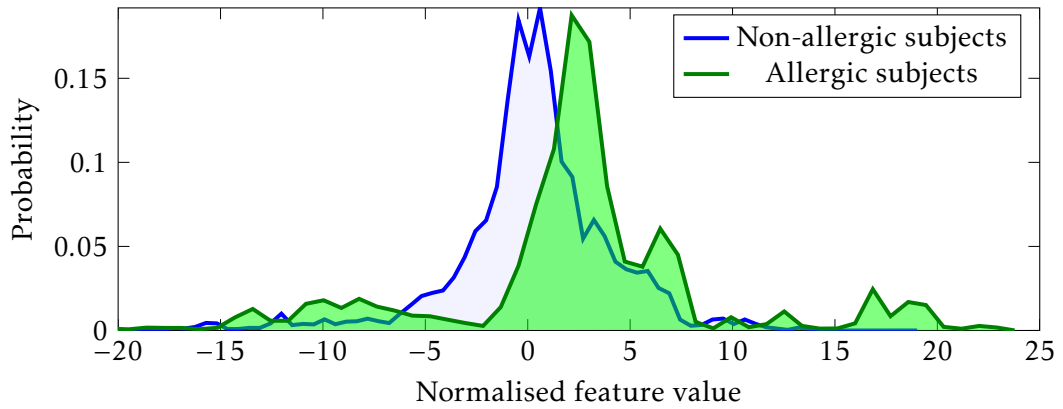
which plots the current R-R interval against the next R-R interval, as is shown by the cluster of $\times$'s in Figure 4.14. This cluster will be dispersed about a line orientated at 45° , which plotted as the solid line in Figure 4.14.

Features are extracted based on the extent of horizontal and vertical distribution of the $\times$'s about the solid line, i.e. perpendicular and parallel to the 45° line. Computer-based computation of these features is simplified if each point is rotated clockwise by 45° about the origin, which can be achieved by the rotation matrix defined in Equation (4.24), with $\theta = -\frac{\pi}{4}$. This results in the cluster of o 's in Figure 4.14 which are dispersed about $y = 0$ on the x -axis.

$$\begin{bmatrix} x' & y' \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \times \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (4.24)$$



(a) PDF of CSI of the heart rate, generated from allergic and non-allergic subjects.



(b) PDF of CVI of the heart rate, generated from allergic and non-allergic subjects.

Figure 4.15: CSI and CVI PDF from Poincaré features.

With the rotated cluster, two measurements from the plot (SD1 and SD2) can easily be computed. SD1 is the standard deviation of the set of rotated x -values, and SD2 is the standard deviation of the set of rotated y -values. Points close to the line of identity indicate a similarity between consecutive beats, and conversely points which are further from the line of identity indicate that a change in the heart rhythm has occurred.

The cardiac sympathetic index (CSI) and the cardiac vagal index (CVI) are computed from SD1 and SD2, and they are defined by Equations (4.25) and (4.26) respectively.

$$CSI = \frac{SD2}{SD1} \quad (4.25)$$

$$CVI = \log(SD2 \times SD1) \quad (4.26)$$

Figures 4.15b and 4.15a show the PDFs of the CSI and CVI features for the allergic and non-allergic subjects. The CVI feature shows a positive mean-shift between allergic and non-allergic subjects, with higher probabilities of allergic subjects at the higher positive side of the chart. The CSI feature for allergic subjects presents a wider distribution of the feature but with a similar mean between the allergic and non-allergic subjects.

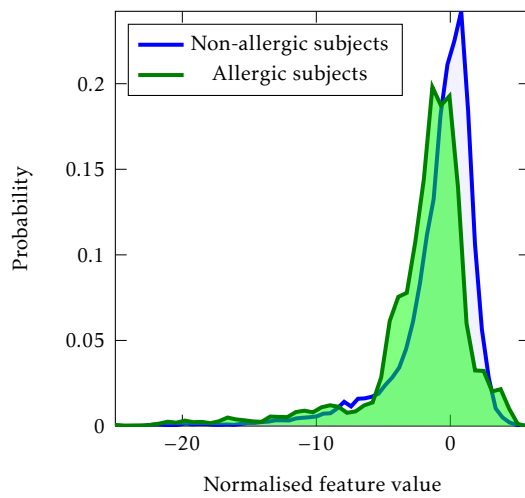
4.6.5 FREQUENCY DOMAIN FEATURES

The frequency spectrum was computed with the Lomb periodogram. The total powers in the very low frequency (VLF), low frequency (LF), and high frequency (HF) bands were extracted from the PSD estimates from the Lomb periodogram. Gombarska and Horicka, 2012, presented Table 4.1 in which the boundaries for VLF, LF, HF and ultra-low frequency (ULF) are presented. These frequency ranges are indicative of physiological and cardiac events which are listed alongside the frequency ranges in Table 4.1.

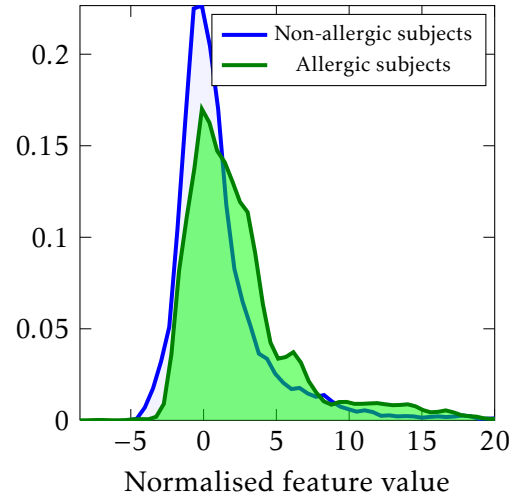
The ULF frequency band was not considered in this work because it is stated by researchers (Tulppo and Huikuri, 2004; Rajendra Acharya et al., 2006) that meaningful determinations of the associated powers can violate the rules governing PSD determinations. This concern

Table 4.1: Table of HRV diagnostic frequency ranges for children.

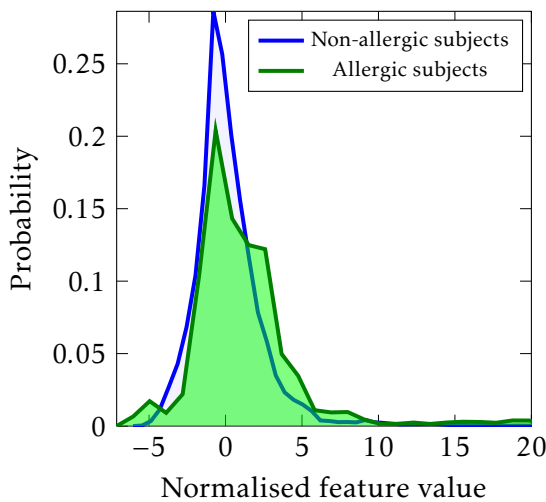
Type	Frequency (Hz)	Origin
HF	0.15 – 0.4	Parasympathetic, respiratory sinus arrhythmia
LF	0.04 – 0.15	Sympathetic + parasympathetic
VLF	0.0033 – 0.04	Sympathetic, chemo-receptors, thermoregulation, endocrine
ULF	< 0.0033	Circadian rhythms



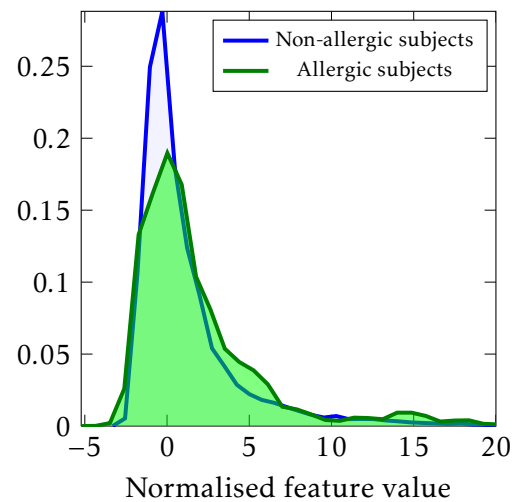
(a) PDF of the power from VLF frequency band of the heart rate, generated from allergic and non-allergic subjects.



(b) PDF of the power from LF frequency band of the heart rate, generated from allergic and non-allergic subjects.



(c) PDF of the power from HF frequency band of the heart rate, generated from allergic and non-allergic subjects.



(d) PDF of the power from LF/HF frequency band of the heart rate, generated from allergic and non-allergic subjects.

Figure 4.16: PDF of the frequency domain features.

is also valid for the VLF frequency band, but it has been clinically reported to be strongly related to the parasympathetic response, even in short duration recordings, (Carney et al., 2001; Kleiger et al., 2005). The ratio between the HF and LF frequency bands was computed as another feature, and it is a measure of the ratio between the sympathetic and parasympathetic response of the ANS.

The PDFs of the frequency-domain HRV features were extracted and are shown in Figures 4.16a — 4.16d. The non-allergic and allergic PDFs overlap significantly, but in all cases there is tendency towards a higher probability of allergy towards the positive and negative extremities of the independent axis. It is expected that these features will perform well in allergy classification because when allergy occurs the ANS should react to combat allergic reactions.

4.7 DISCUSSION

This is the first work which has quantified the effect of allergy on HRV features, and it has shown that in a number of cases a clear separation can be obtained between the allergic and non-allergic classes. In the case of the CVI feature, for example, a mean-shift and wider variance of approximately $3\sigma_{\text{background}}$ was obtained. It is interesting to note that with the accelerometer features no deviation beyond the background distribution was observed while for the HRV features this is not the case. This trait is one that when identified can be exploited by classification platforms. This is a desirable trait for classification purposes as it indicates that separability and classification are obtainable. This is in counterpoint to the accelerometer-based features that were obtained in the previous chapter, and it indicates that HRV-based classification exposes a viable avenue for automated allergy classification.

The sequential domain features also demonstrated significant differences in probability mass and bimodalities were introduced to the allergic PDFs. This modality-distortion also occurred (but to a lesser extent) with the histogram and PNN features. It is interesting to see that the frequency-based histograms are not well-partitioned, with the mean and

variance of the distributions being approximately equal for the allergic and non-allergic classes. These features are well documented by other researchers (Aziz et al., 2012) as being highly correlated to the activity of the ANS.

It should be stated that the figures in this chapter presented histograms in a single dimension. In later chapters it will be shown that employing high dimensional multivariate data modelling can be exploited to yield superior separability. Such high dimensional representations, however, cannot be visualised on the pages of this thesis, and true separability of the classes can only be obtained analytically. The focus of the remaining chapters of this thesis lies in the means of representing and assessing these distributions.

The PDFs shown here also show that a high degree of overlap exists between the allergic and the non-allergic subjects. This is to be expected because OFCs are not temporally annotated, which indicates that supervised class separability is not possible. However, this further supports the argument that machine-based classification of allergy through analysis of HRV features is worthwhile, because even with single dimensional representations of features, class separability has been obtained in many cases.

CHAPTER 5

MACHINE LEARNING FOR ALLERGY CLASSIFICATION

5.1 INTRODUCTION

IN the previous chapter, the eighteen HRV features which are used to quantify the variability of the heart during OFCs were described and these assess whether machine learning algorithms might be utilised for automatic classification of food allergy.

The only temporal label available is that obtained from the ECG data recorded before allergens were administered. Therefore, non-discriminative classification is required for allergy detection as the allergic events are not available. By this process, models are trained on the normalised background heart rate variability features, and allergy will be classified if features anomalous to this distribution are detected. The background data is recorded before any food was ingested by the subjects, so is guaranteed to represent

the normal, non-allergic state. This classification routine is called novelty or abnormality detection.

As the QRS points used here were annotated, this chapter will determine the effectiveness of HRV-based allergy detection, and the manual aspect of QRS identification provides affirmation of this. The results obtained here will also provide the upper-bound of the estimated efficacy of machine-based allergy classification.

5.2 NOVELTY DETECTION FOR OFC

5.2.1 CHOICE OF CLASSIFICATION ROUTINE

Chapter 2 discussed the one-class SVM and GMM classifiers. For the classification work of this thesis, GMMs were chosen over one-class SVMs and other novelty detection algorithms for a number of reasons. GMMs model the distribution of the training data and have been shown to be robust in a number of classification applications. Effectively, this means that the GMMs generate multi-dimensional PDFs which can be used to ascertain the ‘probability’ (or more formally, the likelihood, see later) that new data belong to the background class. GMMs also allow for subject-adaptive procedures, which is important for the classification routine which is employed, and this will be discussed later. One-class SVMs belong to the boundary-of-novelty estimate. This is useful for very high input dimensionality when not many data points are available. Therefore, on the basis of these arguments, one-class GMMs are preferable over other options for OFC classification.

5.2.2 FEATURE TRANSFORMATION

For the classification of allergy, principal component analysis (PCA) (Pearson, 1901) is first performed on the normalised training data. The PCA transformation was performed

in order to de-correlate the feature set. This is a requirement for allergy classification as the allergy database is insufficiently sized to accurately train GMMs with full covariance matrices. Using diagonal covariance matrices dramatically reduces the quantity of samples required for training of the classification models. It is not always possible to assign interpretation to these new components obtained with the PCA transform and this is particularly true in cases where normalisation has been performed on the initial training data (Webb et al., 2011).

For high-dimensional data, the PCA transform is equivalent to minimising the scatter matrix, S_i , over N dimensions, by

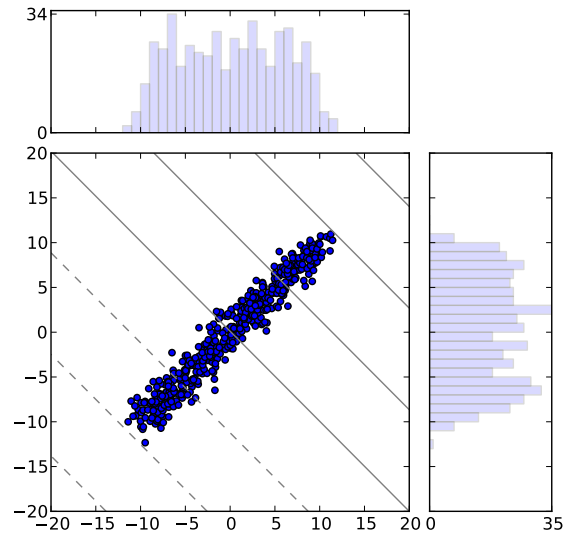
$$S_i = \sum_{j=1}^N (\mathbf{x}_j - \mu)(\mathbf{x}_j - \mu)^T, \quad (5.1)$$

where

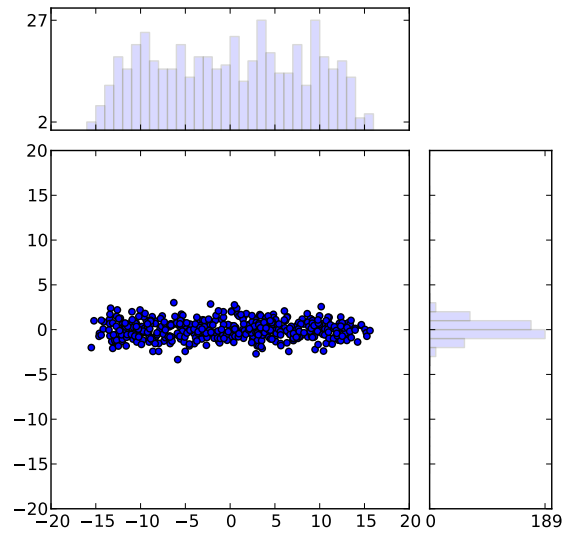
$$\mu = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j. \quad (5.2)$$

The PCA transformation matrix is obtained from the eigenvectors of S_i , and these can be obtained singular value decomposition (De Lathauwer et al., 2000) or other similar procedures.

Figure 5.1 illustrates a two-dimensional example of what the PCA transform accomplishes. Figure 5.1a shows two-dimensional data plotted in feature space. In viewing the histograms associated with each axis it can be seen that the range and variance of feature values are approximately equal in the x - and y -directions. However, once the PCA transform has been performed, the feature data is transformed to the distributions shown in Figure 5.1b which is plotted in ‘component’ space. This is equivalent to viewing the



(a) Two-dimensional feature-space, showing the contours lines of the axis of the first principal component. Histograms show the distribution of features along the x - and y -axes with equal bin ranges.



(b) Two-dimensional component-space after the PCA transformation. Histograms show the distribution of features along the x - and y -axes with a much greater variance range with the x -axis.

Figure 5.1: An illustration of PCA in two-dimensional feature space (subplot a) and two-dimensional component-space (subplot b).

original data along axes shown as the contour lines in Figure 5.1a. In Figure 5.1b, the ranges of the feature data are not equal. Indeed, the computed variance of the first component accounts for 97% of the total variance of the transformed feature data, while in the original distribution the importance of both features is approximately equal.

For higher dimensional cases, if the first C ($C < N$) principal components account for the majority of the variance, they may be used to describe the variance profile of the data accurately. If the remainder of the matrix is discarded in analysis, dimensionality reduction is achieved. It has been shown by other researchers that employing the PCA transform to select a subset of features in a subject-independent manner can improve the accuracy of classification (Thomas, 2010).

5.2.3 GAUSSIAN MIXTURE MODELS

The Gaussian classifier is a density estimation algorithm which is commonly used for modelling data distributions. For classification purposes features are often extracted from the data to obtain more information about the data itself. While the original data may be normally distributed, the extracted features may not preserve this trait. This is a problem for the Gaussian classifier because multi-modal distributions would be poorly modelled by a single Gaussian distribution.

Gaussian mixture models (GMMs) were devised to solve this problem. These employ weighted mixtures of several Gaussian distributions to model arbitrary feature distributions. In a similar manner to how the FFT decomposes a data sequence into a weighted sum of complex exponentials and computes the frequency spectrum, GMMs (through the expectation maximisation algorithm, see later) can be thought of as decomposing a data sequence to a basis of Gaussian distributions and obtaining a model of the density of the data. The mixture of Gaussians, therefore, allows arbitrary and non-trivial distributions to be modelled, provided the algorithm is correctly parameterised.

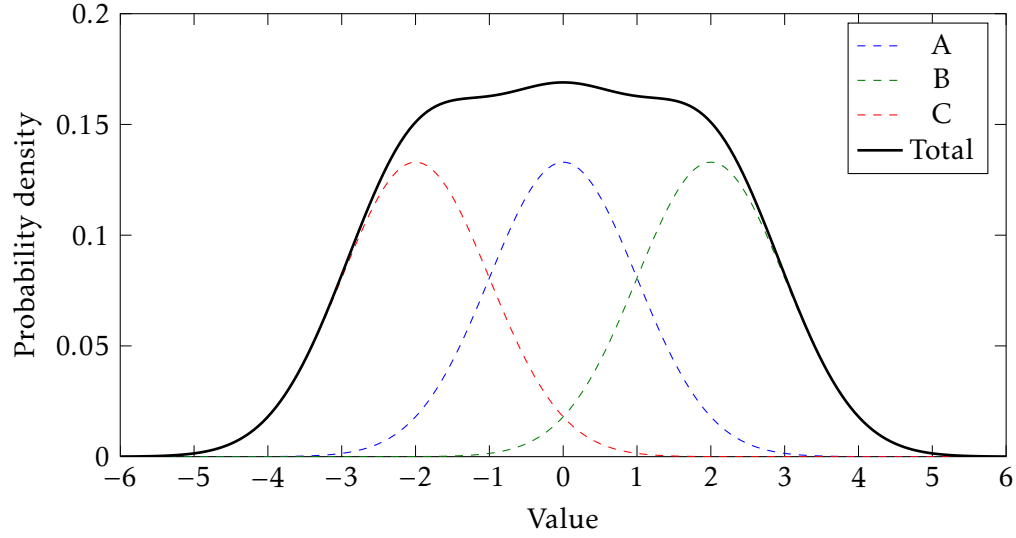


Figure 5.2: A mixture of three equally-weighted Gaussians (dashed lines) which combine to represent a multi-modal non-normal distribution (solid black line).

Figure 5.2 demonstrates how three equally-weighted Gaussian distributions, A , B and C , can be combined in order to represent the non-Gaussian distribution, shown by the solid line in Figure 5.2. The number of Gaussians that are required to represent the distribution depends strongly on the data which are to be represented, and is specific to the application in question.

For GMMs, conditional density is modelled by

$$p(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \sum_{i=1}^m \omega_i N(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (5.3)$$

where m is the number of Gaussians, the weights, ω_i , satisfy

$$\sum_{i=1}^m \omega_i = 1 \text{ and} \quad (5.4)$$

$$\omega_i \geq 0, \forall i, \quad (5.5)$$

and $N(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the multivariate Gaussian function defined by

$$N(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \quad (5.6)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (which are collectively termed the mixture parameters and given the shorthand symbol $\boldsymbol{\theta}$ for conciseness) are mean and variance matrices, and d is the dimensionality of the data, $\mathbf{x}$. $\boldsymbol{\Sigma}$ is a symmetric and positive semi-definite matrix and $|\boldsymbol{\Sigma}|$ is its determinant. The task of training the GMM involves selecting the appropriate mixture parameters to represent the training data accurately. This is performed through the use of k -means clustering and the expectation maximisation algorithm.

5.2.3.1 k -MEANS CLUSTERING

For the GMM algorithm, m Gaussians are required to model the training data set, and the value of m is termed the GMM order. The k -means clustering algorithm (Hartigan and Wong, 1979) can be employed to automatically partition the training data set into k data clusters and discover the mean (or centroid) of each of these clusters. The k -means algorithm is iterative in nature, and will generally have indeterminable computation time (Lewis and Papadimitriou, 1997; Aloise et al., 2009). The following steps outline the procedure of the k -means algorithms:

1. Initialise k initial cluster centroids randomly in feature space.

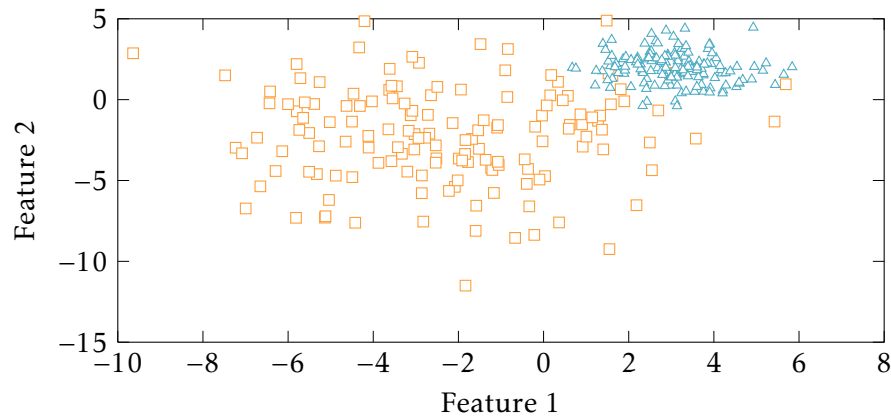
2. Calculate the Euclidean distance (Deza and Deza, 2009) between all data points and the k cluster centroids.
3. Assign membership of each data point to the closest centroid.
4. Compute the new cluster centroids for the k clusters based on the members of each cluster.
5. If no change in centroid coordinates have occurred between steps 2 and 4, the process is terminated. Otherwise repeat step 2 using the new cluster centroid positions that were calculated in step 4.

The k -means algorithm will always segment the data into k clusters there is no guarantee that good partitioning will be obtained. Therefore value of k (which is also termed the Gaussian order, m) must be chosen carefully as the data might be automatically clustered into an inappropriate and unrepresentative number of clusters. However, k -means clustering partitions data without consideration to the density of the data points which are under investigation. To partition data in this manner, expectation maximisation (EM) is required.

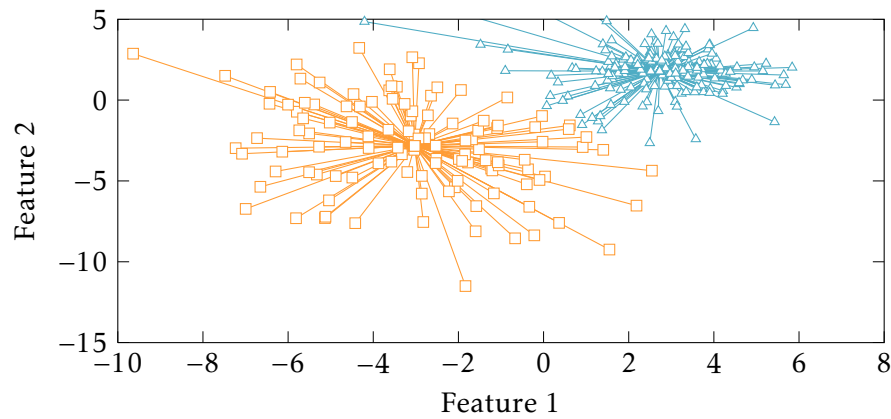
5.2.3.2 EXPECTATION MAXIMISATION

EM takes into consideration not only the means of the clusters, but also the weights and covariance between features and it can provide superior data segmentation than k -means alone (Figueiredo and Jain, 2000). EM is typically initialised by the k -means clustering algorithms, and seeks to select the mixture parameters so as to model the training data accurately. This is achieved with a maximum-likelihood algorithm, and is therefore iterative in nature as no closed form solution exists (Bishop et al., 2006).

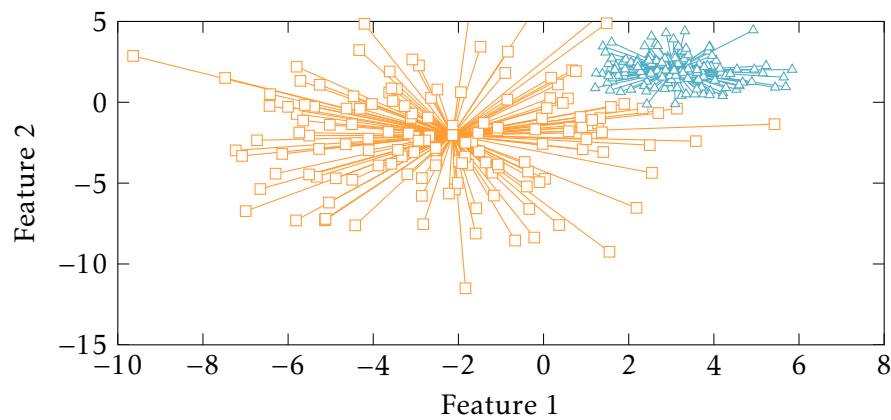
The likelihood is maximised by EM in order to ensure that the distribution of the training data is modelled accurately. In order to simplify computation, and for computer-based floating point precision, the log-likelihood is taken and this requires N floating point



(a) The true relationship between the two Gaussian distributions which make up the non-normal distribution.



(b) The partitioning of the classes obtained by k -means clustering ($k = 2$).



(c) The partitioning of the classes obtained by the expectation maximisation clustering ($m = 2$).

Figure 5.3: A demonstration of the difference in clustering which is obtained by k -means clustering and the expectation maximisation algorithm. With subfigures B and C, a line is drawn from each point to its associated cluster.

additions rather than N floating point multiplications (where N is the number of examples of the data).

$$L(X_i|\Theta_i) = \log \prod_{j=1}^N p(\mathbf{x}_j|\Theta_i) \quad (5.7)$$

$$= \sum_{j=1}^N \log \sum_{l=1}^m \omega_l p(\mathbf{x}_j|\theta_j) \quad (5.8)$$

Like any optimisation algorithm, EM selects parameters so as to minimise a loss function. This process repeats until convergence has been reached or until a pre-specified number of iterations have been performed. EM, therefore, iteratively updates the mixture parameters simultaneously until termination criteria are satisfied. EM is a two-step procedure where an expectation calculation step (E) is first performed. The expectation is the product between the true data distribution and the model of the distribution. EM can therefore be thought of as selecting the mixture parameters so as to maximise the correlation between the true data distribution and the model of the data distribution. The maximisation (M) step selects the next arguments for Equation (5.8) based on the outcome of the previous expectation stages. EM maximises the likelihood because even though the parameters which maximise the likelihood are initially unknown, the knowledge of the distribution of the training data can be used to assess the parameters which were selected on every iteration, and the iterative process guides the algorithm to convergence.

Figure 5.3 shows a distribution which was generated with two Gaussian distributions centred at $(-2, -2)$ (marked with $\square$) and $(3, 2)$ (marked with $\triangle$), with standard deviations of 3 and 1 respectively. While the two distributions are presented in Figure 5.3 as two classes, this is to illustrate the difference between cluster membership of k -means (Figure 5.3b) and EM (Figure 5.3c). It can be seen that k -means partitions the distribution poorly and mis-categorises a number of data points (obtaining accuracy of less than 90%). The EM partitioning algorithm captures much information from the original distribution, and

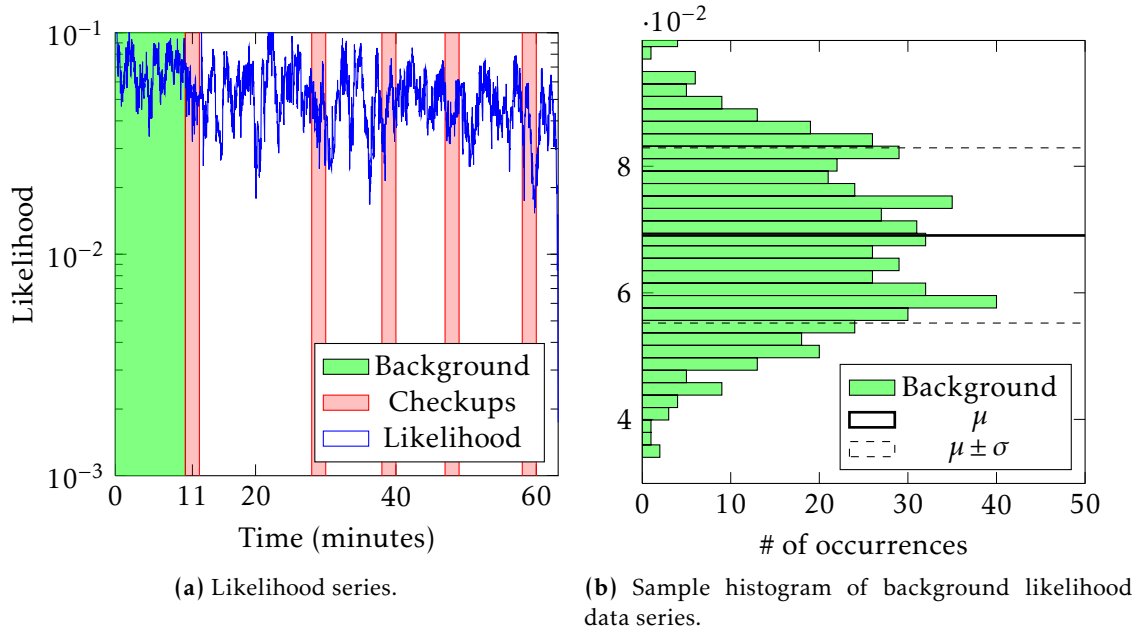


Figure 5.4: Sample likelihood (subplot a) and histogram of the background likelihood (subplot b) of Subject 23.

allows for better modelling of the true data (obtaining over 98% accuracy). While the partitioning is not perfect with either method, EM models the data in a superior manner because of the density-based analysis.

5.2.4 POSTPROCESSING

After learning the mixture parameters, GMMs produce likelihoods for features extracted from new epochs of normalised HRV data. The likelihoods computed are in the range of 0 and 1. The closer the likelihood is to 1 the more likely it is that the data belongs to the background class, and conversely the smaller the likelihood is the less likely it is that the feature vector belongs to the background class. The likelihood will never reach 0, but will become infinitesimally smaller as the data becomes less similar to the background models.

A sample likelihood (for Subject 23) is shown in Figure 5.4a. In this figure, the region highlighted between 0 and 11 minutes represents the background time (before any

allergen was introduced to the OFC). The remaining time segments which are highlighted represent checkup times during which the allergist performed checkups on the subject. The solid trace is the likelihood series computed for the subject. This figure is plotted in a log-scale.

With the goal of machine-based allergy detection, the criterion for novel data (which is classified allergic) is defined as follows: the likelihood must fall below a specific threshold, th , for a specific duration, d . This is defined thusly because allergy should present with non-background features (see Chapter 4), and the more unlike the background new features are, the smaller the computed likelihood will become. However, the PDFs in Chapter 4 also show a significant amount of overlap between the allergic and non-allergic subjects. Therefore, by incorporating the duration parameter, rejection of spurious deviations will be achieved, which will enable superior classification.

