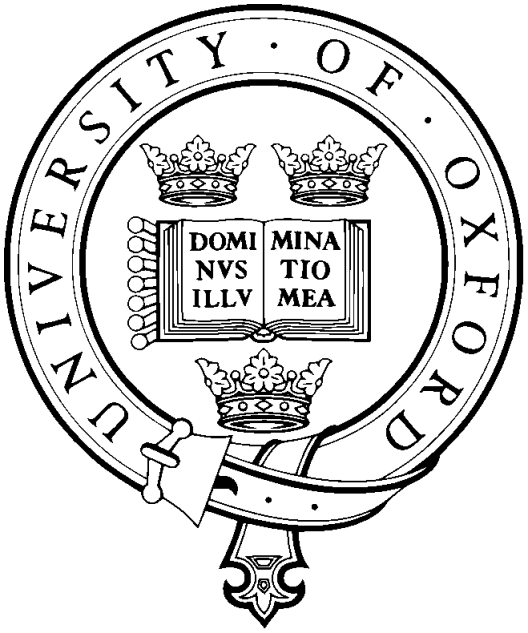
A Perinatal Monitoring System for Low-resource Settings



Lisa Stroux

Wolfson College

University of Oxford

Supervised by
Professor Gari D. Clifford
Professor Stephen Payne
Submitted: Michaelmas Term,

December 11, 2015

This thesis is submitted to the Department of Engineering Science, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research.

L. M. Stroux,

Wolfson College

Abstract

A Perinatal Monitoring System for Low-resource Settings

Lisa Stroux
Wolfson College
Doctor of Philosophy
Michaelmas 2015

The reduction of maternal and child morality has been central to international development efforts over the past twenty years, spurred on by global initiatives such as the Millennium Development Goals. Whereas a significant decrease in mortality rates has been achieved in mothers and the under-five, the fetal and early neonatal period has received less attention resulting in slower progress. Resource-poor settings suffer from the highest stillbirth, maternal and early neonatal death rates, with 99% occurring in low- and middle income countries.

Causes of fetal and maternal compromise can be manifold, both infrastructural and pathophysiological. Common problems remain a scarcity of skilled healthcare personnel leading to low health risk detection, referral and ultimately intervention rates. In the absence of sophisticated equipment fetal compromise may go unnoticed, such as fetal growth restriction, one of the primary causes of poor health outcome.

To work towards improved fetal and maternal risk assessment enabling timely and appropriate referral, this thesis proposes a low-cost mHealth monitoring platform for use in low-resource settings. The analysis of the fetal cardiac activity, prerequisite in hospitals in developed nations, may provide important insights on fetal stress levels and more specifically growth retardation. The fetal cardiac signal is accessible and obtained at low-cost with the help of a Doppler transducer. This thesis in healthcare innovation investigates the feasibility of introducing such technology in communities not familiar with their application, proposes a novel signal quality algorithm suitable for implementation on the phone, develops a classifier based on fetal heart rate variability to distinguish growth restricted from normal babies, and introduces the design of the mHealth platform, results of a pilot deployment in rural Guatemala and the design and implementation of a full randomised control trial.

Acknowledgements

This thesis would not have been possible without the support of many. First and foremost I would like to thank my supervisors Prof. Gari D. Clifford and Prof. Stephen Payne for giving me the opportunity to pursue this DPhil under their guidance. I am grateful for their confidence in allowing me to explore the realm of biosignal processing in the context of fetal health, treading ground that had been, in many ways, unfamiliar to me.

I would also like to thank my colleagues and friends at the Centre for Doctoral Training in Healthcare Innovation and the Institute of Biomedical Engineering who provided such a supportive environment and have made this journey so enjoyable. My thanks to Jo Armitage who was invaluable in navigating the administrative and organisational challenges of the DPhil, and the IT and administration teams never tired of the many inquiries coming their way.

I would like to thank Dr. Antoniya Georgieva and Prof. Christopher Redman in particular at the John Radcliffe Hospital, not only for granting me access to their extensive data archive, but for the valuable feedback and clinical insights into the practice of fetal monitoring; Pawel Szafranski for his help in extracting the data, Dr. Christina Aye and Ali Chevassut for being instrumental in building the first raw fetal Doppler database.

I had the great privilege to travel beyond the borders of Oxford and collaborate with the team of the Maya Health Alliance Wuqu' Kawoq in beautiful Guatemala on the development of the mHealth monitoring platform. My thanks go to Dr. Peter Rohloff, Dr. Boris Martinez, Enma Coyote Ixen and Nora King from whom I learned so much about the realities of healthcare provision in the rural communities they serve. Completing the team, I would also like to thank Dr. Rachel Hall-Clifford at Agnes Scott College and the engineers Andrew El-Masry and Tuong Vi Pham at the Department of Biomedical Informatics at Emory University; it has been truly enjoyable to be exposed to the various disciplines brought together in this project.

Finally, I am extremely grateful for the funding received over the course of my DPhil, provided by the RCUK Digital Economy Programme, Oxford's Centre for Affordable Healthcare Technology, ARM and Wolfson College.

Contents

1 Introduction **1**

1.1 Rational . . . 1

1.1.1 Perinatal mortality and prevalence of intra-uterine growth restriction in resource-limited settings . . . 1

1.1.2 Fetal monitoring modalities . . . 4

1.1.2.1 Fetal monitors for use in low resource settings . . . 8

1.1.3 mHealth . . . 8

1.1.3.1 mHealth applications for maternal and fetal health . . . 10

1.1.4 Field implementation - Guatemala . . . 11

1.2 Thesis outline . . . 14

1.3 Summary of datasets . . . 15

1.4 List of publications and presentations . . . 18

1.5 Contributors . . . 19

2 Literature review **20**

2.1 Introduction . . . 20

2.2 Physiology . . . 20

2.2.1 The fetal cardiac cycle . . . 20

2.2.2 Control of the fetal heart rate . . . 21

2.2.3 Fetal response to hypoxia . . . 24

2.2.3.1 Fetal growth restriction . . . 25

2.3 Fetal heart rate estimation . . . 25

2.3.1 Fetal heart rate estimation from 1D Doppler ultrasound . . . 26

2.3.1.1 Autocorrelation . . . 26

2.3.1.2 Autocorrelation based methods and recent developments 27

2.3.1.3 Databases of fetal heart rate traces . . . 31

2.4 Markers for the detection of IUGR . . . 32

2.5 Summary . . . 36

3 Evaluation of Doppler based fetal heart rate variability measures for IUGR detection **37**
3.1 Introduction . . . 37
3.2 Description of the cCTG dataset . . . 39
3.3 Detection based on trace variability . . . 40
3.3.1 cCTG metrics for IUGR discrimination . . . 40
3.3.2 Statistical analysis methods . . . 42
3.3.3 Logistic regression . . . 43
3.3.4 Results . . . 44
3.3.5 Discussion . . . 45
3.4 Detection based on sleep state dependent variability measures . . . 52
3.4.1 Impact of the fetal behavioural state on heart rate variability metrics 52
3.4.2 Bias in cCTG variability metrics . . . 52
3.4.3 Behavioural state distribution in IUGR and normal traces . . . 54
3.4.4 Behavioural state dependent cCTG metrics for IUGR classification 54
3.4.5 Statistical analysis methods . . . 55
3.4.6 Data pre-processing . . . 56
3.4.7 Results . . . 58
3.4.8 Required trace length for IUGR discrimination . . . 63
3.4.9 Discussion . . . 64
3.5 Conclusion . . . 68

4 Development of an mHealth fetal monitoring system **70**
4.1 Introduction . . . 70
4.2 m-Health monitoring system . . . 71
4.2.1 Design of data acquisition software . . . 74
4.2.2 1D Doppler ultrasound recording . . . 75
4.2.3 Accelerometer . . . 76
4.2.4 Hardware considerations . . . 77
4.3 Prototype testing with skilled personnel . . . 78
4.4 Conclusion . . . 78

5 1D Ultrasound Signal Quality Index **79**
5.1 Introduction . . . 79
5.2 Database . . . 80
5.2.1 Conditions of Recordings . . . 80
5.3 Signal quality annotation and data exclusion . . . 81
5.4 Preprocessing . . . 84

5.5 Feature extraction . . . 84
5.5.1 Sample Entropy features . . . 85
5.5.2 Wavelet features . . . 88
5.5.3 Accelerometer features . . . 93
5.5.4 Feature selection . . . 95
5.6 Classification . . . 96
5.6.1 Support vector machine classifier . . . 96
5.6.2 Statistical measures . . . 97
5.6.3 Balancing and normalising datasets . . . 98
5.6.4 Classification performance assessment . . . 99
5.7 Results . . . 100
5.8 Discussion . . . 103
5.9 Conclusion . . . 105

6 Development and evaluation of a hidden Markov model approach to fetal cardiac signal segmentation on 1D ultrasound data 107
6.1 Introduction . . . 107
6.2 Database . . . 109
6.3 Data selection and preprocessing . . . 109
6.3.1 Spike removal . . . 110
6.3.2 Frequency range of interest . . . 110
6.3.3 Balancing the dataset . . . 114
6.4 Hidden Markov Models . . . 114
6.4.1 Hidden Semi-Markov Model . . . 117
6.4.2 Training HMM parameters - The Baum-Welch algorithm . . . 117
6.4.3 Uncovering the optimal state sequence - The Viterbi algorithm . . . 121
6.5 Signal features . . . 124
6.5.1 Homomorphic envelope . . . 125
6.5.2 Wavelet envelope . . . 126
6.5.3 PSD envelope . . . 127
6.6 The distribution of Doppler state durations . . . 129
6.7 Heart rate estimation from the HSMM state path . . . 130
6.8 Reference signal . . . 133
6.8.1 Manual beat annotation . . . 133
6.8.2 Human reaction time . . . 133
6.9 Results of HSMM Doppler segmentation . . . 134
6.9.1 Performance assessment of heart rate estimation . . . 134

6.9.2 Performance assessment of beat-to-beat interval estimation 140
6.10 Discussion 141
6.11 Conclusion 143

7 Data collection and validation of the mHealth fetal monitoring system in rural Guatemala 145
7.1 Introduction 145
7.2 Feasibility testing 146
7.2.1 Data collection 147
7.2.2 Evaluation tools 149
7.2.3 Feasibility testing outcomes 149
7.2.3.1 Device use 149
7.2.3.2 Staff and patient response 150
7.2.3.3 Doppler signal quality assessment 151
7.2.4 Discussion 152
7.2.5 Conclusion and system implications 154
7.3 System refinement and intervention design 155
7.3.1 Mobile application design 156
7.3.1.1 User interface 158
7.3.1.2 Emergency response 158
7.3.2 Back end infrastructure 160
7.3.3 User manuals 163
7.4 System evaluation 163
7.4.1 Expected study outcomes 164
7.5 Conclusion 166

8 Summary and conclusion 168
8.1 Summary of contributions 168
8.1.1 1D Doppler database 170
8.1.1.1 Fetal Doppler and fECG dataset 171
8.1.2 SQI 172
8.1.3 IUGR risk detection 175
8.1.4 HMM heart rate trace 176
8.1.5 Design of a low-cost perinatal monitoring platform 176
8.1.5.1 Scaling up 178
8.2 Conclusion 179

A Mobile Application 181

B Complications Guide **189**

C Scientific Protocol for Ethical Approval of Study in Guatemala **193**

D Digital Help Manuals for Mobile Application **205**

Bibliography **209**

List of Figures

1.1 Prenatal development stages 2
1.2 Maternal and perinatal mortality rates 3
1.3 Population and health statistics Chimaltenango, Guatemala 12
1.4 Scatter plot of GNI versus low birth weight 13
1.5 Perinatal mortality and stillbirth in Guatemala 13

2.1 The fetal cardiac structures 21
2.2 The cardiac cycle 22
2.3 The vegetative control of the fetal heart rate 24

3.1 Fetal heart rate minute range estimation 41
3.2 Boxplots of STV values in male IUGR and normal fetuses per GA 46
3.3 Boxplots of STV values in female IUGR and normal fetuses per GA . . . 47
3.4 Boxplots STV, LTV and AAC 48
3.5 Univariate ROCs with optimal cut-off values 50
3.6 Distribution of sleep states in traces grouped by duration 53
3.7 Behavioural state distribution 55
3.8 Substitute variability values 58
3.9 ROC of best performing multivariate classifier 59
3.10 Boxplots of STVinHi values in male IUGR and normal fetuses per GA . . 60
3.11 Boxplots of STVinHi values in female IUGR and normal fetuses per GA . 61
3.12 Result of univariate and multivariate logistic regressions models 62
3.13 Onset of the first episode of active sleep in recorded traces per GA 66

4.1 Sound card quality 72
4.2 Bespoke Doppler recording lead 72
4.3 The prototype assembly 73
4.4 mHealth Doppler prototype in use 73
4.5 Android application screens 74
4.6 Continuous Doppler ultrasound 76
4.7 Tri-axial capacitive accelerometer 77

5.1 Signal quality annotation interface 82
5.2 Signal examples of different quality 83
5.3 Data contribution per patient 83
5.4 Sample entropy method 87
5.5 Probability density estimates of sample entropy features 87
5.6 Discrete Wavelet Transform filter bank 89
5.7 Wavelet functions 90
5.8 Probability density estimates of wavelet features 92
5.9 Use of Doppler on the maternal abdomen 94
5.10 The effect of SVM parameters 98
5.11 Training and test sets 99
5.12 Flow chart of classification experiments 100

6.1 Expected frequency range resultant from ventricular motion 112
6.2 Expected frequency range resultant from ventricular motion compared to frequencies recorded at points of highest intensity in training epochs . . . 112
6.3 Example frequency content of interest in training epoch 113
6.4 Training dataset 115
6.5 Test dataset 115
6.6 HMM and HSMM state transition diagrams 118
6.7 Example of the biorthogonal wavelet family 127
6.8 Selected wavelet function for envelope computation 127
6.9 Example epoch and derived feature envelopes 128
6.10 Estimation of the primary cardiac oscillation and the oscillation interval . 130
6.11 Estimation of the cardiac cycle duration by autocorrelation 131
6.12 Duration probability density functions 131
6.13 State path example of intermediate quality epoch 132
6.14 State path example of excellent quality epoch 133
6.15 Processing of manual heart beat annotations 135
6.16 Adjustment of median heart beat annotations 135
6.17 Error in heart rate estimation by autocorrelation 136
6.18 Error in heart rate estimation by HSMM 137
6.19 Hidden Markov model based heart rate estimation performance 138
6.20 Autocorrelation based heart rate estimation performance 139
6.21 Annotator agreement dependent on signal quality 139

7.1 Development process diagram 146
7.2 Prototype in use 148

7.3 Example of applied signal quality score 151
7.4 The relationship between Doppler signal quality and experience 153
7.5 The relationship between Doppler signal quality and gestational age . . . 153
7.6 Flow diagram of mobile perinatal monitoring application 157
7.7 Example application screens 159
7.8 Mobile application backend 162

8.1 Data acquisition in Leipzig 173
8.2 Data acquisition timeline Leipzig 173
8.3 Synchronous data acquisition 174
8.4 Example Doppler recordings with fECG R-peak reference 177

A.1 App screens: patient selection 183
A.2 App screens: patient information 184
A.3 App screens: patient information continued 185
A.4 App screens: health assessment 186
A.5 App screens: physiological signal recording 187
A.6 App screens: physiological signal recording continued 188

B.1 Antenatal complications guide 190
B.2 Birth complications guide 1 191
B.3 Birth complications guide 2 191
B.4 Postnatal complications guide 1 192
B.5 Postnatal complications guide 2 192

D.1 Digital manual for blood pressure 206
D.2 Digital manual for Doppler 207
D.3 Digital manual for pulse oximeter 208

List of Tables

1.1 Fetal cardiac monitoring modalities 6
1.2 Fetal monitors for resource-poor settings 8
1.3 Summary of datasets 16

2.1 Summary of cardiac events 23
2.2 Literature review of Doppler-based beat-to-beat HR estimation 30

3.1 Subject characteristics 40
3.2 Trace features for IUGR classification 42
3.3 Patient characteristics 23-34 wGA 45
3.4 Patient characteristics 35-42 wGA 46
3.5 cCTG metrics correlation 47
3.6 Discrimination quality of individual cCTG risk markers 49
3.7 Discrimination quality of mulitvariate LR classifiers 49
3.8 Features for IUGR classification 56
3.9 Feature correlation in traces with both active and quiet sleep episodes . . 63
3.10 Feature correlation including missing value substitutes 64
3.11 Univariate LR discrimination quality dependent on sleep states 65
3.12 Multivariate LR discrimination quality dependent on sleep states 67

5.1 Subject characteristics 80
5.2 Signal quality label definitions 81
5.3 Classification performance single cardiac signal based features 101
5.4 Classification performance single acceleration based features 101
5.5 Classification performance cardiac feature combinations 103
5.6 Classification performance cardiac and acceleration feature combinations 104

6.1 Cardiac oscillation frequency content 111
6.2 HSMM segmentation performance per subject (SQI 4 and 5) 141
6.3 HSMM segmentation performance per subject (SQI 3) 142
6.4 Gross HSMM segmentation performance (SQI 4 and 5) 142

6.5 Gross HSMM segmentation performance (SQI 3) 142

7.1 Feasibility study participants 148
7.2 List of pregnancy related complications assessed 160
7.3 Physiological measurement alerts 161
7.4 Implementation status of algorithms and study parameters collected . . . 164
7.5 List of evaluation study outcomes 165

8.1 Study parameters Leipzig 172

C.1 Guatemala study enrolment 204

Chapter 1

Introduction

This thesis in Healthcare Innovation focuses on the challenges of fetal, and to some extent maternal health risk screening in resource-limited populations. In particular, this thesis assesses the potential of applying low-cost sensors and utilizing the increasing ubiquity and power of mobile smart phones to address some of the needs identified. In addition to the investigation of engineering solutions the work follows an iterative design process drawing on methods from design theory and practice with the aim of developing an approach appropriate to the use scenario. The thesis is presented with the aspiration of providing a first contribution towards low-cost fetal cardiac monitoring and intelligent decision support, improving the standard of perinatal care in underserved communitites.

1.1 Rational

1.1.1 Perinatal mortality and prevalence of intra-uterine growth restriction in resource-limited settings

While advances in medical care have reduced mortality rates across the globe, perinatal mortality has decreased at a significantly slower pace. This is particularly true of low- and middle-income countries (LMICs), which contribute by far the largest proportion of both perinatal and maternal deaths annually.

Perinatal mortality includes both fetal death in utero (stillbirth) and neonatal death in the first seven days of life. Figure 1.1 illustrates the prenatal development stages, including a definition of the perinatal period.

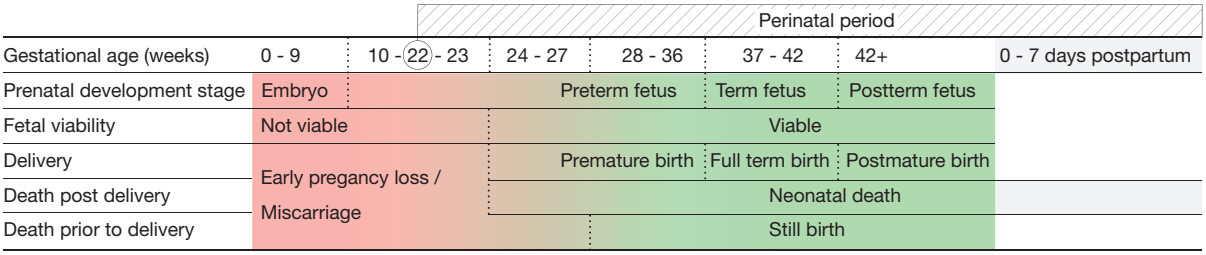


Figure 1.1: The prenatal development stages, including commonly used definitions for fetal and early neonatal death, and the perinatal period [1].

Maternal and child health has long been recognised as a key concern by the international development and healthcare community, as reflected in two of the eight United Nation's Millennium Development Goals (MDGs) [2], which were formulated in 2000 to drastically reduce maternal and child mortality by 2015. Significant progress could be reported in lowering especially child mortality rates over the past 15 years, however targets have not been met with a substantial disparity pertaining between lower resource settings and developed regions, as well as rural and urban areas.

The recent report on MDG outcomes [3] estimated maternal mortality to be 14 times higher in developing regions with only half of pregnant women receiving the recommended number of antenatal visits. Furthermore only 56% of births are attended by skilled healthcare workers in rural communities, as compared to 87% in urban areas.

By far the slowest decline in mortality rates was observed for in utero deaths and neonates in their first days of life [4,5]. Global mortality estimates for 2009 were 2.6 million stillbirths and 2.8 million early neonatal deaths (within the first week of life) among 133 million babies born alive each year [1]. Birth remains an event of extreme risk accounting for 40% of perinatal and maternal deaths (of a total of over 287 000 worldwide [1]). Figure 1.2 summarises recent global estimates of maternal and perinatal mortality rates, assigned to the respective stages of pregnancy and the post-partum period. Close to 99%

of those deaths occur in developing countries [1]. Estimates especially in the perinatal period are prone to inaccuracy due to discrepancies in death records at country and regional levels alike [6]. However, estimates do highlight that risk is prevalent along the continuum of pregnancy and the post-partum period, which has often been neglected.

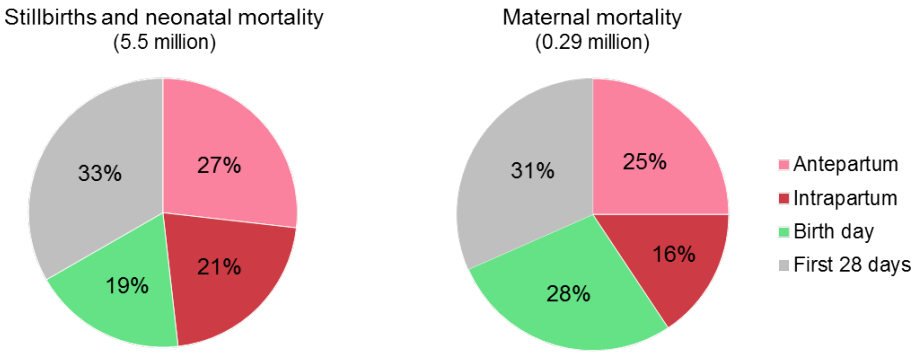


Figure 1.2: Maternal, fetal and early neonatal mortality rates as estimated globally, adapted from [4]. The majority of both maternal and perinatal deaths (close to 99%) occur in developing regions [1]. Perinatal estimates in particular are susceptible to error given discrepancies in death records [6], nevertheless they emphasise the high risk prevalent along the continuum of pregnancy and the often neglected post-partum period.

The little attention that has been given to perinatal compared to child (under 5) and maternal mortality may to some extent be attributed to these substantial discrepancies in registration practices of birth and death including disagreement in definitions of terms such as 'stillbirth' [4,6]. This has resulted in a lack of reliable data to act upon. Relevant data required to effectively address perinatal risk includes harmonised records of fetal and neonatal deaths, information on where deaths occur geographically and with respect to location (in a healthcare facility or at home), in which stage of pregnancy they occur, and their causes if known [7].

In developed regions the steep reduction of perinatal deaths in the past century is often attributed to the advances in skilled birth support, medical technologies and specialised healthcare infrastructure [6]. All of those are often amiss in low-resource and particularly rural regions. Interdependent factors, such as a lack of skilled medical personnel, sparse coverage of specialised care in remote areas or the cost of healthcare provision all play a

role in the slow progress. Among the main direct and indirect causes of perinatal death reported in 2015 were preterm birth (16%), intrapartum related complications (11%) and sepsis (7%) [5]. The largest risk factor, however, for over 80% of neonatal deaths is low birth weight [4]. Babies with low birth weight are either born pre-term, small-for-gestational age (SGA), or both. Some term SGA babies are considered constitutionally small and might develop normally, others are affected by intra-uterine growth restriction (IUGR), a pathological slow-down in growth preventing the fetus from reaching its full growth potential. The IUGR prevalence has been estimated at 11%, six times higher than in developed regions [8,9]. Small-sized babies are equally at substantial risk of post-neonatal mortality and a high morbidity rate in later life [4].

To alleviate the burden of perinatal mortality, improving birth outcomes and registration have been named as high priorities in the post-2015 development agenda [4,6,10], and a particular focus has been suggested on low-birth weight and a reduction in fetal growth-restriction [4,8].

One approach to contribute to the prevention of avoidable deaths and disabilities is accessible perinatal screening of mother and baby, and appropriate medical referral. Fetal wellbeing is most commonly assessed by monitoring of the fetal heart rhythm [11], with clinicians aiming to identify non-reassuring fetal status to initiate an appropriate and timely intervention [12].

1.1.2 Fetal monitoring modalities

In the UK electronic fetal heart rate monitoring became commercially available in the 1960s and was introduced widely into practice in the 1970s [13]. Electronic monitoring succeeded and in many cases superseded intermittent auscultation by Pinard stethoscope[1] and allowed for continuous surveillance with the ultimate goal of decreasing stillbirth and

[1]A Pinard stethoscope, also referred to as Pinard horn and formerly 'fetoscope', is a hollow cone shaped tool used to listen to the fetal heart rate during pregnancy. The stethoscope is named after its inventor, French obstetrician Dr. Adolphe Pinard; introduced in the 19th century, it is still in use by midwives around the world for intermittent auscultation.

injury to the fetal central nervous system [14]. Electronic fetal monitoring has established itself as a requisite element of care for mothers even at low risk in the perinatal period, during labour and birth, with cardiotocography having become the primary method for the assessment of fetal oxygenation, both antepartum and intrapartum [15].

Available monitoring techniques can be categorised into internal or direct (invasive) and external or indirect (non-invasive) methods. Whilst invasive methods are typically employed during labour only, non-invasive methods may find application much earlier in pregnancy. A particular interest of this investigation are techniques suitable for perinatal risk detection in a context where skills as well as financial means are scarce. Further review therefore centres on non-invasive solutions, which are less demanding in skill and generally lower in cost. A summary of the primary non-invasive modalities is provided in Table 1.1 [11,15–17,17–24]. Imaging modalities have not been included in this review as their training and financial barrier are much higher in comparison.

Modality	**Method**	**Application**	**Benefits**	**Limitations**
fECG	The indirect fetal Electrocardiogram (fECG), also referred to as abdominal fECG, measures the electrical activity of the fetal heart with body surface electrodes on the mother's abdomen.	FECG can be applied from >18 to 20 weeks onwards [16]. Signal acquisition is limited and can be impossible during weeks 28 to 32. The modality is used both antepartum and intrapartum, however is not yet routinely used in clinical practice.	From the fECG waveform the baseline heart rate and heart rate variability can be extracted with beat-to-beat accuracy [16]. FECG is an unobtrusive long term monitoring option [11].	FECG can have a low signal to noise ratio, variable signal quality, and the fetal signal is typically small with variable size (becomes smaller around 28th to 32nd week due to insulating effect of the vernix caseosa), extracting the fECG requires separating out the fetal component from the mother's ECG and other interferences [11].
fMCG	Fetal magnetocardiotocography (fMCG) measures the magnetic field of the cardiac cycle with a highly sensitive sensor, the generated cardiac waveform is comparable to the fetal ECG [11,16].	FMCG finds application typically antepartum from >20 weeks [11]. FMCG is not routinely used in perinatal assessment.	The fMCG produces a waveform of good quality. FMCG can be used to classify arrhythmias and to study fetuses with congenital heart diseases [11].	FMCG is usually measured in a shielded room. Signal amplitudes depend on the distance between sensor and the fetal heart and the electrical properties of surrounding tissues, hence amplitudes cannot be used for diagnostic purposes. In comparison to fECG and Doppler/CTG the MCG is expensive with highly skilled personnel required [11].

Doppler US	The transducer isonates the fetal heart with an ultrasound beam, a Doppler shift is imposed by the motion of the heart structures and recorded by the device. Demodulation of the beam allows for the frequency shift to be extracted.	Handheld Doppler finds application antepartum and intrapartum typically from >20 weeks onwards [11]. The handheld Doppler is primarily used intermittently to identify the fetal heart rate (the periodicity of the demodulated signal represents the time between heart beats [18]), heart rate variability, accelerations and decelerations (commonly assessed by auscultation).	The handheld Doppler is a low-cost modality widely used to determine fetal heart rate as the periodicity of the signal can be easily established [11].	The Doppler signal has a complex and changing structure, largely due to variations in the beam orientation. Compared to the transabdominal ECG fetal heart rate detection is generally adequate, albeit averaged. With low or rapidly varying FHR erroneous data can be generated. Typically not designed for continuous monitoring [18].
CTG	(Computerised) Cardiotocography ((c)CTG) simultaneously monitors the fetal heart rate with an ultrasound transducer (as above) and uterine contractions with a pressure sensitive transducer [16].	CTG is applied both antepartum and intrapartum [15,19] from >20 weeks onwards. CTG is the most commonly used method for fetal monitoring during labour and delivery as well as higher risk pregnancies antenatally. The baseline heart rate, long-term fetal heart rate variation, short term variance, episodes of high variation, episodes of low variation, acceleration and decelerations can be extracted from its signal [22].	The CTG is a well-known and established modality designed for continuous monitoring [20].	Interpretation of the cardiotocogram (when not computer-assisted) is subject to high intra- and interobserver variability. Some argue CTG should not be classified as non-invasive due to its inherent ultrasound irradiation [21–23]. This view would also apply to the handheld Doppler described above.
fPCG	The fetal phonocardiogram (fPCG) enables a graphic registration of heart sounds produced by the contracting heart, picked up as vibrations by a microphone [16]. The acoustic waveform capturing a normal, noiseless cardiac cycle contains the significant heart sounds (the mitral and tricuspid valves' closure sound, and the aortic and pulmonary valves) [21].	Application is theoretically possible both antepartum, >24 weeks [25], and intrapartum. The fPCG is the electronic extension of the traditional stethoscope and developed as an alternative to ultrasound techniques. There is currently no commercial device available and it is therefore not used in clinical practice.	PCGs are low in cost, safe and easy to use. Having an omnidirectional pattern it is claimed a passive sensor cannot miss a fetal heart signal [26].	Typically designed for intermittent monitoring only, equivalent to other auscultation methods such as the Pinard stethoscope. The low acoustic energy resulting from the small fetal heart and the noisy environment render the accurate detection of fetal heart sounds difficult and sometimes impossible [21,25].