The specific values of d and n are obtained via subject independent cross validation. The n and d parameters are henceforth referred to as the ‘multiplicative’ and ‘duration’ parameters individually, and collectively are termed the ‘decision making’ parameters.

Figure 5.4b shows the histogram of likelihood values during the initial background state of Subject 23 during OFC, and it can be seen from this example that the values of likelihood follow an approximately normal distribution. Therefore the threshold equation is defined as a function of the mean (μ) and standard deviation (σ) of this background data by

$$th = \mu - n\sigma, \quad (5.9)$$

where n is a multiplicative factor for the standard deviation. This was chosen to make the modelling and decision making subject-adaptive, as it is a function of the subject’s own background distribution. The larger the n parameter the smaller the likelihood must be to surpass the threshold, i.e. the less similar to the background data.

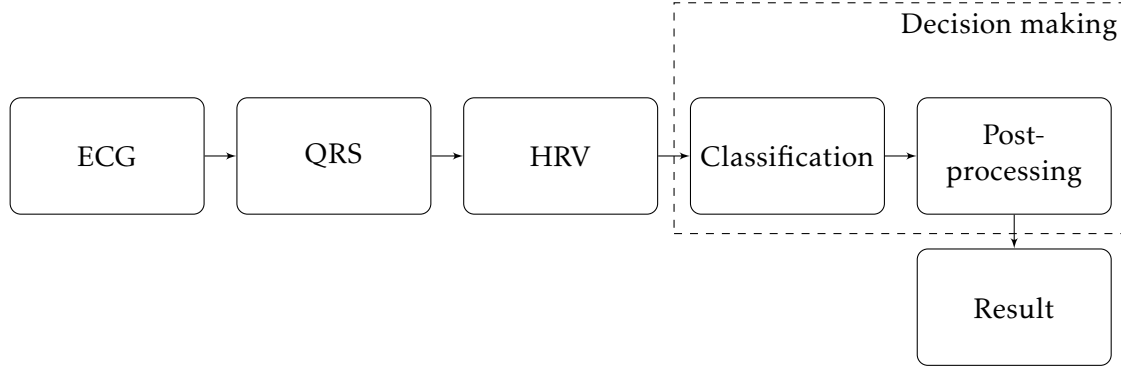


Figure 5.5: Flowchart of classification procedure involving the recording of ECG, annotation of QRS complexes, feature extraction and the classification procedure of OFC.

The second parameter for an allergic decision is d , which is the time for which the likelihood must remain below the threshold th for an allergic decision to be reached. The purpose of this parameter is to reduce the effect of spurious irregularities in the likelihood series which might not be due to allergy, but could be due to the natural variation of the heart. However, the duration parameter will also allow for less extreme threshold values which will facilitate in obtaining better classification results. The signature of abnormality which will be detected by this classification routine can be defined as being a substantiated and sustained departure from the background HRV levels.

5.3 CLASSIFICATION PROCEDURES

Figure 5.5 shows a high-level overview of the allergy classification procedure. The ECG is first recorded. As mentioned in the introduction, analyses in this chapter focus on manually annotated QRS points in order to confirm the presence of allergic signatures in HRV features. The next step in this process is manual annotation of the QRS points. The features described in Chapter 4 are then extracted and background models are learnt. Following this, post-processing and decision making is performed in order to classify the subject in question as being allergic to the allergen they are being tested against. The specifics of the classification and post-processing stages will be discussed in subsequent sections.

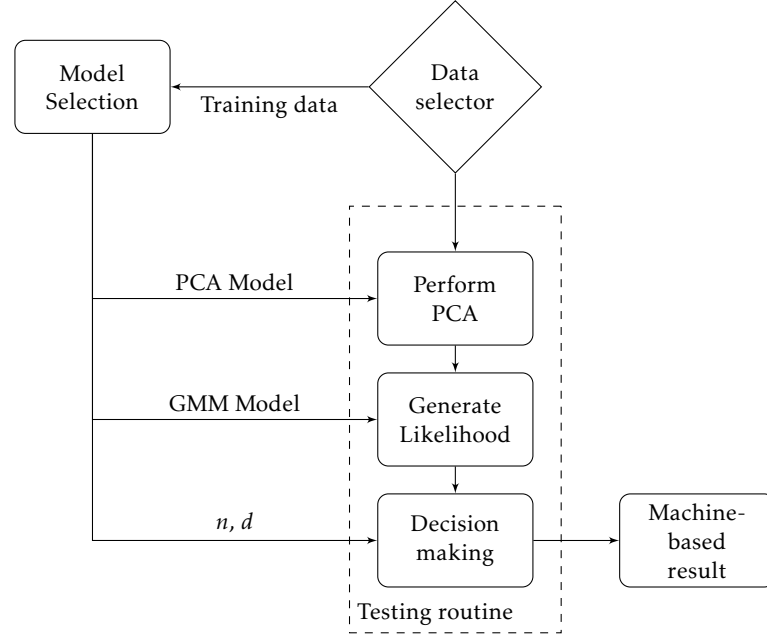


Figure 5.6: Illustration of the data segmentation and testing routines employed in the allergy classification procedure.

For the remainder of this thesis, machine-based detection of allergy is termed **classification** of allergy, while the result of the OFC is termed **diagnosis** of allergy. When a subject is stated as being ‘classified allergic’, there is, therefore, an inherent implication that it was with the statistical modelling and post-processing processes that the classification was made.

The classification and post-processing blocks of Figure 5.5 can be expanded to what is shown in Figure 5.6. In this Figure, model training and testing data are separated. From the training data, PCA and GMM models are selected by the parameter selection routine. The selected multiplicative and duration post-processing parameters are also obtained from the training data and are used by the system to classify allergy on the test subject. This figure shows how the testing and training sections of this classification routine are completely independent, and that the testing data bears no influence in the model and decision making parameter selection routines.

This system only classifies allergy, i.e. it detects abnormal HRV features only. If only normal features are detected tests cannot terminate and will continue as normal.

5.4 CLASSIFIER MODEL SELECTION

5.4.1 PERFORMANCE EVALUATION

There are various performance assessment routines proposed in the literature (Kohavi et al., 1995) such as bootstrapping, split-sample, etc. Their effect on the reported performance for neonatal seizure detection has been compared in previous studies (Temko et al., 2011b). The split-sample method where one fixed partition of the available data is allocated for training and the rest is used as a testing set has several major disadvantages as such a division results in a potentially large bias. Over-optimistic and indeed over-pessimistic results can be obtained depending on what seems an arbitrary partition of the data yielding a ‘good’ or ‘bad’ split.

In this work, leave-one-out (LOO) is used to assess the performance of the developed allergy detector. With this all but one subject are used for training and the remaining subject is used for testing. The process is repeated until every subject was tested, and the average performance is reported. LOO is known to be an almost unbiased estimation of true generalisation error (Vapnik and Kotz, 1982). Additionally, in contrast to randomised re-sampling routines (bootstrapping), the LOO eliminates any subjectivity from the testing protocol, hence it can be repeated and exactly the same results will be obtained (Thomas et al., 2013).

What is examined with the LOO procedure is not a particular model, but the methodology used to obtain such a model. This means that a good modelling system is obtained by this methodology, and the parameters are not fixed for all subjects. Here, 24 data splits of 23 *vs.* 1 are made by the LOO method formed the performance assessment routine, and this means that 24 unique GMM models are obtained by this procedure.

5.4.2 PARAMETER SELECTION

5.4.2.1 SEARCH SPACE

In each of these 24 splits, nested cross-validation model selection on the training 23 subjects' data was performed to choose suitable model parameters. Those include:

- Percentage of information retained by PCA for feature set reduction:

The following set of values was searched: { 80%, 90%, 95%, 99%, 99.9%, 100% }

This range was selected because of the complicated nature of diagnostic HRV, and because preliminary analyses showed that the allergic condition was better identified by preserving more than 80% of the feature variance.

- The number of Gaussians in the GMM model:

The following set of values was searched: { 1, 2, 4, 8, 16, 32, 64 }

This was in order to facilitate modelling the data distributions with simple and complicated models depending on the complexity of the data used in training. This search space has also been successfully employed for EEG and speech processing applications (Reynolds and Rose, 1995; Thomas, 2010).

- The multiplicative factor (n) in decision making:

The integer-rounded values logarithmically distributed over the maximum range were searched.

This range was selected in order to accommodate the entire range of likelihoods which are required. With the logarithmically distributed range more precision is obtained at the lower ranges without affecting the limits of classification.

- The duration parameter (d) in decision making:

The integer-rounded values logarithmically distributed over the maximum range were searched.

5.4.2.2 COST FUNCTION

Fully automated machine learning cannot replace allergists who conduct the OFCs. This is due to the fact that the allergists are required to administer the doses of the problem food to the subjects throughout the challenge, and, should allergy occur, they will be required to administer antihistamines. Therefore, the classification routine which is discussed here is designed as a diagnostic assistance tool, and should complement the diagnosis of allergy in a way that improves diagnoses. The cost function which is defined was designed in collaboration with the clinicians in order to best suit their diagnostic needs.

The consequence of false positive classifications would yield unacceptable effects on the quality of life of the subjects (Chapter 1). With this consideration, the parameters found within the search space which achieve 100% *specificity* in the nested cross validation (i.e. models which correctly classify all non-allergic training subjects) are initially selected. This ensures that the parameters were selected based on obtaining no false positive classification in training. From this reduced set, the parameters which lead to the highest *sensitivity* are selected. If there are more than a single set of parameters that satisfy these conditions, the parameters which achieve the maximum *total time gain* are used. Sensitivity, specificity and time gain are defined in subsequent sections.

The search space was searched in a targeted, but exhaustive manner. This was because only a small portion of the decision making parameters satisfy the criteria mentioned above. Therefore, it was possible to avoid computing the majority of the parameters in the exhaustive enquiry. Other parameter selection routines, such as receiver operating curves, would require the entire search space to be investigated. However, with consideration to the previously defined cost function it is possible to reduce the computation significantly and target the exhaustive search to only the range of values that satisfy the decision making criteria.

		Classification	
		p	n
Diagnosis	p	True positive	False negative
	n	False positive	True negative

Figure 5.7: The confusion matrix showing how sensitivity and specificity are obtained with regard to the ground truth (diagnosis) and predicted (classification) results.

5.5 CLASSIFICATION METRICS

5.5.1 SENSITIVITY/SPECIFICITY

The measurements of sensitivity (Se) and specificity (Sp) were computed in order to measure the overall accuracy of the allergy detection framework. These are commonly referred to as the true positive and true negative rates respectively. They are both bounded between 0 and 100%, and higher values indicated better accuracy. Both sensitivity and specificity measure classification accuracy against the ground truth diagnosis and are defined by

$$Se = \frac{TP}{TP + FN} \times 100\%, \quad (5.10)$$

$$Sp = \frac{TN}{TN + FP} \times 100\%, \quad (5.11)$$

where true positive (TP) are the number of allergic subjects who were classified allergic, false negative (FN) are the number allergic subjects who were misclassified as non-allergic, true negative (TN) are the number non-allergic subjects classified as non-allergic, and false positive (FP) are non-allergic subjects who were classified as allergic. A confusion matrix can be employed to visualise the means by which these metrics are computed, and is shown in Figure 5.7. In this figure, p is a positive result (i.e. allergic) and n is a negative result (i.e. non-allergic). Sensitivity is an assessment of the two upper quadrants while specificity is represented in the lower two quadrants.

5.5.2 TIME GAIN PARAMETERS

Other metrics which provides an insight into the algorithmic performance are related to the time gain. Three time gain metrics are calculated for the allergic subset of subjects only, as no time gain can be obtained for the non-allergic subjects.

5.5.2.1 TIME GAIN

Time gain measures the difference in time between termination of the OFC by the allergists and classification of allergy. In effect it demonstrates whether it is possible to conclude OFC earlier and reduce the overall risk of anaphylaxis and other strong reactions due to early administration of antihistamines.

The average time gain factor is reported in two ways: first as the sum of time gains divided by the number of allergic subjects (always 15 in this study), and second as the total time gain divided by the number of subjects whose allergy was detected by the framework. These two metrics will be equal if perfect sensitivity is obtained, and are termed total time gain (TGT) and specific time gain (TGS) respectively. The remainder of the time gain parameters have associated total and specific results.

The reason for calculating these two time gain metrics is because this allergy detection platform is envisioned as an diagnosis assistance tool. Therefore it is appropriate to quantify its effectiveness with the entire set of allergic subjects, but also its effectiveness on the subset of correctly classified subjects is also of interest.

5.5.2.2 DOSES SAVED

The time gain factor can be converted to another metric which measures the number of doses of the allergen which would not need to be administered if the allergy classification was introduced and the OFC was halted when automatic detection diagnosed allergy. This metric is termed doses saved henceforth.

This doses saved metric is a measure of the risk reduction that can be achieved when allergy detection is employed, as with fewer doses administered there is a smaller likelihood of allergic reactions presenting.

5.5.2.3 ACTIVATION PERCENTAGE

An additional time gain metric is also calculated which determines the percentage of the allergen which was required to be consumed for abnormal HRV-features to be detected by the classification routine. This metric is related to the doses saved metric in a non-linear manner, and, therefore, also gauges the risk reduction that is gained by automatic classification.

For example, if a subject reacts after consuming five doses of an allergen, but the novelty detection framework detects allergy after three doses, it can be stated retrospectively that only 22.5% of the consumed allergen was required to induce signatures of allergy on the HRV features. This percentage is termed the activation percentage. With N total doses administered to a subject, and s doses saved by the classification, the activation percentage is calculated by

$$\text{Activation percentage} = \frac{\sum_{i=1}^{N-s} 2^{i-1}}{\sum_{i=1}^N 2^{i-1}} \times 100\%, \quad (5.12)$$

which, for the previously stated example results in

$$\begin{aligned} \text{Activation percentage} &= \frac{1 + 2 + 4}{1 + 2 + 4 + 8 + 16} \times 100\% \\ &= \frac{7}{31} \times 100\% \\ &= 22.5\%. \end{aligned}$$

5.6 RESULTS

5.6.1 A BRIEF NOTE ON THE STRUCTURE OF THESE RESULTS

This section presents the results which were obtained with an epoch length of 60 seconds. While epoch lengths of 120, 180, and 300 seconds were also investigated, the results which were obtained at an epoch length of 60 seconds will first be demonstrated. Later, results will be presented which show what occurs at the different epoch lengths, before finally discussing a more optimal classification routine which is employed for the remainder of this thesis. The purpose of separating the results is to first introduce the methodology of performance assessment before finally discussing a more optimal routine.

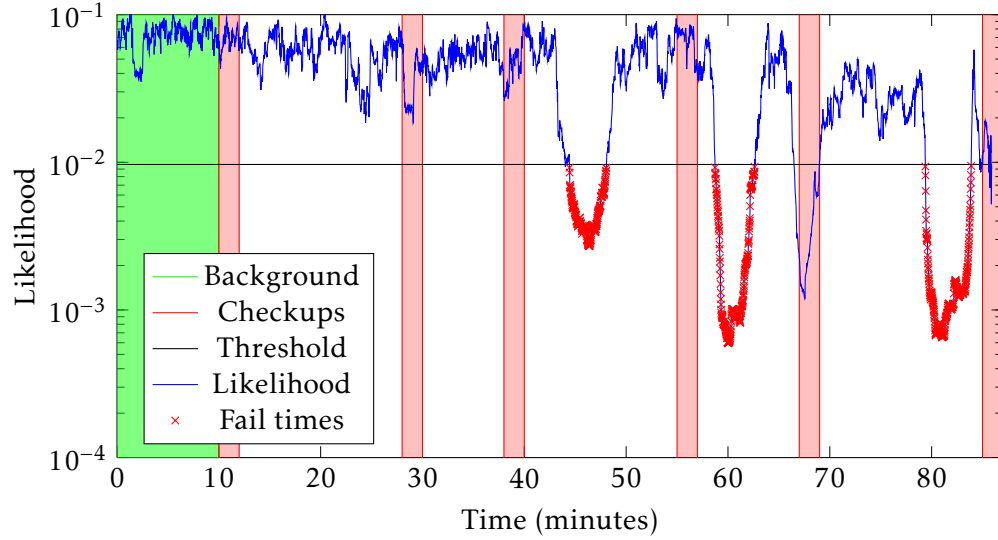


Figure 5.8: A demonstration of early detection of allergy (with Subject 11). The segments at 45, 60 and 80 minutes which fall beneath the threshold were classified as allergy.

5.6.2 RESULTS OBTAINED AT EPOCH LENGTH OF 60 SECONDS

The overall results which were obtained at an epoch length of 60 seconds are presented in Table 5.1. In this table, the first 15 subjects were diagnosed as allergic, and the final 9 subjects were diagnosed as non-allergic. The diagnosis column shows the diagnosis of allergy for the subject, and 1 indicates allergic, and 0 indicates non-allergic. The prediction column presents the results which were obtained by the classification routine, and 1 indicates a classification of allergic and 0 indicates a non-allergic classification. The time gain column presents the time gain results which were obtained. The cells which are marked as ‘—’ did not achieve any time gain. The time gain metric is only applicable for subjects who were diagnosed allergic.

It can be seen from Table 5.1 that all of the non-allergic subjects were classified as non-allergic obtaining 100% specificity. Subjects 1, 3, and 13 were not classified as allergic, however, and this yields sensitivity of 80%.

Figure 5.8 shows how allergy was detected before the OFC was terminated. In this Figure the shaded regions represent the background and checkup times. The signal trace is the

Table 5.1: Classification results obtained with the novelty detection routine at epoch length of 60 seconds.

Subject ID	Diagnosis	Classification	Time gain
1	1	0	—
2	1	1	35.0
3	1	0	—
4	1	1	16.0
5	1	1	34.0
6	1	1	12.0
7	1	1	29.0
8	1	1	12.0
9	1	1	25.0
10	1	1	5.0
11	1	1	40.0
12	1	1	19.0
13	1	0	—
14	1	1	94.0
15	1	1	13.0
16	0	0	—
17	0	0	—
18	0	0	—
19	0	0	—
20	0	0	—
21	0	0	—
22	0	0	—
23	0	0	—
24	0	0	—

likelihood which was calculated and the segments of the likelihood which satisfied the allergic criteria are indicated with $\times$ markers. The likelihoods obtained during checkup times are not considered as allergic. The start of the final checkup period is when the allergist concluded the OFC.

The challenge shown in Figure 5.8 was concluded by the allergist at approximately the 85th minute when symptoms of allergy manifested. It can be seen from the example in Figure 5.8 that the subject is classified allergic by the system developed here approximately 40 minutes sooner than the challenge was concluded by the allergist.

When considering the time gain metrics for the entire database, this achieved TGT of 22.26 minutes, TGS of 27.83 minutes, two of five doses were saved which yields an activation percentage of 22.5%.

5.6.3 OVERALL RESULTS

For all epoch lengths, 100% specificity was obtained. Sensitivities of 80% were obtained at epoch lengths of 60 and 180 seconds, and 73% sensitivity was obtained with epoch lengths of 120 and 300 seconds. It should be noted here that every subject is classified by different modelling parameters and this is due to the LOO procedure which was incorporated. It should also be noted that different modelling and decision making parameters are utilised by the different epoch lengths for the same reason.

5.6.4 INCONSISTENT CLASSIFICATION AT DIFFERENT EPOCH LENGTHS

Subjects 2, 3, 13, 14 and 15 were not consistently classified for different epoch lengths.

5.6.4.1 SHORT-DURATION SIGNATURES OF ALLERGY

Figures 5.9a — 5.9d demonstrate this inconsistency with Subject 2. The highlighted regions in these figures represent the background and checkup periods, and the solid trace is the likelihood which was calculated over the OFC. Satisfaction of the allergy criteria occurred once in these figures and is marked with an arrow in Figure 5.9a at approximately 65 minutes. It can be seen here that the departure which classified Subject 2 as allergic in Figure 5.9a is reduced in significance with longer epochs, until in Figures 5.9c and 5.9d, the anomaly is indistinguishable from the likelihood trace.

With the longer epochs, the extent of this departure is averaged with ‘regular’ features, and as such the likelihoods at these times are less pronounced which reduces the extent

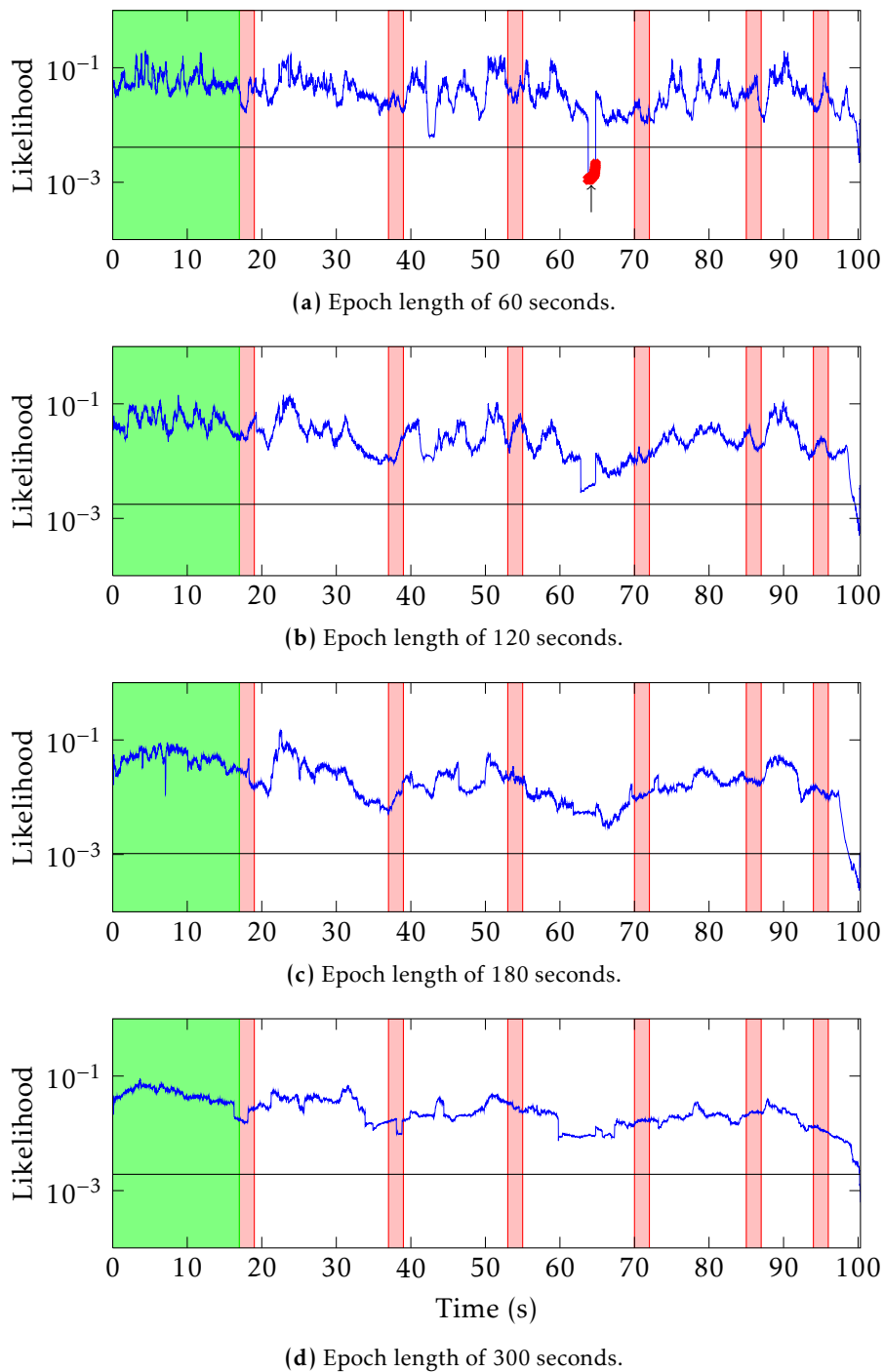


Figure 5.9: Example demonstrating how the generated likelihood for Subject 2 satisfies the allergy criteria at an epoch length of 60 seconds (subplot a) while failing to do so for epoch lengths of 120, 180 and 300 seconds (subplots b — d).

of the departure at these feature lengths. In effect longer epochs will act as a higher-order moving average filter which reduce the significance of these deviations with background HRV metrics, while the shorter epochs allow for classification of the segment due to higher relative importance of novel HRV features.

5.6.4.2 LONGER SIGNATURES OF ALLERGY

The opposite of this phenomenon can also occur where longer feature lengths enhance deviations from the background. Figure 5.10 shows how Subject 13 is misclassified as non-allergic at epoch lengths of 60, 120 and 180 seconds. However, it can be seen that with the longer epochs certain departures from the background likelihood become more pronounced (e.g. at approximately 70 minutes). With an epoch length of 300 seconds Subject 13 is correctly classified as allergic at approximately 35 minutes, which is approximately 70 minutes before the challenge was terminated. This is not an isolated departure from the background levels either, and at 70, 90 and 100 minutes allergy is also detected in the likelihood trace.

5.6.4.3 TOLERANCE TO NON-ALLERGIC VARIANCES

The likelihood trace in Figures 5.9 and 5.10 all surpass the threshold at some instance in time, e.g. at the very end of every trace with Subject 2, and at approximately 90 minutes with Subject 13. Yet, it is only at certain epoch lengths that the allergic criteria are satisfied. This is due to the inclusion of the duration parameter in decision making. Without this parameter, a lower threshold (i.e. a larger departure from the background) would be required to classify allergy in order to reject deviations from the non-allergic subjects.

Figure 5.11 shows the likelihood trace for Subject 16, who was diagnosed non-allergic and also classified as non-allergic at all epoch lengths. In this Figure it can be seen that with longer epoch lengths the baseline of the likelihoods is nearly continuously departed from the background levels of the first checkup. At all epoch lengths (in particular at 50

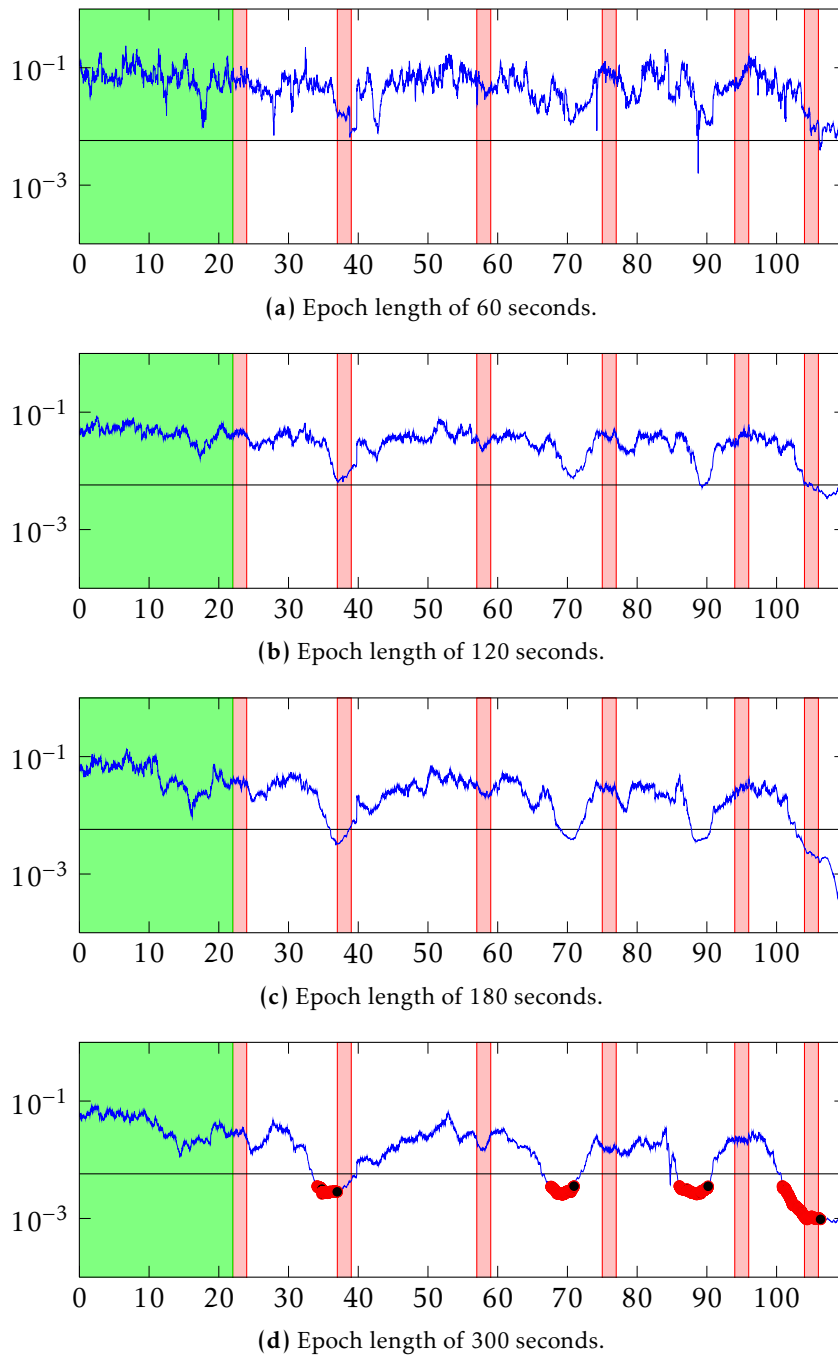


Figure 5.10: Example demonstrating how the generated likelihood for Subject 13 does not satisfy the allergy criteria at an epoch length of 60, 120 and 180 seconds (subplot a — c) but the criteria are then met for the epoch length of and 300 seconds (subplots d).

minutes with an epoch length of 120 seconds in Figure 5.11b) the likelihood surpasses the threshold at multiple occasions. However, with the inclusion of the duration parameter, none of these points are flagged as allergic, and the subject is correctly classified as non-allergic by the classification routine.

It can be seen in Figure 5.11 that with longer epoch lengths, the baseline of the likelihood traces becomes less similar to the background recordings. This is because Subject 16 became agitated during the OFC. This agitation was fueled by the periodic checkups, and this can be seen directly after the administration of the first sub-portion at approximately 10 minutes. This agitation was noted when the ECG was recorded during the OFC. This agitation demanded a number of extra checkups during the challenge as the allergists were required to verify that the agitation was not as a result of allergy.

It is believed that the cause for the agitation was psychological: the subjects who are tested for allergy in this manner have been repeatedly told before the OFC that consuming the allergen may cause them to become sick. Subject 16, for example, was tested against egg which was in the form of a cake. Cakes are instantly recognisable to a six-year-old child, and repeated subjection to the food they live in fear of (Chapter 1) resulted in the subject being continuously agitated which resulted in almost all of the likelihoods calculated during the OFC being departed from 'normal' background. Even though the subject did not react to the food, there is an underlying fear that they might, and this fear is not inhibited by the fact that the challenge is closely monitored by allergists. Without the inclusion of the duration parameter in the classification routine, these non-allergic subjects would increase the value of the multiplicative parameter of the threshold equation in order to satisfy the 100% specificity criteria in training. This would consequently require a larger deviation from the background for allergy classification and might not facilitate as high sensitivities as were obtained.

Figure 5.11a may appear to display a substantiated departure from the background recording, so it is interesting to note that even with this variance, the classification routine did not once classify allergy for this subject, but in all cases correctly classified them

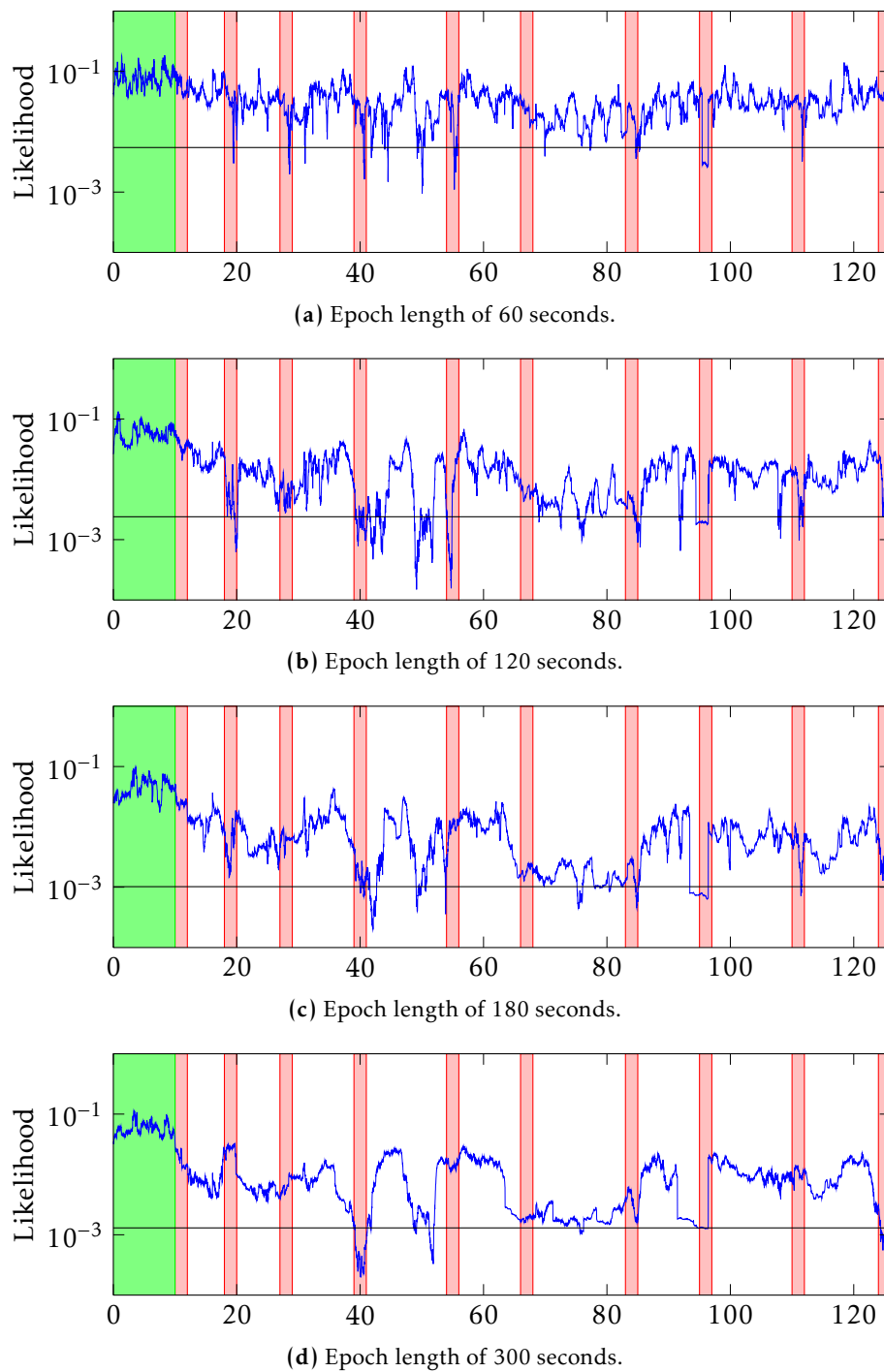


Figure 5.11: Example demonstrating how the generated likelihood for Subject 16 surpasses the threshold, but does not satisfy the allergy criteria due to the inclusion of the duration parameter and for all epoch lengths the subject is correctly classified as non-allergic.

as non-allergic. Even with the presence of ‘agitation’, it is demonstrated that all non-allergic subjects can be classified correctly, which allows justification to the claim that the allergy detection platform demonstrated here is robust in classification, and robust in discrimination between allergy and allergy-like HRV features.

Subject 1 was not classified correctly at any epoch length. This is because this challenge was concluded very abruptly three minutes after the challenge began and because a signature of allergy which was detectable by the means discussed here did not present in the ECG of this subject. Therefore, it is believed that perfect sensitivity is not possible to achieve perfect sensitivity on the ECG of the allergy database.

5.6.5 BOOSTED ALLERGY CLASSIFICATION

Here, ensemble-based result fusion is performed which will be shown to obtain more optimal classification performance.

5.6.5.1 SENSITIVITY/SPECIFICITY

Based on the inconsistent classification results which were obtained, it appears that the detection of abnormal signatures on the HRV is a dynamic process which is not suited to one particular epoch length. Therefore, to obtain more optimal results, the remainder of this chapter and thesis will employ classifier fusion based logically OR-ing the results which were obtained for the individual epoch lengths. This is employed in EEG classification applications where individual channels are OR-ed together. The OR-ing process is justified because it does not violate the subject independent nature of model and post-processing parameter selection.

Table 5.2 shows the tabulation of allergy classification with regard to the epoch length. In this table, only Subjects 1 — 15 (i.e. the allergic subjects) are tabulated. Table

Table 5.2: Tabulation of the classification results of the allergic subjects where ‘1’ represents an allergic classification (TP) whereas ‘0’ represents a non-allergic classification (FN).

Subject ID	Epoch length				Logical OR
	60	120	180	300	
1	0	0	0	0	0
2	1	0	0	0	1
3	0	0	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1
7	1	1	1	1	1
8	1	1	1	1	1
9	1	1	1	1	1
10	1	1	1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	0	1	1	1	1
14	1	0	0	0	1
15	1	1	1	0	1
<i>Sensitivity</i>	80.00%	73.33%	80.00%	73.33%	93.33%
<i>Specificity</i>	100.0%	100.0%	100.0%	100.0%	100.0%

elements signified by ‘1’ indicate that the subject been correctly classified as allergic and ‘0’ represents false negative classification.

The fifth classification result column which is emphasised by the emboldened characters show the results which are obtained by logically OR-ing the classification results obtained at each of the epoch lengths. In the cases of the individual epoch lengths the highest sensitivity obtained was 80% (12/15 correct classification) which were obtained for 60 and 180 seconds, but by considering the logical OR-ing process, 93.33% sensitivity (14/15 correct classifications) is obtained, and only one false-negative classification was achieved.

Table 5.3: Classification result, time gain, doses saved and activation percentages obtained by the classification routine. The results in this table were obtained by fusing the results obtained for the individual epoch lengths together.

Subject ID	Classification	Time gain	Doses saved	Activation percentage
1	0	—	—	—
2	1	35	2	22.58
3	1	55	3	9.68
4	1	17	0	100.0
5	1	39	2	22.58
6	1	20	0	100.0
7	1	30	1	42.86
8	1	48	2	22.58
9	1	27	1	33.33
10	1	32	2	14.29
11	1	42	2	22.58
12	1	22	1	33.33
13	1	73	4	3.23
14	1	94	4	3.23
15	1	14	1	48.39
μ	93.33%	36.53	1.66	38.58
σ	—	23.91	1.30	34.26
μ	93.33%	39.15	1.78	34.18
σ	—	22.50	1.25	30.88

5.6.5.2 TIME GAIN PARAMETERS

Table 5.3 presents the set of time gain results which were obtained by the automatic classification of allergy when employing classifier fusion. As no false positive classifications were obtained here, this table only presents the allergic subjects. This table presents the time gain, doses saved and activation percentage values which were obtained.

5.6.5.2.1 TGT and TGS Column three of Table 5.3 tabulates the set of time gain results which were obtained for the allergic subjects. In this column the ‘—’ symbol is used to mark when the classification routine failed to classify allergy. The mean and

standard deviation of the total and specific time gains (TGT and TGS respectively) are also presented.

The logical OR-ing classification process benefits from selecting the best time gain metrics from the set of time gains which were for every epoch length. Subjects 4 — 12 are classified as allergic at all epoch lengths, and of these, Subjects 4, 5, 7, 9, 11 and 12 allergy were classified at approximately the same time in every instance. However, with Subjects 6, 8 and 10 the OR-ing process adds an additional 60 minutes time gain in comparison to the results obtained from the individual epoch length cases, which contribute overall time gain metrics which are approximately 60 minutes greater than the largest individual epoch length based procedure. For example, between 60 second epoch results and the merged results, both TGT and TGS increased by approximately 12 minutes. The significance of obtaining high time gain metrics is that emergency rescue medication could be quickly administered which could introduce the possibility of reaction-free OFCs for subjects who would suffer from a reaction with the current OFC.

5.6.5.2.2 Portions saved Column four in Table 5.3 presents the number of portions saved which would have been obtained had allergy classification been employed during the OFC recordings.

Approximately 1.8 portions are saved when allergy is classified. In three cases, however, with Subjects 1, 4 and 6, no portions were saved. With Subjects 4 and 6 allergy was classified, but owing to the fact that no additional portions of the allergens were administered between the classification of allergy and the diagnosis of allergy, the full amount of the food was required for the detection.

5.6.5.2.3 Activation percentage The final time gain parameter which is calculated is the activation percentage, and this is shown in the fifth column of Table 5.3. The activation percentage is the percentage of the allergen which was required to be consumed before signatures of allergy in HRV features were detected by the allergy classification

framework. In Table 5.3, values of 100% indicate that the entire dose of the allergen was required.

For Subjects 1, 4 and 6, as no doses were saved from consumption, the activation percentage of 100% was achieved. However, in the case of Subjects 4 and 6, time gains of 17 and 20 minutes were still obtained.

Overall, approximately 40% of the doses administered to the subjects are required for allergic classification based on the HRV features. This figure is reduced to approximately 30% when considering only the subjects who were correctly classified allergic. This value indicates that when machine-based allergy classification is achieved, consumption of less than one third of the dose required for a diagnosis of allergy is required for classification of allergy (or with a 70% reduction in exposure to the problem food).

5.7 DISCUSSION

A close inspection of several subjects is required to provide additional insight into the robust nature of the allergy classification system's behaviour.

5.7.1 SPECIFICITY OF OFC CLASSIFICATION

The importance of obtaining very high specificity was discussed previously. The significance of mis-classifying non-allergic subjects would have a negative effect on their quality of life indefinitely (Sicherer et al., 2006; Cox et al., 2008). This would be an unacceptable consequence, so parameters were selected to not misclassify allergy on the training data. This characteristic was preserved with the unseen testing subjects. It is very significant that this was obtained as the testing data remains unavailable during parameter selection (see Figure 5.6). Indeed, because perfect specificity was obtained in this section, it can be

stated that the classification of allergy is equivalent to the diagnosis of allergy, due to the subject-independent nature of the means in which the parameters were selected.

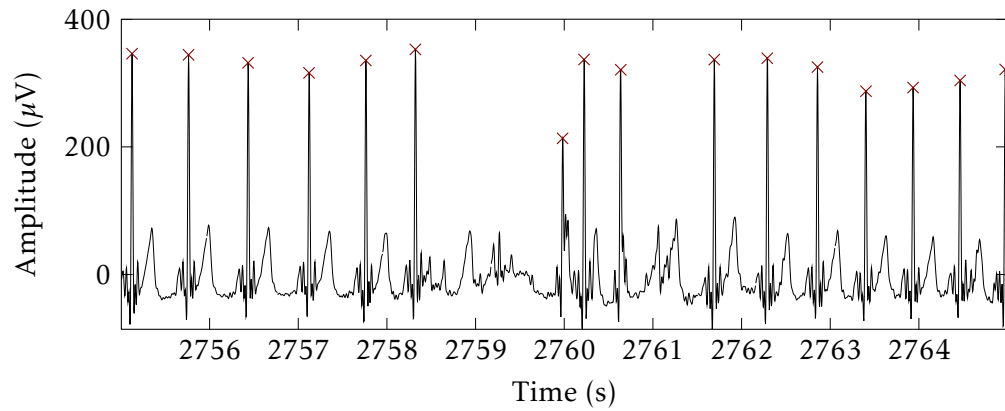
5.7.2 ROBUST CLASSIFICATION

The heart rate of Subject 20 presented with frequent isolated arrhythmia events. This is a condition where the heart rate changes from its resting rate to a much higher value, the heart rate will then relax to its resting value (Clarke et al., 1976). Figure 5.12 shows an example of the arrhythmia for this subject. Figure 5.12a shows the raw ECG signal, with the QRS complexes identified with $\times$, while Figure 5.12b depicts the beat-to-beat heart rate which was derived from the QRS points.

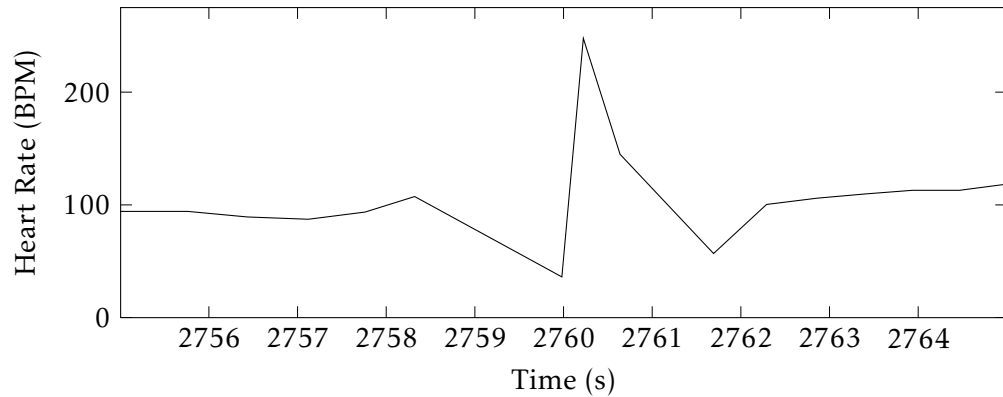
The resting heart rate can be seen to be approximately 100 BPM before the arrhythmia incidents at 2,758 — 2,760 seconds. The heart rate then reduces by over half to approximately 34 BPM at 2,760 seconds as no QRS complexes have occurred, before rising to approximately 250 BPM soon after. The arrhythmia incidents occur between 2,758 and 2,762 seconds. Subject 20 experienced these arrhythmia events in a number of occasions during their food challenge. However, even in light of these abnormal heart beats — which occurred in some instances over five times per minute — Subject 20 was correctly classified as non-allergic at all epoch lengths.

It was previously discussed how Subject 16 presented with unusual HRV features which were due to agitation that the subject felt throughout the challenge. Yet, this subject was also correctly classified as non-allergic by the classification routine presented here at all epoch lengths.

The fact that these two non-allergic subjects who presented with abnormal HRV features were both correctly classified as non-allergic in all routines investigated shows that the classification routine presented in this chapter is robust in discrimination between allergic and non-allergic subjects. Indeed, the system is robust in discrimination between allergic



(a) Raw ECG trace presenting with arrhythmia between seconds 2758 and 2762.



(b) The beat-to-beat heart rate calculated and the effect of arrhythmia beats on this.

Figure 5.12: Example of arrhythmia on the ECG trace (a) and the effect this has on the heart rate (b) on Subject 20.

and allergic-like signatures (i.e. HRV signatures which are affected by agitation and arrhythmia).

It is very common to pre-process extracted QRS points and remove arrhythmia and ectopic beats with HRV. This was not performed here for a number of reasons, principally because later chapters will employ classification routines which discover QRS points automatically. It is possible for non-QRS points to be mis-labelled as QRS points by these algorithms under the influence of artefacts. Therefore, by allowing arrhythmia and ectopic beats in this training set, the developed models should be more tolerant to artefacts, see Chapters 6 and 7.

5.7.3 PARAMETER SELECTION

Many factors can influence parameter selection. In the attempt to achieve the optimal results, post-processing parameters were selected to achieve 100% specificity and the maximum sensitivity on the training data. Yet, the values chosen will be in a region where there is the possibility that mis-classifications might occur. This can be seen in Figure 5.11 where individual cases of the likelihood surpass the threshold at regular intervals, and it is with the selection of appropriate duration parameters that the previously stated robustness is obtained. This is evidenced by the fact that perfect specificity was obtained at all epoch lengths and that very high sensitivity was also obtained.

5.7.3.1 IMPORTANCE OF CORRECT PARAMETER SELECTION

Subjects 1, 2 and 3 presented with HRV features which did not vary greatly due to their allergic reactions, and it is because of this fact, the selection of PCA and GMM parameters is crucial for appropriate emphasis of novel regions. Figure 5.13 shows the likelihood of Subject 2 where the PCA and GMM parameters were selected manually to illustrate how a poor choice of modelling parameters can affect the range of the computed likelihoods. The extent of the anomaly which classified the subject as allergic at the 65th

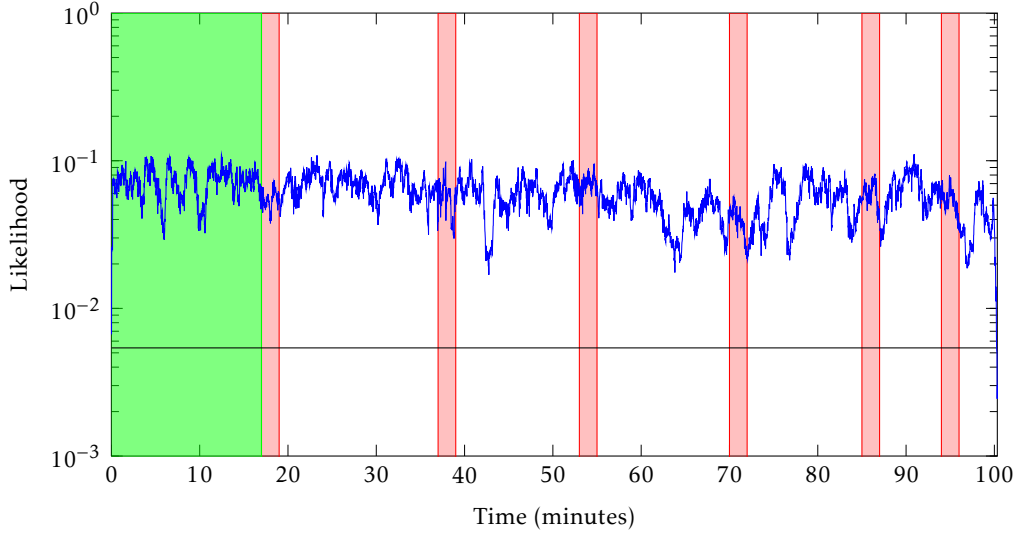


Figure 5.13: The likelihood series chosen for Subject 2 which does not diverge from the background level significantly enough to classify allergy. PCA preserved 80% of the feature variance which was modelled with a GMM order of 32 at an epoch length of 60 seconds.

minute in Figure 5.9a was not equalled in Figure 5.13. Indeed, the deviation shown in this figure is no more varied than the likelihoods obtained in the background.

This example illustrates the importance of parameters selection, and expresses the fact that allergy is non-trivial to detect through statistical HRV feature analysis while also demonstrating the data-driven nature of the classification procedure.

5.7.3.2 ALTERNATIVE PARAMETER SELECTION

Appendix A presents a study in which alternative parameter selection routines to those employed in this chapter were investigated. It shows that while the means by which parameters are selected is important, alternative methods can be utilised which will also achieve 93% sensitivity and 100% specificity. However, the selection routine in this chapter is time gain aware (i.e. parameter selection selects the parameters which obtained the best time gain results in training data) and those are discussed in Appendix A were not. This is reflected in the results which were obtained, and in all cases the parameter

selection routine described in this chapter achieved superior time gain, doses saved and activation percentage results. This provides evidence that the parameter selection and cost function routines which were discussed in this chapter are very well suited for the classification problem.

5.7.4 ROLE OF CLASSIFICATION IN OFCs

The classification routine presented here is well suited as a diagnostic assistance tool. This is because the allergist can never be replaced, as remote monitoring of physiological signals cannot administer allergens or antihistamines when required. Therefore, the allergist will always be present during OFC, and if this classification system is to be used in conjunction with the standard OFC, machine learning algorithms can greatly assist.

Excellent time gain metrics were obtained, and on average, approximately 35 minutes would have been saved by this classification routine. This time could be employed to good advantage for the administration of antihistamines which could reduce — and possibly eliminate! — allergic symptoms and reactions in some cases.

For machine-based allergy classification, false negative classifications (i.e. classifying an allergic subject as non-allergic) can be tolerated, as these challenges reduce to the standard OFC, which is the current state of clinical art. The classification routine here should complement the diagnosis of allergy and should be used alongside the allergists. With having obtained 100% specificity in a subject-independent manner, machine-based allergy detection can introduce many improvements to the clinical diagnosis of allergy.

It is worth noting that allergists diagnose allergy in these patients by having access to more subjective signals and information sources such as mood and temper which were not available for this automated analysis. These are not employed for machine learning purposes because it is difficult to monitor these objectively and in a non-invasive manner. Non-invasive monitoring is important for allergy detection, as the introduction

of discomfort might aggravate the subjects, and may increase the number of non-allergic subjects who presented with likelihood traces similar to Subject 16.

Clinically other signals, such as the blood pressure, blood oxygen saturation and temperature, will be the last physiological metrics which change as a result of allergy. As the goal of this research is to detect allergy as early as possible, only the ECG is recorded. However, the system developed here presents an additional insight into the temporal nature of subjects' states of allergy during OFCs, and even without the extra data which the allergists have access to, excellent results were obtained.

CHAPTER 6

AUTOMATIC QRS DETECTION

6.1 INTRODUCTION

THE features which were extracted for classification in Chapter 5 were extracted from QRS points which were manually annotated. These were used in order to definitively assess whether it was possible to identify allergic subjects through analysis of HRV features which has not been done before. It was discovered that allergy affects HRV in a manner which allows classification, and therefore automated QRS detection is assessed here before fully automated allergy classification is assessed in Chapter 7.

6.2 QRS DETECTION

Software-based QRS detection has been an important research topic for more than 40 years (Kohler et al., 2002). At the core of QRS detection are algorithms which were designed to

specifically enhance the QRS complex of the ECG while diminishing the levels of artefact and other features of the heart beat (i.e. P- and T-waves).

All QRS detection algorithms require a thresholding stage, and the underlying algorithms which are employed cover a wide range of disciplines of DSP. The rich set of algorithms which can be employed reflects how technological ability has evolved over the years in which QRS detection has been researched, and popular methods can involve derivative processing (Okada, 1979; Fraden and Neuman, 1980; Ahlstrom and Tompkins, 1983; Arzeno et al., 2008); digital filter banks and wavelets (Gyaw and Ray, 1994; Di Virgilio et al., 1995; Bahoura et al., 1997; Afonso et al., 1999; Chen et al., 2006; Strang and Nguyen, 1996); template matching (Dobbs et al., 1984); adaptive filtering architectures (Kyrkos et al., 1987; Hamilton and Tompkins, 1988; Thakor and Zhu, 1991); artificial neural networks (Xue et al., 1992; Vijaya et al., 1998; Rajendra Acharya et al., 2003); transformation methods (Bolton and Westphal, 1981a, 1984; Benitez et al., 2000, 2001), and new methods are always being investigated.

The reason that many different areas of signal processing have been employed for QRS detection is because it is very difficult to generalise one algorithm towards the complete set of ECG waveform shapes. The extent of algorithms which exist is also indicative of the fact that it is not appropriate to cater towards 'well behaved' ECG only, because the shape of the ECG and the resulting HRV features can be affected by age (Nunan et al., 2010; Aziz et al., 2012), gender (Antelmi et al., 2004), heart disease (Rajendra Acharya et al., 2006), physical fitness (Hamer and Steptoe, 2007) and even by coffee consumption (Monda et al., 2009). These differences in ECG-shapes can result in missed or extra QRS points which each have the effect of inaccurately reporting the HRV features.

6.2.1 QRS VALIDATION

Of the previously mentioned algorithms, each has their own advantages, but in every case it is necessary to validate the QRS points with hard limits which generally cannot

be exceeded, i.e. the maximum heart rate that can be obtained is age-dependent and is only surpassed in very exceptional circumstances. For this reason, routines have been developed which validate QRS points which are returned by QRS detection algorithms. A set of rules were defined by Hamilton (2002) and are reproduced here:

- Ignore all peaks that precede or follow larger peaks by less than 200 ms.
- If a peak occurs, check to see whether the raw signal contained both positive and negative slopes. If not, the peak represents a baseline shift.
- If the peak occurred within 360 ms of a previous detection check to see if the maximum derivative in the raw signal was at least half the maximum derivative of the previous detection. If not, the peak is assumed to be a T-wave.
- If the peak is larger than the detection threshold call it a QRS complex, otherwise call it noise.
- If no QRS has been detected within 1.5 RR intervals, there was a peak that was larger than half the detection threshold, and the peak followed the preceding detection by at least 360 ms, classify that peak as a QRS complex.

These rules provide the ability to increase the accuracy of the QRS detection which was performed.

6.2.2 VALIDATION DATABASES

Typically, QRS detection algorithms will be validated on databases which provide an abundance of healthy and unhealthy ECG shapes. This strategy allows for objective comparisons between QRS detection algorithms with regard to detection accuracy and computational complexity on challenging data. If QRS detection was assessed on data recorded from healthy volunteers only, quantification of arrhythmias and other

cardiovascular defects would not be addressed, and accuracy would be very subjectively reported.

One of the more popular databases on which QRS detection is assessed is the Massachusetts Institute of Technology Beth Israel Hospital (MIT-BIH) database (Mark et al., 1982; Goldberger et al., 2000; Moody and Mark, 2001). This database consists of 48 half-hour, two-channel ambulatory ECG recordings, which were recorded between 1975 and 1979. Twenty-three of the recordings were manually selected at random from a set of 4,000 in- and out-patients. The remainder were selected from the same set to include clinically significant arrhythmias which were not well represented by the initial selection. Twice-validated expert annotations accompany this database.

In this chapter, QRS detection results will be discussed from the MIT-BIH and allergy databases to verify the assessment of QRS detection. When discussing the MIT-BIH arrhythmia database, individual cases are labelled as **patient** records, while when discussing the allergy database individual cases are labelled as **subjects**. From the medical perspective, this is because participants of the allergy database were not admitted as patients, whereas those of the MIT-BIH database were. This naming convention is also employed to facilitate simple distinction between the two databases without the need to reference the name of the database in question, as references to **patients** refer to persons from the MIT-BIH database, and references to **subjects** refer to persons from the allergy database.

6.2.3 SENSITIVITY AND POSITIVE PREDICTIVITY

The sensitivity of QRS detection measures true positive rate of QRS detection and is defined by

$$Se = \frac{TP}{TP + FN} \times 100\%, \quad (6.1)$$

where, TP and FN are as described previously.

The specificity, which was employed in the previous chapter to measure the true negative classification rate, is an inappropriate measurement for QRS detection, and therefore the precision or positive predictivity (+P) of the automatically extracted points QRS points is computed by

$$+P = \frac{TP}{TP + FP} \times 100\%, \quad (6.2)$$

where FP is also as was described previously.

6.2.4 GOOD DETECTION WINDOW

In some cases the location of the R-wave can be ambiguous, in particular if the patient suffers from cardiovascular disease or arrhythmia. In these cases, automatic QRS detection algorithms might identify the heart beat, but the point might not be localised on the apex of the R-wave. Friesen et al. (1990) and Ganong and Ganong (2005), state that the duration of the QRS complex is approximately 88 ms. Therefore, in this work, if a QRS complex was found to be within this range of the annotated point, it is flagged as a true positive, as the QRS complex has been identified. However, if the identified point is found to be outside of this range of annotated points, the candidate QRS point is flagged as a false positive as the QRS complex was not identified.

6.2.5 FEATURE ACCURACY

It will be seen in later sections that under certain conditions the accuracy of QRS detection is pessimistically reported. The cause for this is that the QRS points reported from the

detection algorithms report times which are outside of the 88 ms ‘good detection window’. These QRS points are therefore reported as false positives by the algorithm. However, while the QRS points which are reported are not within the allowed QRS window, beats are periodically detected, but were localised incorrectly on the S- and T-waves. In order to assess the effect of this, HRV-based feature metrics are computed and these are compared to the same HRV features which were obtained from the manual QRS annotations.

It will be shown that when incorrect QRS localisation occurs, QRS beats are consistently localised at the same part in the wave, i.e. if the QRS is first located on the T-wave, it will generally be located on the T-wave for that recording. Insight into the effect of poor localisation can be ascertained by computing the normalised differences between the HRV features calculated from the manual and automatic sets of QRS points. The difference is computed with the PRD metric which was employed in Chapter 3, and the equation is reproduced below, and has been modified for HRV case.

$$\text{PRD} = \sqrt{\frac{1}{N} \left(\sum_{n=1}^N \left(\frac{f_m(n) - f_a(n)}{f_a(n)} \right)^2 \right)} \times 100\%, \quad (6.3)$$

where f_m is feature vector obtained from the manual annotations, f_a is the vector obtained by automated data, and both of these features are of the same length, N .

6.2.6 BOX-PLOTS

Box-plots are employed in order to graphically present a number of statistical parameters of a distribution. Figure 6.1 shows an example box plot (upper) and its relationship to a normal distribution (lower). Five points are presented with the box plot: the median, the lower and upper quartiles, and the largest and smallest values with 1.5 quartile ranges of the distribution.

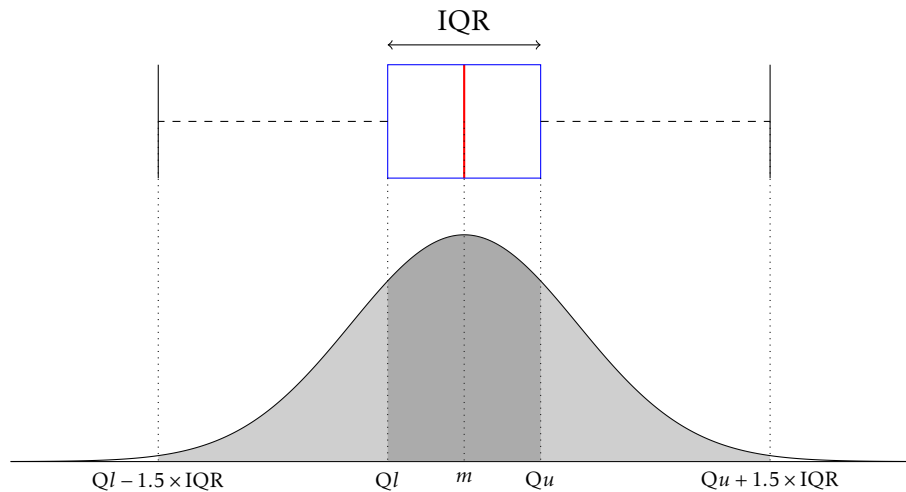


Figure 6.1: Relationship between a box-plot, and quartile ranges with a normal distribution. The locations marked Ql and Qu are the lower and upper quartiles respectively, and the median is marked as m .

The lower- and upper-quartile ranges are shown with the horizontal boundaries of the rectangle in Figure 6.1, and this is termed the inter-quartile range (IQR). The vertical line found within the rectangle is the median of the distribution (which can also be termed the 50th percentile) and is assigned the label m in this figure. In this plot the median is equal to the mean of the distribution shown, however, this will not be the case for general data. To the left and right of the IQR in the box plot, dashed lines radiate until the maximum and minimum values within 1.5 IQRs of the lower and upper quartiles of the distribution are obtained. This region is highlighted in the lighter shade in Figure 6.1.

The box plot is a visual aid which allows the variance of the data to be easily visualised. In Figure 6.1 the box-plot is presented in a horizontal orientation in order to facilitate comparisons to a normal distribution, but box-plots are typically presented vertically. The box-plot is employed in this chapter because it facilitates immediate quantification of the IQR ranges, median, etc, while with a normal distribution a viewer would be required to estimate these values. This also helps in the evaluation of median-based statistics which are useful means of assessing effectiveness of algorithms such as QRS detection.

6.3 CHOICE OF QRS DETECTOR

Two QRS detectors were investigated for this work.

The first QRS detection algorithm which was chosen employs the Hilbert transform (Hilbert, 1912). This algorithm was chosen because the Hilbert transform method of QRS detection has been used by many researchers for over 30 years with very good results (Bolton and Westphal, 1981b; Nygård and Sörnmo, 1983; Bolton and Westphal, 1985). This algorithm has also recently been employed for a robust QRS detector which reduces the effect of baseline drift, muscular and motion artefacts (Benitez et al., 2001). These are necessary traits for use during OFC as the subjects are free to move around the bed they lie on.