Table 1.1: An overview of fetal cardiac monitoring modalities. Devices reviewed only include those which are commonly considered non-invasive. Gestational ages noted in the table refer to the time onwards from which the respective modality is typically used in practice; some devices however, such as Doppler ultrasound, are able to capture fetal cardiac activity as early as 12 weeks.

The introduction of such monitors in lower resource environments however requires careful consideration adressing skill-dependence, cost factors and patient coverage. Reviews on the effectiveness and benefit of CTG have raised concerns regarding its reliability

and reproducibility. Studies have not been able to demonstrate that current antenatal CTG significantly improves perinatal outcome [15,19]. One limitation of CTG is attributed to the difficulty and hence often poor quality of the interpretation of fetal heart rate traces; false negative results may have severe consequences regarding the health outcomes of the fetus and the mother, while false positive results may lead to inappropriate and potentially risk-bearing procedures and an additional burden on resources.

More promising results compared to standard CTG could be demonstrated using computer-assisted interpretation methods [15], as the accuracy may be improved and variation in interpretation may be eliminated. Further evaluation and development of computerised interpretation and decision support has been recommended as no high-quality evidence exists comparing no CTG to computerised interpretation of CTG [15,24]. However, when drawing on evaluations of methods to reduce perinatal morbidity and mortality, it has to be considered that those tools (such as CTG) are predominantly used in clinical environments supported by additional fetal assessment tools such as diagnostic imaging (2D ultrasound). Reliable and affordable fetal heart rate monitoring might have a substantial positive impact where access to other techniques is limited or not available.

The effectiveness of antenatal monitoring using CTG, or equivalent, in lower resource settings has not been studied sufficiently, and no high-quality randomised controlled trials were included in the most recent Cochran review [15,24]. To enhance the effectiveness and efficiency of fetal heart monitoring systems, the World Health Organization's 2010 report *Medical devices: managing the mismatch* [27] suggests research into technologies that eliminate the problem of high false-positive rates often reported in conjunction with the use of electronic fetal monitoring. The report also highlights the need for devices that are easy to use and interpret, circumventing skill-dependence, and which are appropriate for the infrastructure available in lower income countries to achieve higher coverage of care.

1.1.2.1 Fetal monitors for use in low resource settings

A number of efforts have been made recently to introduce lower cost, hand-held fetal monitoring devices for use in LMICs. Devices that are currently available or undergoing development and testing are summarised in Table 1.2. All monitors listed offer fetal heart rate only and no inbuilt intelligence or decision support.

Device	Company	Cost	Functionality	Development status
Moyo Fetal Heart Rate Monitor [28] (Doppler ultrasound)	Laerdal Global Health (not-for-profit)	$198	• Pulsed Doppler transducer, intermittent and continous monitoring • Measures maternal heart rate with dry ECG electrodes • Baseline fetal and maternal heart rate display • Visible and audible alarm for abnormal fetal heart rate	In production
Wind-Up Fetal Heart Rate Monitor [29] (Doppler ultrasound)	PowerFree Education Technology and Philips Healthcare	NA	• Doppler transducer • Baseline fetal heart rate display • Rechareable battery (wind-up) • Audio output	Pre-production prototype
Fetal Heart Rate Monitor for Resource-Limited Settings [30] (Phonocardiography)	Michigan University (student project)	NA	• Microphone and acoustic cone • Baseline fetal heart rate display • Variations from the baseline display • Audio output	Proof of concept prototype

Table 1.2: Fetal monitors designed for use in lower resource settings currently available or under development. The list only includes those devices on which information is available in the public domain. NA indicates information not available.

1.1.3 mHealth

Mobile health (mHealth) is a subcategory of electronic health (eHealth) and related to telemedicine. There is no standardised definition for mHealth available, however, generally accepted descriptions include [31]:

> *Telemedicine (also telecare or telehealth): medical practices mediated remotely, such as over the phone, by video, or asynchronously through a web service.*

mHealth: mobile healthcare, generally referring to telemedicine using a mobile phone.

The World Health Organization defines mobile health as follows [32]:

The practice of medical and public health through the usage of mobile devices.

A mobile device may refer to a mobile phone, a personal digital assistant (PDA), or any medical device enabled to communicate wirelessly with cellular or WIFI networks.
Mobile health interventions can span the entire patient pathway as well as being a tool for improvement in health systems. The functionality of mHealth solutions includes personal wellness monitoring and disease management, information provision and education, decision support, data management and access, or results consultation [33]. Typically, mHealth aims to utilise the inherent capabilities of mobile devices for healthcare purposes. Existing mHealth interventions are to a large extent based on familiar technologies such as text messaging, handset applications, the internet, voice, video, or a combination thereof. Less frequently found, mHealth has the potential to equally exploit integral technologies such as global positioning systems or accelerometers [34,35].
The advances in mobile phone technology, the continuous decrease in cost and the rapidly increasing use of mobile phones worldwide (mobile phone adoption is projected to reach 96.8 and 91.8 per 100 inhabitants by the end of 2015, globally and in developing countries respectively [36]) have driven the development of new mHealth interventions in recent years. Mobile health interventions promise opportunities to extend healthcare delivery and improve outreach, to integrate data more easily with existing eHealth services and to allow for independent and personal health monitoring, to name a few of the promising applications available [32].
Despite mobile health being an increasingly active area of medical product and service development, large scale implementations of mHealth initiatives still lag behind [35,37]. This is particularly true in lower resource environments where mHealth initiatives often are small-scale pilots [38,39]. One aspect influencing the adoption of novel health delivery

methods is the evidence base available to decision makers. It has been argued repeatedly that rigorous evaluations of mobile health interventions that support the claims of developers are still scarce [37,39–42] with progress towards the building of a more robust evidence base being noticeable only recently [34,35,43].

1.1.3.1 mHealth applications for maternal and fetal health

In order to achieve set Millennium Development Goals (MDGs) [2] and further improve maternal and perinatal health it has been acknowledged that information and communication technologies (ICTs) have an increasingly important role to play [44]. The rapidly increasing coverage of smart mobile devices not only in developed regions, but to an astonishing extent in the world's rural communitites and settings where other technologies remain scarce, and ICTs' promising potential in healthcare delivery, have spurred the development of mHealth solutions [45].

Existing applications of mHealth along the continuum of care for perinatal health can be loosely categorised into two areas. Firstly, applications for pregnant women and mothers, including appointment, medication or immunisation reminders, SMS-based health education, improved communication with health workers and improved access to emergency obstetric care. Secondly, applications to improve the capacity of traditional birth attendants (TBAs) and healthcare workers (HCWs), including data collection and sharing, point-of-care remote consultation, referral services and phone-based education [45,46].

A number of mHealth fetal monitoring projects employing a stethoscope or Doppler can be identified [47–49]. The majority however are designed for remote personal monitoring and only one article mentions the potential use in low-resource settings explicitly [47]. None reported in-depth qualitative or quantitative testing, which mirrors the general lack of adequate evaluation of mHealth interventions for maternal, fetal and newborn health [45]. Reviews furthermore found that mHealth projects tend to focus on one aspect of pregnancy only, without acknowledging the importance of the care continuum [45,50].

Key challenges of perinatal mHealth solutions for lower resource environments include not only cost, but also the socio-cultural context and the formal or informal healthcare systems in place. It has been highlighted that, for a successful implementation of mHealth initiatives, their integration into the existing infrastructure, supporting and complementing resources available is crucial [35]. This includes working with frontline community healthcare workers to ensure ownership and committed participation [44].

1.1.4 Field implementation - Guatemala

The successful development of any healthcare intervention requires the early involvement of existing care providers and relevant stakeholders. The non-governmental organisation *Wuqu' Kawoq / Maya Health Alliance*[2] was established in 2007 to provide healthcare solutions for Guatemala's poor, under-served and mostly indigenous communities. Wuqu' Kawoq has a well established midwife program aiming to empower traditional birth attendants (TBAs) to improve care delivery for expectant mothers and newborns in their communities. Wuqu' Kawoq is therefore an excellent partner in the effort to develop and evaluate the appropriateness of anticipated mHealth perinatal monitoring concept presented in this thesis.

Guatemala is likewise an excellent case study for the complexity of the health system problems that underlie high perinatal morbidity and mortality. Health statistics for 2013 estimated 49% of all births to occur in the community as supposed to facilities with only 52% attended by a skilled healthcare worker; the percentage of low birth weight babies ($<$ 2500 g) is high at 11% and the estimated adjusted[3] maternal mortality ratio of 140 (maternal deaths per 100 000 live births) is among the highest in Central America [51] (see Figure 1.3 for population and health statistics). Guatemala will not meet MDG4 or MDG5 by the end of this year. In comparison, most developed nations have reported

[2]Wuqu Kawoq | Maya Health Alliance, 2 calle 5-43 zona 1, Santiago Sacatepéquez, Guatemala 03006, Central America. Web URL: www.wuqukawoq.org

[3]'Adjusted' refers to statistics being adjusted for under-reporting and misclassification [51].

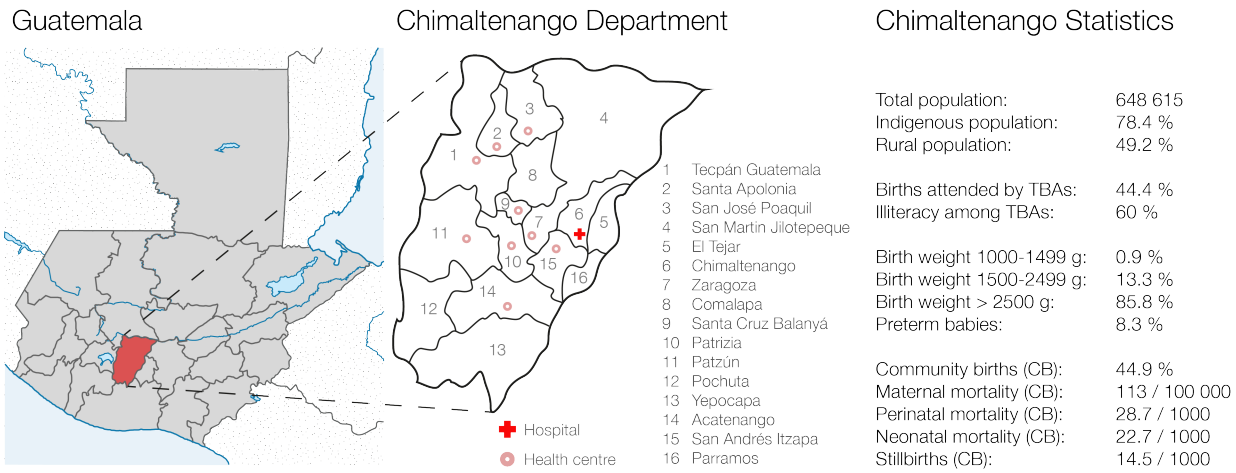


Figure 1.3: The most recent (2013) population [53] and health statistics [52] for the Chimaltenango region in Guatemala, Wuqu' Kawoq's primary area of care provision.

maternal mortality ratios of 10 or less [7] and a low birth weight percentage of under 7% [51]. The relationship between gross national income per capita (GNI) and low birth weight as well as neonatal mortality is shown in Figure 1.4. A significant discrepancy persists in skilled birth attendance between urban (77%) and rural (37%) communities as well as the poorest and richest 20% of the population (20% and 95% respectively) [51]. In 2010 Guatemala introduced policies to promote maternal and neonatal health requiring authorities to provide basic services for women throughout pregnancy and the neonatal period, and initiatives such as obstetrics and paediatric residency programs to increase the availability of skilled staff [52]. In a recent study [52] researchers analysed the progress in reducing maternal, fetal and neonatal mortality rates in the Guatemalan state of Chimaltenango, which is Wuqu' Kawoq's primary area of service delivery. They found maternal mortality lowered from 134 per 100 000 live births in 2010 to 113 in 2013, both stillbirth and perinatal mortality rates significantly decreased from 22.0 per 1000 births to 16.7 and from 43.9 to 31.6 respectively. Notable were the differences in outcomes for facility and home deliveries, illustrated in Figure 1.5. Whereas perinatal mortality could be reduced by 41.4% in facilities, it only reduced by 21.4% for home births over the same period. The same applies to stillbirths with a reduction of 39.0% and only 17.6% respectively.

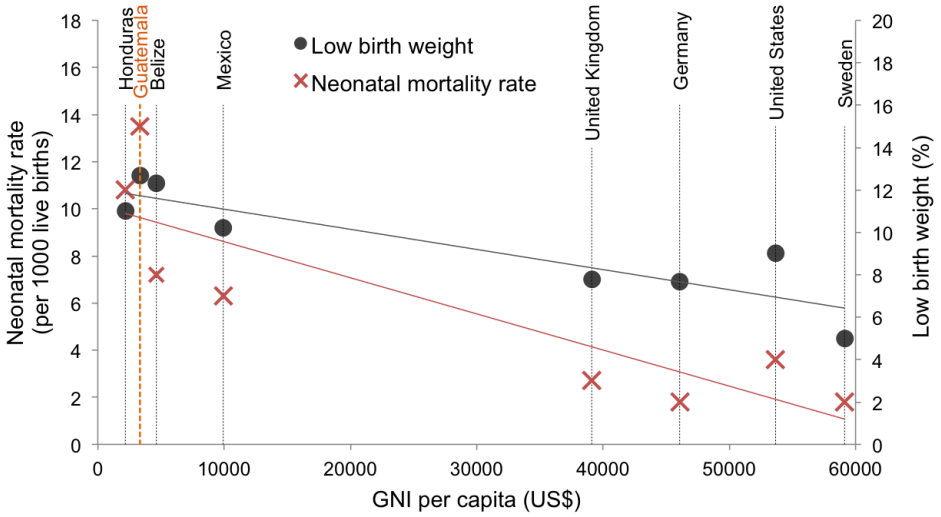
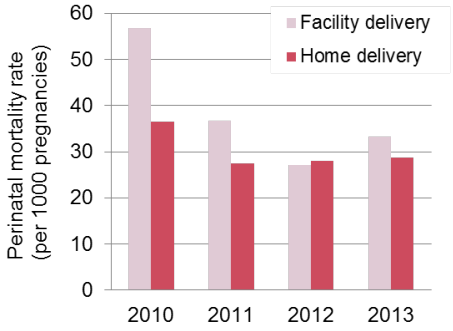
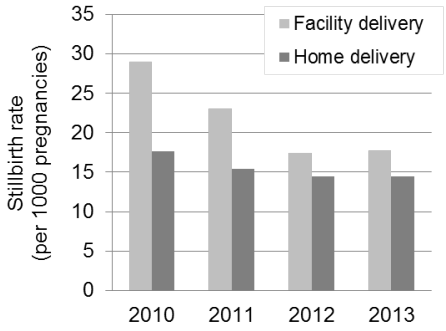


Figure 1.4: The percentage of low birth weight (<2500g) and the rate of neonatal mortality plotted against the gross national income per capita, comparing Guatemala with its neighbouring countries and examples of developed nations. The 2013 country statistics used to produce this graph can be found in [51].



(a) The reduction in perinatal mortality as estimated for Chimaltenango, Guatemala, between 2010 and 2013 [52].



(b) The reduction in stillbirths as estimated for Chimaltenango, Guatemala, between 2010 and 2013 [52].

Figure 1.5: Both perinatal mortality (a) and stillbirths (b) were observed in a recent study [52] to have decreased between 2010 and 2013, coinciding with policy changes to improve access to obstetric care. Whereas mortality could be reduced significantly in healthcare facilities, a slower decrease was noted for home births. Notable, however not discussed in the study report [52], are the overall higher mortality rates reported for facility births. A likely explanation is that the proportion of high risk pregnancies was much greater in facilities, especially before introduced policy changes in 2010, accounting for a higher death rate. Introduced policies heavily promoted facility deliveries leading to almost twice as many facility births in 2013 (2537 births) compared to 2010 (1277 births). It may be hypothesised that the ratio of high risk to normal pregnancies decreased over the study period as a result and therefore contributed, alongside improved antenatal care provision, to the decrease in observed mortality rates.

Given the discrepancy in mortality reduction dependent on birth location, researchers have advocated to phase-out community deliveries led by traditional birth attendants (TBA) [52], which account for at least 59% of births in rural areas of Guatemala [54]. TBAs however, known as 'comadronas', are an integral part of indigenous communities, not only as providers of reproductive care but also as trusted leaders [52]. This may be one of the reasons why expecting women in the rural highlands often choose delivery at home attended by a TBA over delivery at a healthcare facility [52,55]. Equally important are cultural differences, language barriers and high illiteracy rates that deter Mayan mothers from accessing the formal healthcare system. An alternative approach taken by Wuqu' Kawoq is to integrate comadronas into the care delivery model. This approach includes culturally appropriate training to TBAs and education for women in their own language addressing skill and knowledge gaps, to support solutions that empower comadronas to assess perinatal risk early reducing delays in advice seeking, and to include comadronas into the referral process strengthening the linkage to formal healthcare services.

1.2 Thesis outline

The reduction of pregnancy related complications leading to fetal, early neonatal and maternal death and increased morbidity in developing countries deserves particular attention as little progress has been made in the past 15 years. This thesis in Healthcare Innovation aims to address some of the issues related to high perinatal mortality rates in resource-limited environments by proposing an mHealth Doppler approach to fetal cardiac monitoring.

Chapter 2 provides background information on the fetal cardiac physiology, on methods for fetal heart rate estimation and markers to identify fetuses at risk of growth retardation discussed in literature.

Intra-uterine growth restriction (IUGR) is one of the complications significantly contributing to high mortality rates in under-served regions. Chapter 3 investigates the

utility of Doppler-derived fetal heart rate features for IUGR detection to allow for early risk assessment and appropriate referral, if required.

The first prototype of anticipated mHealth system is introduced in Chapter 4, including its hardware and data recording application. The prototype was requisite for the collection of raw Doppler recordings for the development of risk assessment capabilities, and to establish the feasibility of implementing such monitoring solution.

The implementation of novel technologies in low-resource settings creates challenges. Chief amongst these is the likelihood that unskilled or semi skilled healthcare workers are more likely to record lower quality signals and substantially increase the number of false positives or false negatives. Chapter 5 discusses the need for signal quality assessment in order to ensure adequate decision support and introduces a classifier to distinguish Doppler recordings suitable for clinical use from those too poor for further analysis.

Heart rate estimation is the basis for the assessment of fetal cardiac abnormalities. An algorithm to approximate the fetal beat-to-beat heart rate compared to the commonly implemented averaging technique is presented in Chapter 6.

Chapter 7 presents the results of initial feasibility testing conducted in rural Guatemala and discusses the implications for the mHealth system based on the findings. It describes all improvements implemented and introduces the proof-of-concept evaluation study in progress. The thesis ends in Chapter 8 with a summary of contributions, the primary research findings and an outlook of possible avenues for further work.

1.3 Summary of datasets

The work presented in this thesis is based on three datasets, one that comprises variables derived from the fetal heart rate trace, and two comprising raw 1D fetal Doppler data. As there are currently no databases of raw antenatal Doppler available to researchers, the latter two were collected as part of this project. The key characteristics of each dataset are summarised in Table 1.3.

Database	Signals and derived metrics	Patient information	N Patients/Recs	Total recs [hh:mm:ss]	Pathologies	Annotations
JR cCTG DB	LTV STV Recording length	GA Gender Time to delivery Birth weight	2326/2326	NA	IUGR Normal	NA
JR RDDB	1D Doppler (fs=44.1 kHz) Pulse oximetry (fs=75 Hz) Accelerometry (fs≈96 Hz)	GA Maternal age	17/17	01:56:17	Normal	Doppler quality Heart beat instances
WK PDB	1D Doppler (fs=44.1 kHz) Pulse oximetry (fs=75 Hz) Accelerometry (fs≈96 Hz)	GA Maternal age	20/22	04:16:42	Normal	None

Table 1.3: The key characteristics of the databases used in this thesis. (STV: short-term variability of the fetal heart rate, LTV: long-term variability, GA: gestational age of the fetus, IUGR: intra-uterine growth restriction). NA indicates not applicable.

1. *JR antenatal cCTG database (JR cCTG DB)*: The database of the John Radcliffe (JR) hospital in Oxford, UK, contains antenatal 'computerised Cardiotocography' (cCTG) recordings collected between 1990 and 2011, using the Sonicaid Fetal Care system (Huntleigh Healthcare Ltd.). cCTG records contain the patient's fetal heart rate trace (heart rate estimated over 3.75s epochs) computed from the Doppler ultrasound signal of between 10 to 60 minutes, heart rate variables such as short-term variability (STV) and long-term variability (LTV) computed as average per minute, and instances of fetal movement recorded with the help of a handheld patient event marker and/or by an automatic fetal movement detector (fetal movement derived from the ultrasound signal) [56,57]. The database is not publicly accessible.

   A subset of data of this extensive antepartum archive was made available for this thesis to investigate the potential of heart rate variables for IUGR risk detection, described in Chapter 3. The subset contains patient information and metrics from the last CTG trace of the identified 1163 IUGR patients in the database, matched for gender and gestational age with normal controls. The available metrics per patient are: STV and LTV metrics averaged over the duration of the recording as well as over episodes of high and low fetal activity, the fetal gestational age, gender, birth weight and the time from recording to delivery. See Section 3.2 for more details.

2. *JR antenatal raw 1D fetal Doppler database (JR RDDB)*: The 1D fetal cardiac Doppler data were collected at the John Radcliffe (JR) Hospital, Oxford, UK, as part of this thesis (approved by the NHS Health Research Authority, REC reference: 12/SC/0147). Data were recorded onto a mobile phone to test the first mHealth prototype and to lay the basis for a dataset of raw fetal Doppler signals. The database contains 1D fetal cardiac Doppler signals recorded from 17 patients and volunteers using a handheld transducer (AngelSounds Fetal Heart Detector, Jumper Medical Co., Ltd.), concurrent maternal pulse oximetry (WristOx2® 3150, Nonin Medical, Inc.) and the phone's accelerometer data (Samsung Galaxy S2, Samsung Electronics Co., Ltd.). Available also are maternal and gestational age. The Doppler signals were annotated for both signal quality and heart beat location. Data were collected by research midwives and clinicians.

   The dataset was used in Chapter 5 and Chapter 6 for the development of a Doppler signal quality index and a segmentation algorithm for heart rate estimation respectively. The recording equipment used is described in Chapter 4. Further information on the data collection process and database characteristics is included in Chapter 5.

3. *Wuqu' Kawoq antenatal raw 1D fetal Doppler pilot database (WK PDB)*: Using the same equipment as for data collected at the John Radcliffe Hospital, this database was built in Guatemala in collaboration with Wuqu' Kawoq and its network of traditional birth attendants (approved by the Wuqu' Kawoq Institutional Review Board, reference: WK-2014-003). Data were recorded to assess the quality with which untrained healthcare workers can operate the monitoring devices and hence test the feasibility of implementation and establish requirements for system improvement and training. The database builds upon data collected at the John Radcliffe Hospital towards a more substantial archive of raw fetal Doppler data available for future research. The process of data collection and outcomes are discussed in Chapter 7. The Doppler signals in this database were not annotated for quality or heart beat location.

1.4 List of publications and presentations

Work in this thesis has been presented at the following conferences and is in submission to the following journals:

- L. Stroux, G. D. Clifford. "An Affordable Multisensor Perinatal Monitoring Concept", poster presented at the Second WHO Global Forum on Medical Devices: Priority Medical Devices for Universal Health Coverage, Geneva, 22-24 November, (2013).
  (This work appears in Chapter 4)

- L. Stroux, G. D. Clifford. "The Importance of Biomedical Signal Quality Classification for Successful mHealth Implementation", paper presented at the 2014 Tech4Dev International Conference, Lausanne, 4-6 June, (2014).
  (This work appears in Chapter 5)

- L. Stroux, N. E. King, S. Fathima, R. Hall-Clifford, P. Rohloff, G. D. Clifford. "A lowcost perinatal monitoring system for use in rural Guatemala", poster presented at Appropriate Healthcare Technologies for Low Resource Settings - AHT 2014, London, 17-18 September, (2014)
  (This work appears in Chapter 4 and 7)

- L. Stroux, A. Georgieva, C. W. Redman, S. Payne, G. D. Clifford. "Identification of IUGR from antenatal fetal HRV", in preparation for AJOG.
  (This work appears in Chapter 3)

- L. Stroux, S. Payne, G. D. Clifford. "A 1D Ultrasound Signal Quality Measure". In preparation for IOP Phys Meas.
  (This work appears in Chapter 5)

- L. Stroux, S. Payne, G. D. Clifford. "A HMM approach to segmentation and heart rate estimation on 1D ultrasound data". In preparation for IOP Phys Meas.
  (This work appears in Chapter 6)

- L. Stroux, P. Rohloff, R. Hall-Clifford, S. Payne, G. D. Clifford. "An mHealth monitoring system for midwife-led antenatal risk assessment in rural Guatemala". In submission to Journal of Medical Engineering and Technology.
  (This work appears in Chapter 4 and 7)

The work described in this thesis led to NIH funding (Research Project Grant (R01) Number: 1R21HD084114-01) for a full RCT in rural Guatemala, which will be discussed in Chapter 7.

1.5 Contributors

The following people contributed to work presented in this thesis:

- Dr. Rachel Hall-Clifford, Dr. Boris Martinez, Dr. Peter Rohloff and Enma Coyote Ixen helped with identifying some of the pictograms, and the work flow of the intervention, described in Chapter 7.

- Dr. Oliver Gibson and Dr. Carmelo Velardo wrote the Android interface to the NONIN pulseoximeter with input from Arvind Raghu, used in the mobile application described in Chapters 4 and 7. Elements of the application architecture have been contributed by Jonathan Daly.

Unless stated otherwise, described research is my own work.

Chapter 2

Literature review

2.1 Introduction

This chapter provides background information on the physiology of the fetal heart, the control of the fetal heart rate by the autonomic nervous system and its short and long term response to reduced oxygen supply. An overview of fetal heart rate estimation methods and cardiac signal based metrics of fetal growth restriction that are described in the literature concludes this chapter.

2.2 Physiology

2.2.1 The fetal cardiac cycle

The fetal heart starts developing in the third week of gestation and is fully formed by the end of the eighth week [58]. By 10 weeks the cardiac signal can typically be identified by auscultation Doppler. The fetal heart has a number of structural characteristics that adapt it to the unique in utero environment, such as the *Foramen Ovale*, which allows blood to enter the left atrium from the right, and the *Ductus Arteriosus*, which allows blood to flow directly from the pulmonary artery into the aorta, bypassing the lungs. These adaptations allow for the fetal ventricles to pump in parallel driving the

blood to the systemic circulation and placenta, the supplier of oxygenated blood and nutrients. Differences however primarily alter the circulation physiology, whilst cardiac valve and wall motion are the same pre- and postnatally. The primary heart structures are illustrated in Figure 2.1. As in the adult the fetal heart cycle is divided into a diastolic and systolic phase whereby the blood travels from the superior and inferior vena cava and pulmonary veins through the right and left atria respectively into the right and left ventricles. An overview of the heart cycle is provided in Figure 2.2 with a detailed description of the relevant cardiac events listed in Table 2.1.

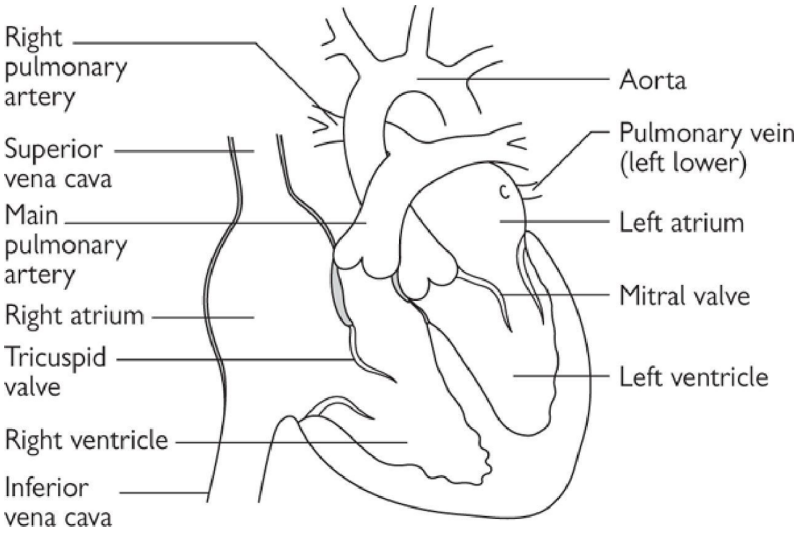


Figure 2.1: Illustration of the normal fetal heart depicting its main structures [58].