The second QRS detector which was selected was presented by Afonso et al. (1999). This method was selected because it incorporates filter banks (which can be thought of as being analogous to wavelet transforms) in order to decompose the signal into uniform bandwidth constituents. Wavelet and filter-bank based signal processing have been shown to be very useful for DSP applications and therefore this was investigated.

Both of these algorithms have reported over 99% sensitivity and positive predictivity when employed on the MIT-BIH database, and the next two sections discuss the procedures and algorithmic properties of these.

Where possible, the difference equations that were used by the QRS detection algorithms will be provided. However, it is not feasible to provide these in every case as the algorithms will often require filters that are approximately of the order of the sampling rate. Where this occurs, the digital design toolbox of Matlab® can be easily used to generate the difference equations required.

6.4 HILBERT TRANSFORM BASED QRS DETECTION

6.4.1 THEORY OF HILBERT TRANSFORM

The first QRS detection algorithm investigated involves the use of the Hilbert transform (Hilbert, 1912). The Hilbert Transform is defined by

$$\widehat{x}(t) = \mathcal{H}[x(t)], \quad (6.4)$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} x(\tau) \frac{1}{t - \tau} d\tau, \quad (6.5)$$

where, $\mathcal{H}$ is the Hilbert transform operator and $\widehat{x}$ is the Hilbert transform of signal, $x(t)$. The Hilbert transform provides a time varying and a linear function of $x(t)$ and Equation (6.4) can be rewritten as the convolution between the signal and $\frac{1}{\pi t}$, i.e.

$$\widehat{x} = x(t) * \frac{1}{\pi t}. \quad (6.6)$$

Convolution can be performed efficiently in the frequency domain as the convolutions theorem states that the Fourier transform (FT) of a convolutions is the pointwise product of the individual Fourier transforms (Katznelson, 2004). Using this, the Hilbert transform can be rewritten as

$$\mathcal{F}\{\widehat{x}(t)\} = \frac{1}{\pi} \mathcal{F}\left\{\frac{1}{t}\right\} \mathcal{F}\{x(t)\}. \quad (6.7)$$

The Fourier transform of $\frac{1}{t}$ is simplified to

$$\mathcal{F}\left\{\frac{1}{t}\right\} = \int_{-\infty}^{\infty} \frac{1}{x} e^{-2\pi f x} dx, \quad (6.8)$$

$$= -j\pi \operatorname{sign}(f), \quad (6.9)$$

where the $\operatorname{sign}(f) = 1$ for $f > 0$, 0 when $f = 0$, and -1 when $f < 0$. With this result, Equation (6.7) is rewritten as

$$\mathcal{F}\{\widehat{x}(t)\} = -j \operatorname{sign}(f) \mathcal{F}\{x(t)\}. \quad (6.10)$$

Taking the inverse Fourier transform of Equation (6.10) yields real and complex time-dependent variables, illustrated in Figure 6.2. The envelope of this function and the original signal is used in many applications and is calculated by

$$B(t) = \sqrt{x^2(t) + \widehat{x}^2(t)}, \quad (6.11)$$

and the instantaneous phase angle can be computed by

$$\theta(t) = \arctan\left(\frac{\widehat{x}(t)}{x(t)}\right). \quad (6.12)$$

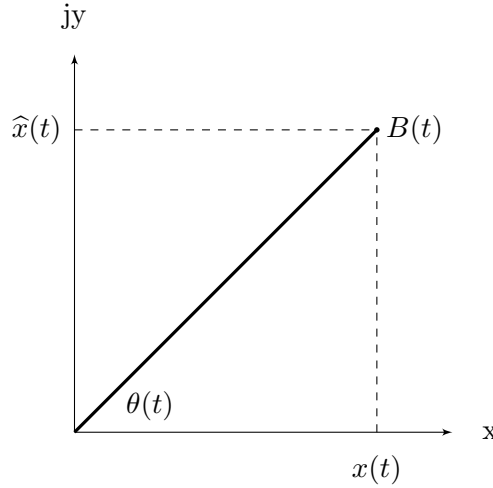


Figure 6.2: The real and imaginary components resulting from the Hilbert Transform of the ECG.

6.4.2 METHOD OF QRS DETECTION WITH HILBERT TRANSFORM

The Hilbert transform was previously formed in the continuous domain. However, discrete-domain equivalents can be obtained by performing the FFT, which was discussed in Chapter 4. The Hilbert transform is an odd function, which means that whenever there is a change in inflection of the signal (i.e. when the signal slope changes from positive to negative, or from negative to positive) the Hilbert transformed signal will cross the horizontal axis. This property is favourable for QRS detection as the R-wave of the complex is characterised by this.

The Hilbert transform was first used for QRS detection by Bolton and Westphal (1981a; 1981b; 1984; 1985) and has since been employed by a variety of ECG researchers (Nygårds and Sörnmo, 1983; Mietus et al., 2000; Benitez et al., 2000, 2001; de Oliveira and Cortez, 2004; Arzeno et al., 2006). However, for computational reasons in some of the early publications, the Hilbert transform was approximated by a band-limited finite impulse response (FIR) filter and the envelope was also approximated by

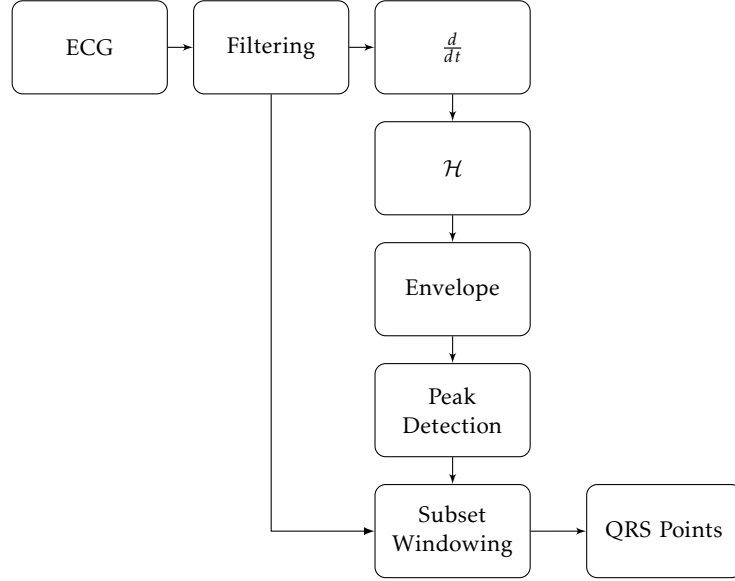


Figure 6.3: The flowchart for QRS detection from the Hilbert Transform.

$$\widehat{B(n)} = |x(n)| + |\widehat{x}_{\text{approx}}(n)|. \quad (6.13)$$

While the original work required optimisation for computational and power reasons, the Hilbert transform was not approximated in this work and was computed by the means outlined in Section 6.4.1. The pipeline for QRS identification is summarised in Figure 6.3. The raw ECG signal is first filtered by a high-order Kaiser-Bessel band-pass FIR filter with the pass-band specified by the bandwidth of typical QRS complexes, i.e. 8 — 20 Hz (Dinh et al., 2001; Kohler et al., 2002; Schlindwein et al., 2006). This frequency band reduces the effect of muscular artefacts while pre-emphasising the QRS complex. After filtering, the ECG signal is differentiated by

$$\frac{d}{dn}x(n) = \frac{1}{2\Delta t} [x(n+1) - x(n-1)], \quad (6.14)$$

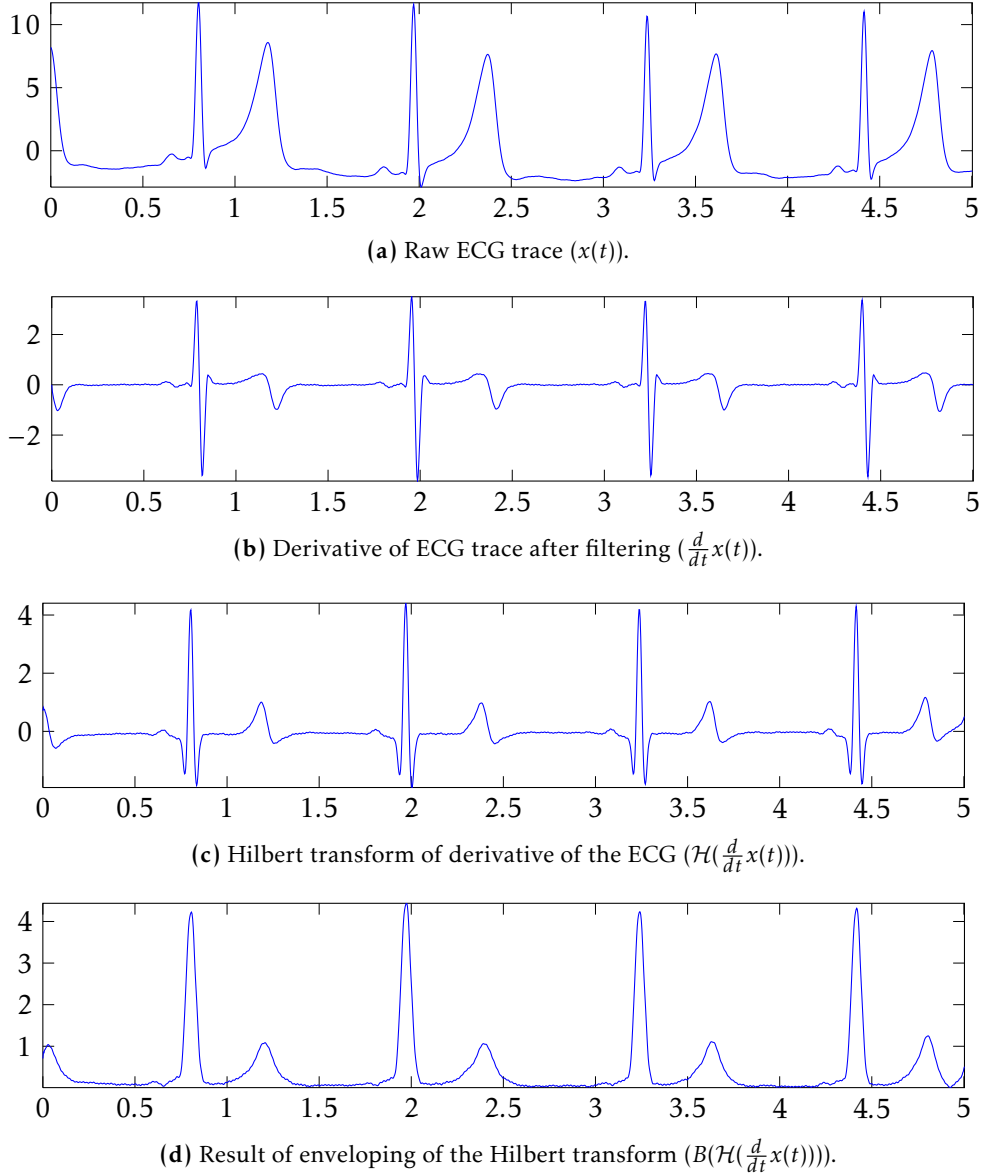


Figure 6.4: The stages employed by the Hilbert Transform QRS detection algorithm (the ECG data was obtained from Patient 113 in the MIT-BIH Database).

where $x(n)$ is the n^{th} sample, and Δt is period of the sampling frequency. The reason for choosing a non-causal filter was not discussed, but it is a popular methodology in the QRS detection algorithms (Kohler et al., 2002).

Figure 6.4 shows the progression of the ECG trace from recording to enveloping. The original ECG (Figure 6.4a) is filtered and the derivative of the result is calculated (Figure

6.4b). The Hilbert transform is then performed on this data (Figure 6.4c) and enveloping is obtained by Equation (6.11) (Figure 6.4d). These figures show how the effect of high-amplitude T-waves are reduced by this algorithm, and how the QRS complex is accentuated whereas other aspects of the ECG are attenuated.

6.4.3 BEAT IDENTIFICATION

Higher values of the envelope indicate a higher probability of a true QRS peaks. The P- and T-waves are typically characterised by a lower frequency bandwidth than the QRS complex, so even with T-waves of comparable amplitude to the QRS complex (Figure 6.4a) the resulting envelope of the QRS complexes is significantly more pronounced than resulting from the P- and T-waves (see Figure 6.4d).

Adaptive thresholding is incorporated to identify QRS complexes as stated by Benitez et al. (2001). Thresholds are dynamically calculated with regard to estimates of signal noise. In this case, noise is any non-QRS ECG signal shapes and is approximated by computing the RMS value of the result of the envelope over a 1024-sample time window. If the RMS value at a particular instance is greater than 18% of the maximum value in the same time window, the level of noise is considered high, and a threshold of 39% of the maximum value over the window is selected and points which exceed this are selected as QRS points. If the noise estimate is less than 18% of the maximum value, the threshold is set low to 1.6 times the RMS noise estimate. If two peaks are detected within 200 ms, one of the peaks is eliminated upon review of both amplitudes and relative position of both peaks to the previous QRS peak. By this process, QRS detection is obtained through the use of the Hilbert transform.

In subsequent sections this algorithm is referred to as the Hilbert transform algorithm. It should be noted, however, that the association refers to the underlying algorithm and not to the author of the publication.

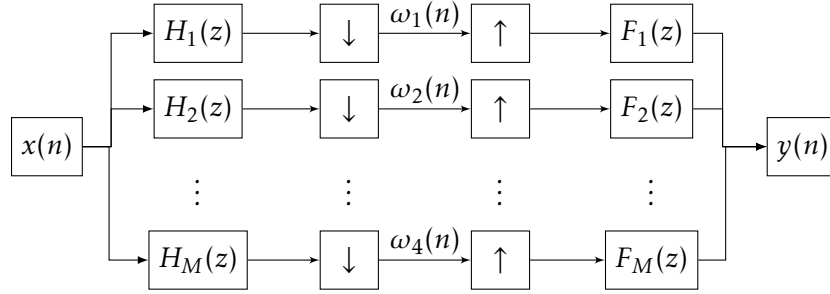


Figure 6.5: The generic filter banks flow chart incorporating both bandpass and synthesis filters. The $\downarrow$ and $\uparrow$ symbols represent down- and up-sampling respectively.

6.5 FILTER-BANKS BASED QRS DETECTION

6.5.1 THEORY OF FILTER BANKS

A filter bank is an array of filters whose purpose is to decompose an input signal into M components (Soman et al., 1993; Saramäki and Bregovic, 2002). In the digital domain, this is achieved by an array of FIR or infinite impulse response (IIR) filters and Figure 6.5 presents the general architecture which can achieve this. In this Figure, the filters H_1 — H_M are applied to the input signal $x(n)$. In some filter bank applications the signal is then down-sampled before analysis is performed. The signal can then be up-sampled, re-synthesised by the synthesis filters F_1 — F_M and combined to generate the representation of the original signal, $y(n)$. The idealised output of the filter banks method is shown in

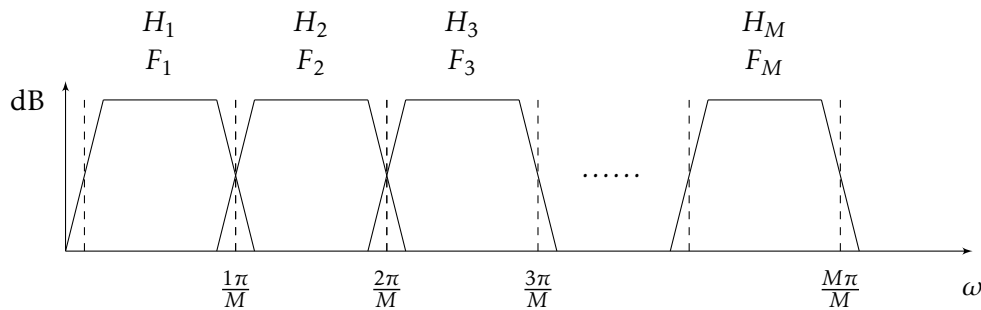


Figure 6.6: The idealised filter response of the filter banks, with M equally-wide sub-bands.

Figure 6.6. In this Figure, M frequency responses are shown, and each band is labelled with the decomposing filters (H_m) which were employed to achieve the response.

6.5.2 QRS DETECTION WITH FILTER BANKS

Afonso et al. (1999), presented a QRS detection algorithm involving filter banks in which the ECG was decomposed into four banks of uniform bandwidths (BWs) of 5.6 Hz (each decomposition filter contains f_s components). Once the signal has been decomposed, the algorithm down-samples the signal on each subband by a ratio R which is calculated by

$$R = \text{round}\left(\frac{f_s/2}{\text{BW}}\right). \quad (6.15)$$

Here, f_s is the signal sampling rate and BW is the bandwidth as before, and $R = 23$ for $f_s = 256$ Hz.

The purpose of down-sampling the subband signals is to reduce the noise and to increase the signal to noise ratio (SNR) of the derived signals (Afonso et al., 1999, 1995). With this algorithm the signals are not reconstructed, so the up-sampling blocks and synthesis shown in Figure 6.5 are not used.

The remainder of this QRS detection involves multi-stage processing incorporating many threshold stages.

6.5.2.1 PRE-PROCESSING

Four features are extracted from the sub-bands, and these features are employed to aid in QRS detection. Each feature combines certain sub-bands of interest and these are defined by Equations (6.16) — (6.19).

$$P_1 = \sum_{l=1}^3 |w_l(z)| \quad (6.16)$$

$$P_2 = \sum_{l=1}^4 |w_l(z)| \quad (6.17)$$

$$P_3 = \sum_{l=2}^4 |w_l(z)| \quad (6.18)$$

$$P_4 = \sum_{l=1}^3 w_l(z)^2 \quad (6.19)$$

where w_l represents the output of the l^{th} bandpass filter. These features effectively measure the energy which is found within the bands of interest by linearly combining the absolute values of the sub-bands. For example, feature P_1 combines the output of the first three bandpass filters, so this is representative of the energy up to 16.8 Hz.

Two-sample moving window integration is performed on each feature.

6.5.2.2 BEAT-CLASSIFICATION LOGIC

Afonso's algorithm involves a complicated series of filtering and thresholding routines. Figure 6.7 attempts to describe the algorithm simply. First, the ECG is filtered into four component bands and downsampled. Then multiple levels of QRS validation are performed, and the result of this is a series of validated QRS points. Subsequently, six

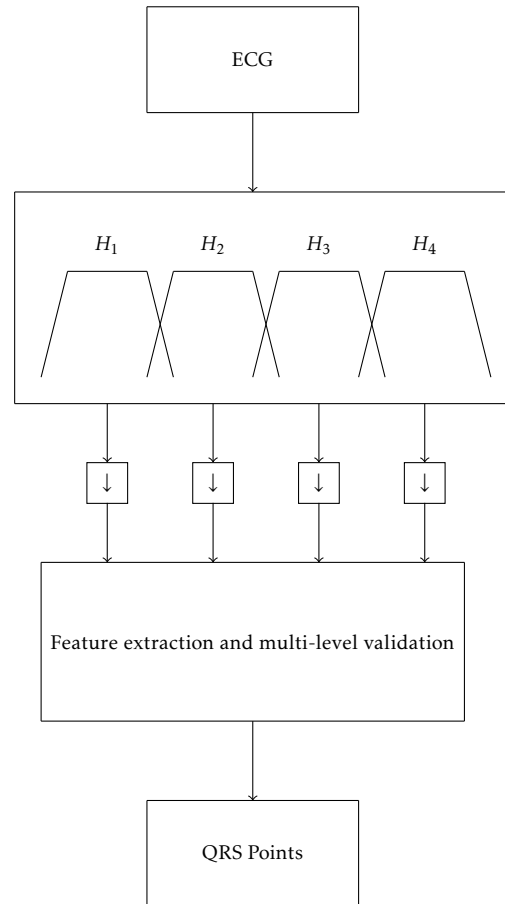


Figure 6.7: Overall simplified flowchart of Afonso's QRS detection method.

levels of QRS validation are executed over all candidate QRS points. These are described here.

6.5.2.2.1 Level 1 The first validation Level involves identifying all the positive peaks of the moving window integrator of feature P_1 , Equation (6.16). This stage is designed to identify as many QRS candidate events as possible. While this process will produce a large number of false-positive points, subsequent validation levels will eliminate inappropriate candidates. Figure 6.8a shows the raw ECG trace which was recorded, and Figure 6.8b shows the events which are selected by this feature. It can be seen that every positive peak which occurred in the time-series is marked as a QRS point. In four instances at t

= {141.75, 144.5, 146, 146.75} seconds it can be seen that T-waves contributed to positive peaks and were identified by this level of QRS detection.

6.5.2.2.2 Level 2 The second Level involves two single-channel threshold stages in which candidate QRS points from level 1 are employed as estimates of ‘signal’ and ‘noise’ of the ECG signal. Here, ‘signal’ refers to the times that Level 1 have identified as QRS candidate points, and ‘noise’ refers to points which were flagged as QRS candidates.

The mean value of feature P_2 at the candidate QRS times from Level 1 is computed. The initial ‘signal’ level is estimated as 10% above this mean value, while the noise level is initially set as 10% below this. The values of the candidate QRS points are compared against these levels giving two sets of QRS events. A parameter is introduced to identify a measure of confidence of whether these are truly signal or noise, and is termed decision strength. This is defined by

$$DS(i) = \frac{\text{feature}(i) - NL}{SL - NL}, \quad (6.20)$$

where $DS(i)$ is the decision strength measure for the i^{th} QRS candidate, NL is the Noise Level estimate, SL is the Signal Level estimate, and $\text{feature}(i)$ is the i^{th} candidate of the feature being considered. This equation is parametrised because the decision strength is used in later levels for validation. The lower the decision strength parameter, the more likely that the candidate point was noise, but the higher the value, the more likely it is a QRS point. The value of the parameter is force-bounded between 0 and 1.

The thresholding which is introduced at this Level is performed on the decision strength parameter. A low and high threshold are utilised, and the low threshold is set at 0.08, while the high threshold is set at 0.7. The low threshold is set so as to identify a large number of candidate points, while the high threshold is set so as to produce a high degree of certainty with a set of candidate points. The signal and noise levels are updated throughout the QRS

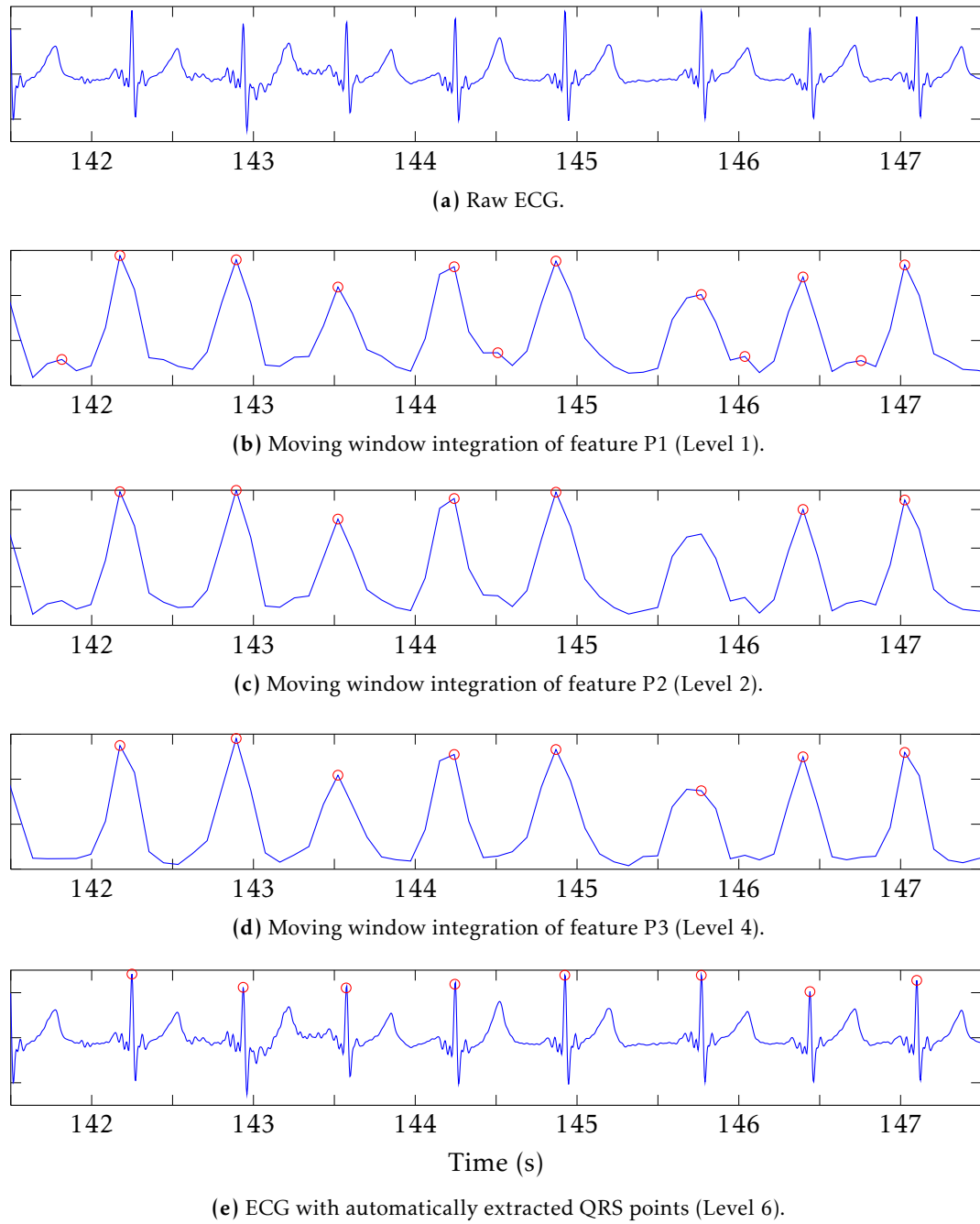


Figure 6.8: The effect of the various QRS validation levels from Afonso's QRS detection algorithm. In these charts, the $\circ$ symbols represent the candidate QRS points. Charts b — d have been down-sampled.

detection stage based on whether the classification of the current event was deemed to be 'signal' or 'noise' with regard to these thresholds and how they are employed later. The events classified in this stage with sample ECG is shown in Figure 6.8c. It can be seen here that all of the false positive QRS points have been removed, but the detection algorithm has also removed a true QRS beat which can be found at 145.75 seconds.

6.5.2.2.3 Level 3 Level three of the beat-classification logic fuses the two candidate QRS points from Level 2 together, i.e. the times at which the decision strengths surpassed the upper and lower thresholds. The results of stage two are termed 'channel 1' (which gives the QRS candidate times with reference to the low threshold) and 'channel 2' (QRS events with the high threshold).

Three possible outcomes can occur:

1. If channel 2 detected an event, a QRS event is always deemed to have occurred as the threshold was set very high. In this situation channel 1 will also have been triggered.
2. If neither channels detect a QRS event, the QRS event from the previous levels is eliminated from consideration.
3. If channel 1 (low threshold) detects an event, but channel 2 does not, the decision strengths of each channel are then computed. Two further parameters are computed from these decision strengths and are defined by

$$\Delta_1 = \frac{DS_1 - th_1}{1 - th_1}, \quad (6.21)$$

$$\Delta_2 = \frac{th_2 - DS_2}{th_2}, \quad (6.22)$$

where th_1 and th_2 are the thresholds which were used when computing the candidate events and Δ_1 and Δ_2 relate to channels 1 and 2 respectively. If Δ_1 is greater than Δ_2 , the event is deemed to have been a QRS event. Otherwise is deemed to have been noise.

6.5.2.2.4 Level 4 The fourth Level uses another one-channel detection block, this time using feature P3. A new threshold of 0.3 (of decision strength) is used here. If Level 3 removed a candidate QRS event, and the decision strength from P3 exceeds the threshold of 0.3, the beat is re-introduced as a candidate. This stage reduces the number of false negative QRS events in this algorithm. This Level only operates on events which were removed Level 3. The events classified in this stage are shown in Figure 6.8d. It can be seen that the point which was removed in Figure 6.8c at 145.75 seconds is re-introduced by Level 4.

6.5.2.2.5 Levels 5 and 6 The fifth stage reviews the points which are still under consideration QRS and performs decision logic based on timing between the QRS events. If the time between two consecutive QRS events is greater than 1.5 times the mean of the previous 100 QRS events, a lower decision strength threshold of 0.2 is employed to accept events which were removed by the validation stages. Furthermore, if the difference between two consecutive QRS events is less than 0.24 seconds, the point which resulted in the smaller peak in the original ECG signal is removed.

6.5.2.3 OVERALL

The set of candidate points which are obtained after Level 6 are considered as true QRS points by the algorithm and no further post-processing is performed. The means by which the points relate to the original ECG trace is shown in Figure 6.8e where the QRS points are shown with the symbol $\circ$. In subsequent sections, this algorithm is referred to as Afonso's algorithm.

Table 6.1: Differences in reported and calculated sensitivity and positive predictivity.

	Afonso		Hilbert Transform	
	Reported	This work	Reported	This work
Sensitivity	99.59%	99.15%	99.87%	99.03%
Positive Predictivity	99.56%	98.16%	99.94%	97.43%

6.6 RESULTS OBTAINED ON MIT-BIH DATABASE

Here, automatic QRS detection is assessed on the MIT-BIH database. The Hilbert transform and Afonso's QRS detection algorithms are employed for this detection process.

6.6.1 SENSITIVITY AND POSITIVE PREDICTIVITY

Table 6.1 tabulates the sensitivity and positive predictivity of QRS detection results which were obtained with the MIT-BIH database. Sensitivity and positive predictivity of 99.15% and 99.16% were obtained with Afonso's algorithm, and the Hilbert transform algorithm yielded sensitivity of 99.03% and positive predictivity of 97.43%. The distribution of these results are presented in Figure 6.10, which shows that the majority of the QRS points which were detected were within the window of acceptance. However, these results are slightly different from the results which were obtained in the literature, and the cause for this was due to incorrect localisation of the QRS points on the ECG signal trace.

Patient 8, whose ECG is shown in Figure 6.9, has a non-standard QRS complex in comparison to the example shown in Chapter 1, where it can be seen that the S-wave drops significantly below the baseline level of the ECG trace. Figure 6.9 illustrates how the bad QRS localisation can occur. In this figure, the solid trace is the raw ECG, the $\times$ markers represent the expert QRS annotations, the $\circ$ markers identify the QRS points from Afonso's algorithm, and the $\square$ markers represent the times which the Hilbert transform identified as the QRS complexes. It can be seen that Afonso's algorithm misclassified the T-wave as

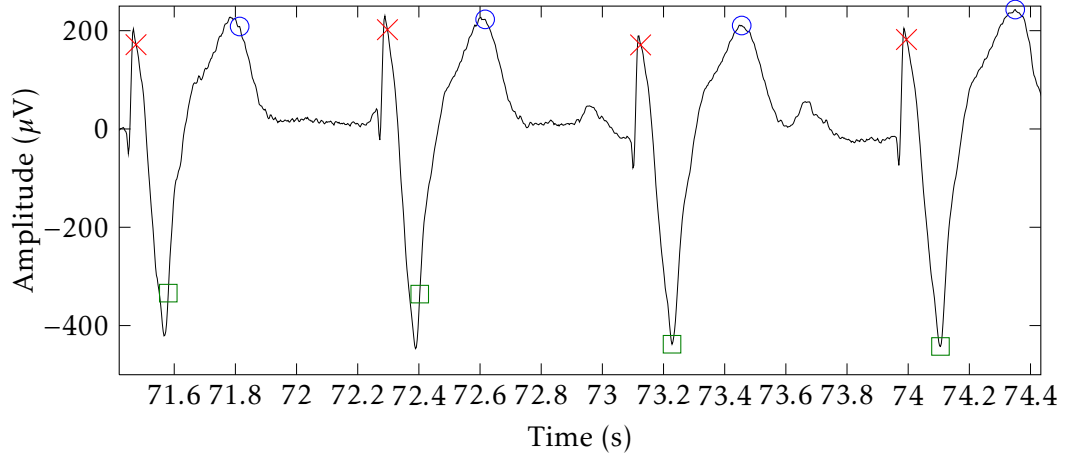


Figure 6.9: Incorrect QRS complex localisation (Patient 8 of MIT-BIH arrhythmia database). Manual QRS annotations are marked with $\times$ and automatic detections are marked with $\square$ (Hilbert transform algorithm) and $\circ$ (Afonso's algorithm).

a R-wave while the Hilbert transform algorithm misinterprets the S-wave as the R-wave. The positions of these mis-localisations are logical given the nature of the two algorithms.

All of these points are flagged as false negative and false positive points as they were outside of the window of acceptance, and therefore the sensitivity and positive predictivity results that were obtained are reduced. For some patients, the incorrect localisation occurred intermittently during their recording. As the majority of the QRS complexes extracted on these patients was consistent with the annotations, it was not deemed appropriate to eliminate these for performance evaluation. Indeed, using median-based statistics, the sensitivity and positive predictivity of detection can all be seen to be above 99.5% from Figure 6.10.

It is believed that the reason for the superior results obtained in the literature is due to a wider window of acceptance. However, it will be shown later that widening the window is not appropriate for this this work.

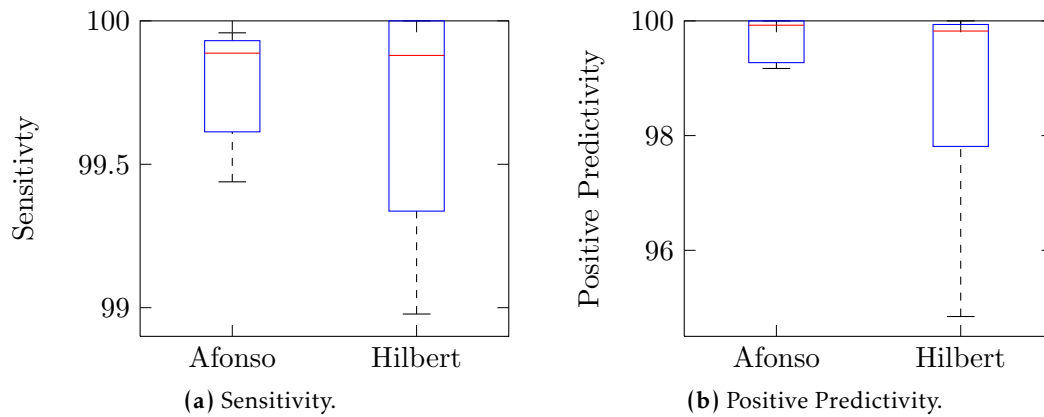


Figure 6.10: Sensitivity and positive predictivity box-plots of QRS detection on the MIT-BIH arrhythmia database.

6.6.2 PERCENTAGE RMS DIFFERENCE

The mean and standard deviation of the heart rate for the patients in the MIT-BIH database were computed from the expert QRS annotations and those extracted automatically by the QRS detectors based on the Hilbert transform and based on Afonso's algorithm. The PRD was computed according to Equation (6.3), and Figure 6.11 shows the box-plot of the PRD of these features over the MIT-BIH database. The closer the PRD is to 0%, the more similar the automatically generated features are to the manual features. Indeed, it can be seen that the median of the PRD in all cases is very low, indicating a high degree of agreement between the manual and automatic QRS points in the majority of cases.

Figure 6.11a shows the PRD values which were calculated from the mean heart rate. This figure shows that even though incorrect localisation of the QRS complex occurred, the features which were obtained from these points do not distort the features, and that the values which were extracted are representative of the heart rates obtained with the manual annotations. This is further supported by the low variance which was computed on the same subjects.

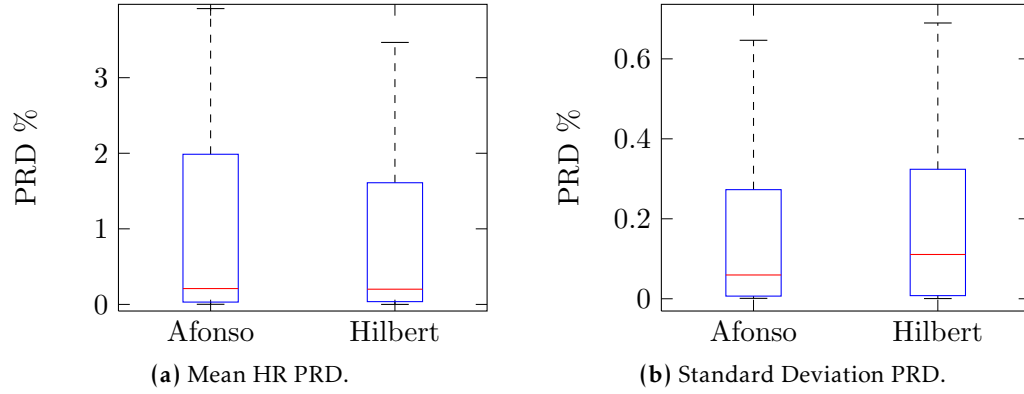


Figure 6.11: PRD box-plots of the mean (μ) and standard deviation (σ) of the heart rate over all subjects in the MIT-BIH arrhythmia database.

6.6.3 CONCLUSIONS ON QRS DETECTION ON MIT-BIH DATABASE

This section has shown that the two QRS detection algorithms which were developed achieved competitive results when considered against the results which were published. While these results have already been obtained in the literature, it is important to verify that the results were reproducible, in particular when the algorithms which were assessed are intended to be employed for diagnostic applications. The two QRS detection algorithms which were employed were found to perform slightly better in the literature.

6.7 REQUIREMENT FOR ARTEFACT DETECTION

6.7.1 INTRODUCTION

In the MIT-BIH database, the patients whose data were recorded were adults that were undergoing important medical examinations. Therefore, these patients were not inclined to move, and while there were a tendencies for some artefacts, their contribution were low in general. However, the subjects who participated in the OFCs were children with a mean age of approximately 5 years, who wanted to play games and move during their tests. As

a result, artefacts were introduced in the ECG that was recorded, and artefact detection is required in order to provide confidence that the QRS points which were extracted were accurate.

6.7.2 ARTEFACT DETECTION ALGORITHM

The bandwidth of the QRS complex ranges from between approximately 10 and 25 Hz (Dinh et al., 2001; Kohler et al., 2002; Schlindwein et al., 2006). For this reason, almost all QRS detectors will filter the ECG with a band-pass (BP) filter so as to attenuate non-QRS complex shapes, such as artefact, P- and T-waves. ECG artefacts are signals which interfere with the signal trace and they arise from physiological (motion, muscular spasm, and touching electrodes), electrical (mains interference) and resistive (poor connection with the skin) conditions. Artefacts can lead to false-detections and non-detections of the QRS complex, in both cases the accuracy of the features which are extracted from the ECG signal are affected.

Even with de-noising filters which reduce the extent of artefacts on the ECG, noise can still affect the fidelity of the QRS complex. Therefore, when considering unconstrained, fully automated biomedical classification platforms, artefact-awareness is an important requirement. An algorithm incorporating energy detection was adopted to detect the signal strength in the ‘noise bands’ of the ECG. ‘Noise bands’ are the frequency bands outside of the bandwidth of interest of the QRS complex. The lower noise band was set to 0 — 10 Hz and the high noise band includes all frequencies greater than 25 Hz.

Energy detection estimation is performed on these frequency bands in order to assess the extent of noise in the ECG. Energy detection is a popular means of signal detection and is used by communications researchers (Horgan and Murphy, 2010). Here, two 50th order FIR filters were designed and are denoted as H_h and H_l . H_h is a high-pass filter with a corner frequency of 25 Hz which was employed in order to detect high-frequency artefacts

resulting from muscular, motion noise, etc. A low-pass filter, H_l , was also designed with a cutoff frequency of 10 Hz. The purpose of this filter was to detect low-frequency artefacts.

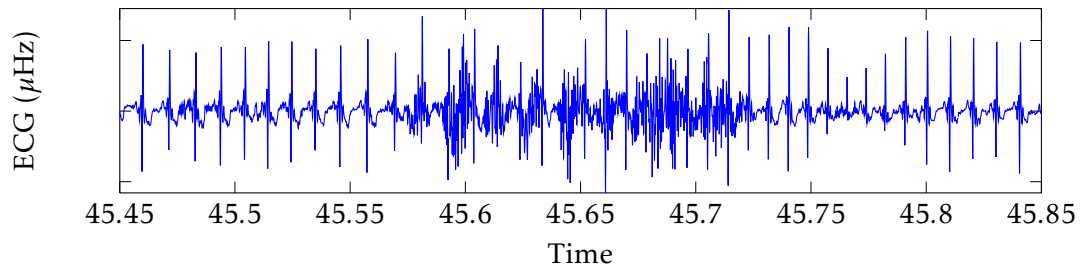
Energy estimates at a time index k are computed from the output of H_h and H_l in a time window of length N , i.e. between $k - N$ and k . In this work, the window length of 0.25 seconds (i.e. 64 points when sampled at 256 Hz) was chosen in order to detect short-duration artefacts. The energy $E(n)$ of a signal x can be computed by a ‘square and accumulate’ algorithm in Equation (6.23).

$$E(k) = \sum_{i=k-N}^k x^2(i) \quad (6.23)$$

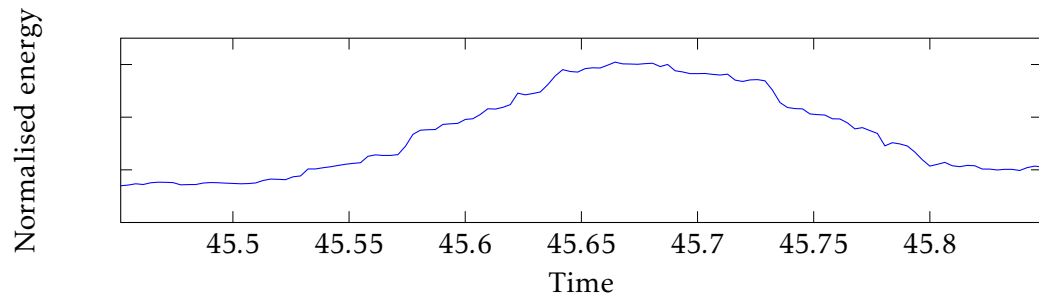
6.7.3 DEMONSTRATION OF ARTEFACT DETECTION

Figures 6.12a and 6.12b show the scaled output of the energy detection algorithm for high frequency artefact detection. At $t = 45.6$ minutes in Figure 6.12a high-frequency muscular artefacts can be observed. The QRS complexes for 6 seconds after this time are obscured by high-frequency artefacts. The output of the high-frequency artefact detection filter is normalised by the standard deviation measurement from all other Subjects in the databases where this algorithm was used by a LOO procedure similar to the one employed for classification in the previous chapter. This output, shown in Figure 6.12b can be seen to rise at the same time as the high-frequency noise presents in the ECG. The artefact output was smoothed with a 2 second moving average filter. The filter was centred so as to remove any phase delay which might occur due to the moving average process.

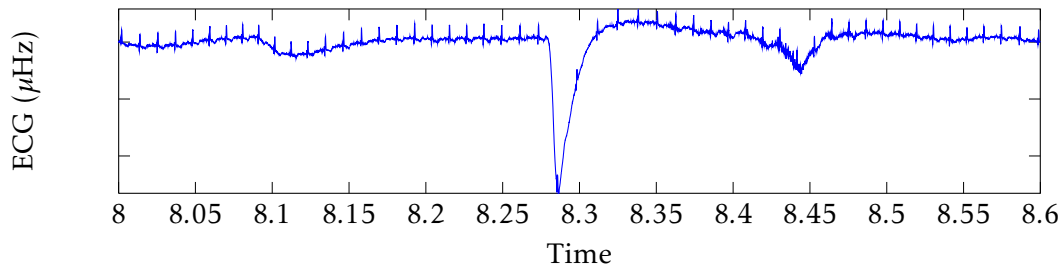
Dynamic thresholding is applied to the output of the energy detection filters and when the result surpasses the threshold, that epoch is considered ‘noisy’ and will not be considered for feature extraction. The threshold for artefact detection is set dynamically for each subject. A moving window of 1 minute of non-artefact energy estimates is selected and the



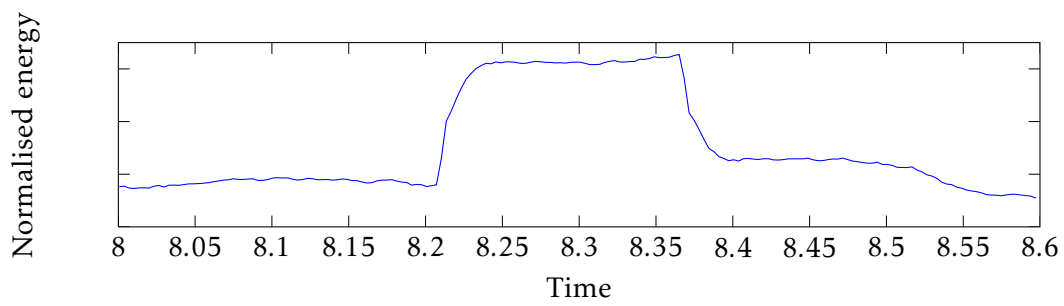
(a) ECG with high-frequency artefacts present.



(b) Output of high-frequency artefact detection filter.



(c) ECG with low-frequency artefacts present.



(d) Output of low-frequency artefact detection filter.

Figure 6.12: Normalised output of high-frequency (a,b) and low-frequency (c,d) energy estimators for artefact detection (data from Subject 8 of allergy database).

mean, μ , and standard deviation, σ , of this window is calculated. If the signal surpassed $\mu + 3\sigma$, the ECG at this segment was labelled ‘noisy.’

Figures 6.12c and 6.12d are similar to Figures 6.12a and 6.12b, but they presents the output of the low-pass artefact detection filter. At $t = 8.28$ minutes a baseline deviation can be observed in the ECG trace. This artefact is likely to be as a result of the subject in question touching the ECG electrode or electrode pop of the sensor. The internal filtering in the hardware ECG daughterboard of the SHIMMER device will compensate for the electrode pop artefact gradually which results in the slower wave which can be seen. Once again the artefact affects the amplitude of the QRS complex to the point where the accuracy of reliable QRS detection is compromised. The normalised output of the energy detection is shown in Figure 6.12b and a change in the artefact signal can be observed at the same time as a change in the ECG trace.

The output of the low and high frequency artefact detection filters are fused together by the logical OR-ing operation.

6.8 RESULTS OBTAINED ON ALLERGY DATABASE

Here, the QRS detection results which were obtained with the automatic QRS detectors on the allergy database are presented.

6.8.1 ARTEFACT DETECTION

The number of artefact events which were was flagged during the OFC recordings of the subjects in the allergy database are presented in the bar chart in Figure 6.13. The subjects indices of the allergy database are presented on the horizontal axis while the number of instances of artefact is presented on the vertical axis.

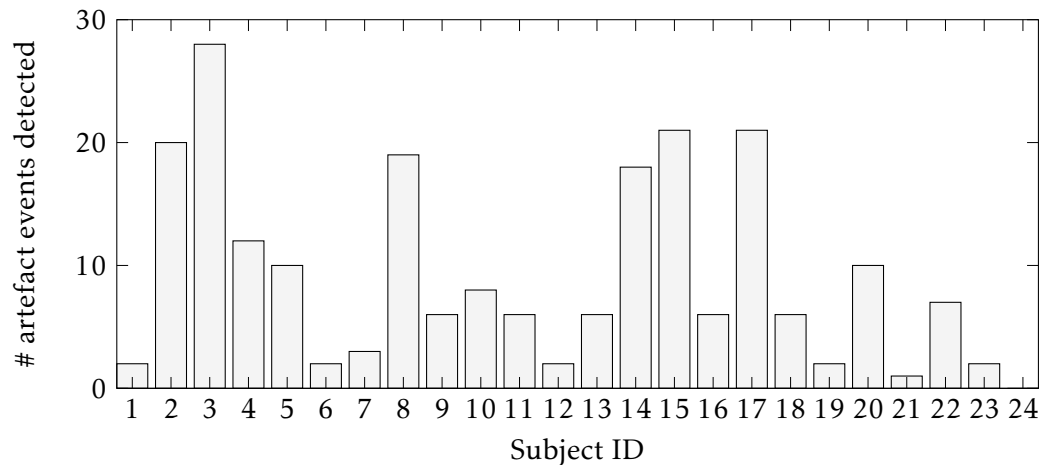


Figure 6.13: The breakdown of the number of artefact events which were detected by the artefact detection algorithm for each subject of the allergy database.

The subject which presented with the greatest number of artefacts was Subject 3 where 28 individual events were identified. The events which were detected were similar to the high and low frequency artefacts which were shown in Figures 6.12a — 6.12d. In earlier chapters, the most stringent requirement for allergy detection was stated to be obtaining 100% specificity (i.e. no false positive allergy classifications) as only with these conditions will the OFC be improved. It will be shown in the next chapter that the presence of artefacts in the ECG will have the capacity to mislabel segments of data as allergy.

Approximately 50% of the artefacts which were detected were recorded during the checkup periods the oral food challenge. When the allergist checks the heart rate, blood pressure, blood oxygen saturation levels and temperature of the subjects during the checkups, the subject is required to move so that the equipment is correctly applied. As a result, high and low frequency artefacts were prevalent during checkups. In allergy classification, the time periods which satisfy the classification criteria and which occur during checkup periods are not classified as allergy. This is because checkups interfere with the heart rate of the subjects, and causes them to move which introduces artefacts to the ECG. It is also the case that artefact detection during the checkup periods is not strictly necessary because if this were used in a clinical setting, the allergists would be aware of the introduction of artefacts, and would likely disregard this data.

Table 6.2: Sensitivity and positive predictivity of QRS detectors on allergy database.

	Afonso	Hilbert Transform
Sensitivity	92.89%	92.17%
Positive Predictivity	93.15%	93.00%

However, in over 50% of cases, artefacts were not due to the influence of the allergist on the subject. Every event which was flagged as an artefact was visually inspected, and it was found that the majority of the artefacts which remained contaminated the ECG points to the degree where QRS detection was unreliable. When QRS annotations were labelled manually for previous chapters, it was possible to identify the QRS points during time periods such as this with careful visual inspection, and it was at times like this when QRS extraction failed.

With approximately 5% of the artefact events, the accuracy of automatic QRS extraction was not affected. While it is possible that short-duration signatures of allergy might not be considered for allergy classification, it was deemed that less than 10% false positive artefact events was an acceptable trade in the attempt to preserve 100% specificity for fully automated allergy classification.

6.8.2 SENSITIVITY AND POSITIVE PREDICTIVITY

Table 6.2 tabulates the mean of the sensitivity and positive predictivity values obtained with QRS detection on the allergy database by the QRS detectors against the manual annotations. The sensitivity and positive predictivity values calculated are of the same overall percentage ($\approx 92 - 93\%$). The majority of these are above 99% which can be seen from the median lines in the box plots of Figure 6.14. In some cases sensitivity and positive predictivity values of approximately 70% were obtained.

Upon investigation of the subjects from whom poor QRS detection was obtained, it was discovered that the quality of the recorded ECG was the primary source of the reduction in

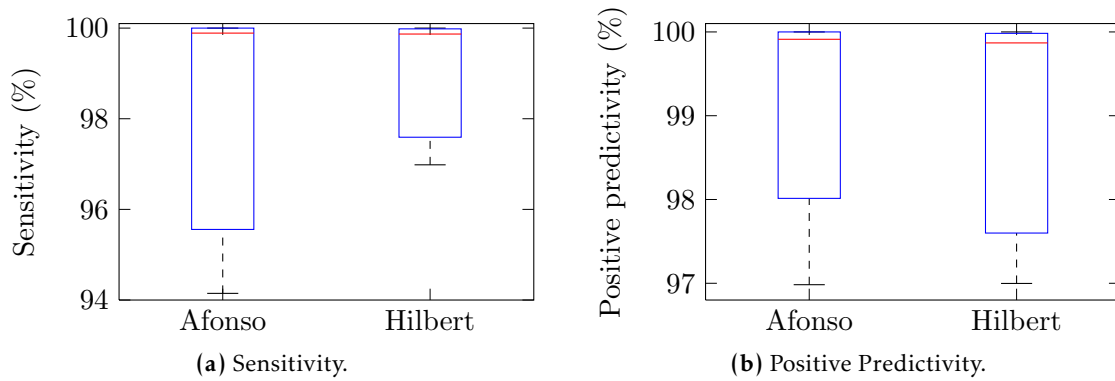


Figure 6.14: Sensitivity and positive predictivity box-plots of QRS detection on the allergy database.

accuracy metrics. It was difficult to determine the exact cause for the poor quality of ECG recorded, but it is believed that it is predominantly due to high-resistance connections between the ECG electrodes and the subject's skin. Figure 6.15 shows an example of the poor quality of ECG which was obtained over a 6 second period after applying filtering (Clifford et al., 2006). Contrasting the ECG in this recording to the ECG found in Figure 6.9, for example, it can be seen that even after filtering, the baseline of the recording in Figure 6.15 contains a significant amount of noise which is of the same bandwidth as the QRS complexes. The amplitude of the QRS complex is also much diminished in comparison to those shown previously in Figure 6.9.

A second cause for the poor performance of QRS detection was the phenomenon of incorrect QRS which was previously discussed for the MIT-BIH database in Section 6.6 and is illustrated in Figure 6.9. This contributed to the lower sensitivity and positive predictivity values which were obtained on the allergy database. Interestingly, while the cause of this in the MIT-BIH database was due to the rich set of arrhythmias which were recorded, the cause for the incorrect localisation on the allergy database was due to poor integrity of ECG signal that was recorded.

The overall sensitivity and positive predictivity could be improved by widening the window of acceptance on the allergy database. As a consequence of doing this, however, many of the QRS complexes would be identified as true positives, even though the true

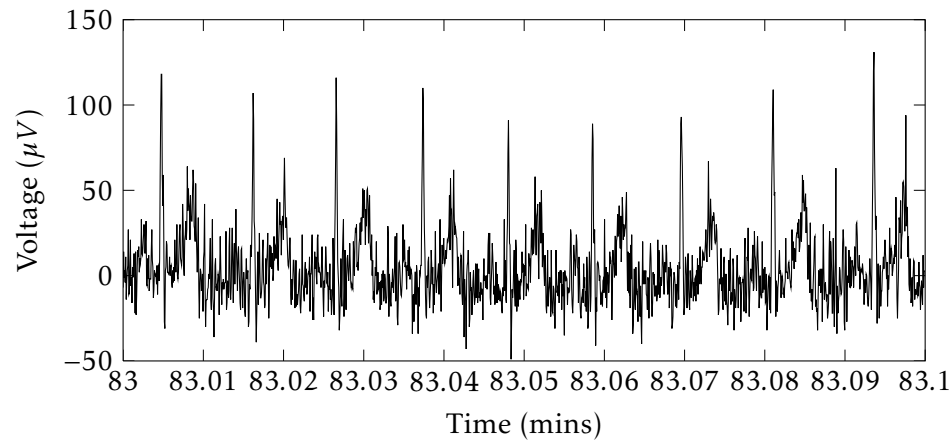


Figure 6.15: Example of poor quality of the ECG signal after the application of denoising filters which contributed to poor sensitivity and positive predictivity values for Subject 3.

source of the point was from artefacts. This would optimistically report the sensitivity and positive predictivity. Therefore, the criteria for true positive and false positive labelling was not relaxed in this work.

6.8.3 PERCENTAGE RMS DIFFERENCE

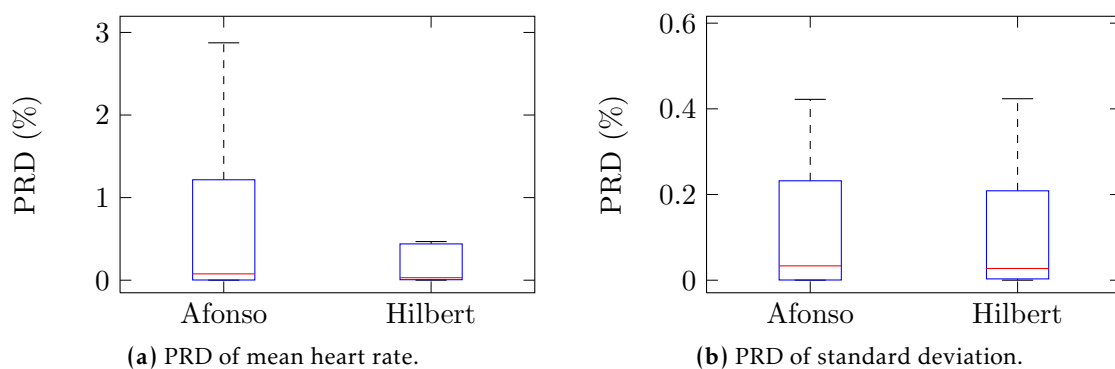
The PRD of the mean and standard deviation features were extracted to assess the effect of obtaining approximately 93% sensitivity and positive predictivity on the HRV metrics. Table 6.3 tabulates the results which were obtained. Excellent mean PRD values were obtained with both Afonso and the Hilbert transform algorithms ($< 1.3\%$ when averaged over all detection algorithms and features). The distribution of the PRD values obtained is outlined in Figures 6.16a and 6.16b. Interestingly, the PRD results obtained here are slightly better than those which were obtained on the MIT-BIH arrhythmia database even though poorer overall sensitivity and positive predictivity results were obtained with the allergy database. The PRD values obtained by the mean heart rate of the Hilbert transform present a significantly lower variance between subjects. Once more, median-based statistics of these indicate that for the majority of the subjects, excellent data are recorded, and as before, it is with some subjects that QRS detection is not as accurate.

Table 6.3: Distribution of mean and standard deviation of the PRD values calculated from automatically extracted QRS points.

	Afonso	Hilbert Transform
Mean Heart Rate	0.68%	0.56%
Standard Deviation	1.26%	1.1%

6.9 OVERALL DISCUSSION

The accuracy of the two QRS detection algorithms was assessed on two databases. Initially, QRS extraction performance assessment was performed on the MIT-BIH arrhythmia database and sensitivity and specificity measurements were computed. It was observed that correct localisation of the QRS complex on the R-wave was not always obtained, and that in certain situations QRS detection algorithms can localise the QRS complex on the S- and T-waves. This resulted in the discrepancies between the results which were reported and those which were obtained here. The slight discrepancies were due to a number of patients' ECG traces toggling between regular ECG and arrhythmia. Further accuracy assessment was performed by extracting HRV features from the QRS points reported by the QRS detection algorithms. The differences between the features were calculated via the PRD function, and average discrepancies of $\approx 1\%$ were obtained, but median-based analysis yielded excellent results.

**Figure 6.16:** Boxplots of the PRD of the mean and standard deviation of the heart rate between the manual and automatic QRS points extracted.

In contrast to the patients of the MIT-BIH arrhythmia database, the subjects of the allergy database were generally of an age where they were unwilling to remain still during their tests. For this reason, artefacts presented frequently in the ECG recordings in allergy database. Artefact detection was performed on the ECGs recorded to instill confidence that the features extracted from the ECG were contributions of the heart beats rather than from artefact noise. The sensitivity, positive predictivity, and the mean and standard deviation of the heart rates resulting from the automatic QRS detection were then extracted from the regions which were not flagged as artefact. The sensitivity and positive predictivity results were $\approx 93\%$, but once more, median-based results of over 99% were obtained for both QRS detectors. This indicates that in the majority of cases, excellent QRS detection is obtained. The extracted features were also very close to the true feature values.

The algorithms that were employed here were adapted directly from the literature which investigated the accuracy of QRS detection for adults. As the subjects under investigation for allergy classification are children under the age of 10 years, it could be argued that in varying certain parameters (e.g. threshold values) that better accuracy would be obtained. However, it was not possible to perform this tuning as no independent database consisting of children's ECG was available to allow objective parameter tuning, and therefore the 'stock' values that were provided in the original publications were employed.

CHAPTER 7

FULLY AUTOMATED ALLERGY DETECTION

7.1 INTRODUCTION

IN Chapter 5, classification of allergy was performed through statistical modelling of background, non-allergic heart rate variability features. Through these means, 93% sensitivity and 100% specificity of classification were achieved. However, these results were obtained on features which were extracted from QRS points that were manually annotated. This process provided an indication of the validity of heart rate variability based allergy classification, but due to the manual intervention required to acquire these annotations, the process would not be suitable for a real-time classification environment.

Therefore, Chapter 6 discussed the accuracy of QRS detection achieved by fully automated QRS detection algorithms. Two automatic QRS detection algorithms were investigated and the accuracy of QRS detection was benchmarked against a number of ECG databases. Based on these results, both of the QRS detection algorithms were deemed suitable for investigation here.

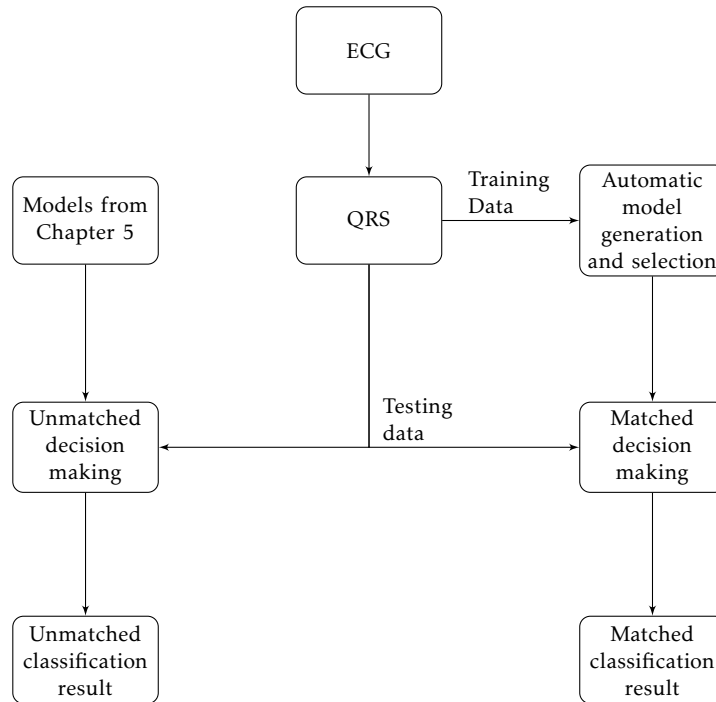


Figure 7.1: The flow of how the matched (right) and unmatched (left) classification results are obtained.

In this chapter, the heart rate variability features extracted with the automatic QRS points are employed in conjunction with the classification framework in order to assess fully automated allergy classification.

7.2 METHODS

Two means of automatic classification of allergy are assessed in this chapter. These are assessed because for applications of this nature where two representations of the same data are available (i.e. manual and automatic HRV features), the accuracy of machine learning algorithms can be assessed in a number of directions. These avenues are termed matched and unmatched classification and the processes of these are shown in Figure 7.1. In this figure, the right hand side shows the process by which the matched classification results are obtained, while the left hand side presents the unmatched process.

The first method assesses the results obtained by employing the classification models which were obtained in Chapter 5 ('manual models') with the normalised HRV features which were extracted from the QRS points which were obtained from the automatic QRS detectors discussed in Chapter 6. These sets of features are termed 'automatic data' to simplify discussions of the results in later sections, and automatic data is also subdivided into 'Afonso' and 'Hilbert' data, and relate to the feature data of each QRS detection algorithm. As these results employ manual models and automatic data, these results are termed 'crossover' and 'unmatched' results henceforth. Because the features which were obtained in Chapter 6 presented low PRD against the features which were extracted from manual QRS annotations, crossover classification sheds light on the sensitivity of the classification framework to input data which is of the same nature, but which is different in origin.

The second method investigated employed the model generation framework from Chapter 5, but trained new classification models and newly selected decision making parameters based only on the automatic data. This process was investigated because the classification framework has previously been stated as being data-driven, and unmatched classification might not be appropriate for fully automated allergy classification. Models generated in this process are termed 'automatic' models. Automatic models are further subdivided into 'Afonso' and 'Hilbert' models in reference to the QRS detection algorithms which were employed in their generation (see Chapter 6). The procedures employed in this chapter utilised the epoch-length fusion which was introduced in Chapter 5.

Matched and unmatched methods are investigated because classification of this variety has not been performed on allergy data previously. Therefore, it is uncertain which procedure will perform best for automatic classification, and this chapter investigates this. Typically, it would be expected to obtain better results with unmatched results as the models which would be employed in this case relate to manual models which were generated without error.