2.2.2 Control of the fetal heart rate

The primary control mechanisms of the fetal heart are the sinoatrial (SA) node, the autonomic nervous system (ANS), and the baroreceptors and chemoreceptors. Figure 2.3 provides an illustration of the fetal heart rate control mechanism. The fetal baseline heart rate is determined by the depolarisation of the SA node, situated in the right atrium. The SA node, also called the pacemaker, has an intrinsic rate of 110 to 160 bpm, which is modulated by the ANS and receptor stimulus as follows [60,61].

The ANS is composed of the sympathetic and parasympathetic nervous system which

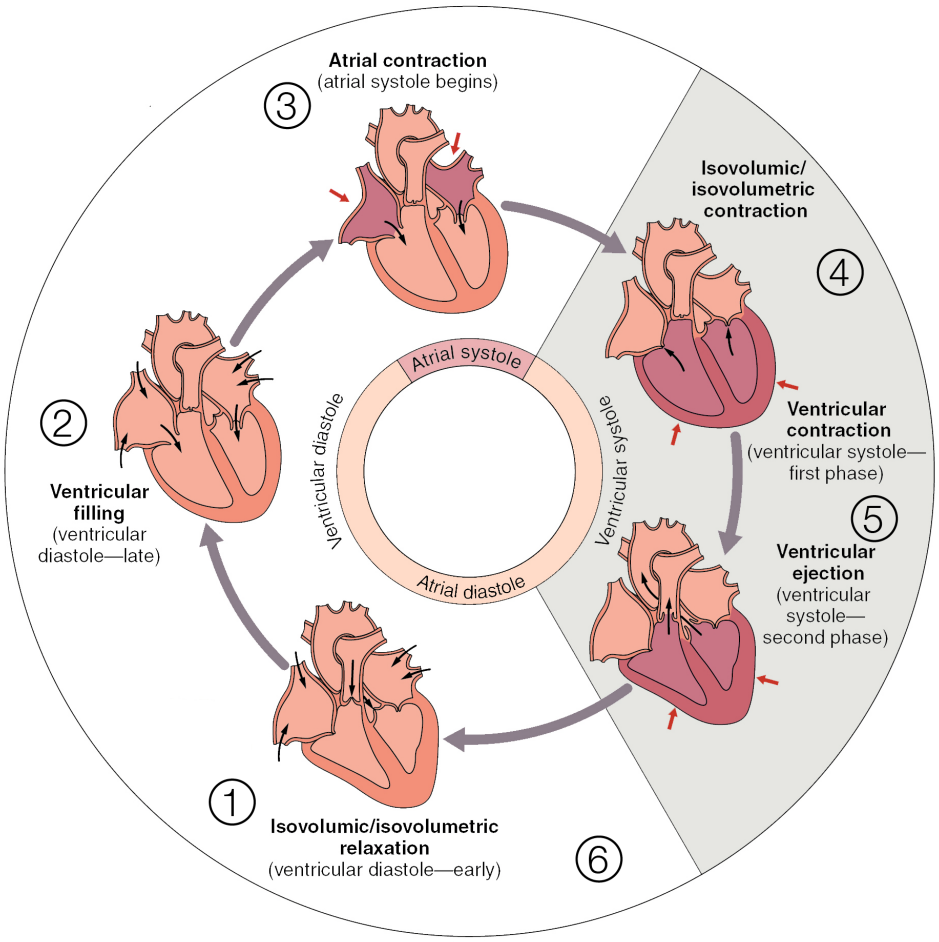


Figure 2.2: The phases of a cardiac cycle illustrated. Adapted from [59]. Table 2.1 contains a more detailed description of the cardiac events.

innervate both atria, ventricles and the SA node respectively. The parasympathetic input (vagal stimulation) decreases the firing rate of the SA node and reduces the speed of transmission down to the ventricles, which causes the heart rate to lower. The sympathetic nervous system has the contrary ability to accelerate the heart rate. The balance between parasympathetic and sympathetic inputs sets the baseline heart rate, and its continuous recalibration generates the heart rate's variability. The sympathetic system matures earlier leading to higher fetal heart rates in the preterm fetus. With advancing gestational age the parasympathetic system matures and its influence on the control of the fetal heart rate increases. The baseline heart rate therefore decreases with advancing gestation. The normal fetal heart rate range is reported as 120 to 160 bpm throughout

	Phase	Right heart	Left heart
1	Isovolumetric relaxation	Blood flows from the inferior and superior vena cava into the right atrium.	Blood flows from the pulmonary veins into the left atrium. (The *Foramen Ovale* connects the atria and allows additional blood flow from right to left).
2	Ventricular filling	The tricuspid valve opens and bloods starts filling the right ventricle.	The mitral valve opens and blood starts filling the left ventricle.
3	Atrial contraction	The right atrium contracts.	The left atrium contracts.
4	Isovolumetric contraction	The right ventricle starts contracting, the tricuspid valve closes. (The pulmonary valve remains closed)	The left ventricle starts contracting, the mitral valve closes. (The aortic valve remains closed)
5	Ventricular ejection	The right ventricle contracts, the pulmonary valve opens and blood ejects into the pulmonary artery towards the lungs.	The left ventricle contracts, the aortic valve opens and blood ejects into the aorta.
6	Isovolumetric relaxation	The pulmonary valve closes, the right ventricle relaxes.	The aortic valve closes, the left ventricle relaxes.

Table 2.1: The blood flow, wall and valve motions during one cardiac cycle. Refer to Figure 2.2 for illustrations of each phase.

gestation [62] and has been found to typically remain between 110 and 180 bpm [63].

The fetal heart rate is furthermore controlled by two sensor systems. Baroreceptors are pressure sensors situated in the aortic arch, carotid arteries and brain stem. Sensing an increase in blood pressure, the vagal nerve receives a stimulus to slow the heart rate and as a result lower the pressure. A fall in blood pressure decreases the parasympathetic 'tone' leading to an increase in both heart rate and blood pressure. The chemoreceptors, likewise found in the aorta, carotid artery and brain stem, sense a decrease in circulating oxygen (hypoxemia). Falling O_2 levels increase the fetal heart rate and cardiac output to improve placental perfusion and therefore oxygenation. An abrupt and severe reduction in ogygen levels (hypoxemia), or an increase in carbon dioxide (hypercapnia), however triggers a vagal response to decrease heart rate and increase blood pressure, a

phenomenon observable in the event of cord compression.

Fetal beat-to-beat intervals change constantly as a result of the fluctuations in vagal impulse and sympathetic reflexes. Normal baseline variability is a sign of balanced parasympathetic and sympathetic control and an indication of a well-oxygenated fetus. The fetal heart rate variability can be affected by several factors such as gestational age, fetal movement, fetal sleep state, acidemia (low blood pH) and hypoxia (decreased oxygen in tissue).

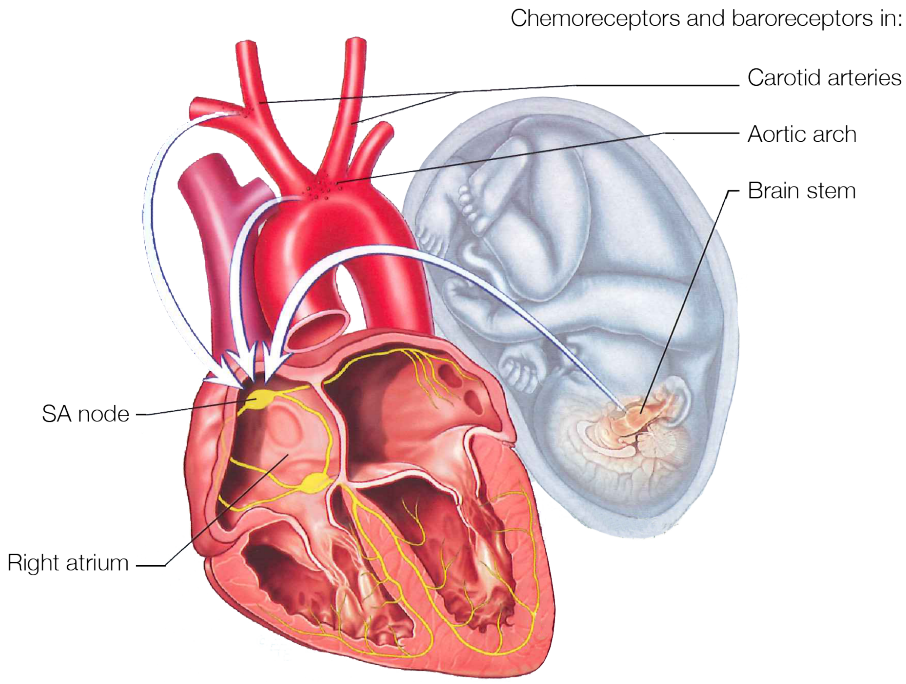


Figure 2.3: The vegetative control of the fetal heart rate. Adapted from [60]

2.2.3 Fetal response to hypoxia

The fetus typically copes well with hypoxic insults such as temporary interruption or reduction in uterine or umbilical blood flow. The fetus can compensate transient hypoxemia by redistributing blood flow to organs vital for survival such as the heart and brain, and by reducing flow to peripheral organs. Fetal movements decrease and the heart rate slows minimising oxygen consumption.

Chronic hypoxia also triggers a slow-down in oxygen consuming activities to direct re-

sources to more vital functions, however, over a prolonged period this may have a direct impact on growth. The fetus is able to adapt to continued reduced delivery of oxygen and nutrients whilst maintaining their metabolism by adjusting its growth rate. With time this reduction in growth rate will manifest itself in the pathological case of fetal growth restriction (FGR), also referred to as intrauterine growth retardation (IUGR) [61,64].

2.2.3.1 Fetal growth restriction

Fetal growth restriction is a pathological process which inhibits the fetus to reach its genetic growth potential. Factors increasing the risk of IUGR can be manifold and include maternal complications, such as malnutrition or vascular disease, fetal complications, such as congenital infection or infectious diseases, as well as placental abnormalities, such as utero-placental vascular insufficiency or infarction [65].

Numerous complications may result from IUGR, ante-, intra-, and postpartum. They include stillbirth, fetal asphyxia (deprivation of oxygen), the need for an emergency Cesarean section and neonatal resuscitation. IUGR may also result in paediatric developmental delay. Preparedness for such perinatal complications is of particular importance where resources are limited and referral to higher level care centres is required for adequate care for mother and baby.

2.3 Fetal heart rate estimation

The most accurate measurement of the duration of the fetal cardiac cycle is by direct ECG whereby the periodicity is estimated as the time interval between successive R-waves in the electrocardiogram. Resultant fetal heart rate trace is a time-event series with each RR-interval T_{RR_i} expressed in beats per minute: $FHR_{RR_i}[bpm] = 60000/T_{RR_i}[ms]$.

The use however of fetal scalp eletrodes is limited to labour when sufficient cervical dilation allows safe application. The abdominal fetal ECG, which has potential to be used antenatally and equally provides beat-to-beat information, is not yet established

in clinical practice. Most wide-spread in use in both antenatal surveillance and intra-partum monitoring remains Doppler ultrasound, applied by intermittent auscultation with a handheld continuous Doppler, or by cardiotocography which allows for continuous monitoring typically with a pulsed Doppler transducer. (See Table 1.1 for a summary of non-invasive monitoring modalities).

2.3.1 Fetal heart rate estimation from 1D Doppler ultrasound

The challenges of non-invasive accurate fetal heart rate estimation and the high inter- and intra-observer error rates in determining the cardiac rhythm and interpreting heart rate patterns by visual inspection were recognised as early as 1950 [66]. With the introduction of electronic ultrasound monitors into routine clinical practice fetal heart rate estimation algorithms emerged, primarily based on direct peak detection from the amplitude signal [48,67]. Given the high susceptibility of those early algorithms to noise leading to erroneous peak detection and high rates of signal loss (resulting in inaccurate heart rate estimation) alternatives were sought. In the 1980s algorithms were largely replaced by correlation methods which had proven more robust whilst maintaining acceptable levels of accuracy compared to direct fetal electrocardiograms [48,67–69]. Their wide-spread implementation in commercial monitors was further made possible by the advent of the micro-processor [69]. Both cardiotocographs and handheld Dopplers in use today still employ autocorrelation based techniques, albeit improved progressively [70]. Current systems have become closed which makes reviewing employed algorithms or accessing raw data for further analysis impossible.

2.3.1.1 Autocorrelation

Autocorrelation is a powerful technique for uncovering regular patterns by comparing the signal of interest with a time-delayed version of itself. Whereas background noise and interference typically do not show with regularity, the fetal cardiac activity has a

periodic signature. The signal and its time-shifted copy are multiplied sample-wise and the products summed over the analysis window as a measure of similarity. Samples of similar amplitude and the same polarity will produce a higher correlation value whereas those of differing amplitude and possibly opposite polarity cancel each other out resulting in a lower correlation value. Repeating computation for lags $k = 1, 2, ...$ produces the sample autocorrelation function r_k, which can be expressed as follows [69,71]:

$$r_k = \frac{c_k}{c_0}, \qquad \text{where} \qquad c_k = \frac{1}{T-1}\sum_{t=1}^{T-k}(y_t - \overline{y})(y_{t+k} - \overline{y}) \qquad \text{for } k = 1, 2, ... \tag{2.1}$$

and c_0 represents the sample variance of the time series, y the analysed waveform. The above equation takes into account implied stationarity with common mean and variance, normalising the output to values between 0 and 1.

In summary, the autocorrelation function decreases the noise contribution and emphasizes the periodicity of the input signal. The periodicity can be easily quantified representing the mean interval between cardiac events in the analysis window, making this method attractive for use in fetal ultrasound processing.

2.3.1.2 Autocorrelation based methods and recent developments

The majority of recent developments present adaptations of the 'classic' autocorrelation approach. Investigators have addressed several aspects of correlation-based methods. They include a reduction in processing time for handheld devices [72], decreasing the number of missed FHR samples [73], exploring signal envelope alternatives for correlation assessment [74,75], assessing optimal parameter settings, such as the auto-correlation window and overlap size [48,76], and the evaluation of a number of different correlation approaches (auto-, cross-correlation, correlation coefficient and YIN estimator[1]) [73].

[1]YIN is a method developed for the estimation of the fundamental frequency of speech or musical sounds, it is based on auto-correlation combining a number of modifications with the aim of lowering error rates [77].

One of the autocorrelation properties most important in the context of this investigation is the averaging of fetal heart rate over the chosen analysis window. Conventional bedside monitors typically readout fetal heart rate every 250 ms, which equates to a sampling rate of 4 Hz. The choice of sampling interval is based on a maximum expected fetal heart rate of 240 bpm [78]. Given that those monitors are proprietary closed systems, the correlation window size applied is not known. Several researchers have however noted the consequential effect of averaging on heart rate variability measures, which are significantly lower compared to variability measures computed from the fetal electrocardiogram [68,79,80]. A number of improvements have been suggested in the literature to progress from the evenly sampled heart rate trace to a more accurate approximation of the true beat-to-beat event series.

Four articles were found describing a method that explicitly aims to estimate beat-to-beat intervals from 1D Doppler recordings. The methods employed follow either of two approaches: the post-processing of the evenly sampled autocorrelation trace by eliminating duplicate samples, with the effect of transforming the trace into a time-event series; or the segmentation of the Doppler signal prior to heart rate estimation to include only two cardiac cycles at a time in the correlation window. Those articles including the methods used and their limitations are summarised in Table 2.2.

Study	Modality	Dataset	Methods	Results	Summary and limitations
Peters et al. [70]	HP 8040A CTG Pulsed transducer Rep. freq: NR Em. freq: 1.024 MHz	Signals: Doppler, fECG N patients: 1 Duration: NR μGA: 40wks, intrapartum Pathologies: None	Preprocessing: 1. Envelope (monitor output) bandpass filtered (100-475 Hz), sampling frequency 1kHz. Beat-to-beat FHR estimation: 1. Low-pass filter (2 Hz) transforming envelope into 'blunt bumps' indicating cycle locations. 2. AC in the frequency domain. Analysis windows defined by estimated beat locations and an empirically chosen window extension (Window start: 125 ms after preceding bump, window end: 125 ms after successive bump). 3. Maximum in ACF (minimum duration 300 ms = 200 bpm) determines HR. HR correction: 1. If outside of range (50-210 bpm) or differing significantly from previous/successive beat HR interpolated from correctly identified beats.	FHR as estimated from the Doppler was compared to RR intervals computed from the direct fECG. For a 100 sec example the following was reported: Time domain: 1. Corr coeff HR: 0.977 ($p<0.001$) 2. Bland Altman plot showing no systematical difference in method. SD of data plotted: 1.18 bpm. Frequency domain: • Corr coeff power spectrum: 0.991 ($p<0.001$)	This approach segments the waveform to identify beat locations and employs ACF subsequently on identified 2-beat windows. Results appear to match closely estimates computed from a reference fECG. Limitations: • Evaluated on one patient only; • Results reported on one 100 sec segment only; • HR validation ranges not reported.
Jezewski et al. [78]	HP M1350 CTG Pulsed transducer Rep. freq: 3.2 kHz Em. freq: 998.4 kHz Sonicaid CTG Pulsed transducer Rep. freq: NR Em. freq: NR	Signals: Doppler N patients: 15 Duration: 35 5 min segments selected from two monitors with no signal loss and a maximum of three per patient. Monitor 1: 375 min total Monitor 2: 483 min total μGA: 37.5wks, antepartum Pathologies: NR	Correction of invalid samples (samples referring to the same cycle with different values): 1. Identification of invalid samples by search for samples of different values and comparison of actual time and summed interval duration. 2. If durations do not match, the sample with minimum difference to next is considered invalid. Removal of duplicate FHR samples (cycles are measured by several estimates): 1. Correction based on the summed time of reported interval samples compared to the actual time.	Algorithms were evaluated on both commercial monitors based on the sum of the sampled interval duration after correction and the actual time (5 min = 300 ms). • Mean duration and SD monitor 1: 300.47 +/- 1.69s • Mean duration and SD monitor 2: 301.47 +/- 2.27s STV indices compared between monitor output and as computed from corrected FHR. • STV increased significantly with corrected FHR leading to the conclusion that traces by commercial monitors are not reliable.	This approach turns the evenly sampled FHR trace into a time-event series (post-processing of the trace). Elimination of duplicate samples approximates the true FHR time-event series more closely leading to a higher accuracy of variability indices. Limitations: • Intervals estimates remain values averaged over an analysis window; • The analysis window has not been defined; • Selection of segments based on quality has not been described.

Roj et al. [80]	Doppler (own device) Pulsed transducer Rep. freq: 3 kHz Em. freq: 1 MHz	Signals: Doppler, fECG N patients: 3 Duration: 82 min, 68 min selected with satisfactory quality. (Total detected beats in ECG: record 1: 6015, record 2: 1188, record 3: 1742, total: 8945) μGA: NR, intrapartum Pathologies: NR	Preprocessing: 1. Bandpassfilter Doppler (50-500 Hz). 2. Computation of envelope (Hilbert transform). 3. Computation of ACF. 4. Estimation of cardiac cycle duration (T_i) by dominant peak selection in ACF. Beat-to-beat FHR estimation: 1. Rough HR estimate by AC (window 3 sec). 2. Envelope downsampled (200 Hz). 3. Estimate used to adjust window to include 2 beats. 4. Window shift of 84 ms (12 Hz). 5. Correction of ACF by Gaussian window with centre $T_i - 1$. 6. Interval validation based on previous estimate. 7. Segmentation by low-pass filtering (2Hz) [70], each maximum detected as one beat. 8. Median of each segment is the final interval value.	FHR and HRV measures compared to synchronous fECG recording (RR-intervals) • Mean absolute HR error: 1.90 ms +/- 3.06 ms • Mean relative error STI (De Haan): -7.61% +/- 20.69	Approach processes the signal into an event series, where each interval is described by its duration and a (rough) marker of its occurence in time. This is achieved by a two-step process of signal cycle segmentation followed by AC. Limitations: • Evaluated on three recordings (assumed patients) only; • No information on signal quality selection provided.
Jezewski et al. [79]	Doppler (own device) Pulsed transducer Rep. freq: 3 kHz Em. freq : 1 MHz Depth range: 3 to 15 cm (three signals are obtained)	Signals: Doppler, fECG N patients: 3 Duration: 82 min, 68 min selected with satisfactory quality. (Total detected beats in ECG: 8945) μGA: NR, intrapartum Pathologies: NR	Preprocessing: 1. Selection of the best quality signal from three different depths recorded. 2. Bandpassfiltered (100-600 Hz). 3. Computation of envelope (Hilbert transform), envelope low-pass filtered (50 Hz). FHR estimation: 1. Dynamic adjustment of AC window based on previous cycle duration and scaling factor (window: 2*T_i). Shift increment adjusted based on previous cycle duration and scaling factor (1/5*T_i, cycle measured 5 times). 2. AC correction with triangular window function centred on $T_i - 1$, if the signal quality (empirically set to amplitude = 0.6 in ACF) is low. 3. Segmentation based on minimum dispersion rate. Each interval estimated as the median of all measurements within segment. 4. Interval validation.	The algorithm was evaluated on direct fECG and compared to one other FHR estimation method (Peters et al.). Comparison with FECG: • Mean absolute interval error equal: 1.91ms (SD 3.48ms) • Mean relative error of short-term variability did not exceed -7% for both methods (presented method and [70]). • Interval corr coeff: r= 0.992, equal for Peters et al. and presented method.	Approach samples HR several times and converts trace into time-event series as a postprocessing step. Good results were reported for both heart rate estimation and short-term variability estimation compared to fetal ECG. Limitations: • Parameters developed on three patients only; • Evaluated on same three patients; • No information on signal quality selection provided; • Bespoke monitor allowing signal selection from several depths for better quality, this is not possible with continuous Doppler mode.

Table 2.2: A summary of published methods of Doppler-based beat-to-beat heart rate estimation. (Abbreviations: R-freq: repetition frequency, E-freq: emission frequency, NR: not reported, GA: gestational age, AC/ACF: autocorrelation (function))

Beat-to-beat heart rate detection based on two beats per analysis window [70] is sensitive to noise. The superiority in accuracy of fetal heart rate averaged over larger window sizes has been noted in the literature before [48], illustrating that there is a potential trade-off between susceptibility to noise and beat-to-beat accuracy when employing correlation based methods. The approach taken by Jezewski et al. [79] combined multiple cycle measurements to improve noise immunity with a segmentation step translating the trace into a beat-to-beat interval series.

2.3.1.3 Databases of fetal heart rate traces

As previously mentioned, commercial monitors used in clinical practice have been designed as closed systems rendering raw fetal Doppler data inaccessible. Furthermore, very few databases of Doppler fetal heart rate data have been made publicly available to date, and few open source algorithms to provide benchmarking.

One notable exception is the CTU-UHB Intrapartum Cardiotocography Database[2] [81]. This database contains 552 CTG traces, which were 'carefully' selected from 9164 recordings collected between 2010 and 2012 at the University Hospital in Brno, Czech Republic. The CTG recordings start no more than 90 minutes before actual delivery, and each is at most 90 minutes long. Each recording contains a fetal heart rate (FHR) time series and a uterine contraction (UC) signal, both sampled at 4 Hz. Each CTG is also accompanied by maternal, delivery, and fetal clinical details. As such, it is an interesting database. However, as data was recorded intrapartum only, it is of little use in the context of this thesis.

No open source CTG algorithms are currently available, although there is some partially complete code with an unspecified license, designed to reproduce the cCTG analysis (based on the Dawes/Redman criteria, discussed further in Section 3.4.4). It is available as appendices in Nyboe [82]. The field of ultrasound-based fetal heart rate assessment

[2]Available from https://physionet.org/physiobank/database/ctu-uhb-ctgdb/

has been limited by the lack of public databases and open source algorithms for many decades.

2.4 Markers for the detection of IUGR

The literature review presented here focuses on work based on Doppler ultrasound as a monitoring modality, and all studies reviewed assessed signals derived from the antenatal fetal cardiac signal recorded by cardiotocography (CTG). This being said, studies based on magnetocardiography (fMCG) and electrocardiography (fECG) have also been included, if deemed relevant. Work analysing the blood flow velocity waveform obtained by arterial and venous Doppler ultrasound were considered outside of the scope of this investigation and are not discussed in this thesis. The indicators based on the fetal cardiac signal discussed in literature can be classified into markers derived in the time domain, frequency domain and measures of the signal's statistical distribution.

In the time domain short-term variability (STV) of the fetal heart rate (FHR) has been used as an indicator of pregnancy outcome. Several authors have assessed the utility of STV as predictor of IUGR [83–90] and as an indicative tool for the timing of delivery [91–94]. Authors investigating IUGR detection found STV to be significantly lower in growth-restricted fetuses compared to normal controls. However, the mean STV values (measured in milliseconds) for IUGR cases computed in studies comparing growth-restricted with normal fetuses [85–87,90] differed substantially, ranging from 4.29±1.6 ms [90] to 8.9±2.7 ms [85]. In the study powered by the largest number of subjects (257 IUGR cases) [93], investigating STV for the timing of delivery, the authors found an STV value $<$3.0 ms to be associated with a higher rate of metabolic acidaemia and early neonatal deaths. STV also predicted the presence or absence of metabolic acidaemia with 81.3% accuracy in a population of growth retarded fetuses.

Other linear time-domain markers that have been investigated are alternative variability measures such as long-term irregularity (LTI) [84,90,95], the interval index (II)[84,90,94,

95], SDNN (standard deviation from normal-to-normal beats) and RMSSD (root mean square of successive differences) [83]. Furthermore, delta (describing the range of the signal in a given time interval) [84,90,94,95], the mean fetal heart rate [84,94,95] and the baseline fetal heart rate [85] have been included in IUGR marker assessment. Ferrario et al. [84,94,95] reported significantly different values for IUGR fetuses in three studies for mean FHR, LTI, II and delta, although based on small sample sizes only (41, 41 (same subjects) and 6 respectively), Fanelli et al. [90] found LTI and delta to be discriminative, but the interval index (II) not to. Buscicchio and colleagues [85] concluded the FHR baseline to be markedly higher in IUGR fetuses (N=100) compared to controls (N=100). From fMCG records Schneider et al. [83] found the variability measures SDNN (standard deviation of normal-to-normal beats) and RMSSD (root mean square of successive differences) both significantly lower in growth-restricted fetuses (N=39) than controls (N=29).

In the frequency domain the total power (TP) and the power content of distinct frequency bands (LF = low frequency power, MF = movement frequency power, HF = high frequency power) have been analysed, as well as the ratios $\frac{LF}{HF}$ and $\frac{LF}{(\frac{MF}{HF})}$. Ferrario et al. [84,94,95] could not find significant differences for LF, MF, HF and the ratio $\frac{LF}{(\frac{MF}{HF})}$ based on CTG data. In contrast to this, working with fMCG signals, Anastasiadis et al. [96] found LF and HF to be discriminative, as well as Schneider et al. [83] who additionally found TP and $\frac{LF}{HF}$ to be significantly lower in IUGR compared to normal fetuses.

In several studies complexity measures were introduced and analysed as potential markers for IUGR. Ferrario et al. [97] reported Multiscale Entropy (MSE) as able to separate IUGR cases (N=23) from healthy cases (N=17). Two further studies on the same data set by the same authors [84,95] however found Sample Entropy (SampEn) and Approximate Entropy (ApEn) not to be contributing to the identification of growth-restriction, whilst a third article [94] reported successful identification of IUGR using the same entropy parameters. The latter included six IUGR subjects only, which strongly

limits the degree to which these results can be generalised. The results presented in the study by Fanelli and colleagues [90] further weaken the proposition of entropy based markers with ApEn found to be non-discriminative. Apart from entropy measures Ferrario et al. [84,95,97] also assessed the Lempel Ziv complexity parameter which was reported to separate IUGR (N=42 (severe and non-severe)) from healthy fetuses (N=17), and severe (N=23) from not severe (N=19) growth-restriction. Increasing with gestational age but not effective in detecting IUGR was the AIF (Autonomic Information Flow) decay marker introduced by [83]. The AIF is a complexity measure of the information transfer in the underlying physiological system (such as the autonomic nervous system) [98].
The use of the maximum Lyapunov exponent (MLE), an estimate of predictability of a dynamic system, was reported with positive results (IUGR fetuses had significantly lower values than normal subjects) by [99]. In the same study, attractor reconstruction and the correlation dimension were assessed, finding different attractor structures and a significant difference in correlation dimension in 31 to 35 weeks between the normal group and the IUGR group. The authors also reported detrend fluctuation analysis (DFA) to discriminate IUGR from normal controls [100]. Ferrario et al. [94] reported similar results for DFA with significantly higher values of long-term scaling exponents in IUGR fetuses. A significantly lower correlation dimension in IUGR compared to normal fetuses was also described by Anastasiadis et al. [96], based on fMCG signals.

A more recent development is the implementation of a signal processing technique called phase-rectified signal averaging (PRSA) to measure heart rate variability, whereby the resultant PRSA curve characterises the cardiac acceleration and deceleration capacity. Lobmaier et al. [87] investigated the averaged acceleration capacity (AAC) as discriminator, and Huhn et al., Graatsma et. al and Fanelli and colleagues [86,88,90] assessed both AAC and the average deceleration capacity (ADC), whilst the latter [90] also computed the acceleration and deceleration phase-rectified slope (APRS and DPRS) compared to both AAC and ADC. All four studies demonstrated the successful use of at least one

PRSA based marker in distinguishing IUGR cases from normal controls, in all cases performing slightly superior to STV.

There was however disagreement as to which of the three markers analysed, AAC, ADC, APRS, were suitable and best performing. Whilst Huhn et al. [86] found AAC (AUC 81.40%) to be superior to ADC (results not reported), this could not be confirmed by Graatsma et al. [88] with results based on fetal ECG data (mean z-scores for STV, AAC, and ADC were lower in growth restricted fetuses by 1.0 SD, 1.5 SD, and 1.7 SD, respectively ($p < 0.0001$)). In the most recent study [90] only APRS ($p = 1.12e-9$) and DPRS ($p = 9.57e-12$) succeeded whereas both AAC ($p = 0.2$) and ADC ($p = 0.06$) failed to distinguish IUGR from healthy fetuses.

Hoyer et al. [89] suggested the use of a newly developed fetal autonomic brain age score (fABAS) to detect growth-retarded fetuses. The score is based on a multivariate model including the following parameters: amplitude, skewness, gMSE(3) (generalised multiscale entropy at scale 3), pNN5[3] and $\frac{VLF}{LF}$ (ratio between very low (0.02-0.08 Hz) and low (0.08-0.2 Hz) frequency band power). The 11 IUGR fetuses showed a reduced fABAS computed from fMCG traces, where the fetuses had been selected to be in an active sleep state[4] by three independent clinicians. The reported inclusion criteria make the score difficult to implement as prior fetal state assessment is required. Furthermore, the temporal FHR resolution obtained by fMCG currently cannot be matched by CTG monitors.

A small number of articles has analysed the change of a marker value with gestational age. Findings in [91] suggest that terminal fetal heart rate variation does not change with gestational age. Schneider et al. [83] reported an increase in complexity and linear fetal

[3] A time domain heart rate variability measure calculated as the percentage of differences between adjacent NN intervals that are greater than 5ms [89].

[4] A fetal behavioural state defined by the authors as: fluctuating FHR with an oscillation bandwidth exceeding +/25 bpm from floating baseline, showing at least 3 accelerations (.15 bpm, .15 sec) per 10 min, and the FHR exceeding 160 bpm only during accelerations [89].

heart rate variability measures correlated with gestational age; the authors also reported a higher $\frac{LF}{HF}$ ratio in the study group that significantly decreased with gestational age. A dependency of the correlation dimension on gestational age was described by [100]; the authors also reported the Lyapunov exponent to be independent of gestational age. The influence of gender on the risk markers has been investigated by one group [101], however the small sample size (IUGR N=15, control N=18) makes assumptions and conclusions beyond the subjects studied difficult.

2.5 Summary

This chapter has set the context for the assessment of fetal wellbeing based on the fetal cardiac signal by providing background information on the fetal cardiac cycle and cardiac control mechanisms. The review of the literature on heart rate estimation gave insight into the current standard correlation method and an overview of published adaptations developed to approximate true beat-to-beat computation. Common characteristics of studies on beat-to-beat estimation presented in literature are the small sample sizes and a lack of implementation detail making replication difficult.

In addition, the literature for IUGR assessment from the fetal heart rate trace was reviewed. Although there is room to be optimistic, small sample sizes and the inevitable resultant conflicting results and identification of significant variables provide an unclear picture. In the following chapter, an extensive database of fetal heart rate traces is analysed to address these issues.

Chapter 3

Evaluation of Doppler based fetal heart rate variability measures for IUGR detection

3.1 Introduction

Intrauterine growth restriction (IUGR) is a pathological decrease in fetal growth rate whereby the fetus cannot reach its full genetic growth potential; it is associated with significantly increased mortality and morbidity [102,103]. Developing countries carry the highest burden of perinatal mortality (98%) [104], of which a significant proportion can be attributed to IUGR-based complications (IUGR prevalence is estimated at over 10%)[8]. Key reasons for this include lack of systematic screening and robust referral. In the context of a well-equipped hospital several measures are typically available to the clinician when screening for, diagnosing and managing IUGR. In addition to maternal history, available tests include symphysis fundal height measurement (SFH), ultrasound biometry, the biophysical profile score and umbilical Doppler studies. The absence of sophisticated equipment in low-resource settings means that IUGR detection is limited to only the identification of maternal risk factors and the measurement of fundal height over

time. The latter has extremely poor sensitivity and specificity, and high inter-observer and intra-observer variability [105]. Given the scarcity of information available to health-care workers, additional insight into the developmental status of the fetus would provide valuable decision support. The fetal cardiac signal is the most accessible source of physiological information on the well-being of the fetus. A low-cost method to record this signal is by hand-held Doppler. This chapter therefore investigates whether analysis of the Doppler-based fetal heart rate signal can contribute to the detection of IUGR risk. This work has been carried out to lay the basis for the development of low-cost automated IUGR screening tools for use in resource-constrained environments.

Growth restriction remains an area of active research due to the complexity surrounding accurate identification and adequate management. Several researchers have investigated either the detection of growth restriction [83–90] or its management [91–94] by assessing its impact on the fetal cardiovascular system. The most extensively researched measures of fetal well-being, and the most commonly used in clinical practice, are heart rate variability measures. Studies in the literature assessing the difference in marker values between IUGR and healthy fetuses have, however, been largely underpowered, rendering any generalisation of results beyond the subjects studied unfeasible. This has motivated an investigation into the utility of variability risk markers for antenatal IUGR detection based on a large cardiotocogram (CTG) dataset of IUGR fetuses and healthy controls, matched for gestational age and gender. The studied dataset is part of the extensive cardiotocography database of the John Radcliffe hospital, Oxford UK, and the largest in subject numbers reported on to date.

The main body of work in this chapter is organised into two parts, a preliminary analysis of variability metrics of the entire fetal heart rate trace (Section 3.3: Detection based on trace variability), followed by an in-depth investigation (Section 3.4: Detection based on sleep state dependent variability) into recomputed variability measures for select trace segments and related trace characteristics as a response to initial observations.

Three risk markers have been included in the preliminary investigation: short-term vari-

ability (STV), long-term variability (LTV) and the acceleration capacity of the fetal heart rate (AAC). STV has been most extensively tested and used as a parameter for the assessment of fetal wellbeing and outcome. Given the positive results documented in the literature [85–87,90], it therefore seems a logical feature to include in the investigation of an IUGR prediction model. The inclusion of a measure of slow fluctuations in the fetal heart rate is proposed in the form of LTV as comparable markers, such as long term irregularity and the range of the signal in a given time interval (delta), have shown promise [84,90,94,95]. A more recent development is the implementation of the signal processing technique called phase-rectified signal averaging (PRSA) to measure heart rate variability, whereby the resultant PRSA curve characterises the cardiac acceleration and deceleration capacity [106]. AAC has been reported in recent studies [86,87] to have superior performance compared to STV in the discrimination of growth retarded fetus and normal controls; it has therefore also been included in the investigation.
The subsequent analysis explores the utility of STV and LTV for IUGR discrimination when taking into account the impact of the fetal behavioural state on the heart rate variability. Metrics characterising the state distribution within a trace have also been studied.

3.2 Description of the cCTG dataset

To confirm the utility of selected markers for the antepartum detection of IUGR fetuses, an extensive computerised cardiotocography (cCTG) archive of fetal heart rate variables was utilised, collected at the John Radcliffe hospital in Oxford, UK, previously introduced in Section 1.3. In this retrospective study all fetuses with a birth weight under the third percentile (adjusted Yudkin birth weight centile)[107] were considered growth-restricted and included in the dataset for analysis. The control group of normal fetuses was matched for both gender and gestational age. All cCTG traces included were the last recording of each patient prior to delivery. The resultant study dataset '*JR cCTG DB*' of 1163

IUGR cases and 1163 controls is the largest of its kind by more than a factor of 10 (its attributes are summarised in Table 3.1).

	IUGR	Control
Number of patients	1163	1163
Gestational age at CTG (weeks)	34.84 [23-42]	34.84 [23-42]
CTG recording length (minutes)	34.46 [10-60]	23.85 [10-60]
Time to delivery (days)	6.34 [0.1-86]	23.98 [0.1-126.6]

Table 3.1: Subject characteristics of the study dataset. Data are presented as quantity or the mean with the range provided in brackets.

3.3 Detection based on trace variability

3.3.1 cCTG metrics for IUGR discrimination

Short-term variability (STV) and long-term variability (LTV) values for each trace were readily available in the database, computed by the cCTG system according to Dawes and Redman [108,109]. In the Dawes/Redman system[1] STV is calculated as the epoch-to-epoch difference of the averaged pulse interval. Each minute is split into epochs of 3.75 seconds and the average fetal heart rate is calculated as a beat interval, RR_m, in milliseconds. The change in the mean values from one epoch to the next is computed and averaged over one minute to obtain:

$$STV_{minute} = \frac{1}{16} \sum_{m=1}^{16} |RR_{m+1} - RR_m| \tag{3.1}$$

[1]The Dawes/Redman system provides numeric fetal heart rate analysis of the CTG trace (nonstress test) based on criteria developed by Professors Dawes and Redman at the Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital, in Oxford. The normality of a CTG trace is evaluated against the Dawes/Redman criteria repeatedly during a reading until criteria are met and the operator is informed that it is safe to stop monitoring. The test criteria include thresholds for short- and long-term variability and the presence of high variability episodes, further discussed in Section 3.4.2. The system is described in Dawes et al. [110] and Pardey et al. [108], including a complete list of criteria. The analysis is implemented in Sonicaid Fetal Care monitors by Huntleigh Healthcare Ltd. and used routinely in antenatal care.

The STV values per minute are then averaged over the entire length of the recording. LTV is measured as the average minute range of the recording, whereby the range of each minute is the difference between the highest and the lowest fetal heart rate relative to the baseline (see Figure 3.1).

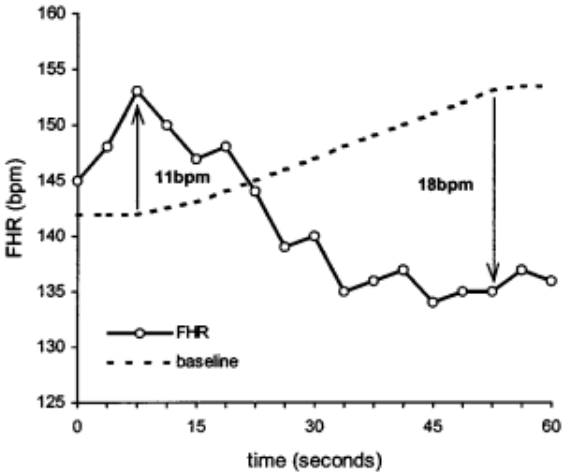


Figure 3.1: Estimation of the minute range from the fetal heart rate trace. Adapted from [108].

AAC was computed from the PRSA signal according to Huhn et al. [86]:

$$AAC(S) = \frac{1}{2S}\left[\sum_{i=0}^{S-1}\overline{y(i)} - \sum_{i=-S}^{-1}\overline{y(i)}\right] \tag{3.2}$$

where S is the time scale and i the index of the PRSA signal y. Two parameters, T and L, need to be specified to obtain the PRSA signal. T selects the number of samples averaged prior to anchor detection. $2L$ defines the length of the windows (centred on each anchor point) which are aligned and averaged to derive the PRSA series. Both T and L were set to 1. The time scale S was chosen equal to T. The cCTG metrics and patient information used as feature inputs for IUGR discrimination are summarised in Table 3.2.

Feature [unit]	Definition
LTV [ms]	The average long−term variability
STV [ms]	The average short−term variability
AAC [bpm]	The average acceleration capacity
RecGest [weeks]	The gestational age as estimated at time of recording
Gender	The gender of the fetus

Table 3.2: Summary of the selected features for IUGR classification, including the three metrics reflecting trace variability and available patient information.

3.3.2 Statistical analysis methods

As the distributions of the selected cCTG features, LTV, STV and AAC, are non-Gaussian their correlation was assessed using Spearman's rho. Whether feature values are statistically significantly different in the growth retarded and normal population or not was investigated with the non-parametric Mann-Whitney U-test and visualised using box plots. A p-value of <0.05 was selected as the level of significance.

The data were randomly divided into a training and test set (70/30 split). Binary logistic regression classifiers were trained for all three individual single risk markers over a 5-fold cross-validation, as well as for every possible combination including the additional features 'gestational age at recording' (GA) and gender (Logistic regression is further defined in the following Section 3.3.3). The associated average area under the receiver operator curve (AUROC or simply AUC) was computed for comparison of the predictive power of the risk markers and their combinations. The AUC statistic of the overall best performing feature combination was compared for statistical significance with the AUC of the remaining classifiers achieving the best result for each number of features. Significance values were adjusted with the Holm-Bonferroni method to avoid the problem of multiple testing (Type 1 errors) [111]. The best performing univariate and multivariate classifiers were tested on a held-out test set and the AUC reported for comparison with the training results.

Optimal threshold values were estimated for the three cCTG markers using Youden's J statistic [112] defined as

$$J = max_c(Se(c) + Sp(c) - 1) \tag{3.3}$$

where Se and Sp are the sensitivity and specificity respectively given a threshold, and the value c, which maximises the expression, is the optimal cut-off.
The cut-off thresholds estimated on the training data were tested on the test set and the resulting sensitivity and specificity reported.

3.3.3 Logistic regression

Prediction in logistic regression is based on a regression line fit to the training data points. In a binomial case LR predicts the probability that an observation belongs to one of two classes, which is defined as follows:

$$\pi(x) = E(Y|x) \tag{3.4}$$

where $x = (x_1, x_2, ..., x_p)$ represents an observation and $Y \in \{1, 0\}$, dependent on the observation's class affiliation. The logistic regression for Y can be denoted as [113]:

$$logit(Y) = ln(\frac{\pi}{1-\pi}) = \alpha + \beta_1 x_1 + ... + \beta_p x_p \tag{3.5}$$

and the probability that an observation belongs to one of the two classes as:

$$\pi(x) = \frac{\exp(\alpha + \beta_1 x_1 + ... + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + ... + \beta_p x_p)} \tag{3.6}$$

where α is the Y intercept, β are the regression coefficients and exp indicates exponentiation using the natural base. The regression coefficients are commonly estimated using the maximum likelihood method [114].

3.3.4 Results

The CTG traces analysed (last trace of each included fetus) were recorded between the 23rd and 42nd week of gestation. By visualising the distribution of cCTG marker values for IUGR and healthy fetuses according to gestational age, it became apparent that the values of the two populations lie further apart earlier in gestation; this is true for both male and female fetuses (as an example, the STV value distribution per GA is shown in Figures 3.2 and 3.3). Risk markers computed from the last traces of compromised fetuses were lower earlier on in gestation and increased with gestational age. The rate of change in markers of the normal population with increasing gestational age was markedly lower than in the IUGR group (regression slopes b: $b_{IUGR} = 0.113$, $b_{CONTROL} = 0.0493$). Observation hence suggests that early-onset IUGR (<34 weeks of gestation) and late-onset IUGR (>34 weeks of gestation) do not impact the fetal cardiovascular system in the same way. The study dataset was therefore divided into two subsets to confirm the initial assumption that IUGR fetuses can be better distinguished from healthy fetuses earlier in gestation. The first subset includes fetuses with their last trace recorded at or before 34 weeks, the second subset includes all fetuses with their last trace recorded after 34 weeks of gestation. The characteristics of the two populations are summarised in Tables 3.3 and 3.4. The risk markers were found to be significantly lower in the IUGR group compared to the control group, both before and after 34 weeks of gestation. Boxplots of all three cCTG risk marker distributions are shown in Figure 3.4. The investigated cCTG risk markers were found to be highly correlated (Table 3.5).

In the subset of fetuses recorded before 35 weeks of gestation the best performing individual cCTG risk marker tested on the held-out data (test set) was found to be LTV with an AUC of 74%, followed by STV and AAC. The AUC of LTV was significantly higher than both the estimates for STV and AAC. For fetuses with the last trace recorded after 34 weeks LTV was equally best performing; the estimated AUC for LTV however dropped substantially to 57% and was not significantly higher compared to STV and AAC (Table 3.6).

	IUGR	Control
Number of patients	463	463
Gestational age at CTG (weeks)	31.17 [23-34]	31.17 [23-34]
CTG recording length (minutes)	44.44 [10-60]	25.97 [10-60]
Time to delivery (days)	9.17 [0.1-86]	45.35 [0.1-126.6]

Table 3.3: Patient characteristics of the study subgroup of 23 to 34 weeks gestation. Data are presented as quantity or the mean with the range provided in brackets.

For the population between 23 and 34 weeks of gestation, the three-feature combination of LTV, AAC and gestational age gave the highest AUC of 76% on the training set and 74% when the logistic regression model was tested on the held-out test set. For traces recorded after 34 weeks of gestation the combination of LTV, STV and AAC achieved the best result. The classifier returned an AUC of 60% on the training and 59% on the test set (Table 3.7). However, for neither of the two populations was the best performing multivariate classifier significantly superior to the best combination for each number of features, including single features.

The optimal thresholds estimated for LTV, STV and AAC were 35.10 ms, 6.82 ms and 3.35 bpm respectively for the dataset including gestational ages 23 to 34 weeks (Figure 3.5a). Cut-off values for 35 to 42 weeks gestation were higher at 47.80 ms, 9.51 ms and 3.73 bpm respectively (Figure 3.5b). When applying the thresholds to the held-out test data, the sensitivity was calculated to be 63% for LTV, 63% for STV and 67% for AAC in a population of 23 to 34 weeks gestation. The specificity values were 81% for LTV, 78% for STV and 70% for AAC. The sensitivity values for the data subset of 35 to 42 weeks gestation were 55% for LTV, 63% for STV and 52% for AAC; specificity was calculated to be 70% for LTV, 60% for STV and 74% for AAC.

3.3.5 Discussion

The preliminary analysis demonstrated some discrimination between growth restricted and normal fetuses, which was most effective for traces recorded before 34 weeks of gestation. Of the markers evaluated, long-term variability of the fetal heart rate trace

	IUGR	Control
Number of patients	700	700
Gestational age at CTG (weeks)	37.27 [35-42]	37.27 [35-42]
CTG recording length (minutes)	27.86 [10-60]	22.45 [10-60]
Time to delivery (days)	4.47 [0.1-41.3]	9.84 [0.1-50.1]

Table 3.4: Patient characteristics of the study subgroup of 35 to 42 weeks gestation. Data are presented as quantity or the mean with the range provided in brackets.

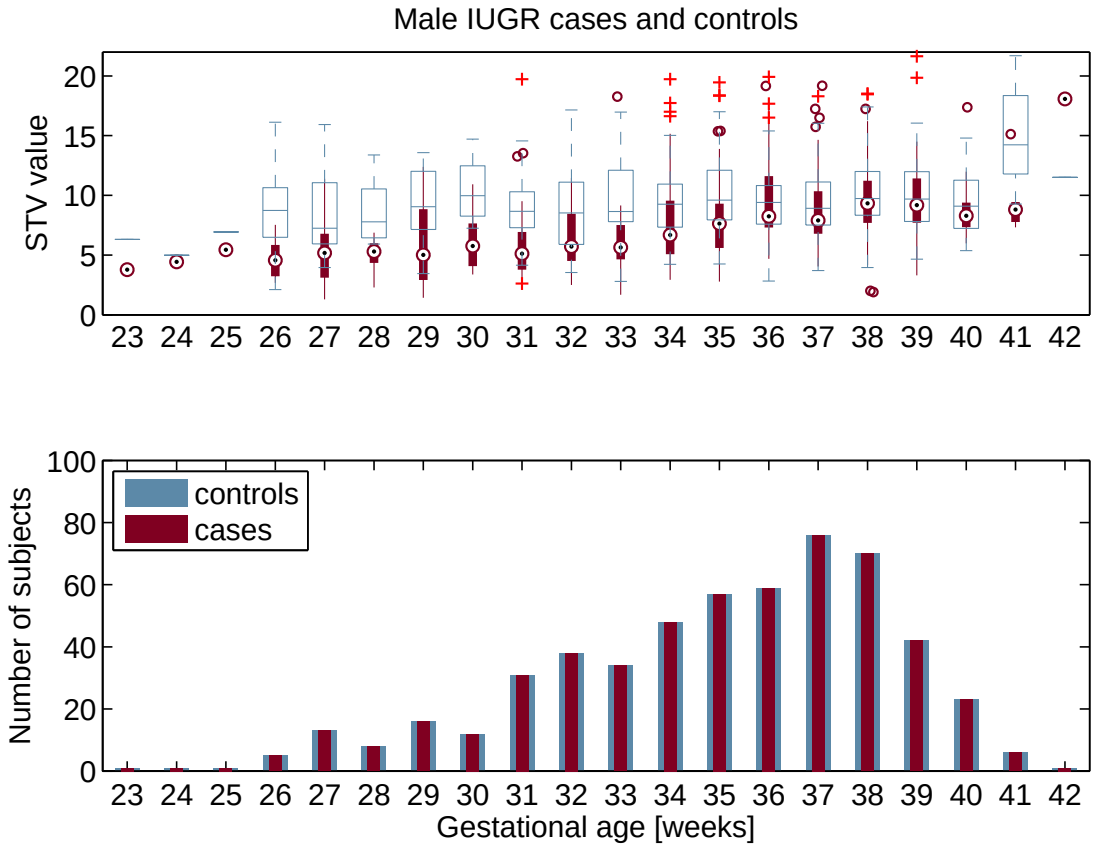


Figure 3.2: Top graph: Visual assessment of the difference of STV values in the male IUGR population against normal male controls per gestational age group. Bottom graph: The number of male cases and controls per gestational age group. Of all age groups with N$\geq$15 for both cases and controls the two study populations differ significantly at 29, 31, 32, 33, 34, 35 and 37 weeks ($p<0.05$) (Significance levels based on the Mann-Whitney U test).

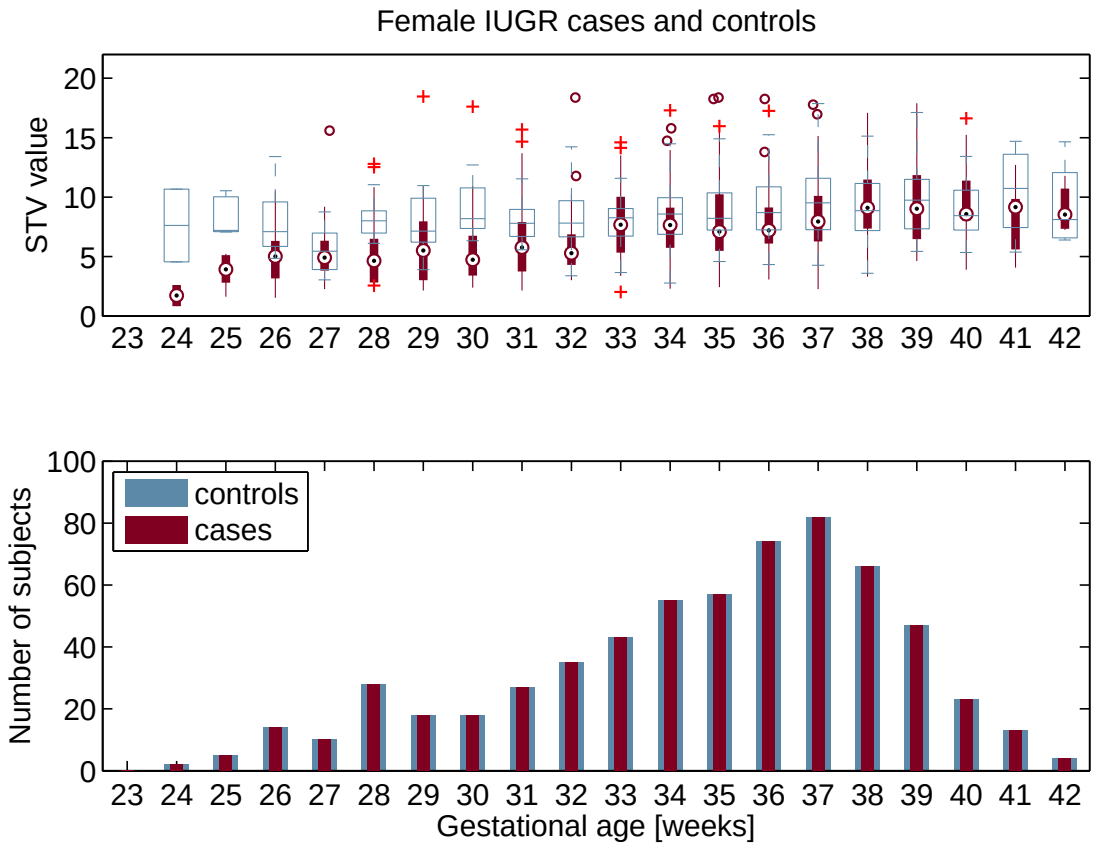
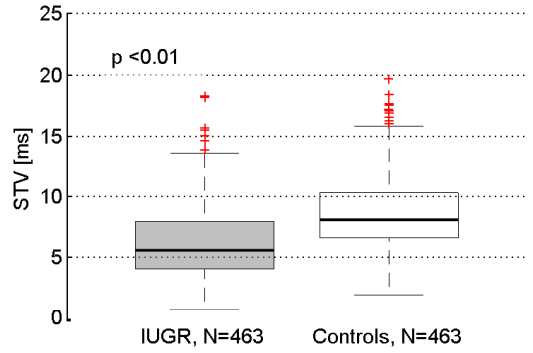


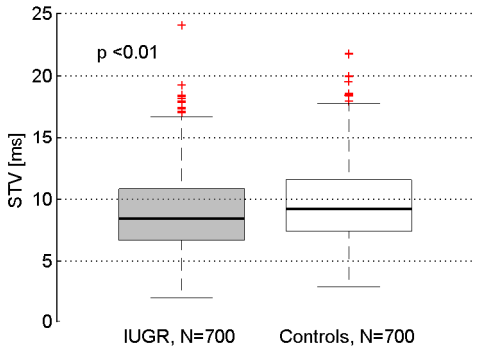
Figure 3.3: Top graph: Visual assessment of the difference of STV values in the female IUGR population against normal female controls per gestational age group. Bottom graph: The number of female cases and controls per gestational age group. Of all age groups with N≥15 for both cases and controls the two study populations differ significantly at 28, 30, 31, 32, 36 and 37 weeks ($p<0.05$) (Significance levels based on the Mann-Whitney U test).

	LTV	STV	AAC
LTV	—	+0.942, p>0.001	+0.897, p>0.001
STV	+0.942, p>0.001	—	+0.953, p>0.001
AAC	+0.897, p>0.001	+0.953, p>0.001	—

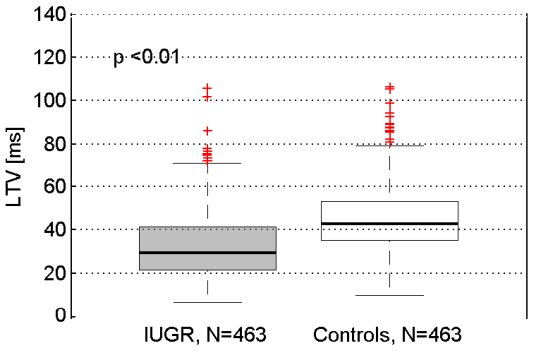
Table 3.5: The relationship between the cCTG variables as measured by Spearman's rho, assessed on the entire dataset.



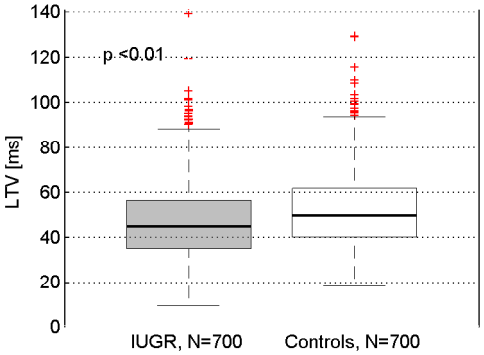
(a) STV 23–34 wGA. Median IUGR:5.70 ms IQR (4.19–8.05), Median Controls: 8.18 ms IQR (6.72–10.41)



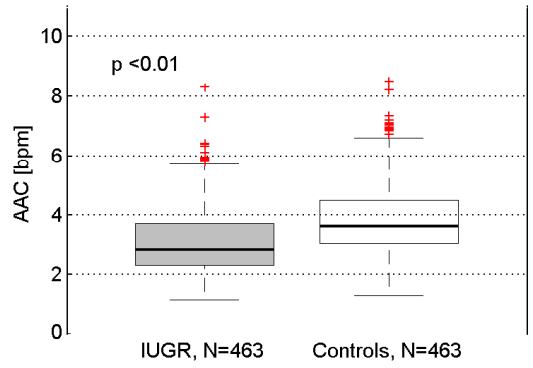
(b) STV 35–42 wGA. Median IUGR:8.35 ms IQR (6.60–17.75), Median Controls: 9.16 ms IQR (7.34–11.48)



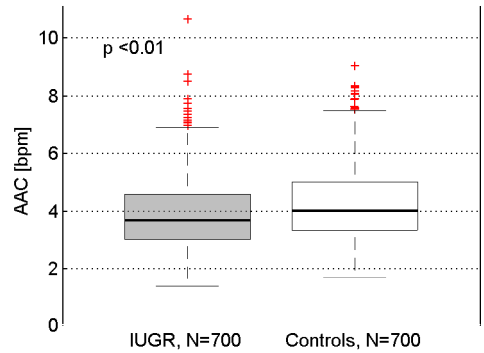
(c) LTV 23–34 wGA. Median IUGR: 30.2 ms IQR (22.20–42.05), Median Controls: 43.6 ms IQR (35.74–53.88)



(d) LTV 35–42 wGA. Median IUGR: 44.45 ms IQR (34.70–56.00), Median Controls: 49.40 ms IQR (39.75–61.30)



(e) AAC 23–34 wGA. Median IUGR: 2.85 bpm IQR (2.32–3.72), Median Controls: 3.64 bpm IQR (3.06–4.51)



(f) AAc 35–42 wGA. Median IUGR: 3.65bpm IQR (2.98–4.54), Median Controls: 3.97 bpm IQR (3.30–4.96)

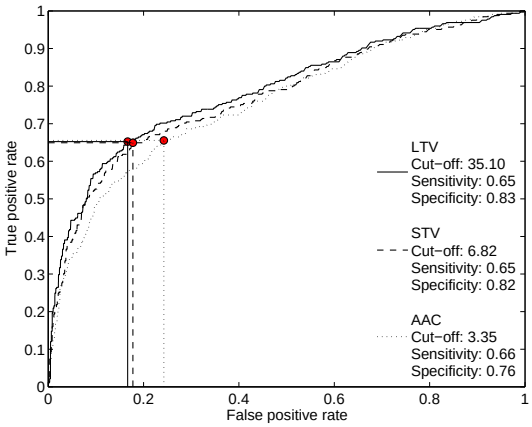
Figure 3.4: Boxplots showing the distribution (median and interquartile range) of the short-term variability (top), long-term variability (centre), and average acceleration capacity (bottom) values of the growth-restricted (shaded boxplot) compared to the normal control group. Boxplots in the left column include all fetuses recorded at gestational ages equal to or under 34 weeks, whilst the right column shows the dataset including fetuses recorded at gestational ages of and over 35 weeks. The level of difference between the populations (IUGR and healthy) was tested using the Mann-Whitney U test.