Table 7.1: Sensitivity and specificity of classification results obtained with the manual classification models on the automatically extracted HRV features (i.e. crossover classification results).

QRS detector	Sensitivity	Specificity
Afonso	60%	88.88%
Hilbert	66.66%	88.88%
Manual	93.33%	100%

7.3 UNMATCHED CLASSIFICATION RESULTS

This section presents the crossover classification results, and then discusses the applicability of this routine for autonomous classification of allergy.

7.3.1 RESULTS OF UNMATCHED CLASSIFICATION

Table 7.1 presents the sensitivity and specificity results which were computed by the crossover classification process. It can be seen from this table that in both cases specificity of 88.88% was calculated, indicating that false positive classifications were made. The sensitivity which was obtained is also quite poor in comparison to Chapter 5, with Afonso data achieving six false negatives (60% sensitivity) and Hilbert transform obtaining five false negatives (66.66% sensitivity).

The time gain metrics which were obtained with crossover classification will not be discussed here. This is because in previous chapters it was repeatedly stated that the most important aspect of allergy classification is obtaining 100% specificity, and this metric was not obtained with the crossover classification results. To employ an analogy, obtaining suboptimal specificity and discussing time gain metrics is equivalent to discussing a novel feature of a vehicle, when its braking mechanism does not function correctly. The priority in these cases should be safety, and this must be addressed before additional features warrant analysis. Without obtaining perfect specificity, machine-based classification of

allergy should not be trusted because the impact on the quality of life of a subject would elicit in unacceptable consequences.

7.3.2 DISCUSSION ON UNMATCHED CLASSIFICATION RESULTS

Subtle dynamics control the applicability of models for classification of allergy. This is evidenced by the fact that both poor sensitivity and poor specificity were obtained with crossover classification. From a decision making point of view, poor sensitivity implies that the duration and threshold parameters that were employed were too large as the allergy classification criteria were not satisfied in a number of cases. However, as this process also obtained suboptimal specificity, these decision making parameters were also, in fact, too low in other cases. Therefore, it can be stated that classification of allergy is a very strong function of the modelling parameter selection routine, and the novelty of certain segments can be under-represented with poor choices in these parameters. Indeed, this is perhaps not surprising. In Chapter 5, for example, it was shown that even with matched classification that the extent of the ‘novelty’ of certain regions of the HRV can be under-represented with some choices of parameters.

This suggests that the model and parameter selection is data driven, and that optimal automated allergy classification is not obtained by crossover classification. Every aspect of the classification framework is subject to the data which was employed to generate the models, and even with slight variations with these data, different results, model and decision making parameters are obtained. This, then, explains that while 100% specificity and good sensitivity was preserved between training and testing in Chapter 5, it was not preserved between unmatched results.

7.4 MATCHED CLASSIFICATION RESULTS

The results achieved in this section employed the boosting ensemble-based classification framework and learnt new classification models from the HRV data obtained from the QRS detection algorithms. The sets of PCA, GMM, duration and multiplicative parameters utilised for the generation of these results are all new to these methods and are entirely independent of the manual models, see Figure 7.1.

7.4.1 SENSITIVITY AND SPECIFICITY

Table 7.2 presents the overall results which were obtained by the two QRS algorithms which were investigated. In both cases, specificities of 100% were obtained by the automatic classification algorithms. This means that no false positive classifications were encountered. While 100% specificity has already been obtained in Chapter 5, obtaining 100% specificity with fully automated models is a very significant result as it means that a subject, having been classified allergic, can immediately be diagnosed as allergic at this time without the risk of false positive classifications. The same confidence which is attributed to the diagnosis of allergy can, therefore, be attributed to classification of allergy as classification of allergy is equivalent to a diagnosis of allergy.

Automatic classification obtained 80% sensitivity with both QRS detection algorithms. The subjects who were misclassified are Subjects 1, 2 and 3. Subject 1 was not classified as allergic by either the manual or automatic classification routines. The reason for this

Table 7.2: Sensitivity and specificity of classification results obtained by Afonso's and the Hilbert transform QRS detectors.

QRS detector	Sensitivity	Specificity
Afonso	80%	100%
Hilbert	80%	100%
Manual	93.33%	100%

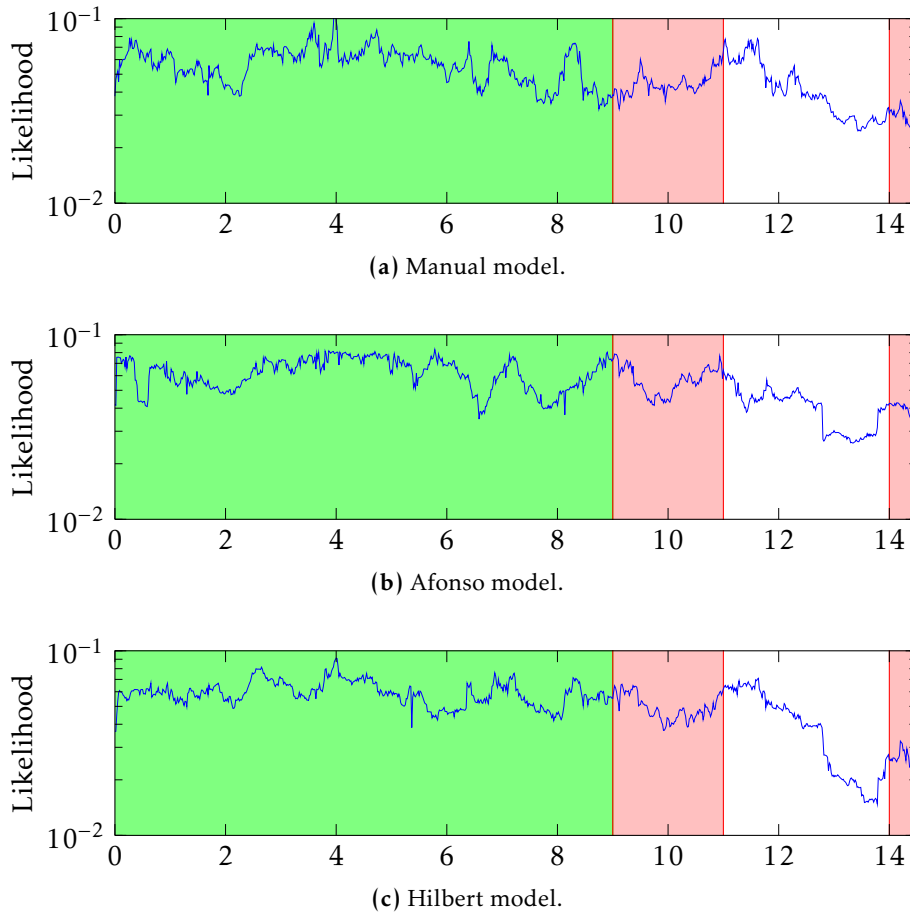


Figure 7.2: Likelihood plots of Subject 1 for manual and automatic models at epoch lengths of 60 seconds. Subplots (a) — (c) show the likelihoods which were obtained with manual, Afonso and Hilbert models respectively. In all cases the threshold for allergy classification is off the scope of the figures.

is because the Subject reacted to the food three minutes after consuming it. Within this short time frame, the features which were extracted did not change sufficiently for the allergy classification criteria to be satisfied. Figure 7.2 shows the likelihood traces which were obtained for Subject 1 with manual (Figure 7.2a), Afonso (Figure 7.2b) and Hilbert transform (Figure 7.2c) models at epoch lengths of 60 seconds. It should be noted that while the length of the recording for Subject 1 is approximately 14 minutes, the first dose of the allergen was administered 11 minutes after the challenge began.

It can be seen in this Figure that the set of three likelihood traces tend to deviate away from the background likelihood levels before the challenge was terminated. This is due to the

allergic reaction which the subject would experience that induced the subject to vomiting at approximately 14 minutes. While there is a visible departure from the background likelihood range with manual, Afonso and Hilbert likelihoods, the extent of this departure was insufficient for correct classification of this subject and in all cases the threshold is outside the boundary of the vertical axes. This is because some subjects (Subject 16, for example) presented with non-background-like HRV features throughout the OFC and in order to obtain 100% specificity in training, larger thresholds are selected. The slight differences in the likelihoods also support the data-driven nature of the classification process, and models selected for each process will display the same trends (i.e. departure from background as allergic reaction manifests) but the extent of the trends will vary depending on the models selected.

Subjects 2 and 3 were correctly classified by the manual models, but were not classified correctly with automatic models. In Chapter 5 the likelihoods which were computed for Subject 2 were shown (Figure 5.9, page 126) and it was seen that at approximately 65 minutes a period of the likelihood was classified as allergic. It was also shown that these departures can disappear with poor choices of classifier model with matched classification (see Chapter 5). The extent of the departures obtained with fully automated methods was not sufficient for machine-based classification of allergy for these subjects.

7.4.2 ARTEFACT DETECTION

The requirement for artefact detection in the ECG was discussed in Chapter 6. It will be shown in this section that artefact detection did not reduce the sensitivity, but rather allowed Subjects 7 and 10 to be correctly classified as allergic when, without artefact detection, they would have been incorrectly classified as non-allergic.

Figure 7.3a shows an example of of Subject 7's likelihood trace with epoch lengths of 60 seconds. The threshold is also presented in this figure when artefact detection was not employed. The shaded regions indicate the background and checkup times as

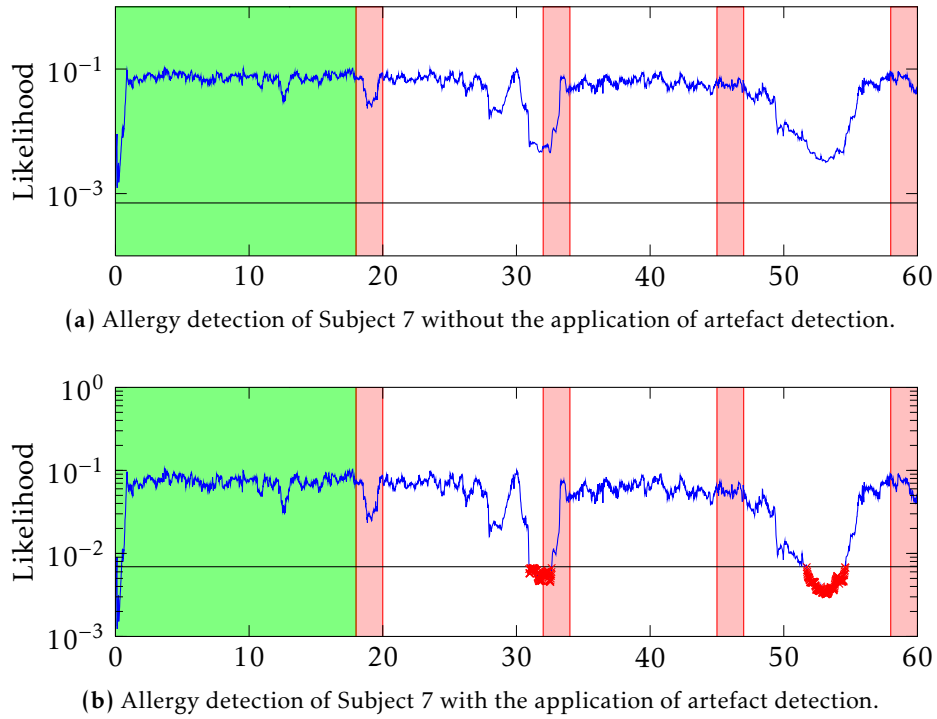


Figure 7.3: Likelihood plots of Subject 7. Subplot (a) shows the threshold which was computed without the aid of artefact detection and how allergy classification does not classify allergy. Subplot (b) shows the threshold which was computed when artefact detection was incorporated and how allergy classification is successful with artefact-aware classification.

before. In this Figure, the threshold can be seen to have been selected at approximately 7×10^{-4} , and this threshold was not surpassed once during the OFC even though two substantiated deviations from the background levels can be seen in the likelihood traces at approximately 30 and 50 minutes.

The reason that a large threshold was selected is because the mean and standard deviation of the background likelihood was contaminated with artefact, which can be seen near $t = 0$. This increased the mean and standard deviation of the background data, and in the calculation of the threshold, this increase was scaled by the multiplicative parameter to the point where the allergic criteria were unable to be satisfied.

Figure 7.3b shows the same likelihood trace, but when the region at the beginning of the challenge was dismissed by the artefact detection algorithm. As a result, the mean

and standard deviation of the background was calculated with only ‘normal’ data and the threshold which was calculated allowed allergy to be classified for this subject with a time gain of approximately 30 minutes.

It should be stated that it is uncertain what occurred at 31 minutes in Figure 7.3. It is possible — owing to the close proximity of this to the checkup at the 34th minute — that the subject became agitated and that the reduced likelihood is as a result of the impending checkup, but it is also possible that the deviation is due to the subject fighting the allergic reaction. It is believed, however, that the latter option occurred in this case due to the fact that no false positives were obtained in any of the experiments that were performed on the allergy database (with the manual and automatic models).

7.4.3 TIME GAIN

Table 7.3 tabulates the time gain metrics which were obtained with fully automated allergy classification. The metrics which are presented are the specific time gain metrics, see Chapter 5. The reason for choosing the specific time gain is because when comparing the total time gain metrics against those obtained by the manual models the comparison will favour the manual models. This is because the total time gain obtained by manual models would be diluted with only one single 0 minute time gain (as 93% sensitivity was obtained), whereas automatic results would contend with three (as 80% sensitivity was obtained), so specific time gain results present a fairer comparison of the performance.

It can be seen from Table 7.3 that in all cases the time gain which was obtained by manual models outperformed the time gain metrics which were obtained from the automatic models. It can also be seen that the time gain metrics which were obtained with the Hilbert transform QRS detection algorithm consistently outperformed those obtained by Afonso’s algorithm even though both algorithms obtained identical sensitivity of detection. However, the mean difference between Hilbert and Afonso time gains is

Table 7.3: Specific time gain metrics obtained from fully automatic allergy classification based on Afonso and Hilbert transform QRS detectors.

QRS detector	Time gain (mins)	Doses saved (portions)	Activation percentage
Afonso	34.62	1.47	53.77%
Hilbert	35.47	1.67	50.63%
Manual	39.15	1.78	31.91%

approximately one minute, which is not a significant difference. The total time gain is a linear scaling of the sensitivity and with both QRS detectors is approximately 30 minutes.

7.5 DISCUSSION

The unmatched classification results that were obtained were not acceptable as suboptimal specificity was obtained. This was due to the data-driven nature of the classification and model generation process, and, with new types of data (i.e. automatic QRS data), poor sensitivity and specificity were obtained with the use of manual models. Because of the disparity between the crossover results and fully manual results, it is stated that crossover classification is not appropriate for the classification of allergy and fully automated classification must be employed.

Fully automatic models (i.e. automatic models generated with automatic data) obtained 100% specificity. This is a very significant result as it means that (due to the subject-independent means in which the results were obtained) equivalence can be attributed to machine-based classification of allergy and clinical diagnosis of allergy. 80% sensitivity was also obtained, and this corresponds to three false negative classifications. While this sensitivity is lower than the sensitivity that was obtained with manual models, automatic classification of allergy can still introduce a significant improvement to the state of the art of allergy diagnosis as OFCs could have been terminated on average over 30 minutes sooner with machine-based classification. This time gain can be used to good advantage by

administering antihistamines, such as Zyrtec, to counter the symptoms of allergy. Zyrtec begins to take effect between 10 — 20 minutes and by achieving time gain of over 30 minutes it might therefore be possible to eliminate the effects of allergy for some subjects who underwent OFC.

The thresholds which were selected for some subjects were too large to obtain classification. Figures 7.2a — 7.2c shows the likelihood traces for Subject 1, and the threshold is outside the limits of the axes in each case. The reason for this is due to the subject-independent means by which the threshold is selected, and the requirement to obtain 100% specificity in training as some subjects contributed to larger deviations from the background. However, OFC will be tolerant to false negative classifications. This is because the same vigilance which is employed during OFC would be employed during monitored OFC, so in the cases of false negatives, the current state of clinical art is not impaired, and safe diagnosis of allergy will always be obtained.

For Subjects 2 and 3 the models which were selected did not exemplify the signatures of allergy that were obtained by the manual models and data. It was stated in Chapter 5 and earlier in this chapter that the choice of an appropriate classification model is important for the detection of signatures of allergy in likelihood traces, and indeed that a poor choice in these parameters can have the effect of under-representing the ‘novelty’ of the HRV data. This occurred with Subjects 2 and 3 with the automatic models, where, upon manual investigation of alternative likelihoods, larger deviations from background levels could have been chosen with automatic models. It is inappropriate to infer whether different model choices would have yielded correct classification, as this would introduce manual analysis, and the aim of this chapter is to assess automated classification of allergy.

7.6 CONCLUSION

This chapter has demonstrated the means by which fully automated classification of allergy can be achieved. Because of the subject-independent nature of the procedure

employed here, the results which were obtained should be representative of results that will be obtained with future data. Importantly, these results obtained no false positive classifications on the unseen data in a fully objective manner, and therefore the machine-based classification of allergy which is presented here is equivalent to clinical diagnosis of allergy. Consequently an important question to consider is whether this procedure might be employed for oral food challenges. Having demonstrated the ability of obtaining 100% specificity and 80% sensitivity, the application of machine-assisted classification is appropriate for automated detection of allergy without compromising the quality of medical diagnoses. Manual annotations (which provide an upper bound of the expected performance) have demonstrated that it is possible to achieve higher sensitivity, but even in light of this, the time gain results which were obtained allow the state of clinical art to be significantly improved by over 30 minutes, which introduces the capability of reducing the effect of allergic reactions on subjects through the early administration of antihistamines. Therefore, the state of the clinical art of food allergy diagnosis can be significantly advanced with digital signal processing and artificial intelligence based monitoring of the heart rate variability features during oral food challenge.

CHAPTER 8

OVERALL SUMMARY, FINAL CONCLUSIONS AND FUTURE WORK

8.1 SUMMARY OF THIS THESIS

THIS thesis has presented an investigation of the applicability of machine-based analyses of non-invasive on-body sensors for the purpose of automatically detecting signatures of food allergy. This work was inspired by two observations of allergists who have conducted food challenges. The observations were that there is both a tendency for subjects to become quiet before the onset of allergic reactions, and that there is also a tendency for the heart rate of the subjects to change before allergic reactions occur (Bindeslev-Jensen et al., 2002).

It was shown that the accelerometer-based approaches investigated in this thesis did not provide a viable means of assessing the state of allergy, even though this is one of the

observations identified by the allergists as being indicative of the onset of allergy. Indeed, this is, perhaps, the more identifiable of the two observations for the allergists as it is easily observed. The reason for the poor efficacy of acceleration-based analysis is that the range of activities — and therefore the movements and related energy expenditure — that the subjects can undertake is limited by the fact that subjects are required to remain on a bed for the duration of the oral food challenge. As a result of this, no separability was achieved between allergic and non-allergic subjects with activity-based metrics and with energy expenditure estimation algorithms. It is possible that by allowing subjects to play freely that a means of allergy detection based on activity and energy analyses could be formed. Furthermore, it is possible that with placebo-based challenges that more non-allergic data would be available for model generation. However, these new procedures would require a new brand of oral food challenge and were not investigated.

The second observation was then investigated. In order to confirm the existence of signatures of allergy in the ECG, the QRS points were manually annotated. Heart rate variability features were then extracted from these annotations, and the features were separated based on whether the data originated from an allergic subject. By these means, it was possible to assess the characteristic differences between the allergic and non-allergic heart rate variability features. This is the first work which has investigated this, and indeed it was shown that, even in single-dimensional representations, many of the features extracted from the ECG obtained separability between the allergic and the non-allergic classes. In comparison to the separation capabilities which were obtained with the accelerometer data, the heart rate variability data provides a more suitable platform on which to base machine learning algorithms.

GMM-based novelty detection was employed on manually annotated HRV feature data with nested, subject-independent parameter selection and performance assessment routines. This type of classification was performed because the only set of labelled data available was the background data which was recorded before the administration of the first dose of the allergen. By these novelty detection means, 93% sensitivity and 100% specificity were obtained. These metrics also facilitated in obtaining approximately

39 minutes time gain after consumption of just 30% of the dose which can drastically reduce the risk of reactions. This is the first research which has been performed for the classification of allergy based on the heart rate variability features. The results show that this approach is very suitable and that through these means excellent classification, perfect specificity and very high time gain results can be obtained. Manual QRS annotations yield undeniable affirmation of this as there is no uncertainty about the validity of the QRS points and therefore the features.

These results were obtained with manually annotated QRS points, and therefore represent the upper-bound of the expected results that could be obtained with automatically extracted heart rate variability features. Two automatic QRS detection algorithms were therefore investigated, and these were validated against the well-known MIT-BIH arrhythmia database. These algorithms were then tested against the ECG which was recorded during the oral food challenges. Both of these algorithms performed satisfactorily against both databases, and were therefore employed for the fully automated and autonomous classification of allergy.

From the automatically identified QRS points, heart rate variability features were extracted. These were employed with the subject-independent classification framework mentioned previously. It was found that 100% specificity can be obtained by this process. While this was previously obtained with manual classification models, this is a very significant result as they were obtained in a fully automated manner. For this fully-automated classification of allergy framework, therefore, it can be stated that classification of allergy is equivalent to a diagnosis of allergy because no false positive classifications were obtained.

Sensitivity of 80% was obtained by the fully automated routine. This value is lower than that which was obtained with manual annotations, but the time gain metrics which were obtained with this data were very strong, and approximately 30 minutes would have been saved on average. While this time gain is also lower than that which was obtained with the manual classification models, achieving this in a fully automated manner is a

very significant result. This time gain could be employed to administer antihistamines to the subjects which could potentially eliminate allergic reactions for some subjects. Alternatively, this time gain could be employed to allow the subjects to recover from the allergic condition naturally, as with no additional doses of the allergen administered it may be possible that the subjects could ‘fight’ the reaction and overcome the symptoms naturally.

8.2 PRIMARY CONTRIBUTION OF THIS THESIS

Allergy is a chronic disorder which can only be diagnosed in a clinically invasive manner. This thesis has been the first to investigate two approaches to the classification of allergy. It was shown previously that accelerometer-based assessment of activity (Twomey et al., 2010b) during oral food challenges is unable to characterise the allergic and non-allergic classes in a manner which allows separability. This thesis has ruled out a number of avenues for its applicability with allergy classification.

However, it was discovered that allergy affects the heart rate variability in a detectable manner where it can be exploited to classify the condition (Twomey et al., 2010a, 2011, 2013b). It was also discovered that signatures of allergy are sometimes better exemplified at different epoch lengths. By employing novel result fusion in a subject-independent manner, subtle characteristics of the signatures of allergy can be identified and classification of allergy can be achieved in an objective manner that obtains 100% specificity. This allows equivalence to be drawn between the automatic classification and clinical diagnosis of allergy.

Signatures of allergy were repeatedly identified on the heart rate variability features before the onset of physical reactions. In fact, this thesis has shown how signatures of allergy can be classified approximately 30 minutes earlier than allergists with the consumption of approximately two-thirds less of the problem foods. The significance of this is that

subjects could be diagnosed with the same certainty and accuracy, but with a reduced risk of allergic reactions and anaphylaxis (Twomey et al., 2013a).

The chapters of this thesis described the means to significantly and safely advance the current state of clinical art of allergy diagnosis in an objective and automatic manner, and this work, therefore, supports the case for the inclusion and use of intelligent monitoring for the diagnosis of allergy during oral food challenges.

8.3 POSSIBLE AVENUES OF FUTURE WORK

This thesis has explored a number of approaches to classifying food allergy. There are other avenues which deserve examination that could not be investigated in this thesis. These are briefly described below.

8.3.1 DATA COLLECTION

The results obtained excellent and consistent results, which, due to the subject independent nature of the classification routine, is indicative of the relationship between allergy and the HRV. In order to further define the effect of allergy on the HRV features, a larger database on which to test this classification routine should be obtained. This should be the primary focus of any researcher who takes on this research, and with this, other classification methods, such as bootstrapping, can be investigated with ‘fixed’ models. In contrast to the leave one out procedure that was used in previous chapters, bootstrapping utilises multiple test subjects and would validate the robustness of one model over multiple participants in the test set.

8.3.2 ALTERNATIVE NOVELTY DETECTORS

Chapters 2 and 5 discussed the area of novelty detection, and the reasons that GMM-based classification was selected. However, while GMM-based novelty detection was stated as being preferable for the initial analysis, investigation into alternative classification routines should also be performed. Two alternative novelty detection routines which might be analysed include:

One-class SVM: The principal alternative is the one-class SVM. These classifiers are the single-class equivalent to the popular SVM discriminative classifier (Schölkopf et al., 2000, 2001) and these estimate the boundary of support of the singly-labelled data, efficiently making this a discriminative novelty detection classifier. Whereas GMM-based novelty detection computed the likelihood of new data belonging to the background class, one-class SVMs compute the distance of new data to the surface of a hypersphere which surrounds the background distribution. One-class SVMs can also employ the ‘kernel trick’ (Aizerman et al., 1964) which can perform non-linear mapping of features to higher-dimensional feature space. This has proven to be useful for many applications, and may prove to be a useful feature for allergy classification.

Elliptic envelope: Minimum covariance determinant estimator (Rousseeuw and Van Driessen, 1999) can also be employed to detect novelty as described here. This algorithm effectively fits elliptic curves about training distributions, and benefits from rotational symmetry which one-class SVMs and GMMs may not encompass. This method works well for normal (and normal-like) data, but with multimodal distributions, elliptic envelopes have a tendency to poorly model the underlying structure as it cannot generalise to multimodal distributions unlike GMMs and GMMs. Yet, it might yield interesting results with allergy detection as it should result in a simpler model representation.

8.3.3 REAL-TIME AND PORTABLE IMPLEMENTATION

All of the classification and analyses which were performed here were performed off-line, i.e. the data was recorded and stored, and the data interrogation was performed on a computer. While the results which were presented at the end of this thesis are fully automated, it would be interesting to verify that the same classification results are obtained in a real-time solution. Preliminary work for this has been performed already (Gutiérrez et al., 2013).

For a fully-wireless solution, a mobile device would record the ECG data, and compute features and their novelty. However, as it might be necessary to review the decisions that the mobile solution obtained, it would be necessary to log the ECG data electronically and transfer this via a wireless link. These two operations are heavy consumers of power in mobile platforms. With lossy compression, however, the data which needs to be transmitted can be compressed to a high degree. Twomey et al. (2010c) demonstrated that with a compression ratio of 30, reliable heart rate variability features can be extracted, and the area of ECG compression has also been investigated by other researchers (Olmos et al., 1996; Craven et al., 2012). Analysis of the effect of compression on classification accuracy and overall time gain would unveil the applicability of the lossy compression solution for long-term monitoring situations.

8.3.4 FEATURE AND EPOCH ANALYSIS

It was demonstrated that in some cases different epoch lengths display the effects of certain characteristics of allergy better than others. It would be useful to investigate if features which are employed with speech classification (Temko et al., 2010, 2011a) or EEG classification (Doyle et al., 2010) might be suitable in obtaining accurate classification of allergy with fusion of fewer epoch lengths.

Additional insight could be gained from quantifying the effectiveness of the individual features which were employed for the classification. Preliminary work which investigates the importance of the categories of features was performed on this in Appendix B, and it is concluded that the full set of features is required for the best classification and time gain results. However, fuller investigation is justified, and analysis of the results of algorithms such as recursive feature elimination could be assessed.

8.4 PUBLICATIONS RESULTING FROM THIS WORK

The publications which have resulted from this research and which have been published, are currently in review and preparation are listed below.

Twomey, Niall and Faul, Stephen and Marnane, William P (2010); Comparison of accelerometer-based energy expenditure estimation algorithms; *Pervasive Computing Technologies for Healthcare*; pages 1–8.

Twomey, Niall and Walsh, Noel and Doyle, Orla and McGinley, Brian and Glavin, Martin and Jones, Edward and Marnane, WP (2010); The effect of lossy ECG compression on QRS and HRV feature extraction; *Engineering in Medicine and Biology Society*; pages 634–638.

Twomey, Niall and Faul, Stephen and Daly, Deirdre and Hourihane, JO and Marnane, William P (2010); Classification of biophysical changes during food allergy challenges; *International Symposium on Applied Sciences in Biomedical and Communication Technologies*; pages 1–5.

Twomey, Niall and Temko, Andrey and Hourihane, Jonathan O'B and Marnane, William P (2011); Allergy detection with statistical modelling of HRV-based non-reaction baseline features; *International Symposium on Applied Sciences in Biomedical and Communication Technologies*; pages 134–138.

Twomey, Niall and Temko, Andrey, and Cullinane, Claire, and Daly, Deirdre, and Marnane, William P, and Hourihane, Jonathan O'B (2013); Detection of heart rate variation could improve patient safety and diagnostic yield during oral food challenge; *European Academy of Allergology and Clinical Immunology*; *In press*.

Twomey, Niall and Gutiérrez, Raquel, and Marnane, William P, and Campos-Garcia, Jesús (2013); Real-Time Allergy Detection; *IEEE Society of Intelligent Signal Processing*; *In press*.

Twomey, Niall and Temko, Andrey and Hourihane, Jonathan O'B and Marnane, William P (2013); Fully automated allergy detection from paediatric ECG; *IEEE Transactions on Information Technology in Biomedicine*; *In press*.

APPENDIX A

ALTERNATIVE PARAMETER SELECTION ROUTINES

A.1 INTRODUCTION AND METHODS

THIS chapter investigates the effect of alternative model selection routines. Previously, decision making parameters were chosen based on a cost function and candidates which did not satisfy the cost function were eliminated from consideration. This process resulted in the elimination of many tens of thousands of parameters. While this process yields excellent classification and time gain results, this chapter investigates whether the optimal results are obtained by employing alternative parameter selection routines based on less-austere cost functions.

In Chapter 5, the means by which the parameters were chosen was discussed. The number of parameters which are selected is strongly influenced by the constraint of selecting the

parameters which obtain the maximum time gain in the training data, and in general only one parameter is selected from the entire search space. The effect of relaxing this restriction on the data, and alternative means of selecting post-processing parameters was investigated here.

The statistical mode (i.e. the most frequently occurring parameter) was the means by which parameters were selected in Chapter 5. The parameter search space is reduced by means of a cost function which assesses the entire search space and reduces the quantity by the following three factors (listed in order of importance):

1. Eliminate parameters which fail to obtain 100% specificity.
2. Eliminate parameters which fail to obtain the maximum sensitivity from the subset resulting from step 1.
3. Eliminate parameters which fail to obtain the maximum time gain from the subset resulting from step 2.

While item 1 is the most important (followed by items 2 and 3) the number of parameters which are eliminated by each step are ordered in descending order of importance of each step, i.e. step 3 eliminates the majority of the parameters while step 1 eliminates the fewest. Approximately ten thousand parameters are removed by step 3*, and it is by considering the set of ten thousand parameters that this alternative parameter selection is performed.

Figure A.1 shows an example density of parameters obtained before the time gain criteria eliminated parameters. The density distribution is plotted against the multiplicative and duration parameters, and darker regions indicate areas of higher density. Parameter selection which terminates at item two of the itemisation is termed 'relaxed parameter selection' henceforth, as the time gain criteria is not considered by this process and this is

*This number is dependent on many factors, but 10,000 was calculated based on the average number of parameters which were obtained over all of the internal leave-one-out stages from the simulations which were performed.