DB: 23-34 GA	Training data	Test data	
LTV - AUC[%]	75 (95% CI = 71-79)	74	p<0.01
STV - AUC[%]	73 (95% CI = 70-77)	72	
AAC - AUC[%]	71 (95% CI = 67-77)	70	
DB: 35-42 GA	Training data	Test data	
LTV - AUC[%]	60 (95% CI = 56-63)	57	
STV - AUC[%]	58 (95% CI = 55-62)	55	
AAC - AUC[%]	59 (95% CI = 56-63)	57	

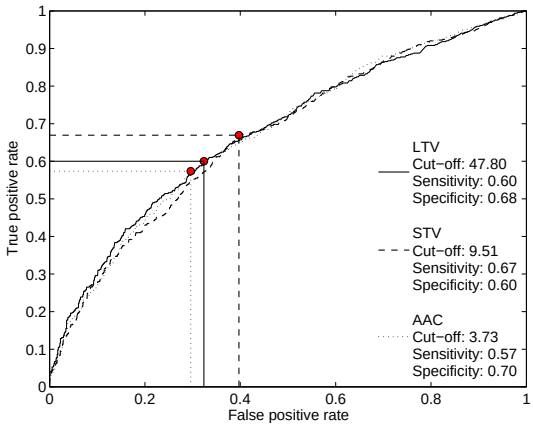
Table 3.6: The AUCs of each individual cCTG risk marker for populations including 23-34 weeks (top) and 35-42 weeks of gestation (bottom). Results are reported as achieved over a 5-fold cross-validation on the training data and as returned when tested on the held-out test data (70/30 split). The quality of the best performing marker was compared to the two remaining, a significant difference was found in the dataset including traces between 23 and 34 weeks gestation (DB: 23-34 GA), however not in DB: 35-42 GA.

DB: 23-34 GA	Training data	Test data
AUC [%] (Features: LTV, GA, AAC)	76 (95% CI = 72 80)	74
DB: 35-42 GA	Training data	Test data
AUC [%] (Features: LTV, STV, AAC)	60 (95% CI = 57 - 64)	59

Table 3.7: The best performing multivariate logistic regression classifier for 23-34 weeks (top) and 35-42 weeks of gestation (bottom). Results are reported as achieved over a 5-fold cross-validation on the training data and as returned when tested on the held-out test data (70/30 split). The quality of the classifiers was compared to the best-performing of each number of features, they were not found significantly superior in performance.



(a) The underlying database includes gestational ages from 23 to 34 weeks.



(b) The underlying database subset includes gestational ages from 35 to 42 weeks.

Figure 3.5: Receiver operator characteristic of LTV, STV and AAC with optimal cut-off values as derived from the Youden Index.

appeared significantly, yet marginally, better than short-term variability in early-onset IUGR, which has not been reported in the literature before. Both LTV and STV were superior to the average acceleration capacity of the fetal heart rate. Since the cCTG risk markers analysed were found to be highly correlated, combining each of the markers did not improve the quality of the classifiers significantly.

Recent investigations into the discrimination of IUGR and healthy fetuses based on heart rate variability markers were conducted by Huhn and colleagues [86] and Lobmaier and colleagues [87]. The age characteristics of the study subset 23 to 34 weeks of gestation resemble those of the populations reported on by Huhn et al. [86] and Lobmaier et al. [87]. Both computed STV according to Dawes and Redman, averaged over the entire recording length (trace length: 39 min IUGR, 42 min control; and 60 min both IUGR and control respectively). Comparing results the following could be observed: both the AUC for AAC (70%) and for STV (72%) were considerably lower than the results presented by Huhn et al.[86] with an AUC of 97% and 85% for AAC and STV respectively. Similarly the AUC for AAC was lower than the value of 76% found by Lobmaier et al.[87]; the AUC statistic for STV however was greater compared to the 70.50% reported. Both computed optimal thresholds for AAC and STV. The AAC cut-off estimated in this

investigation (3.35 bpm for the gestational age group of 23-34 weeks) is approximately 1 bpm above the values reported in literature (2.4 bpm and 2.23 bpm respectively). The STV thresholds reported by Huhn at al.[86] and Lobmaier et al.[87] differed greatly (8.9 ms and 5.66 ms respectively), both computing STV according to Dawes and Redman (following instructions provided by Pardey at al.[108]). The estimate resulting from the investigation presented here was found to lie in between at 6.82 ms.

The presented results could not confirm the superiority of AAC over STV as reported in the literature by Huhn et al.[86] and Lobmaier et al.[87]. It is important to note however that only epoch-to-epoch heart rate data were available for the computation of AAC. The poor temporal resolution (of 0.25Hz) may explain the inferior performance of the AAC risk marker in the investigation presented here. Disparities in outcomes may to some extent be explained by differences in the underlying study populations, influenced for example by the respective inclusion criteria. Both Huhn et al.[86] and Lobmaier et al.[87] defined IUGR as a birth weight $<$10th percentile for gestation (estimated from 2D ultrasound) combined with the umbilical artery resistance index $>$95th percentile. Huhn et al. restricted inclusion further by a required abdominal circumference $<$10th percentile and a cerebral artery index $<$10th percentile for gestational age.

There are a number of limitations to the investigation presented here. Fetal growth restriction was solely defined by the newborn's birth weight percentile ($<$ 3rd percentile). Antenatal measures to assess IUGR, such as Doppler velocimetry of the umbilical artery (pulsatility index) [65,115] were not accounted for, as they were not available. Similarly information on maternal demographics, health and medication were also not taken into account. Most importantly, variability metrics computed by the CTG were averaged over the duration of the recording, which can lead to a significant bias dependent on the trace composition of fetal sleep states. This motivated a more in-depth analysis taking into account the behavioural pattern of the fetal trace, described in the following Section 3.4.

3.4 Detection based on sleep state dependent variability measures

3.4.1 Impact of the fetal behavioural state on heart rate variability metrics

The fetal heart rate shows change corresponding to the respective fetal behavioural state. The two distinctive biological sleep states identified by research are active sleep (rapid-eye-movement (REM)) and quiet sleep (no rapid-eye-movement (non-REM)) [116]. Sleep state patterns evolve over the duration of pregnancy and become easier to identify as gestation progresses. A clear distinction of the sleep states is possible from 28 to 31 weeks of gestation onwards [116,117], however fetal rest-activity cycles have been observed in fetuses as early as 23 weeks [116]. The frequency of REM phases has been reported to increase with gestational age whilst ambiguous phases (neither REM nor non-REM) appear to decrease [117]. Periods of active sleep are associated with fetal movement and high heart rate variability. Quiet sleep in contrary is characterised by the absence of movement and low heart rate variability [118].

3.4.2 Bias in cCTG variability metrics

An episode of active sleep is an indicator of fetal wellbeing and is one of the key criteria used in the Dawes/Redman system to establish normality.

In order to assign fetal behavioural states researchers have established fetal heart rate patterns as criteria used when visually inspecting a trace post recording [89]. The Dawes/Redman system on the contrary screens a trace for active and quiet periods as part of its automated analysis. Its classification mechanism is based on the long-term variability of the fetal heart rate.

A low episode is identified if the LTV in at least 5 of 6 consecutive minutes is equal to or less than the established threshold (30 ms). The cCTG indicates an episode of high

activity provided that the fetal long-term heart rate variability in at least 5 of 6 consecutive minutes is equal to or exceeds the established threshold (32 ms). A high episode is confirmed by the system if its mean LTV is above the first centile of LTV values derived from the extensive JR CTG database. The long-term heart rate variability is defined as the minute range, the sum of the maximum heart rate excursion above and below the baseline within a one minute window as described in Section 3.3.1 and visualised in Figure 3.1. The Dawes/Redman system is described in detail by Pardey et al. [108]. This, amongst other indicators, is considered by the Dawes/Redman system to advise on normality. If criteria are met, the system can be switched off which results in recordings of dissimilar length, typically between 10 and a maximum of 60 minutes, where short traces contain a high percentage of active sleep and longer traces can be composed of, for example, 54 minutes of quiet sleep with low variability followed by only a short period of high variation (see Figure 3.6). Overall measures of short and long term variability as analysed above therefore contain a bias due to a potentially large sampling error dependent on the recording length. To investigate this issue the database was revisited to compute the long- and short-term variability for periods of active and quiet sleep, as well as associated metrics.

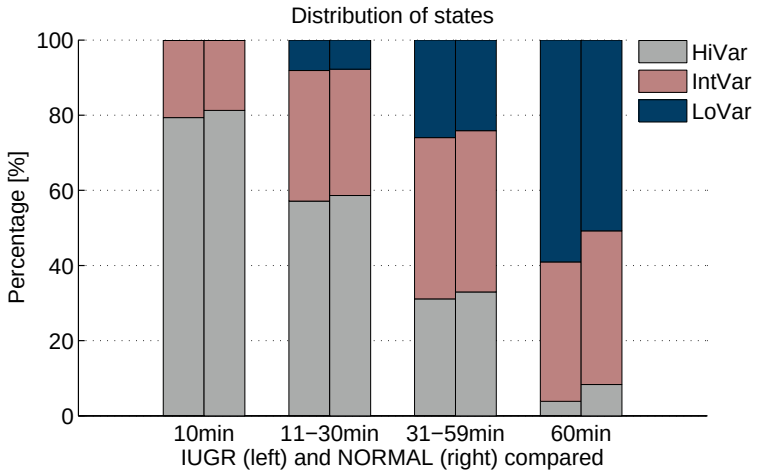


Figure 3.6: The average percentage of time in high (HiVar), low (LoVar) and undetermined (IntVar) heart rate variability recorded in short, medium-length, long and maximum-length traces.

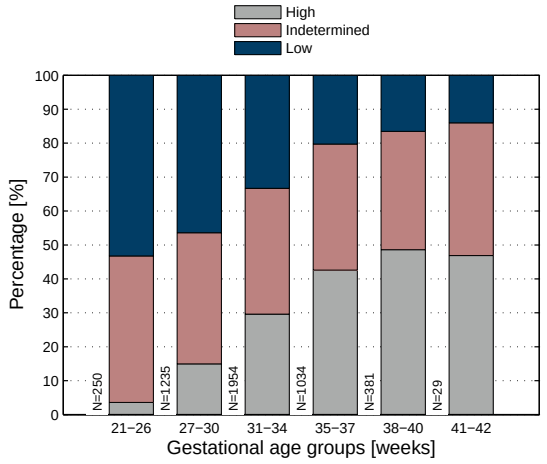
3.4.3 Behavioural state distribution in IUGR and normal traces

To observe the occurence of behavioural states in the database their average distribution per gestational age group was plotted (Figure 3.7). The quantitative comparison of percentages with those from available studies [116,119,120]) is not feasible due to the differences in state identification, recording protocol and monitoring modalities. However, the increase of active periods with maturation of the fetus, concomitant with its sleep organisation, as reported in literature, is clearly visible. This is true for both IUGR and normal traces. IUGR fetuses show a lower percentage of high variability (active sleep) compared to the normal population, in particular under 35 weeks of gestation. This may be explained by delayed or compromised sleep state organisation as observed by [121, 122]. Given these differences in the state distribution of quiet, active and indeterminate sleep between compromised fetuses and healthy controls, metrics characterising the trace composition, such as the number of state episodes and their average duration, have been included as features in the discrimination model. Those features are described further in the following section and are listed in Table 3.8.

3.4.4 Behavioural state dependent cCTG metrics for IUGR classification

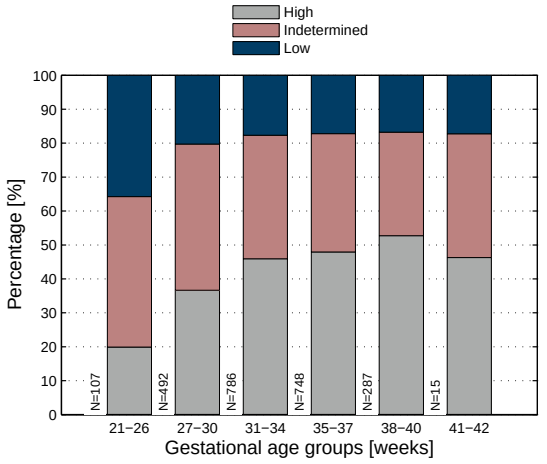
All cCTG metrics for identified episodes of high and low variability, associated trace characteristics, and patient information available for classification input are summarised in Table 3.8. The primary metrics of interest are long-term variability *LTVinHi* and *LTVinLo*, and short-term variability *STVinHi* and *STVinLo*. Both LTV and STV metrics were computed by the cCTG system according to Dawes and Redman [108,109], as described in Section 3.3.1. *STVinHi* and *STVinLo* are STV_{minute} (see Equation 3.1) averaged over episodes identified as of high or low variability. *LTVinHi* and *LTVinLo* are the minute range (see Figure 3.1) averaged over high and low variability episodes respectively. Furthermore the total number of high and low episodes in the trace (*No-*

HiEpi, NoLoEpi), their average duration (*AvgDurHi, AvgDurLo*), the number of minutes in high or low variability (*HiMin, LoMin*), and the onset of the first high variability episode (*HiEpiStrt*) were included.



(a) Average behavioural state distribution in IUGR traces.

	Mean trace duration (min)	SD (min)
21-26 wGA	53	15
27-30 wGA	49	17
31-34 wGA	39	20
35-37 wGA	30	17
38-40 wGA	28	15
41-42 wGA	40	17



(b) Average behavioural state distribution in normal traces.

	Mean trace duration (min)	SD (min)
21-26 wGA	41	20
27-30 wGA	33	18
31-34 wGA	28	16
35-37 wGA	27	15
38-40 wGA	26	14
41-42 wGA	32	19

Figure 3.7: The average percentage of minutes of low, indeterminate and high heart rate variation in IUGR (a) and normal (b) traces. Fetal traces from the entire database (8783 traces of 2326 patients) over 10 minutes in length were grouped into the gestational age ranges 21-26 wGA (extremely premature), 27-30 wGA (moderately premature), 31-34 and 35-37 wGA (mildly premature), 39-40 wGA (term) and 41-42 wGA (post-term) (Range of recording length: 11-60 min).

3.4.5 Statistical analysis methods

The same statistical methods were applied as outlined in Section 3.3.2. Correlation was assessed using Spearman's rho, statistical differences in features between growth retarded and normal fetuses were investigated with the non-parametric Mann-Whitney U-test and visualised using box plots (chosen level of significance: p-value <0.05).

Binary logistic regression classifiers were trained for all individual risk markers over a

Feature [unit]	Definition
LTVinHi [ms]	The long−term variability averaged over periods of high heart rate variation.
STVinHi [ms]	The short−term variability averaged over periods of high heart rate variation.
LTVinLo [ms]	The long−term variability averaged over periods of low heart rate variation.
STVinLo [ms]	The short−term variability averaged over periods of low heart rate variation.
NoHiEpi [ph]	The number of high variation episodes normalised by the length of the recording.
NoLoEpi [ph]	The number of low variation episodes normalised by the length of the recording.
HiMin [ph]	The number of minutes in high variation normalised by the length of the recording.
LoMin [ph]	The number of minutes in low variation normalised by the length of the recording.
AvgDurHi [min]	The average duration of episodes of high variation.
AvgDurLo [min]	The average duration of episodes of low variation.
HiEpiStrt [min]	The onset of the first high variation episode.
RecGest [weeks]	The gestational age as estimated at time of recording.
RecLen [min]	The recording length.
Gender	The gender of the fetus.

Table 3.8: Selected features from cardiotocograph metrics, trace characteristics and patient information for IUGR classification.

5-fold cross-validation, as well as for every possible combination of features. The dataset was split 70/30 into training and test sets, randomly assigned and maintaining a balanced event rate. The AUCs were computed for comparison of the predictive power of features and their combinations. The AUC statistic of the overall best performing feature combination was compared for statistical significance with the AUC of the remaining classifiers achieving the best result for each number of features (adjusted with the Holm-Bonferroni method). The performance of univariate and the best multivariate models given the number of features (2-12) was tested on a held-out test set and the AUC reported for comparison with the training results.

3.4.6 Data pre-processing

The dataset was scrutinised for anomalous values and corrected where required. Corrected data entries were those with missing gender information (3 patients) and erroneous inputs in cases of non-existent episodes of low variability (10 patients).

The dataset contained 387 traces (314 IUGR and 73 healthy controls) with no episodes of high variability and 1410 traces (580 IUGR and 820 healthy controls) with no episodes of low variability. Traces with no high variability indicate that the monitored fetus did not enter a period of high heart rate variation within 60 minutes of recording. Fetuses that do not enter a phase of active sleep within one hour may be suspected of being compromised; they are therefore of particular interest in this investigation and should be included in the analysis. The majority of traces with no low variation are traces that have met criteria after an episode of high variability and have been terminated before a period of low activity could occur. The absence of low variation in a trace does therefore not allow for assumptions concerning fetal wellbeing to the same degree. However, short- and long-term variability during episodes of low heart rate variation may contain information on the severity of compromise. Substitute long-term variability and short-term variability values were therefore estimated for both 'missing' high and low periods.

Substitution estimates were established based on the data available. For long-term variability in active sleep episodes (LTVinHi) the lowest value of existing *LTVinHi* measures was identified in each gestational age group with ten or more traces available. A regression line was fitted and *LTVinHi* substitute values computed for each week of gestation. This approach was based on the assumption that the minute ranges in traces without an identified episode of high variability would have always been equal to or lower than this constructed threshold, as they would have otherwise been classified as of high variation. The same approach was used to estimates substitutes for *LTVinLo* and *STVinHi*; a visual representation is shown in Figure 3.8. No clear threshold could be identified in short-term variability in quiet sleep episodes; substitute values were inferred from a regression line fitted to the mean of existing *STVinLo* measures instead.

Compared to the thresholds used in the cCTG analysis to classify periods of varying variability, as described by [108], the substitute values appear to be a reasonable choice. *LTVinHi* substitute values lie above the 32ms and *LTVinLo* below the 30ms reported.

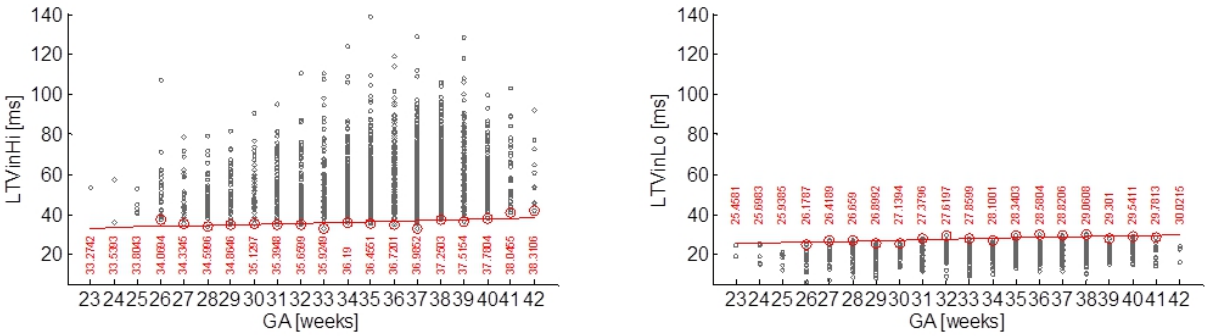


Figure 3.8: Example of the approach to substitute value estimation. *LTVinHi* and *LTVinLo* were inferred from a regression line fitted to the lowest value in each gestational age group containing more than 10 traces with an episode of high or low variability respectively.

3.4.7 Results

As observed in the preliminary analysis, variability features appear to discriminate better earlier on in gestation in both male and female traces. This can be observed in the example boxplots showing short-term variability averaged over episodes of active sleep (Figures 3.10 and 3.11). As expected the quality of risk markers declines with increasing gestational age (Figure 3.12a). Based on the visual inspection of the discriminative ability of variability markers (Figures 3.10 and 3.11), the difference in behavioural pattern under 35 weeks (Figure 3.7) and the univariate models of both *STVinHi* and *LTVinHi* achieving an acceptable AUC value of 0.7 (Figure 3.12a), the previously chosen cut-off at 34 weeks of gestation appears reasonable. The performance of individual features for 23-34 weeks of gestation is listed in Table 3.11. For population characteristics see Table 3.3. The number of minutes in high variation per hour (*HiMinph*) was most predictive with an AUC of 75% on the training set and 73% on the test set. *HiMinph* was statistically not significantly superior to the variability risk markers in active sleep, *LTVinHi* (74% training, 72% test) and *STVinHi* (71% training, 71% test). Both *LTVinHi* and *STVinHi* were found to be significantly different in IUGR traces compared to normal controls ($p<0.01$). The AUC of multivariate models also declines with an increasing gestational age range (see best performing models per number of features in Figure 3.12b). The model including the features gestational age, long-term and short-term variability in high variation episodes,

the average duration in high variation and the number of high episodes in the trace, could improve the discriminative performance to 77% on the training set and 76% on the test set. The results are summarised in Table 3.12, including those for 35 to 42 weeks of gestation for comparison. The ROC curve for the best performing multivariate model is shown in Figure 3.9 as a graphical representation of the performance of the classifier as its discrimination threshold is varied.

The correlation between heart rate variability markers was assessed in those traces containing periods of both low and high variability (Table 3.9) and all traces, includ-

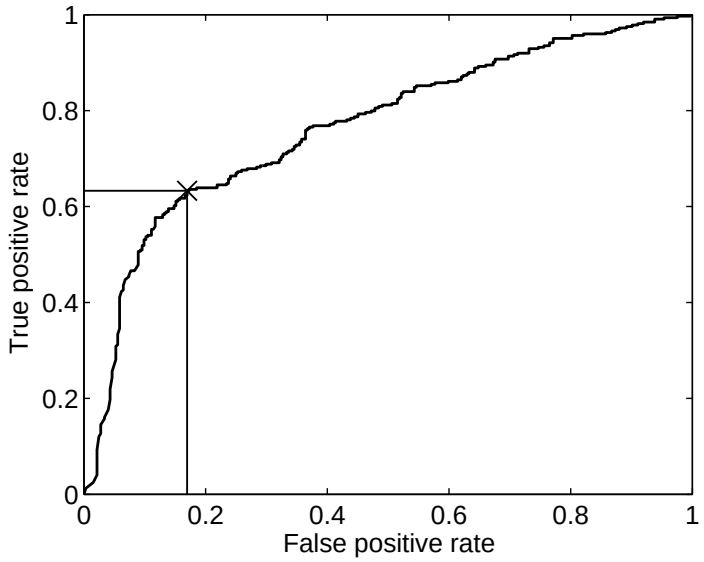


Figure 3.9: The ROC curve of the best performing multivariate classifier illustrating how the performance changes when the test threshold is varied. Sensitivity (true positive rate), specificity and positive predictive value (PPV) are important measures of a classifier's suitability for use in the field and dependent on the appropriate choice of test threshold. As an example, a threshold of 57% (optimal according to Youden's J statistic (Equation 3.3) results in a sensitivity of 63% and a specificity of 83% as shown in the figure. The PPV is dependent on the IUGR prevalence in the population. The dataset discussed in this chapter is matched, the prevalence of IUGR therefore 50%, resulting in a PPV (probability that subjects with a positive screening test have the disease) of 79%. The average prevalence in Guatemala is 11% which would reduce the PPV significantly Taking into account the characteristics of the target population and the use environment is hence important when choosing a cut-off value. Furthermore, triaging women based on socio-economic factors or clinical assessment prior to screening for IUGR (discussed further in Section 8.1.3) may be considered to increase the prevalence in the test population and therefore the classifier's PPV, reducing the number of false positives.

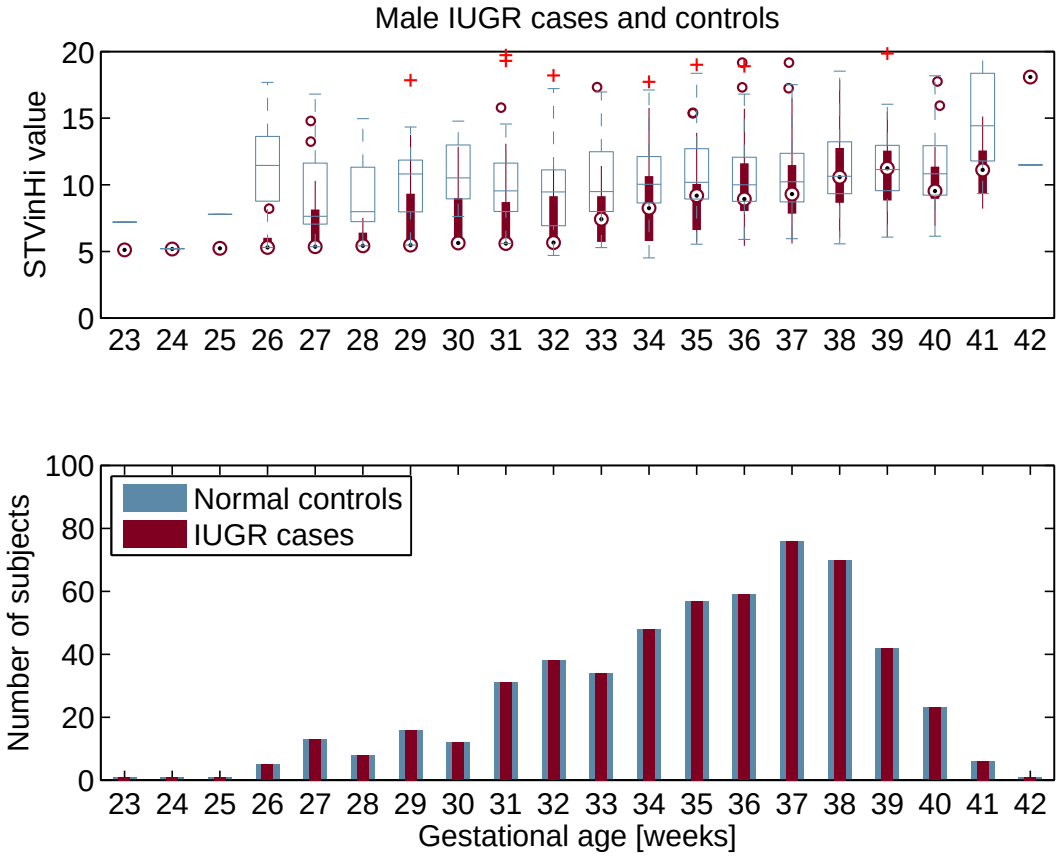


Figure 3.10: Top graph: Visual assessment of the difference of STVinHi values in the male IUGR population compared to normal male controls per gestational age group. Bottom graph: The number of male IUGR cases and controls in each gestational age group. Of all age groups with $N >= 15$ for both cases and controls the two study populations differ significantly at 29 and 31 to 36 weeks ($p < 0.05$).

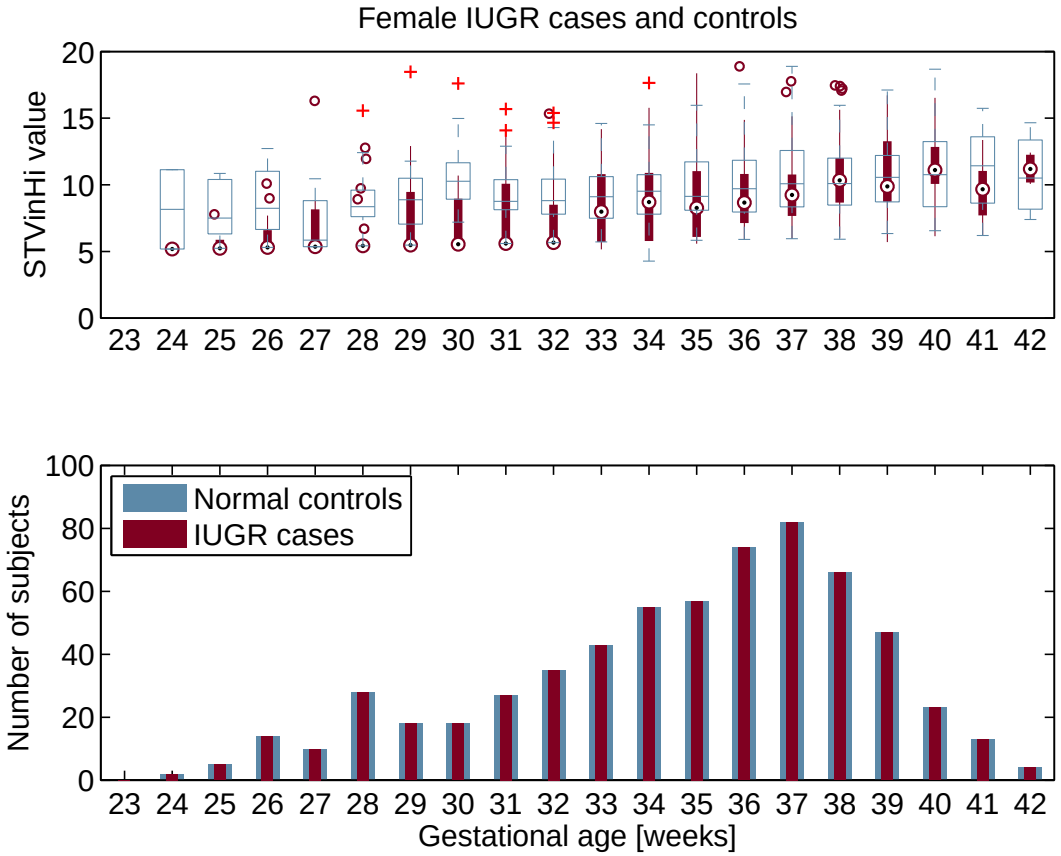
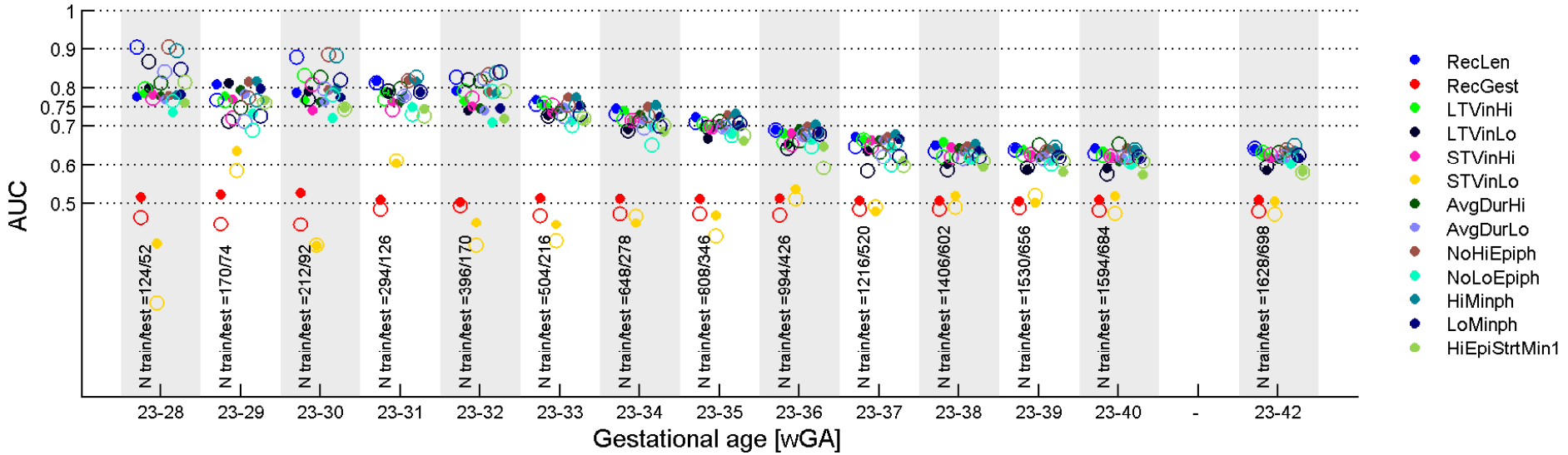
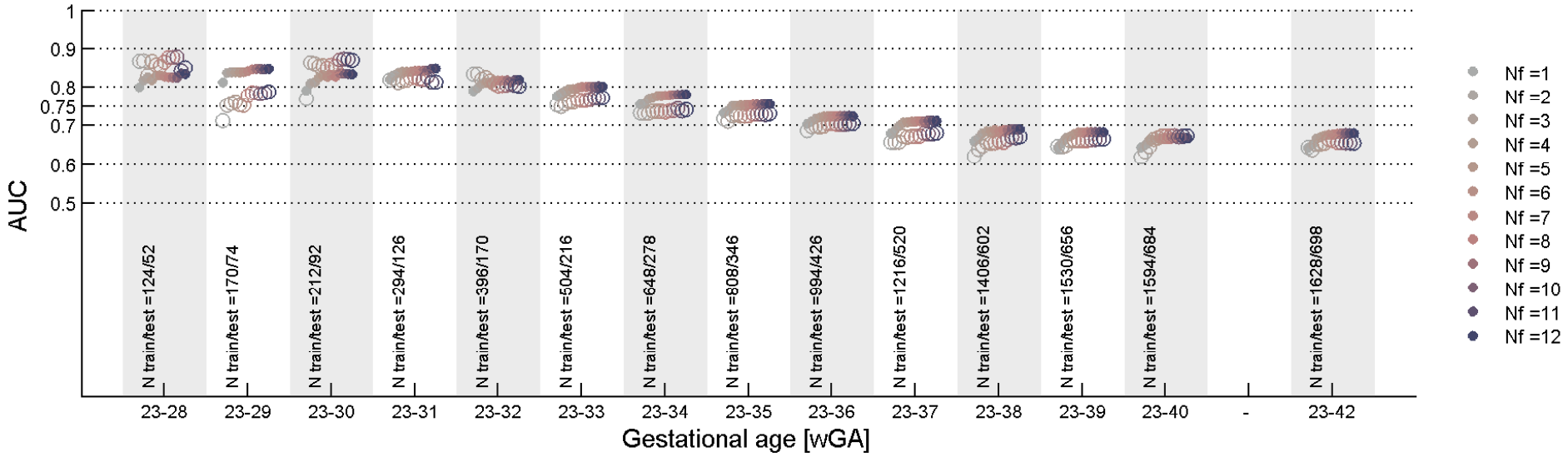


Figure 3.11: Top graph: Visual assessment of the difference of STVinHi values in the female IUGR population compared to normal female controls per gestational age group. Bottom graph: The number of female IUGR cases and controls in each gestational age group. Of all age groups with $N >= 15$ for both cases and controls the two study populations differ significantly at 28, 30, 31, 32, 36 and 37 weeks ($p < 0.05$).



(a) Univariate logistic regression results for increasing gestational age groups



(b) The best multivariate logistic regression results for each number of features (1-12)

Figure 3.12: AUC results of uni- and multivariate logistic regression models. Performance for the age group 23-41 wGA was not computed as the additional number of subjects was negligible. Filled circles are AUC values achieved over a 5-fold CV, empty circles are test results.

	LTVinHi	STVinHi	LTVinLo	STVinLo
LTVinHi	-	+0.627, p<0.001	+0.052, p>0.05	+0.144, p<0.05
STVinHi	+0.627, p<0.001	-	+0.070, p>0.05	+0.283, p<0.001
LTVinLo	+0.052, p>0.05	+0.070, p>0.05	-	+0.690, p<0.001
STVinLo	+0.144, p<0.05	+0.283, p<0.001	+0.690, p<0.001	-

(a) The relationship between the cCTG variables was measured by Spearman's rho, assessed on all IUGR traces with a high and low variability episode (N=271).

	LTVinHi	STVinHi	LTVinLo	STVinLo
LTVinHi	-	+0.765, p<0.001	-0.038, p<0.001	+0.145, p<0.001
STVinHi	+0.765, p<0.001	-	+0.021, p<0.001	+0.251, p<0.001
LTVinLo	-0.038, p<0.001	+0.021, p<0.001	-	+0.585, p<0.001
STVinLo	+0.145, p<0.001	+0.251, p<0.001	+0.585, p<0.001	-

(b) The relationship between the cCTG variables was measured by Spearman's rho, assessed on all normal traces with a high and low variability episode (N=267).

Table 3.9: Variability risk marker correlation in traces with both active and quiet sleep episodes.

ing those with substitute values in the absence of high or low variability episodes (Table 3.10). The correlation in IUGR and normal traces was investigated separately to uncover whether relationships depend on the developmental state of the fetus. For traces with high and low variability episodes, *LTVinHi* and *STVinHi* were found to be correlated in both IUGR and normal traces, however to a greater extent in the control group. When including traces with substitute values the correlation increased, particularly in IUGR traces which can be explained by the greater number of 'missing' high variation periods and therefore substitutes. *LTVinLo* and *STVinLo* showed a correlation in traces of growth restricted fetuses and healthy controls, including and excluding substitute values. Adding substitute values introduced a correlation between LTVinHi and LTVinLo in IUGR traces.

3.4.8 Required trace length for IUGR discrimination

Given the context of this investigation, i.e., to develop a model for IUGR risk detection in low-resource environments, the trace length required in order to make meaningful predictions is important. Recordings of 60 minutes, the standard maximum trace length in

	LTVinHi	STVinHi	LTVinLo	STVinLo
LTVinHi	-	+0.908, p<0.001	+0.608, p<0.001	+0.292, p<0.001
STVinHi	+0.908, p<0.001	-	+0.593, p<0.001	+0.305, p<0.001
LTVinLo	+0.608, p<0.001	+0.593, p<0.001	-	+0.533, p<0.001
STVinLo	+0.292, p<0.001	+0.305, p<0.001	+0.533, p<0.001	-

(a) The relationship between the cCTG variables was measured by Spearman's rho, assessed on IUGR traces, including substitute values in absence of an episode of high or low variability (N=1163).

	LTVinHi	STVinHi	LTVinLo	STVinLo
LTVinHi	-	+0.847, p<0.001	+0.383, p<0.001	+0.275, p<0.001
STVinHi	+0.847, p<0.001	-	+0.366, p<0.001	+0.264, p<0.001
LTVinLo	+0.383, p<0.001	+0.366, p<0.001	-	+0.625, p<0.001
STVinLo	+0.275, p<0.001	+0.264, p<0.001	+0.625, p<0.001	-

(b) The relationship between the cCTG variables was measured by Spearman's rho, assessed on normal traces, including substitute values in absence of an episode of high or low variability (N=1163).

Table 3.10: Variability risk marker correlation with missing values substituted.

cCTG monitors, may not be feasible due to both examination time and space limitations. Results have shown that variability measures are better predictors of IUGR when derived from active sleep episodes. The time of onset of the first high heart rate variability episode in recorded traces was therefore reviewed and is shown in Figure 3.13. An easily made assumption that a high episode only occurs much later in a majority of IUGR recordings, given the higher number of long recordings found in the database, did not apply. In 61% of IUGR and 84% of normal traces an episode of high heart rate variability occurred within 30 minutes of starting the recording. When taking into account the percentage of fetuses that did not enter a period of active sleep at all within 60 minutes (27% IUGR, 6% normal), monitoring for only 30 minutes would capture 83% of the expected high variability episodes in IUGR and 89% in healthy fetuses.

3.4.9 Discussion

Models based on measures computed for periods of quiet and active sleep resulted in a similar performance when compared to the initial investigation (variability metrics averaged over the entire traces). The bias, however, introduced by the varying length of

AUC[%] Features	DB: 23-34 GA Training data	Test data
LTVinHi	74 (95% CI = 70–78)	72
LTVinLo	69 (95% CI = 65–73)	69
STVinHi	71 (95% CI = 67–75)	71
STVinLo	46 (95% CI = 41–50)	47
AvgDurHi	73 (95% CI = 69–76)	71
AvgDurLo	71 (95% CI = 66–75)	69
NoHiEpiph	75 (95% CI = 71–79)	73
NoLoEpiph	69 (95% CI = 65–73)	65
HiMinph	**75** (95% CI = 71–79)	**73**
LoMinph	71 (95% CI = 67–75)	70
HiEpiStrtMin	63 (95% CI = 59–67)	70
RecGest	48 (95% CI = 43–52)	47
RecLen	74 (95% CI = 70–78)	73

Table 3.11: The AUCs of each individual risk marker for populations including 23-34 weeks. Results are reported as achieved over a 5-fold cross-validation on the training data and as returned when tested on the held-out test data (70/30 split). The quality of the best performing feature early in gestation, the minutes per hour in high variation (*HiMinph*), was compared to the remaining. *HiMinph* was significantly superior than gestational age and metrics associated with low variation episodes ($p<0.01$), with the exception of (*LoMinph*). No difference was found when comparing to features also derived from high variability episodes. Gender has been excluded from the analysis as it will not be available in anticipated use context. The recording length has been evaluated as a feature solely to highlight its potential to bias outcomes. The recording length can to some extend be viewed as a proxy for clinical judgement at the time of recording. It has not been included in multivariate models.

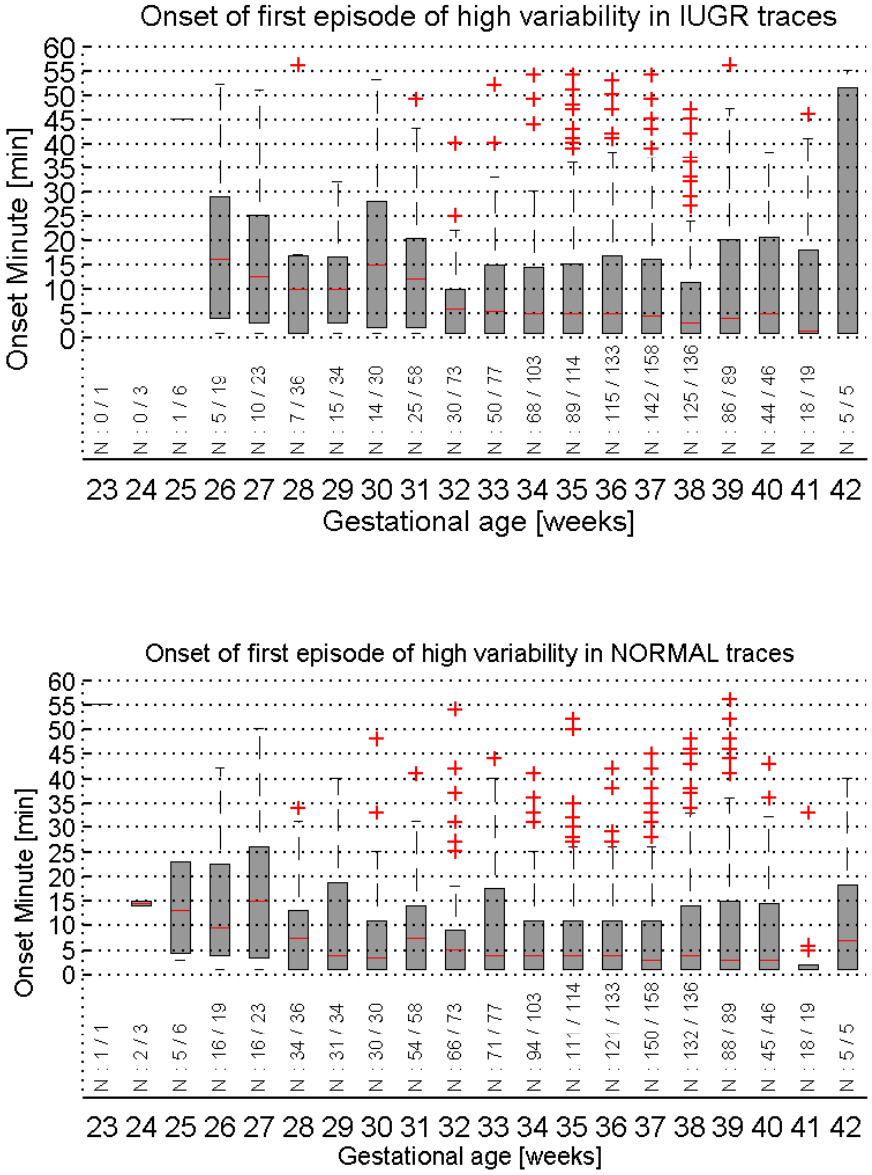


Figure 3.13: The median and interquartile range of the onset time of the first episode of high heart rate variation in recorded traces (top: IUGR traces, bottom: healthy controls). Boxplots are shown per gestational age, below each boxplot the number of traces with a high variation episode out of the total number of traces for the gestational age group is indicated.

DB: 23-34 GA	Training data	Test data
AUC [%] (Features: RecGest, LTVinHi, STVinHi, AvgDurHi, NoHiEpiph)	77 (95% CI = 73–80)	76
DB: 35-42 GA	**Training data**	**Test data**
AUC [%] (Features: LTVinHi, NoHiEpiph)	57 (95% CI = 54–61)	61

Table 3.12: The best performing multivariate logistic regression classifier for 23-34 weeks (top) and 35-42 weeks of gestation (bottom). Results are reported as achieved over a 5-fold cross-validation on the training data and as returned when tested on the held-out test data (70/30 split). The quality of the classifiers was compared to the best-performing of each number of features, including single features; as in preliminary study they were not found significantly superior in performance.

cCTG recordings, possibly leading to very different behavioural state compositions of the traces, has been confirmed in section 3.4.2 (see Figure 3.6). In addition, the univariate model outcomes of metrics computed for individual sleep states has revealed the superior discriminative power of metrics associated with active compared to quiet sleep episodes. This indicates that the fetal behavioural state may confound heart rate variability analysis in general. This has also been observed by [123]. Two investigations could be identified in literature that took sleep state information into account when analysing the difference in heart rate variability metrics in IUGR and normal fetuses ([123] and [89]). Both found their respective metrics under investigation (the developed 'brain age score' [89] and the HRV measures SDNN (standard deviation of normal to normal), RMSSD (root mean square of successive differences) and PPA (phase plane area)) to be lower in growth restricted compared to normal fetuses.

The fact that (*HiMinph*) was the best performing single feature and *AvgDurHi* and *NoHiEpiph* were part of the best performing multivariate model allow the conclusion that the sleep pattern provides an important indication on the fetal developmental state, and that knowledge about the trace pattern can aid discrimination.

Active sleep periods could be observed much earlier on in IUGR traces than anticipated, which supports the notion of point-of-care risk assessment. Taking into account the

prevalence of fetuses not showing a high variability episode within 60 minutes, an active sleep episode can be expected to occur in the first 10 minutes in 64% and 70%, and within 30 minutes in 83% and 89% of IUGR and normal traces respectively.

3.5 Conclusion

The aim of the research presented in this chapter was to evaluate the utility of promising Doppler-based fetal heart rate variability markers to identify fetal growth restriction antenatally. Analysis of the extensive dataset confirmed that the distribution of marker values differs significantly in growth-restricted fetuses compared to normal controls, also when averaging over periods of active and quiet sleep respectively. The importance of fetal behavioural states in the analysis of variability markers was suspected and could be confirmed in the investigation. The discriminative ability of the risk markers evaluated here was considerably better when excluding fetuses recorded at a gestational age greater than 34 weeks. Under 35 weeks of age the best multivariate classifier is of acceptable clinical utility [124] with an AUC of greater than 0.75. The discrimination quality improves to over 0.8 for gestational ages ≤ 31 weeks. In summary, the results suggest that fetal heart rate variability markers together with trace characteristics can contribute to the detection of early-onset IUGR; however, they are not suitable for the classification of late-onset IUGR. The difficulty of detecting late-onset IUGR has been previously noted in relation to other monitoring modalities, such as umbilical artery Doppler [88,125].
This investigation could confirm the dependence of heart rate variability markers on gestational age, which has been reported in the literature before [83,88]. Risk marker values increased with gestational age in both growth-restricted and normal cases. The rate of change in compromised fetuses was, however, much higher with marker values being considerably lower earlier in gestation and approximating to control values towards term.
The investigation presented here furthermore allows the assumption that in a majority of cases monitoring of the fetus for a maximum of 30 minutes will be sufficient for IUGR

risk assessment. In a setting where time is precious as midwives often have to travel far to provide antenatal care, and where parallel assessment is not feasible as patient encounters typically occur at home, this is a very valuable insight.

Chapter 4

Development of an mHealth fetal monitoring system

4.1 Introduction

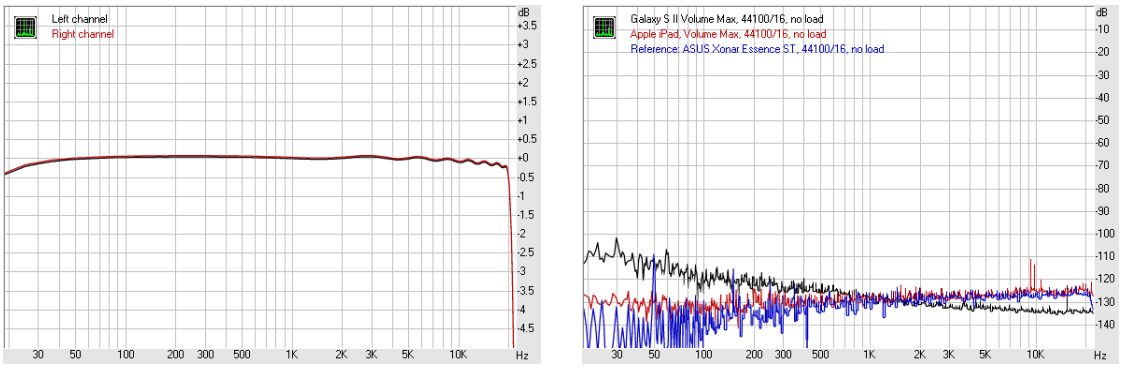
Auscultation (1D) Doppler has been identified as one of the fetal monitoring modalities most widely used by healthcare professionals (primarily midwives) world-wide. Next to monitoring equipment used in clinical practice several consumer models have entered the market over the past three years, offering mothers-to-be a device to listen to the fetal heart beat at home. The clinical utility of such devices for personal use is debatable; other than the promise of reassurance to parents which may equally turn into concern if the fetal signal could not be identified or misinterpreted; such monitors typically do not offer any form of analysis. This development however means that extremely low-cost transducers, which can easily be adapted to connect to recording and processing devices such as smart phones, are available for purchase. This has motivated the use of such transducers and the testing of its appropriateness for risk screening in low-resource environments.

The primary objective of the anticipated mHealth perinatal monitoring system was to enable the recording of fetal cardiac activity onto a mobile phone for further processing

and fetal risk analysis. In its wider context the system was also conceived as a platform for a more comprehensive maternal-fetal health assessment. The prototype should therefore include not only a sensor to capture fetal cardiac activity, but also a pulse oximeter to record maternal pulse and oxygen levels, both interfacing with a smart phone.
This chapter provides an overview of the system prototype including its hardware elements, their assembly, as well as an introduction to the signals recorded from its three sensor components.

4.2 m-Health monitoring system

To build the prototype for an affordable mHealth perinatal monitoring system the Doppler lowest in cost at time of project start was selected. The CE-marked, hand-held 1-D Doppler transducer, *AngelSounds Fetal Heart Detector* (Jumper Medical Co., Limited, Shenzhen, Guangdong Province, P.R. China) (cost: $17 in bulk) has two audio outputs, one of which was connected to the mobile phone (Samsung Galaxy S2, Samsung Electronics Co., Ltd., Suwon, South Korea) via the microphone input for recording. The phone's sound card demonstrates a reasonably flat frequency response and good performance assessing noise levels as illustrated by Figure 4.1. In order to allow for continuous recording and to match the voltage levels the phone can tolerate a custom recording lead was built for this purpose (see Figure 4.2 for a diagram of the bespoke lead). Earphones or speakers were attached to the second audio output to enable the operator to listen to the ultrasound signal at the same time as recording. For a compact design and more importantly in order to potentially utilise the phone's accelerometer data to infer information about the Doppler's position and movement with respect to the maternal abdomen, a mechanism was added by which a smart phone holder could be slid onto the upper casing of the Doppler sensor and maintained there steadily throughout recording. To capture the maternal pulse and oxygen levels the wrist-worn pulse oximeter $WristOx_2$ 3150 (Nonin Medical Inc., Plymouth, Minnesota, USA) was integrated into the system



(a) Frequency response

(b) Noise levels

Figure 4.1: Quality metrics of the Samsung Galaxy S2 mobile phone[1]. (a) The sound card's frequency response, illustrating the range of sounds the device can produce. The range should ideally span what the human ear can perceive (20Hz - 20kHz) and be flat throughout showing no significant spikes or troughs. (b) The signal to noise ratio, comparing the desired output signal to the level of background noise. The Galaxy S2 has overall low noise levels in the frequency range of interest (see Section 6.3.2) and performs particularly well in high frequencies.

via Bluetooth connection. As previously noted the pulse oximeter is part of a wider concept including antenatal maternal health screening; this platform extension however is not central to this thesis and will only be discussed further in the context of the field trial of the system (Chapter 7) and future work (Chapter 8). For illustrations of the prototype assembly and the prototype in use, see Figure 4.3 and Figure 4.4.

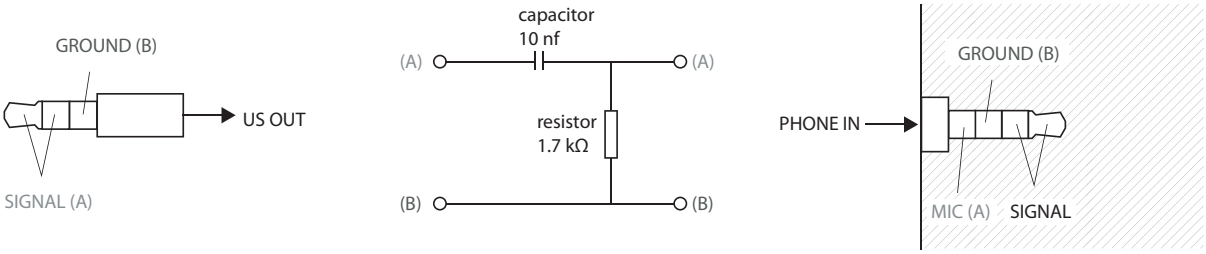


Figure 4.2: The smart phone's audio jack is designed for both headphone (audio output) as well as microphone (audio input) connections. In order to detect the connection type the phone sends out a direct current (DC) detecting a reduction in voltage. To connect the Doppler device a capacitor and resistor had to be added to eliminate the DC and adjust the voltage input to levels tolerated by the smart phone's microphone input.

[1]Graphs adapted from AnandTech: http://www.anandtech.com/show/4686/samsung-galaxy-s-2-international-review-the-best-redefined/13

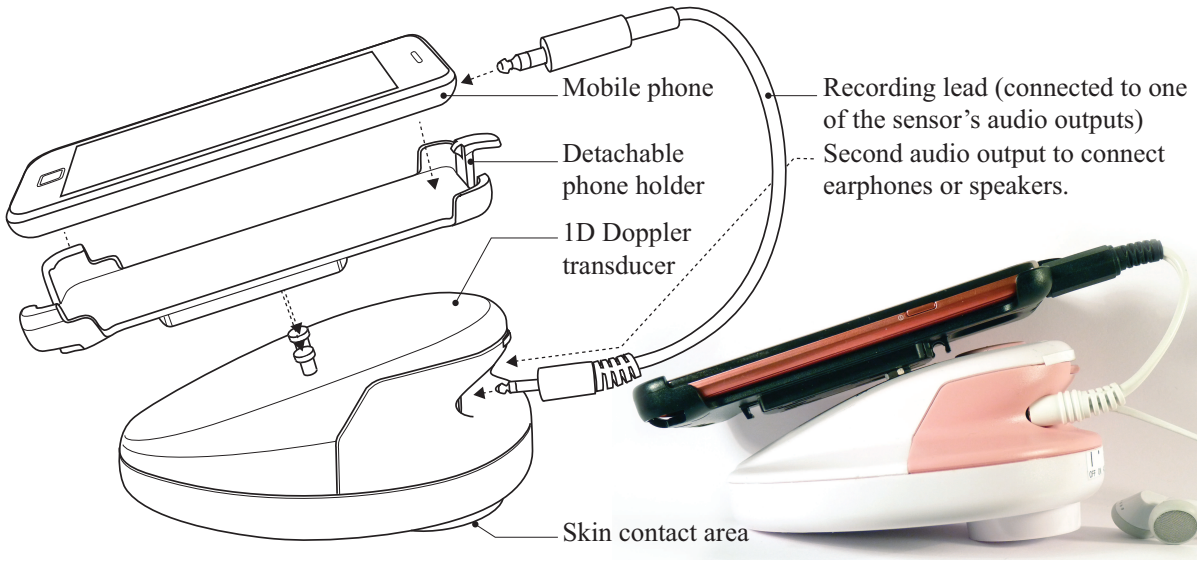


Figure 4.3: The prototype was assembled from the off-the-shelf 1D Doppler, a smart phone holder, the smart phone, custom made recording lead and peripherals such as a set of speakers or headphones. The Doppler was modified to allow the phone holder to be slid on and to remain steadily attached to its upper casing when recording.