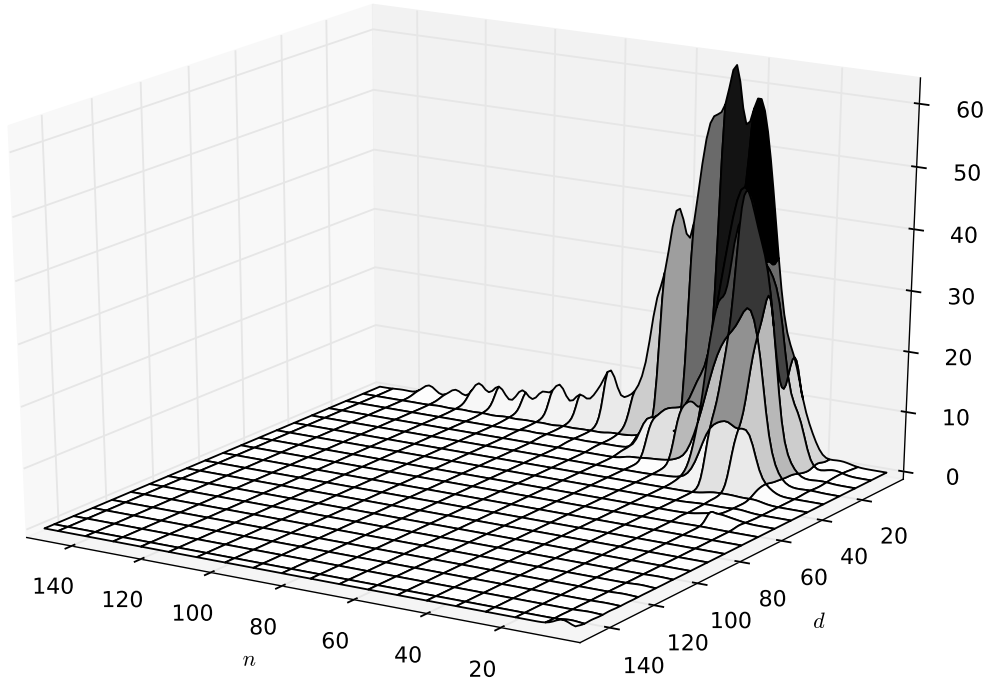


Figure A.1: The estimated distribution of duration and multiplicative post-processing parameters which achieve 100% specificity and maximum sensitivity on the training dataset. The image is limited to d and n parameters of 75. The darker regions indicate a higher density of suitable parameters.

viewed as a relaxation of the routine. The parameters which result from this are termed the ‘relaxed parameter set’. For clarity, the original parameter selection routine from Chapter 5 is termed the ‘original parameter selection routine’ in this section and the ‘original parameter set’ are the parameters which result from this.

It can be seen in Figure A.1 that at a duration of 1 yields the highest density of parameters. This is an intuitive value as with a constant multiplicative parameter and an increasing duration parameter, there will be a larger quantity of points which satisfy a smaller duration parameter than would a larger parameter.

It is also expected that with a constant duration parameter and an increasing multiplicative parameter that the density will increase to a maximum value and will then reduce, as is shown in Figure A.1. This is because low multiplicative parameters will

not satisfy the 100% specificity criteria and are rejected from step 1 and few parameters are chosen. As the multiplicative parameter increases an increasing number of points are selected. However, as the parameter continues to rise the number of points which satisfy the maximum sensitivity criteria begins to roll off, and fewer points are selected as the maximum sensitivity is not attained.

Ripples which are seen on the chart for increasing multiplicative values are due to the distribution of the duration and multiplicative parameters. The search space consists of the unique, integer-rounded values which are logarithmically distributed between 1 and 300. The density chart was obtained by KDE and in Figure A.1 the x - and y -axes are continuously distributed while the search space is not. Therefore, when the density of the parameters is estimated between the discrete values, the density reduces until the next parameter has been encountered.

This section investigates the effect of selecting the mean, median and statistical mode of the distribution of the relaxed parameter set.

A.2 RESULTS

Table A.1 tabulates the sensitivity, specificity and time gain metrics which were obtained when employing the relaxed parameter selection constraints. The results which were presented in this table were calculated by the boosting classification procedure described in Chapter 5.

The mean selection routine achieved the same sensitivity as the results from the original parameter selection routine. Subject 16 was misclassified as allergic by this parameter selection routine, however. This is the same subject who was stated in Chapter 5 as being physically opposed to consuming the food type. Even though the subject ended up not being allergic to the food type, given the subject's attitude during the OFC, it is likely that allergists performing the challenge would not utilise automatic classification.

Table A.1: Tabulation of sensitivity, specificity, and the time gain metrics which were obtained by selecting the mean, median and mode of the set of post-processing parameters from the training data. In the case of the mean method, imperfect specificity was obtained.

Selection type	Sensitivity %	Specificity %	Time gain mins
Mean	93.33	88.88	28.36
Median	86.66	100.00	21.75
Mode	93.33	100.00	20.29
Chapter 5	93.33	100.00	36.5

The cause for the misclassification of Subject 16 is that the duration and multiplicative parameters are typically biased in a region where the threshold can be surpassed by both allergic and non-allergic subjects. The duration parameter will reject spurious deviations from the background and only substantiated departures from the background will be classified as allergy. However, upon investigation of Figure A.1, it can be seen that a high proportion of duration parameters at a value of 1 are obtained. As a result, when computing the mean parameter of the distribution, a lower duration parameter was selected, and Subject 16 was misclassified as allergic as a result.

The median and mode selection methods obtained perfect specificity, and sensitivity of 87% and 93% respectively; results which are competitive with those obtained by the original parameter selection criteria. It is interesting to note, however, that for both full and relaxed selection criteria, the parameter selection routine based on the mode of the distributions obtained the same sensitivity and specificity.

Table A.1 also shows the mean specific time gain which was calculated from these parameter selection methods. Of the values obtained with the relaxed parameter selection, parameter selection based on the mean of the distribution gain obtained the largest time gain. This is also due to the high percentage of low-value duration parameters in the distribution. The selection routines based on the median and mode functions achieved approximately equal time gain. In all cases, however, the total time gain obtained from the original parameter selection criteria outperform the relaxed criteria.

A.3 DISCUSSION

Alternative parameter selection routines can yield classification results which are as accurate as the parameter selection routine which was described in Chapter 5 (sensitivity of 93.33% and 100% specificity). This is because the full parameter selection routine selects parameters which have been ‘known’ (from training data) to obtain the best time gain results, and this high accuracy was preserved between training and unseen testing data.

In many cases the signature of allergy is detected a number of times during the OFC, as is shown by Figure 5.8 (Chapter 5, page 123) where the allergy is classified at approximately 45, 60 and 80 minutes. With the original parameter selection routine, the multiplicative and duration parameters were selected in such a way as to specialise towards finding the first signature of allergy from the training data. With a plurality of departures from the background likelihoods, it was observed that there is a tendency that the extent of departure from the background levels increases in comparison to the previous departure. Therefore, smaller multiplicative parameters will yield better time gain results as they provide the greatest likelihood of satisfying the failure criteria.

However, the relaxed parameter selection routine gives no consideration to the time gain values when selecting parameters. Hence the sensitivity and specificity results which were obtained were excellent and in some cases matched the sensitivity and specificity results obtained in Chapter 5.

A.4 CONCLUSION

This section investigated whether the parameter selection routine utilised previously selected the optimal parameters in comparison to other parameter selection routines.

It was shown that alternative selections can yield equal sensitivity and specificity to what was obtained in Chapter 5. However, as the selection routine is not time-gain aware, the time gain results were approximately one half of those which were obtained with time gain aware parameter selection.

Therefore, the parameter selection routine which was incorporated previously is the best methodology of all obtained, as it achieved the best accuracy and time gain results and this process should be used for parameter selection .

APPENDIX B

INVESTIGATION INTO THE IMPORTANCE OF FEATURES

B.1 INTRODUCTION AND METHODS

FOR the classification of allergy, PCA is performed on normalised training data. This transformation was performed in order to de-correlate the features which was a requirement for allergy classification as the allergy database is insufficiently sized to train GMMs with full covariance matrices.

With the transformed features, GMMs were generated. The EM algorithm is employed to compute the optimal means, covariances and weights of the mixture model. It is possible to investigate the explained variances of the PCA matrices, (Wold et al., 1987) and the weights which are attributed to the features in order to assess the relative significance of the features. However, these values relate to the PCA components rather than the features

themselves. It is not appropriate to assign interpretation to the transformed features, in particular when normalisation was performed (Webb et al., 2011; Bishop et al., 2006) so explained variances and Gaussian weights cannot be utilised to assess the importance of the features.

It is also difficult to employ the probability density functions which were presented in Chapter 4 as a means of quantifying the effectiveness of individual features. This is because these will contain a significant amount of non-allergic data (as non-allergic HRV features will present in an allergic subject's challenge until an allergic reaction occurs).

Therefore, in order to quantify the importance of feature categories, the allergy classification framework utilised previously was employed on subsets of features, where each subset is representative of a feature type. The full set of features can be loosely grouped into time, frequency, Poincaré and sequential groups (see Chapter 4 for the members of each group). Classification was then performed once for each subset of feature category. Based on the results of this process, higher feature importance is attributed to feature sets which obtain superior results. The order of importance of the results are as follows:

1. 100% Specificity.
2. Sensitivity.
3. Time gain.

and ranking is achieved first by descending specificity, then by descending sensitivity and finally by descending time gain.

Two key differences between the classification framework outlined in Chapter 5 and this process outlined here exist. Firstly, as the process outlined here performs classification based on a subset of features, the dimensionality of the input to the classification framework is reduced from 18 to N , where N is the number of features associated with a particular category. The second difference is that the full set of N features is then preserved after the PCA transformation, and dimensionality reduction is not

performed. This procedure was chosen because this process was executed to assess the significance of feature categories. Allowing feature reduction on the feature categories would not illuminate the importance of the feature categories as only the importance of the components could be attained.

The models which were generated are termed time, frequency, Poincaré and sequential domain models and these relate directly to the category of the features on which the models were trained.

B.2 RESULTS

Table B.1 tabulates the results which were obtained. This table tabulates the sensitivity, specificity, total time gain, doses saved and activation percentage metrics for all models investigated (for definitions of these metrics, see Chapter 5). The time gain metrics which are presented are the total time gain results which were obtained by the experiment. This metric was chosen over the specific time gain parameter because the overall performance of classification is of interest rather than the specific performance on certain correctly classified subjects.

Table B.1: Classification metrics which were obtained with the time–, frequency–, Poincaré– and sequential-domain classification models ranked by order of importance of the feature category in question.

Feature Category	Specificity	Sensitivity	Time gain (mins)	Doses saved (portions)	Activation percentage
Time domain	100.00%	80.00%	30.25	1.27	54.44%
Poincaré domain	100.00%	53.33%	15.63	1.0	69.19%
Frequency domain	77.77%	6.66%	0.65	0	100.00%
Sequential domain	66.66%	13.33%	0.06	0.13	95.55%
Chapter 5	100%	93.33%	36.5	1.67	38.57%

It can be seen that time-domain features achieved good classification metrics in all cases, obtaining 100% specificity and 80% sensitivity. Interestingly, Subjects 1, 3 and 4 were misclassified by the time domain models. It is not surprising that Subjects 1 and 3 were misclassified as Subject 1 wasn't classified correctly by any classification routine and Subject 3 was inconsistently classified at different epoch lengths in Chapter 5. However, this is the first instance over all classification runs up to this point where Subject 4 is misclassified as non-allergic. It is interesting to note that while the time domain models achieve very good overall performance, one of the most consistently classified subjects of the allergy database is misclassified by the highest ranked feature category.

The frequency domain models achieved very poor results. This was an unexpected result because the frequency domain parameters can estimate the sympathetic and parasympathetic response of the ANS (Kamath et al., 1993; Stein et al., 1999; Hedman et al., 2008). It was postulated that quantification of these metrics would be good indicators of allergic signatures in Chapter 4, as a subject's ANS should react to the onset of an allergic reaction. It is uncertain why such poor classification of allergy was achieved with frequency domain features. However, upon investigation of the histograms which were obtained in Chapter 4, it can be seen that poor separation was obtained between the allergic and non-allergic categories. It is interesting to note that Subject 4 was the only subject who was correctly classified by these models when this subject was one of the subjects who was misclassified by the time domain models. The frequency domain models also misclassified two non-allergic subjects as allergic, and achieved very low average time gain as so few subjects were correctly classified.

The Poincaré models performed adequately, yielding 53% sensitivity and 15 minutes average time gain. However, Poincaré features also measure the response of the ANS (Brennan et al., 2002; Mourot et al., 2004; Piskorski and Guzik, 2005) but the models that were generated based on these features outperformed the frequency-domain models which can also measure the response ANS. It is also interesting to note that the Poincaré features identified allergy in Subject 3, who has been inconsistently classified. Indeed,

this is the only feature category which correctly classified Subject 3 as allergic. Poincaré features obtained 100% specificity.

The sequential domain models performed poorest of all, misclassified three non-allergic subjects, obtained the lowest specificity (66.66%), and only correctly classified two allergic subjects (13.33% sensitivity). It is very interesting to note that one of the two subjects who was classified correctly as allergic was Subject 1 (obtaining a time gain of approximately 15 seconds). This is the only case in which the parameters which were selected achieved correct classification of this subject. However, while correct classification of Subject 1 occurred, it should be noted that the majority of the remaining allergic subjects were incorrectly classified.

B.3 DISCUSSION

It must be noted that the results which were obtained in this section are not intended to compete with the classifiers discussed in previous chapters. The purpose of this section is to assess the importance of each feature category and to investigate whether the optimal results were obtained in Chapter 5 with the full set of features.

When ranking the feature categories based off the importance of the results, the most important feature category was found to be the time-domain features. These features yielded 100% specificity, 80% sensitivity and an average of 30 minutes time gain. Three subjects were misclassified as non-allergic by these features because the likelihoods which were selected did not deviate from background levels to the same degree as they did with manual models. Upon investigation of these subjects, it was seen that two were correctly classified by the manual models. The cause for misclassification of these subjects is that the time domain features alone did not exhibit signatures of allergy for these subjects.

The specific time gain of the time domain features, approximately 37 minutes ($30.25 \times \frac{15}{12}$ minutes), is very close to the specific time gain obtained by manual models (39.15

minutes). Therefore, the time domain models are the biggest contributor of time gain to allergy classification as well as the best performing individual feature category.

Interestingly, classification based on the sequential domain features correctly classified Subject 1 as allergic. This is the first and only instance where this Subject was correctly classified as allergic. While this subject was correctly classified, the sensitivity and specificity which were obtained with the sequential domain classification models was quite poor. Therefore, it appears that the sequential domain features are suitable for the classification of quick-onset allergic reactions, but these features do not appear to be able to well distinguish between allergic and non-allergic subjects. The number of quick-onset allergic reactions is not sufficient to substantiate this assertion with certainty. However, evidence based on previous models support the argument as these correctly classified the entire set of allergic subjects except for Subject 1. Indeed, as this feature achieved poor specificity it is likely that contributions of features in this category were removed by feature reduction during PCA model selection.

Subject 16 was one of the three non-allergic subjects who were misclassified as allergic by the sequential features. This subject was discussed in Chapter 5 and it was stated that they presented with allergy-like signatures in the likelihoods which were obtained. As a result, the likelihood values computed for this subject resembled those which might be obtained by allergic subjects.

The Poincaré models yielded 100% specificity and approximately 50% sensitivity. However, while this feature category did not yield as high a sensitivity as the time domain features, Subjects 3 and 4, who were both misclassified by the time domain models, are correctly classified by the Poincaré models. If one considered the fusion of the results of the time domain and Poincaré models, one would achieve the same accuracy of classification that was obtained in Chapter 5.

The fact that the combination of the time domain and Poincaré domain features matches the sensitivity obtained previously is a retrospective result which sheds light on the importance of considering combinations of features, and should not be interpreted as

the best of features which will obtain the optimal features results. The reason for this is that the choice of time domain and Poincaré domain features together might yield results which are comparable to those obtained by the manual models, but they may not yield the optimal results, and inclusion of less important features may yield better results.

One of the Poincaré features computes the CSI which is a measure of the sympathetic response of the autonomic nervous system. However, while the frequency domain features also measure this response, the frequency domain parameters did not achieve good classification results, achieving the second lowest specificity, and the lowest sensitivity.

B.4 CONCLUSION

This section performed allergy classification on subsets of features for the purpose of assessing the importance of each category of feature in allergy classification and identifying whether the results obtained in Chapter 5 could be improved.

It was discovered that the time domain features account for the majority of the sensitivity and time gain that was obtained in the Chapter 5. However, the sensitivity obtained with time gain models alone did not account for the entirety of the results obtained in previous chapters. For the optimal results, the inclusion of other features complement the classification process and yield the best results.

However, correct classification of Subject 1 was obtained by the sequential domain features. This indicates that there might be a possibility of obtaining perfect sensitivity of classification. It is uncertain whether it might also be possible to obtain perfect specificity as the sequential domain features achieved very poor specificity. However, as the most important criterion for classification of allergy has consistently been stated as obtaining 100% specificity, it is believed that the overall results which were obtained by utilising the full set of features cannot be improved upon, and that the full set of features should be employed for allergy classification.

REFERENCES

- Afonso, V., Tompkins, W., Nguyen, T., and Luo, S. (1999). ECG beat detection using filter banks. *IEEE Transactions on Biomedical Engineering*, 46(2):192–202.
- Afonso, V., Tompkins, W., Nguyen, T., Trautmann, S., and Luo, S. (1995). Filter bank-based processing of the stress ECG. In *Engineering in Medicine and Biology Society*, pages 887–888.
- Ahlbom, A., Backman, A., Bakke, J., Foucard, T., Halken, S., Kjellman, N., Malm, L., Skerfving, S., Sundell, J., and Zetterstrom, O. (1998). Pets Indoors—A Risk Factor For or Protection Against Sensitisation/Allergy. *Indoor Air*, 8(4):219–235.
- Ahlstrom, M. and Tompkins, W. (1983). Automated high-speed analysis of Holter tapes with microcomputers. *IEEE Transactions on Biomedical Engineering*, 30(10):651–657.
- Ainsworth, B., Haskell, W., Herrmann, S., Meckes, N., Bassett, D., Tudor-Locke, C., Greer, J., Vezina, J., Whitt-Glover, M., and Leon, A. (2011). 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values. *Medicine and Science in Sports and Exercise*, 43(8):1575–1581.
- Ainsworth, B., Haskell, W., Leon, A., Jacobs, D., Montoye, H., Sallis, J., and Paffenbarger, R. (1993). Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise*, 25(1):71–80.

- Ainsworth, B., Haskell, W., Whitt, M., Irwin, M., Swartz, A., Strath, S., O'Brien, W., Bassett Jr, D., Schmitz, K., Emplaincourt, P., et al. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, 32(9):498–504.
- Aizerman, A., Braverman, E. M., and Rozoner, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25(1):821–837.
- Almqvist, C., Egmar, A., Hedlin, G., Lundqvist, M., Nordvall, S., Pershagen, G., Svartengren, M., Hage-Hamsten, M., and Wickman, M. (2003). Direct and indirect exposure to pets—risk of sensitization and asthma at 4 years in a birth cohort. *Clinical and Experimental Allergy*, 33(9):1190–1197.
- Aloise, D., Deshpande, A., Hansen, P., and Popat, P. (2009). NP-hardness of Euclidean sum-of-squares clustering. *Machine Learning*, 75(2):245–248.
- Andersen, R., Borgs, C., Chayes, J., Hopcroft, J., Jain, K., Mirrokni, V., and Teng, S. (2008). Robust PageRank and locally computable spam detection features. In *Adversarial information retrieval on the web*, pages 69–76.
- Annadhorai, A., Guenterberg, E., Barnes, J., Haraga, K., and Jafari, R. (2008). Human identification by gait analysis. In *Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, page 11.
- Antelmi, I., De Paula, R. S., Shinzato, A. R., Peres, C. A., Mansur, A. J., and Grupi, C. J. (2004). Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease. *The American Journal of Cardiology*, 93(3):381–385.
- Arshad, S., Kurukulaaratchy, R., Fenn, M., and Matthews, S. (2005). Early Life Risk Factors for Current Wheeze, Asthma, and Bronchial Hyperresponsiveness at 10 Years of Age*. *Official Journal of American College of Chest Physicians*, 127(2):502–508.
- Arzeno, N., Deng, Z., and Poon, C. (2008). Analysis of first-derivative based QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 55:478–484.

- Arzeno, N., Poon, C., and Deng, Z. (2006). Quantitative analysis of QRS detection algorithms based on the first derivative of the ECG. In *Engineering in Medicine and Biology Society*, pages 1788–1791.
- Avery, N., King, R., Knight, S., and Hourihane, J. (2003). Assessment of quality of life in children with peanut allergy. *Pediatric Allergy and Immunology*, 125(6):1327–1335.
- Aziz, W., Schlindwein, F. S., Wailoo, M., Biala, T., and Rocha, F. C. (2012). Heart rate variability analysis of normal and growth restricted children. *Clinical Autonomic Research*, 261(5):480–487.
- Azpiri, A., Alonso, E., Gamboa, P., Jauregui, I., Antepará, I., Fernandez, E., Fernández de Corres, L., Audicana, M., Muñoz, D., and Escobar, A. (1999). Prevalence of pollinosis in the Basque Country. *European Journal of Allergy and Clinical Immunology*, 54(10):1100–1104.
- Badilini, F. and Blanche, P. (1996). HRV Spectral Analysis by the Averaged Periodogram. *Annals of Noninvasive Electrocardiology*, 1(4):423–429.
- Bahoura, M., Hassani, M., and Hubin, M. (1997). DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis. *Computer Methods and Programs in Biomedicine*, 52(1):35–44.
- Bailón, R., Laguna, P., Mainardi, L., and Sörnmo, L. (2007). Analysis of heart rate variability using time-varying frequency bands based on respiratory frequency. In *Engineering in Medicine and Biology Society*, pages 6674–6677.
- Bailón, R., Mainardi, L., Orini, M., Sörnmo, L., and Laguna, P. (2010). Analysis of heart rate variability during exercise stress testing using respiratory information. *Biomedical Signal Processing and Control*, 5(4):299–310.
- Baldzer, K., Dykes, F., Jones, S., Brogan, M., Carrigan, T., and Giddens, D. (1989). Heart rate variability analysis in full-term infants: spectral indices for study of neonatal cardiorespiratory control. *Pediatric research*, 26(3):188–195.

- Barold, S. S. (2003). Willem Einthoven and the birth of clinical electrocardiography a hundred years ago. *Cardiac Electrophysiology Review*, 7(1):99–104.
- Benbasat, A. and Paradiso, J. (2002). An inertial measurement framework for gesture recognition and applications. *Gesture and Sign Language in Human-Computer Interaction*, 2298(1):9–20.
- Benitez, D., Gaydecki, P., Zaidi, A., and Fitzpatrick, A. (2000). A new QRS detection algorithm based on the Hilbert transform. In *Computers in Cardiology*, pages 379–382.
- Benitez, D., Gaydecki, P., Zaidi, A., and Fitzpatrick, A. (2001). The use of the Hilbert transform in ECG signal analysis. *Computers in Biology and Medicine*, 31(5):399–406.
- Bergmann, R., Diepgen, T., Kuss, O., Bergmann, K., Kujat, J., Dudenhausen, J., and Wahn, U. (2002). Breastfeeding duration is a risk factor for atopic eczema. *Clinical and Experimental Allergy*, 32(2):205–209.
- Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., and Sleight, P. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *Journal of the American College of Cardiology*, 35(6):1462–1469.
- Biala, T., Dodge, M., Schlindwein, F. S., and Wailoo, M. (2010). Heart rate variability using Poincaré plots in 10 year old healthy and intrauterine growth restricted children with reference to maternal smoking habits during pregnancy. In *Computing in Cardiology*, pages 971–974.
- Bindslev-Jensen, C., Briggs, D., and Osterballe, M. (2002). Can we determine a threshold level for allergenic foods by statistical analysis of published data in the literature? *Allergy*, 57(8):741–746.
- Bishop, C. et al. (2006). *Pattern recognition and machine learning*. springer New York.
- Black, P., Udy, A., and Brodie, S. (2000). Sensitivity to fungal allergens is a risk factor for life-threatening asthma. *European Journal of Allergy and Clinical Immunology*, 55(5):501–504.

- Boardman, M. A., Schlindwein, F. S., Thakor, N. V., Kimura, T., and Geocadin, R. (2002). Detection of asphyxia using heart rate variability. *Medical and Biological Engineering and Computing*, 40(6):618–624.
- Bock, S., Muñoz-Furlong, A., and Sampson, H. (2001). Fatalities due to anaphylactic reactions to foods. *Journal of Allergy and Clinical Immunology*, 107(1):191–193.
- Bock, S., Muñoz-Furlong, A., and Sampson, H. (2007). Further fatalities caused by anaphylactic reactions to food, 2001–2006. *The Journal of Allergy and Clinical Immunology*, 119(4):1016–1018.
- Bock, S. and Sampson, H. (1994). Food allergy in infancy. *Pediatric Clinics of North America*, 41(5):1047.
- Bolton, R. and Westphal, L. (1981a). Hilbert Transform Processing of ECG's. In *IREECON*.
- Bolton, R. and Westphal, L. (1981b). Preliminary results in display and abnormality recognition of Hilbert transformed ecgs. *Medical and Biological Engineering and Computing*, 19(3):377–384.
- Bolton, R. and Westphal, L. (1984). On the use of the Hilbert Transform for ECG waveform processing. *Computers in Cardiology*, 19:533–536.
- Bolton, R. and Westphal, L. (1985). ECG display and QRS detection using the Hilbert Transform. *Computers in Cardiology*, 31(1):399–406.
- Bořil, H., Boyraz, P., and Hansen, J. (2012). Towards Multimodal Driver's Stress Detection. *Digital Signal Processing for In-Vehicle Systems and Safety*, 132(1):3–19.
- Bouten, C., Koekkoek, K., Verduin, M., Kodde, R., and Janssen, J. (1997a). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3):136–147.
- Bouten, C., Sauren, A., Verduin, M., and Janssen, J. (1997b). Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Medical and Biological Engineering and Computing*, 35(1):50–56.

- Bouten, C., Westerterp, K., Verduin, M., and JANSSEN, J. (1994). Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine and Science in Sports and Exercise*, 23(1):21–27.
- Bradley, T. D. and Floras, J. S. (2003). Sleep apnea and heart failure. *Circulation*, 107(12):1671–1678.
- Braun-Fahrlander, C., Vuille, J., Sennhauser, F., Neu, U., Kunzle, T., Grize, L., Gassner, M., Minder, C., Schindler, C., Varonier, H., et al. (1997). Respiratory health and long-term exposure to air pollutants in Swiss schoolchildren. SCARPOL Team. Swiss Study on Childhood Allergy and Respiratory Symptoms with Respect to Air Pollution, Climate and Pollen. *American Journal of Respiratory and Critical Care Medicine*, 155(3):1042–1049.
- Brennan, M., Palaniswami, M., and Kamen, P. (2002). Poincaré plot interpretation using a physiological model of HRV based on a network of oscillators. *American Journal of Physiology-Heart and Circulatory Physiology*, 283(5):1873–1886.
- Bryld, L., Hindsberger, C., Kyvik, K., Agner, T., and Menne, T. (2003). Risk factors influencing the development of hand eczema in a population-based twin sample. *British Journal of Dermatology*, 149(6):1214–1220.
- Burns, A., Greene, B. R., McGrath, M. J., O’Shea, T. J., Kuris, B., Ayer, S. M., Stroiescu, F., and Cionca, V. (2010). SHIMMER—a wireless sensor platform for noninvasive biomedical research. *IEEE Journal of Sensors*, 10(9):1527–1534.
- Call, R., Smith, T., Morris, E., Chapman, M., and Platts-Mills, T. (1992). Risk factors for asthma in inner city children. *The Journal of Pediatrics*, 121(6):862–866.
- Carney, R., Blumenthal, J., Stein, P., Watkins, L., Catellier, D., Berkman, L., and Freedland, K. (2001). Depression, heart rate variability, and acute myocardial infarction. *Circulation*, 104(17):2024–2028.
- Castro, W., Schilgen, M., Meyer, S., Weber, M., Peuker, C., and Wörtler, K. (1997). Do “whiplash injuries” occur in low-speed rear impacts? *European Spine Journal*, 6(6):366.

- Catal, C. and Diri, B. (2009). Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem. *Information Sciences*, 179(8):1040–1058.
- Chang, K., Monahan, K., Griffin, M., Lake, D., and Moorman, J. (2001). Comparison and clinical application of frequency domain methods in analysis of neonatal heart rate time series. *Annals of Biomedical Engineering*, 29(9):764–774.
- Chatagnon, M. and Busso, T. (2006). Modelling of aerobic and anaerobic energy production during exhaustive exercise on a cycle ergometer. *European Journal of Applied Physiology*, 97(6):755–760.
- Chen, K. and Bassett Jr, D. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and Science in Sports and Exercise*, 37(11):490.
- Chen, K. and Sun, M. (1997). Improving energy expenditure estimation by using a triaxial accelerometer. *Journal of Applied Physiology*, 83(6):2112–2122.
- Chen, S., Chen, H., and Chan, H. (2006). A real-time QRS detection method based on moving-averaging incorporating with wavelet denoising. *Computer Methods and Programs in Biomedicine*, 82(3):187–195.
- Cherkassky, V. and Mulier, F. (2007). *Learning from data: concepts, theory, and methods*. Wiley-IEEE Press.
- Chon, K. H., Dash, S., and Ju, K. (2009). Estimation of respiratory rate from photoplethysmogram data using time–frequency spectral estimation. *IEEE Transactions on Biomedical Engineering*, 56(8):2054–2063.
- Churchill, W. (2008). The Magnificent Century of Cardiothoracic Surgery. *History of Medicine*, 4(3):187–191.
- Clarke, J., Shelton, J., Venning, G., Hamer, J., and Taylor, S. (1976). The rhythm of the normal human heart. *The Lancet*, 308(7984):508–512.

- Clifford, G. and Tarassenko, L. (2005). Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE Transactions on Biomedical Engineering*, 52(4):630–638.
- Clifford, G. D., Azuaje, F., McSharry, P., et al. (2006). *Advanced methods and tools for ECG data analysis*. Artech house London.
- Cogdell, J. W. and Piatetski-Shapiro, I. (1990). *The arithmetic and spectral analysis of Poincaré series*. Academic Press Boston, MA.
- Cole, C. R., Foody, J., Blackstone, E. H., Lauer, M. S., et al. (2000). Heart rate recovery after submaximal exercise testing as a predictor of mortality in a cardiovascularly healthy cohort. *Annals of Internal Medicine*, 132(7):552–555.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301.
- Cosmet (2013). CPET Homepage. [Online; accessed March-2013] — <http://www.cosmed.it>.
- Cox, L., Williams, B., Sicherer, S., Oppenheimer, J., Sher, L., Hamilton, R., and Golden, D. (2008). Pearls and pitfalls of allergy diagnostic testing: report from the American College of Allergy, Asthma and Immunology/American Academy of Allergy, Asthma and Immunology Specific IgE Test Task Force. *Annals of Allergy, Asthma and Immunology*, 101(6):580–592.
- Craven, D., Glavin, M., Kilmartin, L., and Jones, E. (2012). Potential for Extended Battery Life in Mobile Healthcare with Bluetooth Low Energy and Signal Compression. *Irish Signals and Systems Conference*, 42:151–165.
- Crouter, S. E., Clowers, K. G., and Bassett, D. R. (2006). A novel method for using accelerometer data to predict energy expenditure. *Journal of Applied Physiology*, 100(4):1324–1331.
- Culhane, K., OConnor, M., Lyons, D., and GM., L. (2005). Accelerometers in rehabilitation medicine for older adults. *Age and Ageing*, 34(6):556–560.

- de Carvalho, J., da Rocha, A., de Oliveira Nascimento, F., Neto, J., and Junqueira Jr, L. (2002). Development of a Matlab software for analysis of heart rate variability. In *Signal Processing*, pages 1488–1491.
- De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278.
- de Oliveira, F. and Cortez, P. (2004). A QRS detection based on hilbert transform and wavelet bases. In *Machine Learning for Signal Processing, 2004*, pages 481–489.
- Dekker, J. M., Crow, R. S., Folsom, A. R., Hannan, P. J., Liao, D., Swenne, C. A., and Schouten, E. G. (2000). Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes: the ARIC Study. *Circulation*, 102(11):1239–1244.
- Dekker, J. M., Schouten, E. G., Klootwijk, P., Pool, J., Swenne, C. A., and Kromhout, D. (1997). Heart Rate Variability from Short Electrocardiographic Recordings Predicts Mortality from All Causes in Middle-aged and Elderly Men The Zutphen Study. *American Journal of Epidemiology*, 145(10):899–908.
- Deza, M. and Deza, E. (2009). *Encyclopedia of distances*. Springer.
- Di Virgilio, V., Francaiancia, C., Lino, S., and Cerutti, S. (1995). ECG fiducial points detection through wavelet transform. In *Engineering in Medicine and Biology Society*, pages 1051–1052.
- Dinh, H., Kumar, D., Pah, N., and Burton, P. (2001). Wavelets for QRS detection. In *Engineering in Medicine and Biology Society*, pages 1883–1887.
- Dobbs, S., Schmitt, N., and Ozemek, H. (1984). QRS detection by template matching using real-time correlation on a microcomputer. *Journal of Clinical Engineering*, 9(3):197–212.
- Dorland, W. A. N. (1901). *The American illustrated medical dictionary: a new and completed dictionary of the terms used in medicine, surgery, dentistry, pharmacy, chemistry, and the kindred branches with their pronunciation, derivation, and definition*. Saunders.