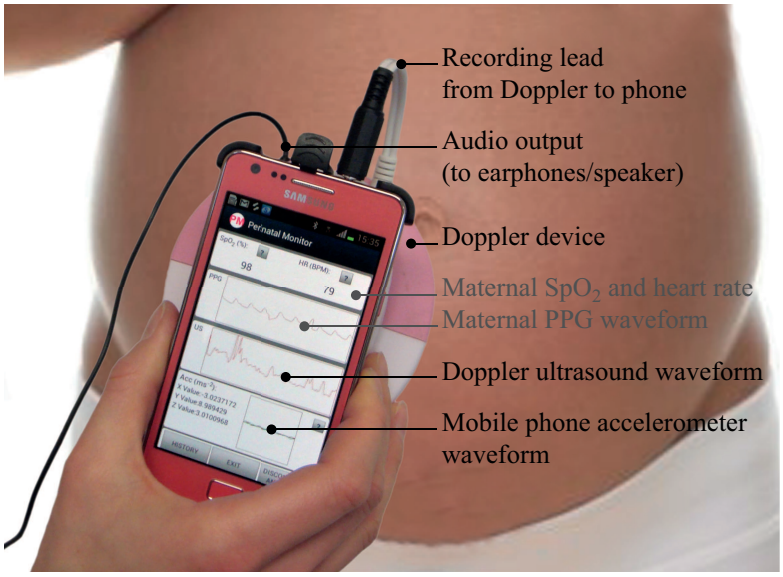
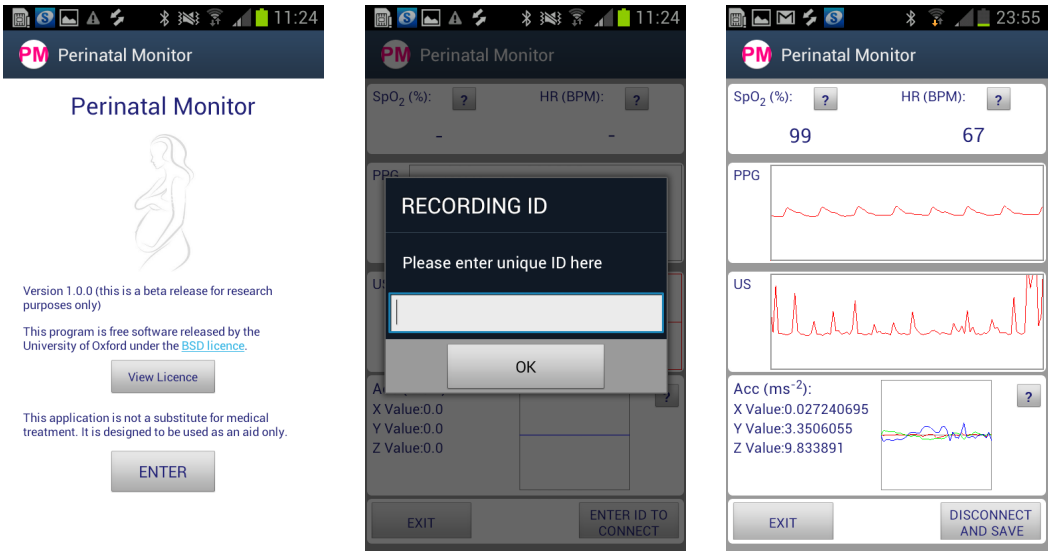


Figure 4.4: The initial 1D Doppler prototype and recording application in use. The maternal arterial oxygen saturation (SpO_2) heart rate and photoplethysmogram (PPG) were recorded by the Nonin WristOx$_2$ 3150 pulse oximeter concurrently via Bluetooth for later use, such as maternal health assessment and as a reference signal for improving fetal heart rate estimation.



(a) Start screen (b) Patient ID screen (c) Recorder screen

Figure 4.5: The initial mobile application programmed for Android operating systems. (a) The application was developed for research purposes only, to collect data and to test the feasibility of using an mHealth platform in resource constraint settings. (b) Each patient encounter is recorded with a unique patient identifier. (c) The recording screen visualises all signals recorded for real-time feedback.

4.2.1 Design of data acquisition software

A recording application was designed to concurrently capture the Doppler auscultation of the fetal heart activity, the phone's accelerometer data and the maternal pulse and blood oxygenation (Figures 4.4 and 4.5). Of primary interest in this investigation are the Doppler ultrasound data, which were written by the phone to an uncompressed WAV file format at the maximum sampling rate of 44,100 Hz. The principles of Doppler ultrasound and technical specifications of the selected transducer are summarised in the following section.

The recording application was programmed to capture the tri-axial accelerometer data at maximum sampling frequency, which resulted in an average sampling frequency of 96 Hz. As outlined above the accelerometer data were extracted to obtain feedback on the device's movement over the duration of a recording, which may serve as input to assess signal quality. A brief introduction to accelerometry and the phone's sensor can be found

below. The development of a signal quality metric and use of the accelerometer data are subjects of the following chapter (Chapter 5).

4.2.2 1D Doppler ultrasound recording

An ultrasound probe houses a piezoelectric transducer emitting and receiving ultrasonic sound waves by transforming electrical charge into mechanical energy and vice versa. Ultrasound waves are waves in a frequency range higher than the audible limit of human hearing, which is typically 20 kHz. Two types are common in clinical use: pulsed and continuous wave transducers. Several modes are available in medical applications, including the well-known B-mode (brightness mode) producing an image of a selected scanned plane in the body, also known as 2D mode, and M-mode which emits pulses in rapid succession producing an ultrasound video. The *AngelSounds* transducer operates in continuous Doppler mode with a frequency of 3.3 MHz. This means the probe continuously generates and receives, which is accomplished by the use of a two-crystal transducer to fulfill both functions.

Doppler mode refers to the probe's capability to measure the change in frequency between the emitted and the observed signal, reflected from the moving structures in the ultrasound beam's path. This allows for estimation of the velocities of those structures, and in the context of this investigation the identification of heart valve and wall motion (an illustration of the cardiac application is provided in Figure 4.6). The shift in frequency is called the *Doppler effect* after the Austrian mathematician and physicist Christian Doppler (1803-1853), and is written as:

$$f_D = \frac{2f_o}{c} V cos\theta \tag{4.1}$$

where f_D is the measured change in frequency (Hz), f_o the frequency of emitted ultrasound (Hz), c the speed of sound in soft tissue (m/s), V the velocity of the reflecting interface and θ its angle with the ultrasound beam [126].

The Doppler mode for fetal heart rate assessment is also described as auscultation Doppler given that the resultant cardiac signal is translated into audible cardiac sounds[2]. To distinguish clearly from the 2D B-mode, which is commonly associated with medical ultrasound, the signal is also referred to as one-dimensional, or simply 1D, in this thesis.

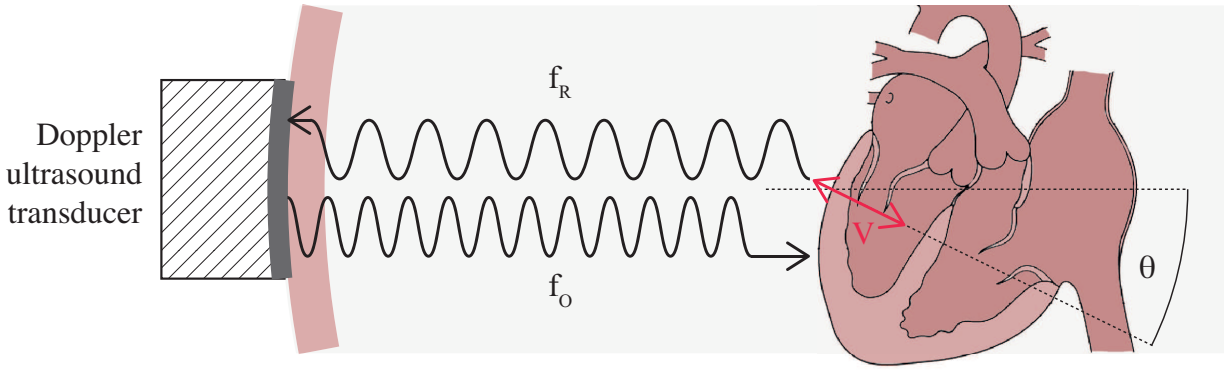


Figure 4.6: The continuous Doppler transducer emits with frequency f_o and receives the reflected signal with frequency f_R. In this example the heart motion is away from the transducer, since it is contracting, and so the reflected wave has a longer frequency ($f_R < f_o$). The change in frequency, the Doppler shift, is dependent on the velocity V of the moving target, the angle θ between ultrasound beam and direction of motion and the speed of sound in soft tissue c. To monitor fetal cardiac activity the transducer is placed against the maternal abdomen and directed towards the fetal heart.

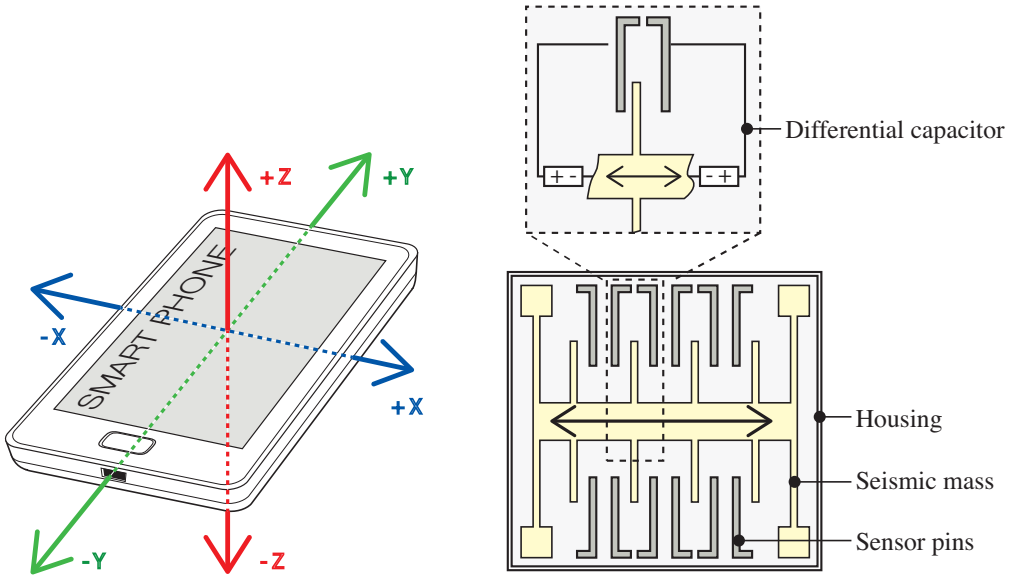
4.2.3 Accelerometer

The accelerometer is a built-in electromechanical sensor measuring acceleration. As part of a smart phone it records tilt and motion with the primary function of activating screen rotation when the phone's orientation changes to support phone features based on gesture control. Most smart phone accelerometers are capacitive tri-axial sensors (refer to Figure 4.7 for an illustration of the sensor's working mechanism). They consist of a housing, a fixed metal pin structure and a seismic mass (silicone 'comb' structure) moving in between according to the applied forces such as gravity or phone displacement. Ac-

[2] The term 'cardiac sounds' is used to differentiate these recordings from classic phonocardiogram auscultation where the term 'heart sound' refers to specific acoustic events such as valve openings and closings. Refer to Section 1.1.2 for more details on the phonocardiogram.

celeration is measured based on the change in capacitance occurring when the distance between metal sensor pins and moving mass alters.

When attached to the Doppler transducer, the phone's accelerometer can provide information on the Doppler's location in relation to a set reference and any movement together with its intensity over the duration of a recording.



(a) The phone's accelerometer records acceleration in X, Y and Z direction.

(b) The capacitive sensor houses stationary metal pins and a flexible silicone structure (seismic mass) which moves in between according to the forces applied.

Figure 4.7: (a) The in-built accelerometer in the majority of smart phones is a tri-axial sensor. (b) The sensor works based on measuring the capacitance between the fixed pins and 'fingers' of the moving mass.

4.2.4 Hardware considerations

It is important to note that signal characteristics are dependent on selected hardware and sensors, which requires consideration when implementing algorithms presented in this thesis for use across different monitoring platforms. To provide an example, the recorded Doppler signal is not only determined by the transducer specifications but also by pre-processing steps, such as demodulation and filtering of the received signal, of

which details are not available given most devices' proprietary nature. The built-in accelerometer similarly may differ across mobile phones, as does the maximum sampling frequency the phone is set to. To limit the effect of device-specific pre-processing a Doppler probe providing no in-built analysis was selected for introduced prototype.

4.3 Prototype testing with skilled personnel

To test the usability of the prototype and to start building a database of fetal and maternal recordings the prototype was applied in practice by skilled midwives at the John Radcliffe Hospital in Oxford, UK. This was part of a study approved by the NHS Health Research Authority (REC reference: 12/SC/0147). The signals as described above were collected from 17 participating patients and volunteers to enable the design of processing tools for the fetal cardiac signal in particular. The development of a 1D Doppler signal quality metric is discussed in Chapter 5. Chapter 5 also provides further details on the study, patient and data characteristics. The development of a 1D Doppler segmentation algorithm for beat-to-beat fetal heart rate estimation (based on collected data) is laid out in Chapter 6.

4.4 Conclusion

The system was able to capture useful data from all subjects. Trained midwives are highly experienced in the use of Doppler probes for intermittent auscultation as well as palpation to locate the fetus; staff participating in the first testing and building of database *JR RDDB* (a summary of the database was provided in Section 1.3) were therefore comfortable using the system, which was designed to be as compact as possible for ease-of-use.

The prototype was later trialled in a low-resource setting and adapted to suit the specifics of the use context. This is described in Chapter 7.

Chapter 5

1D Ultrasound Signal Quality Index

5.1 Introduction

One of the most powerful characteristics of mHealth solutions is their potential to reach large numbers of patients even in remote areas and to address the shortage of skilled healthcare personnel by equipping local resources with intelligent decision and communication support. A key consideration for successful mHealth implementation should be the unusual circumstances that arise when previously out-of-reach technologies are made available to users with limited experience in handling them. Ensuring the quality of data is a primary concern and best controlled at the point of data collection. Feedback on the usability of the data captured for analysis is therefore critical. The user is then able to retake recordings if required and review the system's output alongside a confidence rating, which is related to the quality of underlying data.

This chapter addresses this challenge and describes the development of a Doppler Signal Quality Index (SQI), one crucial building block in the anticipated mHealth fetal monitoring system.

		Gestational Age (weeks)		Maternal Age (years)	
Pregnancy Status	N	Mean	Range	Mean	Range
High Risk*	14	37	32-40	33	24-38
Low Risk	3	31	20-38	36	35-37

Table 5.1: Summary of the demographics of the study subjects who contributed to the recording database.
(*Risk concerns maternal complications such as preeclampsia and gestational diabetes, there were no fetuses with known health compromise.)

5.2 Database

For the purpose of this research, recordings of fetal heart activity using the described hand-held Doppler-phone recorder were collected at the John Radcliffe Hospital in Oxford, UK, as previously mentioned in Chapter 4. The study was approved by the NHS Health Research Authority (REC reference: 12/SC/0147). Women with singleton pregnancy, over the age of 18 and scheduled for a routine cardiotocography (CTG) recording [15] were invited by their clinician or midwife to take part in the study. Potential candidates received detailed information on the study protocol and their right to withdraw from the study at any stage of the recording session. Written consent was obtained from each participant prior to data collection. The fetal heart activity was recorded from three volunteers and 16 recruited patients; two recording sessions had to be terminated, one due to health concerns and one due to technical problems (low battery). The demographics of included study subjects are summarised in Table 5.1.

5.2.1 Conditions of Recordings

All women recorded were in the supine position, as is common for the assessment of fetal heart activity. Ultrasound transmission gel was used to ensure good contact of the probe with the mother's skin. The probe with attached phone (see Figures 4.3 and 4.4) was directed by listening to the ultrasound output until fetal cardiac activity was audible and then held in place for up to 10 minutes. In cases where fetal heart sounds were lost, the device was realigned.

Signal Quality Score		Definition provided to annotators
1	NIA	No information available in the signal.
2	BAD	The signal may contain heartbeats, but it is too 'noisy' to identify them.
3	INTERMEDIATE	Difficult to hear fetal beats, but can be done with some effort. Heart rhythm detection may be possible.
4	GOOD	Some background noise, but heartbeats can be heard clearly. Heart rhythm detection is possible.
5	EXCELLENT	Almost no background noise, fetal heartbeats are easy to identify, heart rhythm detection is possible.

Table 5.2: Signal quality labels and their definitions used by the annotators. Annotators could differentiate further by choosing label values on the continuous scale 1-5.

All recordings were anonymised by assigning a unique numeric identifier and transferred from the mobile phone onto a secure server within the Institute of Biomedical Engineering at the University of Oxford.

5.3 Signal quality annotation and data exclusion

The start and end of recordings were edited to remove leading and trailing noise by removing data up until the point the transducer was placed on the abdomen and the moment after it was lifted. The recordings were split into 1 minute episodes for ease of annotation. Three annotators evaluated the recordings through a custom-built Matlab-based annotation interface, which allowed for both visual and audio assessment of the Doppler signals. An illustration of the interface is shown in Figure 5.1. All annotators used equivalent PCs and high-quality head-phones. Annotators sub-segmented each 1 minute recording into periods of homogeneous quality and assigned signal quality labels as described in Table 5.2. All three annotators had relevant experience in analysing cardiac audio recordings. Examples of different quality signals in both the time and frequency domain are shown in Figure 5.2.

For the purpose of the classification task described in this chapter, assigned labels were rounded to the next integer and recordings split into episodes of homogeneous quality. Episodes were chosen to be 3.75 seconds, as used in the computerised analysis of fetal

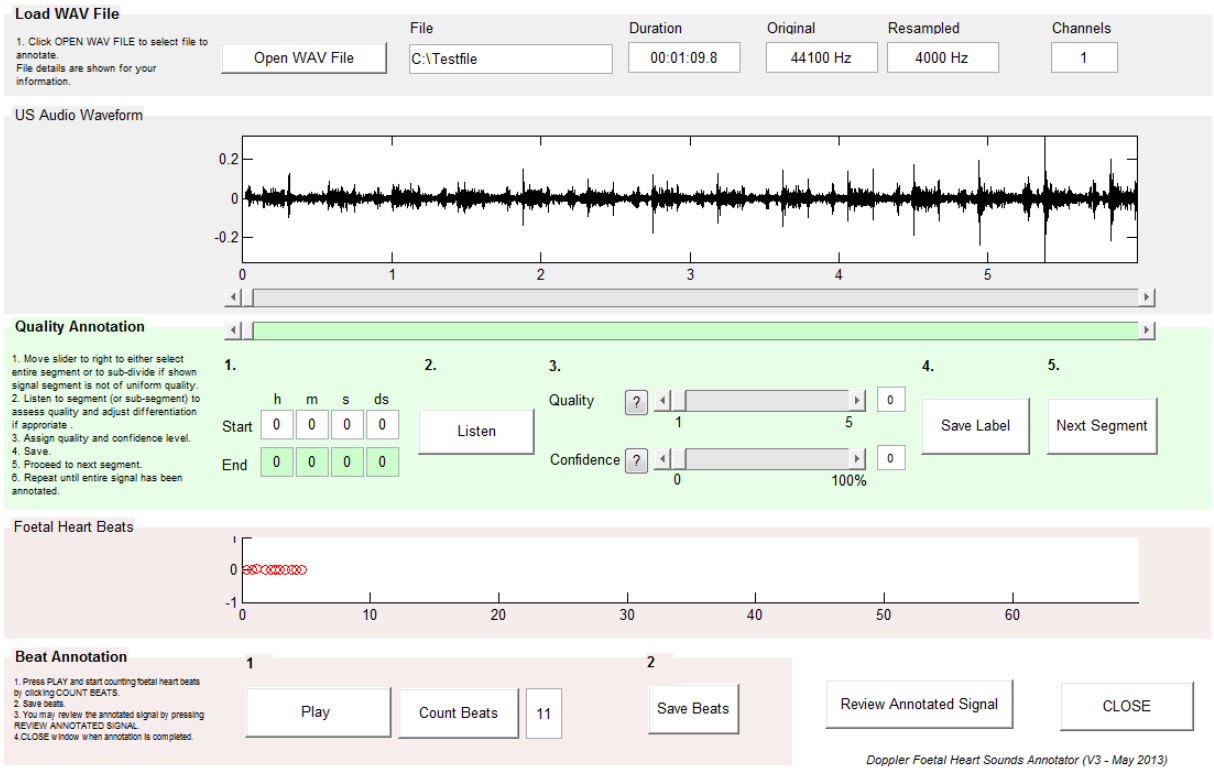
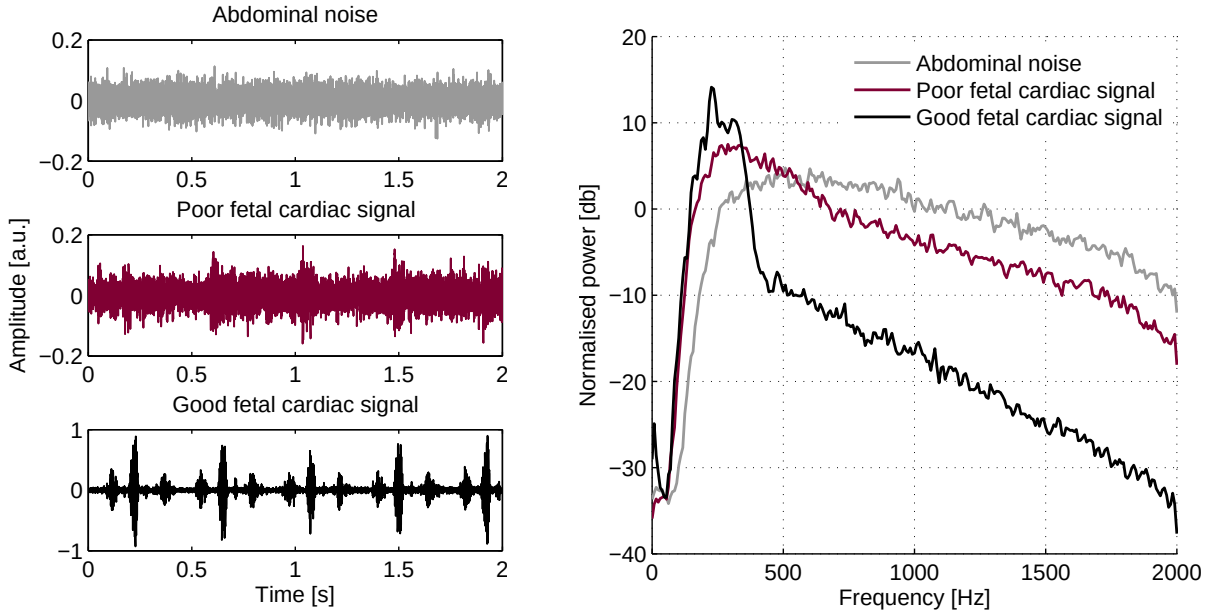


Figure 5.1: A bespoke Matlab annotation interface was programmed to facilitate the annotation of recorded Doppler signals. Signals can be reviewed both visually and aurally and a quality score assigned for selected segments. In a second step annotators were also asked to indicate beat locations by mouse click. Resulting beat annotations were collected as a reference signal and are discussed further in Chapter 6.

nonstress tests based on the Dawes/Redman criteria [108,127]. All episodes with complete agreement on quality between annotators were selected for inclusion into the classification data set. To investigate the classification challenge the data set was then split into groups of 'poor' and 'good' quality signals and data labels were allocated accordingly. 'Poor' quality epochs included epochs rated as 1, 2 or 3, 'good' quality epochs those that scored 4 and 5. Intermediate quality signals (score 3) may be ambiguous, however they contained predominantly barely audible beats and were therefore considered of poor quality. A total of 1043 poor and 439 good quality epochs were selected, the percentage of data contribution of each subject to the dataset being illustrated in Figure 5.3.



(a) (b)

Figure 5.2: Two second example recordings of abdominal noise (interference with the ultrasound beam), a poor and a good quality fetal cardiac signal are shown in (a). The example recordings' frequency content (the normalised power spectral density using Welch's method) is visualised in (b). Both (a) and (b) illustrate the temporal and frequency overlap of noise and regions of interest.

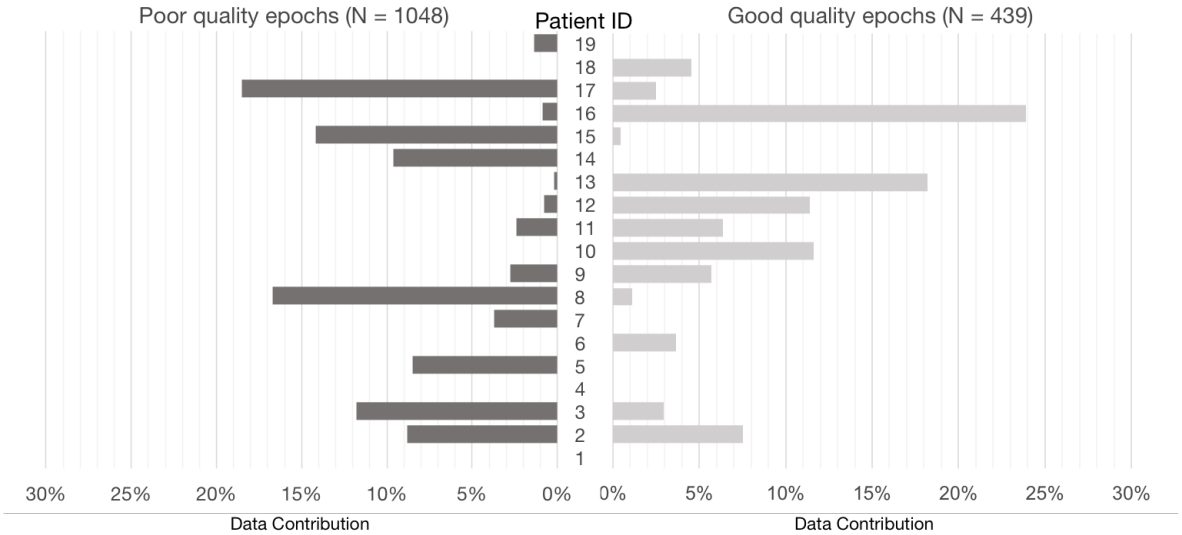


Figure 5.3: The percentage of data contributed by each patient to the poor and good quality data sets. 1048 poor and 439 good quality epochs were identified, which equates to approximately 65.5 and 27.4 minutes of recordings respectively.

5.4 Preprocessing

The ultrasound recordings were downsampled to 4000 Hz using a least-square linear phase anti-aliasing filter. The information content of fetal heart activity in Doppler signals has been observed in frequency ranges dependent on the moving structures [18]. Valve motion has been reported to produce the highest response, observed at around 1000 Hz with a 2 MHz transducer [128] which translates to 1650 Hz in this investigation given Equation 4.1, the transducer used (see transducer specifications in Section 4.2.2) and the assumption that the angle between direction of valve motion and the ultrasound beam is zero. The Nyquist-Shannon sampling criterion [129] is hence expected to be satisfied.

As outlined in Section 4.2.1, the average phone sampling frequency is 96 Hz. The phone however does not sample the accelerometer values at a constant rate; the accelerometer data were therefore resampled by linear interpolation to 100 Hz.

5.5 Feature extraction

Doppler audio signals are characterised by an often high and fluctuating Signal to Noise Ratio (SNR). Modality-extrinsic sources of interference are the movement of the probe against the skin, and maternal and fetal movements. The transducer used in this investigation operates in continuous Doppler mode (see also Section 4.2.2 for device details). In contrast to pulsed Doppler, the continuous probe constantly transmits and receives from its two crystals and therefore does not provide the ability for depth discrimination. Consequently the Doppler output contains information from every reflecting structure along its beam's path, which can also significantly increase the signal's noise levels. Changes in the position of the probe or the fetus alter the alignment of the ultrasound beam with the fetal heart, with the potential to change its morphology substantially. It is for this reason that the Doppler signal is therefore considered non-stationary.

Signals of good quality, however, clearly exhibit the relative periodicity of the fetal heart beats. The signal quality indices described below are based on the above-mentioned as-

sumptions, assessing the presence of reoccuring signal patterns, the levels of in utero noise contamination and the degree of external interference.

5.5.1 Sample Entropy features

Sample entropy can be used to analyse signals for reoccurring patterns to evaluate regularity. The sample entropy depends on three parameters, the length of the signal N, the pattern length m and the matching tolerance r. Sample entropy H_s is defined as the negative logarithm of the probability that a time series of length N, that has repeated itself for data points of length m within set tolerance r, will also repeat itself for $m+1$ data points within the same tolerance constraints, with self-matches not included in the probability computation [130].

There are no explicit guidelines on the choice of parameters m and r. Based on the assumption that the accuracy of the entropy measure increases with the number of pattern matches, a short template length is typically suggested. The selection of threshold r has been recommended as a fraction of the standard deviation of the individual time series in order to relate r to given time series achieving scale invariance and therefore allowing measurements on data sets with different amplitudes to be compared [130–133].

The sample entropy H_s was computed for the most widely used values, $m = 2$ and r between 0.1 and 0.25 times the standard deviation of the input time series $x(n) = x(1), x(2), ..., x(N)$, as follows [130]:

- For $1 \leq i \leq N - m + 1$, define template vectors $X_m(i) = [x(i), x(i+1), ..., x(i+m-1)]$, vectors of m consecutive values of x, starting at the ith sample.