- Doyle, O., Temko, A., Marnane, W., Lightbody, G., and Boylan, G. (2010). Heart rate based automatic seizure detection in the newborn. *Medical Engineering and Physics*, 32(8):829–839.
- Dublin Institute of Technology (2013). Healthy, annotated ECG trace. [Online; accessed March-2013] — <http://eleceng.dit.ie>.
- Duda, R., Hart, P., and Stork, D. (1995). *Pattern Classification and Scene Analysis 2nd ed.* springer New York.
- DunnGalvin, A., Cullinane, C., Daly, D., Flokstra-de Blok, B., Dubois, A., and Hourihane, J. (2010). Longitudinal validity and responsiveness of the Food Allergy Quality of Life Questionnaire–Parent Form in children 0–12 years following positive and negative food challenges. *Clinical and Experimental Allergy*, 40(3):476–485.
- Ebden, M. (2002). A Comparison of HRV Techniques: The Lomb Periodogram versus The Smoothed Pseudo Wigner-Ville Distribution. *A report submitted to Prof. Lionel Tarassenko*, 23(1):325–364.
- Elliot, S. (2010). A Strategy When Times Are Tough. New York Times.
- Falkner, B., Onesti, G., Angelakos, E., Fernandes, M., and Langman, C. (1979). Cardiovascular response to mental stress in normal adolescents with hypertensive parents. Hemodynamics and mental stress in adolescents. *Hypertension*, 1(1):23–30.
- Faundez-Zanuy, M. and Monte-Moreno, E. (2005). State-of-the-art in speaker recognition. *IEEE Aerospace and Electronics Systems Magazine*, 20(5):7–12.
- Ferrannini, E. (1988). The theoretical bases of indirect calorimetry: a review. *Metabolism*, 37(3):287–301.
- Figueiredo, M. and Jain, A. (2000). Unsupervised selection and estimation of finite mixture models. In *Pattern Recognition*, pages 87–90.
- Fine, S., Navratil, J., and Gopinath, R. A. (2001). A hybrid GMM/SVM approach to speaker identification. In *Acoustics, Speech, and Signal Processing*, pages 417–420.

- Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press.
- Flannery, B. P., Press, W. H., Teukolsky, S. A., and Vetterling, W. (1992). *Numerical recipes in C*. Press Syndicate of the University of Cambridge, New York.
- Fraden, J. and Neuman, M. (1980). QRS wave detection. *Medical and Biological Engineering and Computing*, 18(2):125–132.
- Freedson, P. S., Melanson, E., Sirard, J., et al. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, 30(5):777–781.
- Friesen, G., Jannett, T., Jadallah, M., Yates, S., Quint, S., and Nagle, H. (1990). A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1):85–98.
- Frigo, M. and Johnson, S. (1998). FFTW: An adaptive software architecture for the FFT. In *Acoustics, Speech and Signal Processing*, pages 1381–1384.
- Fuchs, R. M., Achuff, S., Grunwald, L., Yin, F., and Griffith, L. (1982). Electrocardiographic localization of coronary artery narrowings: studies during myocardial ischemia and infarction in patients with one-vessel disease. *Circulation*, 66(6):1168–1176.
- Furlow, B. (2009). Contrast-enhanced ultrasound. *Radiologic Technology*, 80(6):547–561.
- Galvin, G. J., Davis, T. J., and MacDonald, N. C. (2000). Micromechanical accelerometer for automotive applications. US Patent 6,149,190.
- Ganong, W. F. and Ganong, W. (2005). *Review of medical physiology*. McGraw-Hill Medical ^ eNew York New York.
- Gardner, A., Krieger, A., Vachtsevanos, G., and Litt, B. (2006). One-class novelty detection for seizure analysis from intracranial EEG. *The Journal of Machine Learning Research*, 7(1):1025–1044.
- Gardner, A. B. (2004). *A novelty detection approach to seizure analysis from intracranial EEG*. PhD thesis, Georgia Institute of Technology.

- Giddens, D. and Kitney, R. (1985). Neonatal heart rate variability and its relation to respiration. *Journal of Theoretical Biology*, 113(4):759–780.
- Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C., and Stanley, H. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):215–220.
- Gombarska, D. and Horicka, M. (2012). Evaluation of heart rate variability in time and Frequency domain. In *ELEKTRO*, pages 415 , 418.
- Google Inc. (2013). Google Scholar. [Online; accessed March-2013] — <https://scholar.google.com>.
- Gutiérrez, R., Twomey, N., Marnane, W. P., and Campos-Garcia, J. (2013). Real-Time Allergy Detection. In *IEEE Society of Intelligent Signal Processing*.
- Gyaw, T. and Ray, S. (1994). The wavelet transform as a tool for recognition of biosignals. *Biomedical Sciences Instrumentation*, 30:63–68.
- Hamer, M. and Steptoe, A. (2007). Association between physical fitness, parasympathetic control, and proinflammatory responses to mental stress. *Psychosomatic Medicine*, 69(7):660–666.
- Hamilton, P. (2002). Open source ECG analysis. *Computers in Cardiology*, 10:101–104.
- Hamilton, P. and Tompkins, W. (1988). Adaptive matched filtering for QRS detection. In *Engineering in Medicine and Biology Society*, pages 147–148.
- Hartigan, J. and Wong, M. (1979). Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108.
- Hayton, P., Scholkopf, B., Tarassenko, L., and Anuzis, P. (2001). Support vector novelty detection applied to jet engine vibration spectra. *Advances in Neural Information Processing Systems*, 13(1):946–952.

- Healey, J. A. and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166.
- Hedman, A., Hartikainen, J., Tahvanainen, K., and Hakumäki, M. (2008). The high frequency component of heart rate variability reflects cardiac parasympathetic modulation rather than parasympathetic. *Acta Physiologica Scandinavica*, 155(3):267–273.
- Hemokinetics, I. (1993). Tritrac-R3D Research Ergometer Operations. *Journal of Applied Physiology*, 7:149–159.
- Hendelman, D., Miller, K., Baggett, C., Debold, E., and Freedson, P. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Medicine and Science in Sports and Exercise*, 32(9):442–449.
- Hernando, D., Bailón, R., Laguna, P., and Sornmo, L. (2011). Heart rate variability during hemodialysis and its relation to hypotension. In *Computing in Cardiology*, pages 189–192.
- Hester, T., Hughes, R., Sherrill, D. M., Knorr, B., Akay, M., Stein, J., and Bonato, P. (2006). Using wearable sensors to measure motor abilities following stroke. In *Wearable and Implantable Body Sensor Networks*, pages 4–7.
- Hilbert, D. (1912). *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Berlin.
- Horgan, D. and Murphy, C. C. (2010). Voting rule optimisation for double threshold energy detector-based cognitive radio networks. In *Signal Processing and Communication Systems*, pages 1–8.
- Huikuri, H. V., Castellanos, A., and Myerburg, R. J. (2001). Sudden death due to cardiac arrhythmias. *New England Journal of Medicine*, 345(20):1473–1482.

- Jan, K., Nagel, J., Hurwitz, B., and Schneiderman, N. (Oct-3 Nov). Decomposition of heart rate variability by adaptive filtering for estimation of cardiac vagal tone. In *Engineering in Medicine and Biology Society*, pages 660 – 661.
- Järvinen, K. M., Amalanayagam, S., Shreffler, W. G., Noone, S., Sicherer, S. H., Sampson, H. A., and Nowak-Wegrzyn, A. (2009). Epinephrine treatment is infrequent and biphasic reactions are rare in food-induced reactions during oral food challenges in children. *Journal of Allergy and Clinical Immunology*, 124(6):1267–1272.
- Jeppesen, J., Beniczky, S., Fuglsang-Frederiksen, A., Sidenius, P., and Jasemian, Y. (2010). Detection of epileptic-seizures by means of power spectrum analysis of heart rate variability: A pilot study. *Technology and Health Care*, 18(6):417–426.
- Jones, A. M. and Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences*, 14(4):321–327.
- Jovanov, E., Frith, K., Anderson, F., Milosevic, M., and Shrove, M. T. (2011). Real-time monitoring of occupational stress of nurses. In *Engineering in Medicine and Biology Society*, pages 3640–3643.
- Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A., and Chellappa, R. (2003). Gait analysis for human identification. In *Audio-and Video-Based Biometric Person Authentication*, pages 706–714.
- Kamath, M. V., Fallen, E., et al. (1993). Power spectral analysis of heart rate variability: a noninvasive signature of cardiac autonomic function. *Critical reviews in Biomedical Engineering*, 21(3):245–311.
- Katznelson, Y. (2004). *An introduction to harmonic analysis*. Cambridge University Press.
- Kemp, A. H., Quintana, D. S., Gray, M. A., Felmingham, K. L., Brown, K., and Gatt, J. M. (2010). Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. *Biological Psychiatry*, 67(11):1067–1074.

- Kilpelinen, M., Terho, E., Helenius, H., and Koskenvuo, M. (2000). Farm environment in childhood prevents the development of allergies. *Clinical and Experimental allergy*, 30(2):201–208.
- Kleiger, R., Stein, P., and Bigger Jr, J. (2005). Heart rate variability: measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, 10(1):88–101.
- Kliegman, R. et al. (2007). *Nelson textbook of pediatrics*. Saunders Elsevier Philadelphia.
- Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Artificial Intelligence*, pages 1137–1145.
- Kohler, B., Hennig, C., and Orglmeister, R. (2002). The principles of software QRS detection. *IEEE Engineering in Medicine and Biology Magazine*, 21(1):42–57.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89–109.
- Kostis, J., Moreyra, A., Amendo, M., Di Pietro, J., Cosgrove, N., and Kuo, P. (1982). The effect of age on heart rate in subjects free of heart disease. Studies by ambulatory electrocardiography and maximal exercise stress test. *Circulation*, 65(1):141–145.
- Kwapisz, J. R., Weiss, G. M., and Moore, S. A. (2010). Cell phone-based biometric identification. In *Biometrics: Theory Applications and Systems*, pages 1–7.
- Kyrkos, A., Giakoumakis, E., and Carayannis, G. (1987). Time recursive prediction techniques on QRS detection problem. In *Engineering in Medicine and Biology Society*, pages 13–16.
- Laguna, P., Moody, G., and Mark, R. (1998). Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. *IEEE Transactions on Biomedical Engineering*, 45(6):698–715.
- Lanningham-Foster, L., Foster, R. C., McCrady, S. K., Jensen, T. B., Mitre, N., and Levine, J. A. (2009). Activity promoting games and increased energy expenditure. *The Journal of Pediatrics*, 154(6):819–823.

- Leon-Garcia, A. and Leon-Garcia, A. (2009). *Probability, statistics, and random processes for electrical engineering*. Pearson/Prentice Hall.
- Lewis, H. and Papadimitriou, C. (1997). *Elements of the Theory of Computation*. Prentice Hall PTR.
- Licht, C. M., de Geus, E. J., van Dyck, R., and Penninx, B. W. (2009). Association between anxiety disorders and heart rate variability in The Netherlands Study of Depression and Anxiety (NESDA). *Psychosomatic Medicine*, 65(12):508–518.
- Licht, C. M., de Geus, E. J., Zitman, F. G., Hoogendijk, W. J., van Dyck, R., and Penninx, B. W. (2008). Association between major depressive disorder and heart rate variability in the Netherlands Study of Depression and Anxiety (NESDA). *Archives of General Psychiatry*, 65(12):508–518.
- Lindh, W., Pooler, M., Tamparo, C., and Dahl, B. M. (2009). *Delmar's comprehensive medical assisting: administrative and clinical competencies*. Cengage Learning.
- Lobstein, T., Baur, L., and Uauy, R. (2004). Obesity in children and young people: a crisis in public health. *Obesity*, 5(1):4–85.
- Lomb, N. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462.
- Lombardi, F., Mäkilä, T. H., Myerburg, R. J., and Huikuri, H. V. (2001). Sudden cardiac death: role of heart rate variability to identify patients at risk. *Cardiovascular Research*, 50(2):210–217.
- Mannini, A. and Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10(2):1154–1175.
- Mark, R., Schluter, P., Moody, G., Devlin, P., and Chernoff, D. (1982). An annotated ECG database for evaluating arrhythmia detectors. *IEEE Transactions on Biomedical Engineering*, 42:205–210.
- Markov, K. and Nakamura, S. (2008). Improved novelty detection for online GMM based speaker diarization. In *Interspeech*, pages 363–366.

- Martegani, A., Meairs, S., Nolsøe, C., Piscaglia, F., Ricci, P., Seidel, G., Skjoldbye, B., Solbiati, L., Thorelius, L., Tranquart, F., et al. (2008). Guidelines and Good Clinical Practice Recommendations for Contrast Enhanced Ultrasound (CEUS)–Update 2008. *Ultraschall in der Medizin*, 39(2):187–210.
- Matasar, M. and Neugut, A. (2003). Epidemiology of anaphylaxis in the United States. *Current Allergy and Asthma Reports*, 3(1):30–35.
- McSharry, P. and Cifford, G. (2004). Open-source software for generating electrocardiogram signals. *Medical Engineering and Physics*, 1:2–10.
- Mietus, J., Peng, C., Ivanov, P., and Goldberger, A. (2000). Detection of obstructive sleep apnea from cardiac interbeat interval time series. In *Computers in Cardiology*, pages 753–756.
- Miles, S., Fordham, R., Mills, C., Valovirta, E., and Mugford, M. (2005). A framework for measuring costs to society of IgE-mediated food allergy. *European Journal of Allergy and Clinical Immunology*, 60(8):996–1003.
- Minnen, D., Starner, T., Ward, J., Lukowicz, P., and Troster, G. (2005). Recognizing and discovering human actions from on-body sensor data. In *Multimedia*, pages 1545–1548.
- Monda, M., Viggiano, A., Vicidomini, C., Viggiano, A., Iannaccone, T., Tafuri, D., and De Luca, B. (2009). Espresso coffee increases parasympathetic activity in young, healthy people. *Nutritional Neuroscience*, 12(1):43–48.
- Moody, G. (1993). Spectral analysis of heart rate without resampling. In *Computers in Cardiology*, pages 715–718.
- Moody, G. and Mark, R. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50.
- Mourot, L., Bouhaddi, M., Perrey, S., Rouillon, J.-D., and Regnard, J. (2004). Quantitative Poincare plot analysis of heart rate variability: effect of endurance training. *European Journal of Applied Physiology*, 91(1):79–87.

- Nolan, J., Batin, P. D., Andrews, R., Lindsay, S. J., Brooksby, P., Mullen, M., Baig, W., Flapan, A. D., Cowley, A., Prescott, R. J., et al. (1998). Prospective study of heart rate variability and mortality in chronic heart failure: results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-heart). *Circulation*, 98(15):1510–1516.
- Nunan, D., Sandercock, G., and Brodie, D. (2010). A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults. *Pacing and Clinical Electrophysiology*, 33(11):1407–1417.
- Nygårds, M. and Sörnmo, L. (1983). Delineation of the QRS complex using the envelope of the ECG. *Medical and Biological Engineering and Computing*, 21(5):538–547.
- O’Brien, I., O’Hare, P., and Corral, R. (1986). Heart rate variability in healthy subjects: effect of age and the derivation of normal ranges for tests of autonomic function. *British Heart Journal*, 55(4):348–354.
- Obrist, P. A., Gaebelin, C. J., Teller, E. S., Langer, A. W., Grignolo, A., Light, K. C., and McCubbin, J. A. (2007). The relationship among heart rate, carotid dP/dt, and blood pressure in humans as a function of the type of stress. *Psychophysiology*, 15(2):102–115.
- Oh, C., Sohn, H., and Bae, I. (2009). Statistical novelty detection within the Yeongjong suspension bridge under environmental and operational variations. *Smart Materials and Structures*, 18(12):125–132.
- Okada, M. (1979). A Digital Filter for the ORS Complex Detection. *IEEE Transactions on Biomedical Engineering*, 42(12):700–703.
- Olmos, S., MillAn, M., Garcia, J., and Laguna, P. (1996). ECG data compression with the Karhunen-Loeve transform. In *Computers in Cardiology*, pages 253–256.
- Omenaas, E., Bakke, P., Elsayed, S., Hanoa, R., and Gulsvik, A. (1994). Total and specific serum IgE levels in adults: relationship to sex, age and environmental factors. *Clinical and Experimental Allergy*, 24(6):530–539.

- Oude Elberink, J., de Monchy, J., van der Heide, S., Guyatt, G., and Dubois, A. (2002). Venom immunotherapy improves health-related quality of life in patients allergic to yellow jacket venom. *Journal of Allergy and Clinical Immunology*, 110(1):174–182.
- Oude Luttikhuis, H., Baur, L., Jansen, H., Shrewsbury, V. A., O'Malley, C., Stolk, R. P., and Summerbell, C. D. (2009). Interventions for treating obesity in children. *The Cochrane Database of Systematic Reviews*, 1(1):1.
- Parvin, B., Yang, Q., Fontenay, G., and Barcellos-Hoff, M. (2002). BioSig: an imaging bioinformatic system for studying phenomics. *Computer*, 35(7):65–71.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.
- Pereira, B., Venter, C., Grundy, J., Clayton, C., Arshad, S., and Dean, T. (2005). Prevalence of sensitization to food allergens, reported adverse reaction to foods, food avoidance, and food hypersensitivity among teenagers. *Journal of Allergy and Clinical Immunology*, 116(4):884–892.
- Phua, C., Alahakoon, D., and Lee, V. (2004). Minority report in fraud detection: classification of skewed data. *ACM SIGKDD Explorations Newsletter*, 6(1):50–59.
- Piskorski, J. and Guzik, P. (2005). Filtering Poincaré plots. *Computational Methods in Science and Technology*, 91(2):201–208.
- Pober, D. M., Staudenmayer, J., Raphael, C., Freedson, P. S., et al. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and Science in Sports and Exercise*, 38(9):16–26.
- Primeau, M., Kagan, R., Joseph, L., Lim, H., Dufresne, C., Duffy, C., Prhcal, D., and Clarke, A. (2000). The psychological burden of peanut allergy as perceived by adults with peanut allergy and the parents of peanut-allergic children. *Clinical and Experimental Allergy*, 30(8):1135–1143.

- Radon, K., Ehrenstein, V., Praml, G., and Nowak, D. (2004). Childhood visits to animal buildings and atopic diseases in adulthood: An age-dependent relationship. *American Journal of Industrial Medicine*, 46(4):349–356.
- Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C., and Suri, J. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing*, 44(12):1031–1051.
- Rajendra Acharya, U., Subbanna Bhat, P., Iyengar, S., Rao, A., and Dua, S. (2003). Classification of heart rate data using artificial neural network and fuzzy equivalence relation. *Pattern Recognition*, 36(1):61–68.
- Rautava, S., Kalliomaki, M., and Isolauri, E. (2002). Probiotics during pregnancy and breast-feeding might confer immunomodulatory protection against atopic disease in the infant. *Journal of Allergy and Clinical Immunology*, 109(1):119–121.
- Ravi, N., Dandekar, N., Mysore, P., and Littman, M. L. (2005). Activity recognition from accelerometer data. In *Proceedings of the national conference on artificial intelligence*, pages 1541–1546.
- Rawenwaaij-Arts, C., Kallee, L., Hopman, J., et al. (1993). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *European Heart Journal*, 52(17):1353–1365.
- Reynolds, D. and Rose, R. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83.
- Riese, H., Van Doornen, L. J., Houtman, I. L., and De Geus, E. J. (2004). Job strain in relation to ambulatory blood pressure, heart rate, and heart rate variability among female nurses. *Scandinavian Journal of Work, Environment and Health*, 61(3):387–396.
- Roberts, G., Patel, N., Levi-Schaffer, F., Habibi, P., and Lack, G. (2003). Food allergy as a risk factor for life-threatening asthma in childhood: a case-controlled study. *Journal of Allergy and Clinical Immunology*, 112(1):168–174.

- Roberts, S. (2000). Extreme value statistics for novelty detection in biomedical data processing. *IEE Proceedings Science, Measurement and Technology*, 147(6):363–367.
- Roberts, S. J. (1999). Novelty detection using extreme value statistics. *IEE Proceedings Vision, Image and Signal Processing*, 146(3):124–129.
- Robinson, B. F., Epstein, S. E., Beiser, G. D., and BRAUNWALD, E. (1966). Control of Heart Rate by the Autonomic Nervous System Studies in Man on the Interrelation Between Baroreceptor Mechanisms and Exercise. *Circulation*, 19(2):400–411.
- Rousseeuw, P. J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223.
- Sabiston Jr, D. C. (1981). Heart disease: a textbook of cardiovascular medicine. *Annals of Surgery*, 194(1):116.
- Sampson, H. (1999). Food allergy. Part 2: diagnosis and management. *Journal of Allergy and Clinical Immunology*, 103(6):981–989.
- Sampson, H., Mendelson, L., and Rosen, J. (1992). Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *The New England Journal of Medicine*, 327(6):380–384.
- Sampson, H. A., Muñoz-Furlong, A., Campbell, R. L., Adkinson Jr, N. F., Allan Bock, S., Branum, A., Brown, S. G., Camargo Jr, C. A., Cydulka, R., Galli, S. J., et al. (2006). Second symposium on the definition and management of anaphylaxis: summary reportSecond National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *Annals of emergency medicine*, 117(2):391–397.
- Sandercock, G., Bromley, P. D., Brodie, D. A., et al. (2005). Effects of exercise on heart rate variability: inferences from meta-analysis. *Medicine and Science in Sports and Exercise*, 37(3):433–439.
- Saramäki, T. and Bregovic, R. (2002). Multirate systems and filter banks. *Multirate Systems: Design and Applications*, 2:27–85.

- Schechtman, V., Raetz, S., Harper, R., Garfinkel, A., Wilson, A., Southall, D., and Harper, R. (1992). Dynamic analysis of cardiac RR intervals in normal infants and in infants who subsequently succumbed to the sudden infant death syndrome. *Pediatric research*, 31(6):606–612.
- Schlindwein, F. S., Yi, A., Edwards, T., and Bien, I. (2006). Optimal frequency and bandwidth for FIR bandpass filter for QRS detection. In *Advances in Medical, Signal and Information Processing*, pages 1–4.
- Schoeller, D. A. et al. (1988). Measurement of energy expenditure in free-living humans by using doubly labeled water. *The Journal of Nutrition*, 118(11):1278–1289.
- Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., and Williamson, R. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471.
- Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., and Platt, J. (2000). Support vector method for novelty detection. *Advances in Neural Information Processing Systems*, 42(1):582–588.
- Seccareccia, F., Pannozzo, F., Dima, F., Minoprio, A., Menditto, A., Lo Noce, C., and Giampaoli, S. (2001). Heart rate as a predictor of mortality: the MATISS project. *American Journal of Public Health*, 91(8):1258–1263.
- Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, 53(3):683–690.
- SHIMMER research (2010). SHIMMER research. [Online; accessed March 2013] — <http://www.shimmer-research.com>.
- Sicherer, S., Noone, S., and Muñoz-Furlong, A. (2001). The impact of childhood food allergy on quality of life. *Annals of Allergy, Asthma and Immunology*, 87(6):461–464.
- Sicherer, S. and Sampson, H. (2006). 9. Food allergy. *Journal of Allergy and Clinical Immunology*, 117(2):470–475.
- Sicherer, S., Sampson, H., et al. (2006). 9. Food allergy. *The Journal of Allergy and Clinical Immunology*, 117(2):470–475.

- Soman, A., Vaidyanathan, P., and Nguyen, T. (1993). Linear phase paraunitary filter banks: Theory, factorizations and designs. *IEEE Transactions on Signal Processing*, 41(12):3480–3496.
- Sörnmo, L. and Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press.
- Sörnmo, L. and Laguna, P. (2006). Electrocardiogram (ECG) signal processing. *Wiley Encyclopedia of Biomedical Engineering*.
- Srikanth, T., Napper, S., and Gu, H. (1998). Assessment of resampling methodologies of electrocardiogram signals for feature extraction, statistical and neural networks applications. In *Computers in Cardiology*, pages 537–540.
- Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., and Freedson, P. (2009). An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 107(4):1300–1307.
- Stein, P., Kleiger, R., et al. (1999). Insights from the study of heart rate variability. *Annual Review of Medicine*, 50(1):249–261.
- Strang, G. and Nguyen, T. (1996). *Wavelets and filter banks*. Cambridge University Press.
- Swartz, A. M., Strath, S. J., Bassett Jr, D. R., O’Brien, W. L., King, G. A., Ainsworth, B. E., et al. (2000). Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and Science in Sports and Exercise*, 32(9):450–456.
- Tanaka, H., Monahan, K., and Seals, D. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156.
- Temko, A., Boylan, G., Marnane, W., and Lightbody, G. (2010). Speech recognition features for EEG signal description in detection of neonatal seizures. In *Engineering in Medicine and Biology Society*, pages 3281–3284.
- Temko, A., Nadeu, C., Marnane, W., Boylan, G. B., and Lightbody, G. (2011a). EEG signal description with spectral-envelope-based speech recognition features for detection

- of neonatal seizures. *IEEE Transactions on Information Technology in Biomedicine*, 15(6):839–847.
- Temko, A., Thomas, E., Boylan, G., Marnane, W., and Lightbody, G. (2009). An SVM-based system and its performance for detection of seizures in neonates. In *Engineering in Medicine and Biology Society*, pages 2643–2646.
- Temko, A., Thomas, E., Marnane, W., Lightbody, G., and Boylan, G. (2011b). Performance assessment for EEG-based neonatal seizure detectors. *Clinical Neurophysiology*, 122(3):474–482.
- SHIMMER-research (2010). CE Certification. [Online; accessed March 2013] — <http://www.shimmer-research.com>.
- Thakor, N. and Zhu, Y. (1991). Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Transactions on Biomedical Engineering*, 38(8):785–794.
- theonlineallergist.com (2013). Epinephrine autoinjector. [Online; accessed 13-March-2012] — <http://www.theonlineallergist.com>.
- Thomas, E. (2010). *A machine learning framework for neonatal seizure detection*. PhD thesis, University College Cork.
- Thomas, E., Temko, A., Lightbody, G., Marnane, W., and Boylan, G. (2009). A gaussian mixture model based statistical classification system for neonatal seizure detection. In *Machine Learning for Signal Processing*, pages 1–6.
- Thomas, E., Temko, A., Marnane, W., Boylan, G., and Lightbody, G. (2013). Discriminative and generative classification techniques applied to automated neonatal seizure detection. *IEEE Journal of Biomedical and Health Informatics*, 31(7):1047.
- Tsuji, H., Larson, M. G., Venditti, F. J., Manders, E. S., Evans, J. C., Feldman, C. L., and Levy, D. (1996). Impact of reduced heart rate variability on risk for cardiac events: the Framingham Heart Study. *Circulation*, 94(11):2850–2855.

- Tulppo, M. and Huikuri, H. (2004). Origin and significance of heart rate variability. *Journal of the American College of Cardiology*, 43(12):2278–2280.
- Tulppo, M. P., Makikallio, T., Takala, T., Seppanen, T., and Huikuri, H. (1996). Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American Journal of Physiology-Heart and Circulatory Physiology*, 271(1):244–252.
- Twomey, N., Faul, S., Daly, D., Hourihane, J., and Marnane, W. (2010a). Classification of biophysical changes during food allergy challenges. In *Applied Sciences in Biomedical and Communication Technologies*, pages 1–5.
- Twomey, N., Faul, S., and Marnane, W. P. (2010b). Comparison of accelerometer-based energy expenditure estimation algorithms. In *Pervasive Computing Technologies for Healthcare*, pages 1–8.
- Twomey, N., Temko, A., Cullinane, C., Daly, D., Marnane, W. P., and Hourihane, J. O. (2013a). Detection of heart rate variation could improve patient safety and diagnostic yield during oral food challenge. *European Academy of Allergology and Clinical Immunology*.
- Twomey, N., Temko, A., Hourihane, J., and Marnane, W. (2011). Allergy detection with statistical modelling of HRV-based non-reaction baseline features. In *Applied Sciences in Biomedical and Communication Technologies*, pages 134–138.
- Twomey, N., Temko, A., Hourihane, J. O., and Marnane, W. P. (2013b). Fully automated allergy detection from paediatric ECG. *IEEE Transactions on Information Technology in Biomedicine*.
- Twomey, N., Walsh, N., Doyle, O., McGinley, B., Glavin, M., Jones, E., and Marnane, W. (2010c). The effect of lossy ECG compression on QRS and HRV feature extraction. In *Engineering in Medicine and Biology Society*, pages 634–637.
- University of Nottingham (2013). Einthoven ECG configuration. [Online; accessed March-2013] — <http://www.nottingham.ac.uk>.

- Uswatte, G., Foo, W. L., Olmstead, H., Lopez, K., Holand, A., Simms, L. B., et al. (2005). Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke. *Archives of Physical Medicine and Rehabilitation*, 86(7):1498–1501.
- Uswatte, G., Giuliani, C., Winstein, C., Zeringue, A., Hobbs, L., and Wolf, S. L. (2006). Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. *Archives of Physical Medicine and Rehabilitation*, 87(10):1340–1345.
- van Ravenswaaij-Arts, C., Kollee, L., Hopman, J., Stoeltinga, G., and van Geijn, H. (1993). Heart rate variability. *Annals of Internal Medicine*, 81(6):1803–1810.
- Vapnik, V. N. and Kotz, S. (1982). Estimation of dependences based on empirical data. *Springer-Verlag New York*, 89(12):5675–5679.
- Viinanen, A., Munhbayarlah, S., Zevgee, T., Narantsetseg, L., Naidansuren, T., Koskenvuo, M., Helenius, H., and Terho, E. (2007). The protective effect of rural living against atopy in Mongolia. *European Journal of Allergy and Clinical Immunology*, 62(3):272–280.
- Vijaya, G., Kumar, V., and Verma, H. (1998). ANN-based QRS-complex analysis of ECG. *Journal of Medical Engineering and Technology*, 22(4):160–167.
- Wang, Y. and Lobstein, T. (2006). Worldwide trends in childhood overweight and obesity. *International Journal of Pediatric Obesity*, 1(1):11–25.
- Webb, A., Copsey, K., and Cawley, G. (2011). *Statistical pattern recognition*. Wiley.
- WHO (2000). Obesity: preventing and managing the global epidemic. *World Health Organization Technical Report Series*, 70(3):510.
- Wilson, F. N., Johnston, F. D., et al. (1946). On Einthoven's triangle, the theory of unipolar electrocardiographic leads, and the interpretation of the precordial electrocardiogram. *American Heart Journal*, 32(3):277–310.

- Winter, E., Jones, A., Davidson, R., Bromley, P., and Mercer, T. (2006). *Sport and Exercise Physiology Testing Guidelines: Volume I-Sport Testing. The British Association of Sport and Exercise Sciences Guide*. Routledge, UK.
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52.
- Xue, Q., Hu, Y., and Tompkins, W. (1992). Neural-network-based adaptive matched filtering for QRS detection. *IEEE Transactions on Biomedical Engineering*, 39(4):317–329.
- Yamamoto, Y., Hughson, R. L., and Peterson, J. C. (1991). Autonomic control of heart rate during exercise studied by heart rate variability spectral analysis. *Journal of Applied Physiology*, 71(3):1136–1142.
- Yanishevsky, Y. and Hourihane, J. O. (2010). Differences in treatment of food challenge induced reactions reflect physicians’ protocols more than reaction severity. *The Journal of Allergy and Clinical Immunology*, 126(1):182.