- Compute the Chebyshev distance, also known as the maximum value distance, between $X_m(i)$ and $X_m(j)$, defined as $d[X_m(i), X_m(j)] = max_{k=0,...,k=m-1}(|x(i+k) - x(j+k)|)$, where j ranges from 1 to $N - m$, and $j \neq i$ to exclude self-matches.

- For each $X_m(i)$ count the number of j ($1 \geq j \geq N-m, j \neq i$) for which $d[X_m(i), X_m(j)] \leq$

r, denoted as B_i. For $1 \geq i \geq N - m$:

$$B_i^m(r) = \frac{1}{N-m-1} B_i \tag{5.1}$$

and $B^m(r)$, the probability that two sequences match for m points:

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{5.2}$$

- Increase the dimension to $m+1$ and repeat above computation to obtain A_i, the number of $X_m + 1(i)$ within r of $X_m + 1(j)$ $(1 \geq j \geq N - m, j \neq i)$:

$$A_i^m(r) = \frac{1}{N-m-1} A_i \tag{5.3}$$

and the probability that sequences match for length $m + 1$:

$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \tag{5.4}$$

- Compute the sample entropy from derived probabilities given as:

$$H_s(m, r, N) = -ln \frac{A^m(r)}{B^m(r)} \tag{5.5}$$

The method of computing sample entropy is illustrated in Figure 5.4.

Given that $A^m(r)$ is always lower than $B^m(r)$, their ratio will be smaller than 1, and the resultant entropy measure is thus always positive. With respect to the classification task, the sample entropy of signals which contain mostly or solely noise is expected to be high. Conversely it is expected to be low for time-series with clearly identifiable heart beats. The threshold parameter $r = 0.1$ x SD was selected based on the extent of separation between sample entropy distributions of poor compared to good quality data, assessed visually on the probability density estimates of training epochs, as shown in Figure 5.5

(The data split into training and test epochs are described further in Section 5.6.3).

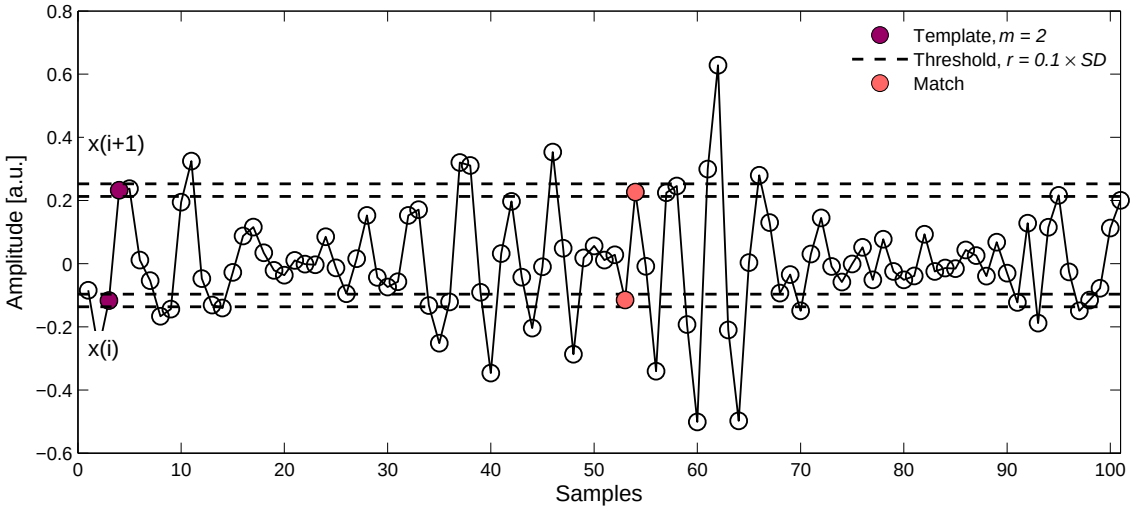


Figure 5.4: Illustration of the procedure for computing sample entropy for a template length of $m = 2$ and a threshold $r = 0.1 \times SD$. The template is shown in dark red, the matching pattern within threshold limits in light red.

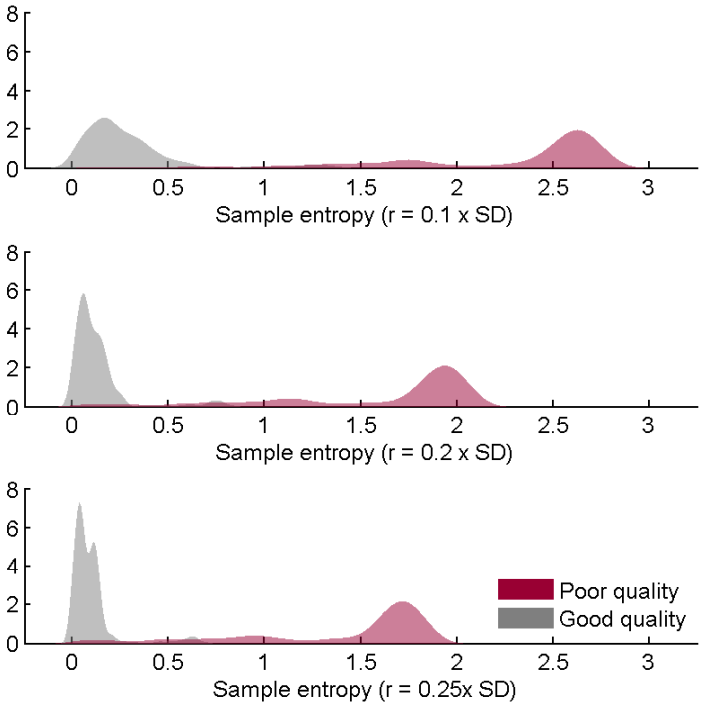


Figure 5.5: Probability density distributions for poor and good quality training epochs (Data split into training and test epochs are described in Section 5.6.3). The separation of the two distributions provides an indication of the feature's potential to distinguish poor from good quality data.

5.5.2 Wavelet features

Wavelet analysis is often described as an alternative to the short-term Fourier transform and particularly suited to the analysis of localised power within non-stationary signals, as it has simultaneously high temporal and frequency resolution [134,135]. The wavelet transform assesses the input signal's frequencies in different resolutions by shifting a scalable function along the signal, computing the spectrum at every position. This is repeated with shorter window lengths for every cycle resulting in a collection of time-scale representations of the signal (also referred to as multiresolution analysis).

Two types of wavelet transforms are typically considered for signal processing applications, the continuous and the discrete wavelet transform (CWT and DWT). The CWT provides finer sampling, however as a result is highly redundant; in the context of this investigation the DWT is assumed to provide sufficient information for analysis whilst being computationally far less intense [134,135]. When analysing a signal's energy content using the DWT the result is a set of coefficients and the process is referred to as wavelet decompositon.

The wavelet coefficients are computed through the convolution of the signal with high and low-pass filters that separate the frequency content into frequency bands of the same width. Each decomposition step halves the frequency bandwidth (doubles the frequency resolution) while its output is downsampled by a factor of two (time resolution is halved). Analysis follows a step-wise filter bank approach, the resultant sub-band containing the upper half of the frequency content yields the detail coefficients, the sub-band containing the lower half being referred to as the approximation coefficients. Each new level decomposes the approximation coefficients further, which allows the analysis of the frequency content at each scale of the wavelet decomposition.

The wavelet filters, associated with high-pass filtering, can be described as scaled versions of the chosen wavelet function $\psi(t)$, referred to as the mother wavelet [136]. The mother wavelet $\psi(t)$ in the DWT is shifted and scaled in discrete steps by powers of two (dyadic

sampling) resulting in:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t - k2^j}{2^j}\right) \tag{5.6}$$

where j is the scale parameter and k is the shift parameter, both of which are integers. A scaling function $\phi_{j,k}$ is used in the DWT to span the low frequencies of the signal [137]. The scaling function is associated with low-pass filtering.

The coefficients are computed as follows [135]:

$$y_{high}(k) = \sum_n g(n - 2k)x(n) \tag{5.7}$$

$$y_{low}(k) = \sum_n h(n - 2k)x(n) \tag{5.8}$$

where y_{high} denotes the detail coefficients, y_{low} the approximation coefficients, g the high-pass (wavelet) filter, h the low-pass (scaling) filter, and $x(n)$ the filter input. A diagram of the wavelet decomposition procedure is shown in Figure 5.6.

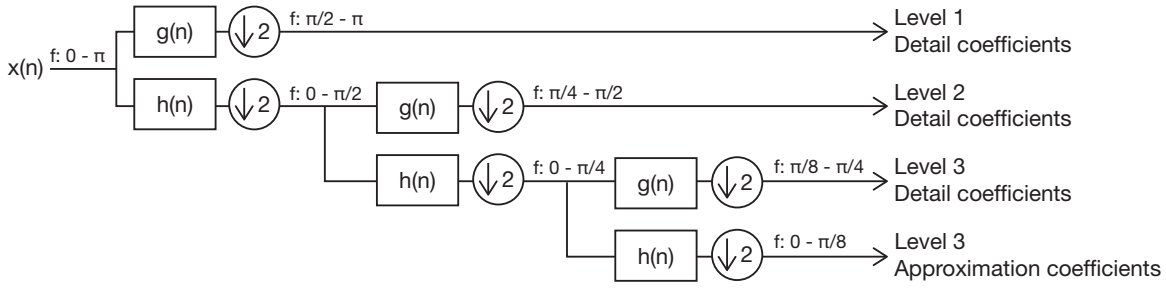
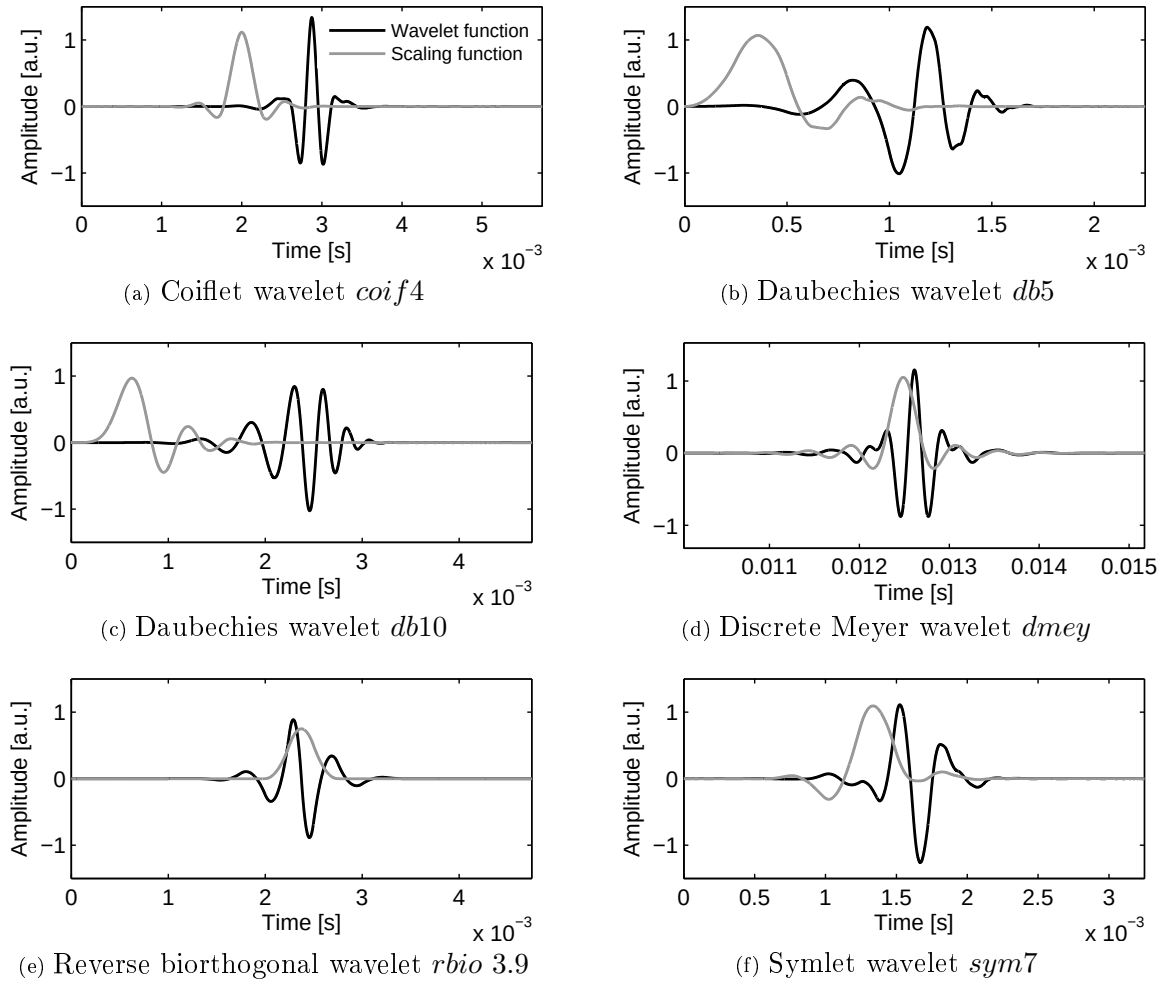


Figure 5.6: Shown is the procedure for a 3-level filter bank discrete wavelet decomposition, where $x(n)$ is the input signal, convolved with low-pass filter $h(n)$ and high-pass filter $g(n)$, followed by dyadic decimation (downsampling), to compute the approximation and detail coefficients respectively. The frequency band is denoted by f and expressed in radians.

The choice of mother wavelet influences the decomposition quality and is usually guided by 'similarity in shape' with the key features of the signal to be investigated. Based on their qualities and previous successful use working with similar signals (fetal phonocardiograms) [138] mother wavelets from the Symlet, Daubechies, Coiflet and Meyer

families as well as a reverse biorthogonal wavelet were assessed for suitability. The wavelet and scaling functions used in this preliminary exploration are shown in Figure 5.7.



(a) Coiflet wavelet $coif4$

(b) Daubechies wavelet $db5$

(c) Daubechies wavelet $db10$

(d) Discrete Meyer wavelet $dmey$

(e) Reverse biorthogonal wavelet $rbio$ 3.9

(f) Symlet wavelet $sym7$

Figure 5.7: Wavelet (black) and scaling (grey) functions selected for preliminary feature analysis. Wavelets were chosen based on 'shape similarity' with the signal of interest and their previous use in literature [138].

In order to select a sensible decomposition level information on the velocity ranges of the fetal myocardium were reviewed. During a cardiac cycle several heart structures are in motion at different times and at different velocities, which therefore manifest in the 1D Doppler recording at different frequencies. Peak axial cardiac wall velocities have been measured between 60 and 100 mm/s [18], which translates to a frequency range of 257-429 Hz, given the Doppler transducer used in this investigation (see transducer specifications in Section 4.2.2). Valve motions have been reported to produce a response in a

much higher range as observed by Raymond and Whitfield [128] (valve activity recorded at 1000 Hz with a 2 MHz transducer). Murata and Martin describe the use of a 600-2000 Hz bandpass filter to monitor valve activity [139], which suggests its presence in this frequency band. The frequency range of 600-2000 Hz translates to 990-3300 Hz in recordings investigated in this thesis.

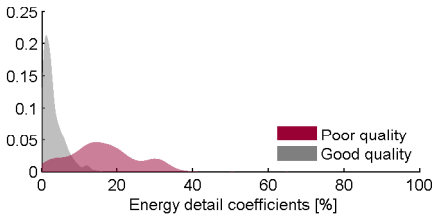
Taking reported observations as guidance, cardiac activity was expected to be located predominantly in frequency bands (decomposition levels) 3 and 4, spanning 500 Hz to 1 kHz (both wall and valve motion is expected to occur here) and 250 Hz to 500 Hz (wall motion is expected to occur here). The epochs selected for signal quality analysis were therefore decomposed to 4 levels, the energy content of each band being computed as input feature as follows:

- E_{dc}: The percentage of energy present in the respective frequency band (computed for detail coefficients on decomposition level 1, 2, 3 and 4), calculated as follows:

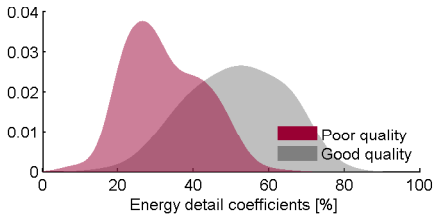
$$E_d(k) = 100\frac{\sum_{i=1}^{n} c_d(i)^2}{E_t} \tag{5.9}$$

  where k denotes the level of decomposition, $c_d(i)$ the detail coefficients computed at level k, and E_t the sum of the squared coefficients over all levels. The energy present on level 1 and 2 was expected to be greater in poor quality epochs due to high frequency noise interference compared to good quality epochs, where a high energy percentage was expected on levels 3 and 4 instead.

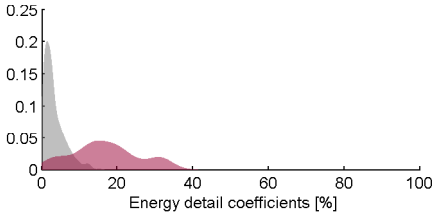
In order to select a mother wavelet the classification potential of features was reviewed based on their probability density estimate. The distributions of energy present on levels 1 and 3 are shown as an example in Figure 5.8 for assessed mother wavelets. The reverse biorthogonal wavelet *rbio*3.9 appeared to be most promising for levels 1 and 3, the Daubechies wavelet *db*5 for levels 2 and 4, given the distributions' earth mover's distance (EMD) [140] as the indicator for the feature's ability to separate good from poor epochs.



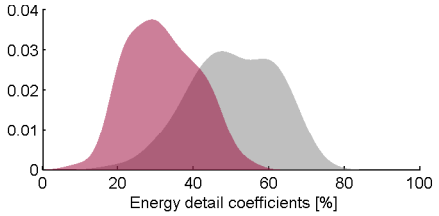
(a) *coif*4 level 1 - EMD: 9.7



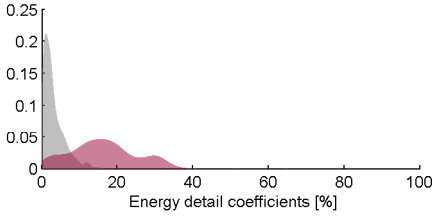
(b) *coif*4 level 3 - EMD: 16.3



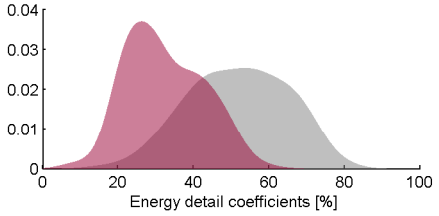
(c) *db*5 level 1 - EMD: 10.1



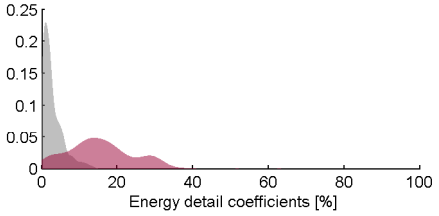
(d) *db*5 level 3 - EMD: 16.0



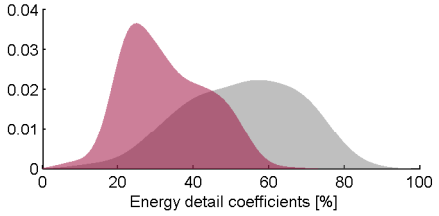
(e) *db*10 level 1 - EMD: 9.6



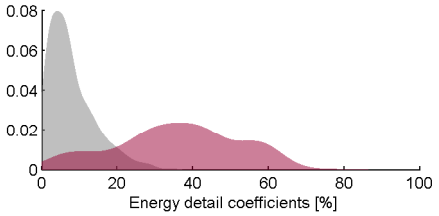
(f) *db*10 level 3 - EMD:16.2



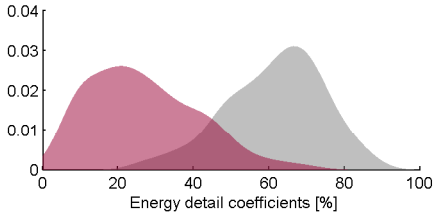
(g) *dmey* level 1 - EMD: 9.1



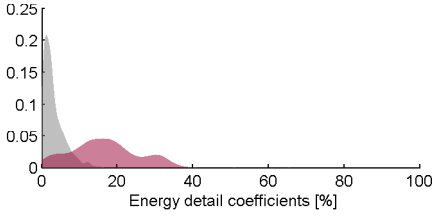
(h) *dmey* level 3 - EMD: 16.0



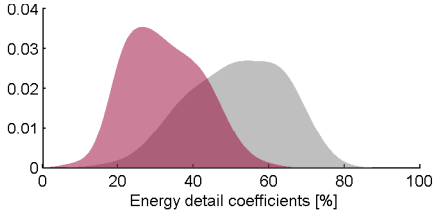
(i) *rbio* 3.9 level 1 - EMD: 22.0



(j) *rbio* 3.9 level 3 - EMD: 33.4



(k) *sym*7 level 1 - EMD: 9.8



(l) *sym*7 level 3 - EMD: 16.5

Figure 5.8: Probability density estimates for energy features computed for level 1 (left column) and level 3 (right column) of the wavelet decomposition (training data). The earth mover's distance (EMD) provides an indication for the feature's potential to separate poor from good quality data.

5.5.3 Accelerometer features

The loss of fetal heart sounds in the signal or a more general reduction in signal quality (decrease of SNR) can have several origins, some of which are related to movement of the probe. Probe movement may occur due to the operator searching for the fetal heart position or an unsteady probe position, whether hand-held or held in place on the abdomen by an elastic belt. The phone's triaxial accelerometer output was therefore recorded concurrently with the Doppler auscultation, with the assumption that there would be an increase in recorded acceleration when a decrease in quality occurs.

The accelerometer records both linear acceleration as well as the effect of the local gravitational field. Prior to deriving features the gravity component in each axis was eliminated. This was achieved by high pass filtering the acceleration in the x, y and z direction with a cut-off frequency of 2.5 Hz (finite impulse response filter). The distance travelled over the duration of an epoch was approximated by cumulatively integrating twice (trapezoidal method) to obtain velocity v from linear acceleration a_l:

$$v(t) = v_0 + \int_0^t a_l(t)dt \tag{5.10}$$

and ultimately displacement d:

$$d(t) = d_0 + \int_0^t v(t)dt \tag{5.11}$$

In order to eliminate drift the velocity was detrended prior to its integration.

The following features were computed for classification input:

- $A_{x,y,z}$: The composite linear acceleration in all three axes, computed as the sum of the absolute acceleration recorded in the x, y and z direction.
- $\widetilde{A}_{x,y,z}$: The median of the absolute value of the linear acceleration in all three axes combined.

- P_a: The percentage of the total linear acceleration in the x axis, P_{ax}, and z axis, P_{az}. Due to the manner in which the Doppler and phone assembly are operated, it can be assumed that the greatest difference between poor and good quality signals can be observed in the x and z direction (See Figure 5.9).

- $D_{x,y,z}$: The composite displacement in all three axes, computed as the sum of the absolute displacement computed in the x, y and z direction.

- P_d: The percentage of the total displacement in the x axis, P_{dx}, and z axis, P_{dz}. The manner in which the Doppler and phone assembly is operated suggests that the greatest difference between poor and good quality signals can be observed in the x and z direction (See Figure 5.9).

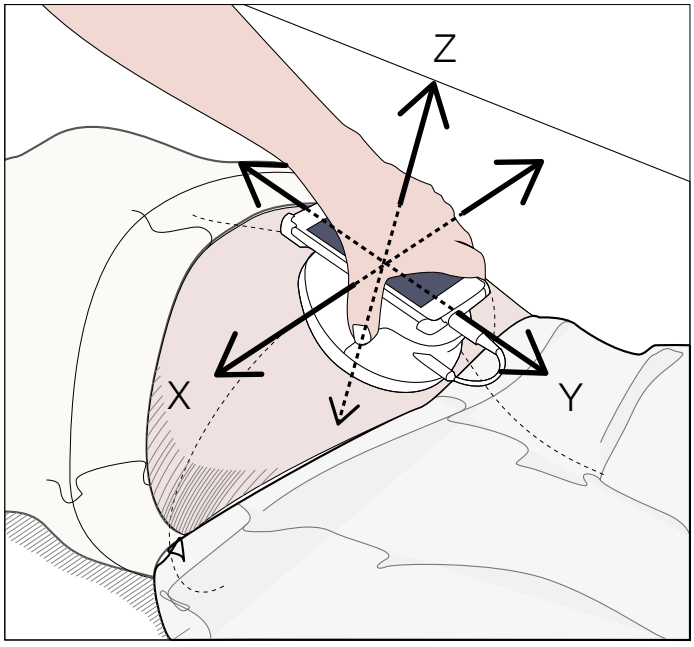


Figure 5.9: The Doppler and phone assembly is typically held as shown and moved over the maternal abdomen when searching for the fetal cardiac signal. Observation suggests that users move the probe primarily in the x-z plane given the way of holding the Doppler and the most likely search path (across the abdomen).

5.5.4 Feature selection

Appropriate feature selection may minimise classification error and avoid unnecessary computation time. The challenge of feature selection is to identify a subspace of the given feature set that optimally characterises the target classification variable. In this investigation the method applied to rank features was the *maximum Relevance Minimum Redundancy (mRMR)* algorithm that has been proven comparable to but more efficient than other approaches [141]. Optimal characterisation in mRMR refers to achieving maximal relevance (based on mutual information) balanced with minimal redundancy to exclude features that do not further improve the discriminative power of the classifier, as they are highly dependent on an already selected feature. An estimate of the optimal order of features was calculated on the training set.

For two variables x and y the mutual information is defined as [141]:

$$I(x;y) = \iint p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{5.12}$$

where $p(x)$, $p(y)$ and $p(x,y)$ are their probabilistic density functions.

The maximal relevance D and minimal redundancy R are then computed as follows:

$$max\ D(S,c),\ D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i, c) \tag{5.13}$$

$$min\ R(S),\ R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{5.14}$$

where S is the feature set, x_i and x_j are individual features, I is the mutual information and c is the target class. Optimising for maximal relevance D and minimal redundancy R simultaneously, $mRMR$ is denoted as:

$$max\ \phi(D,R),\ \phi = D - R \tag{5.15}$$

5.6 Classification

Logistic regression (LR) is one of the most widely used methods for binary classification problems that are linearly separable. Logistic regression is a computationally simple method, relatively robust to noise, and straightforward to interpret given that its output is expressed as a probability. Considering the anticipated implementation of the classification algorithm on a mobile phone, logistic regression appears to be a sensible choice given its low complexity. However, an alternative method that has shown some advantages over logistic regression, such as being superior in high correlation data and data with non-linear relationships, as well as requiring fewer features to achieve equivalent results, is the Support Vector Machine (SVM) [142]. In addition the solution to an SVM optimisation problem is global and unique (convex optimisation problem). Logistic regression was previously defined in Section 3.3.3. The SVM classifier is introduced in the following section. Results for both methods are compared thereafter to assess their suitability for a given classification task.

5.6.1 Support vector machine classifier

An SVM's underlying principle is to map input vectors to a higher dimensional feature space using a non-linear function ϕ, turning a linear classifier into a non-linear one. Provided with a training set of pairs (xi, yi), $i = 1, ..., n$, where $xi \in R^n$ are the observation and $y_i \in \{1,\text{-}1\}$ are the associated class label, training of an SVM constructs a hyperplane in the feature space separating the two classes with equal margin [143,144]. The optimal hyperplane is given by $\mathbf{w^T x} + b = 0$, where $\mathbf{w}$ is the vector of hyperplane coefficients and b a bias term (see Figure 5.10a). Given the scenario where classes are not separable without error, the optimisation problem (maximising the margin) is to minimise the

following, introducing the slack variable ξ, where $\xi_i \geq 0$, $i = 1, ..., n$:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^\mathbf{T}\mathbf{x}_i + b) \geq 1 - \xi_i \end{aligned} \tag{5.16}$$

The parameter C governs the trade-off between maximising the margin and minimising the slack. A small value of C ignores data points close to the decision boundary whilst a larger C value results in a higher error penalty (see Figure 5.10b). The mapping in a SVM classifier occurs with the help of a kernel function ensuring that dot products can be computed easily. The (Gaussian) Radial Basis Function (RBF) used in the binary SVM classifier (libSVM library [145]) was chosen here based on its successful use in related work [146]. The RBF kernel is defined by $K(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2)$, where γ ($\gamma > 0$) influences the flexibility of the classifier by adjusting the width of the Gaussian (see Figure 5.10c). Low values of γ lead to near linear boundaries whilst increasing values render boundaries more adaptive; high values may hence lead to over-fitting [147].

An SVM using the RBF kernel therefore requires the setting of two parameters, the penalty parameter C and the kernel parameter γ. The best generalisation performance is achieved if there is a balance between the accuracy generated with a specific training set of finite size and the SVM's capacity to learn any training set without error. If the capacity is too high the SVM will be too restrictive, if too low it will be too lax.

In the following classification experiments the parameters were optimised in a 5-fold cross-validation grid search on the training data set. The grid space was defined by C, $C \in \{2^{-3}, 2^{-1}, ..., 2^5\}$, γ, $\gamma \in \{2^{-5}, 2^{-3}, ..., 2\}$ and the number of features [148].

5.6.2 Statistical measures

The performance of both LR and SVM classifiers were evaluated by resultant accuracy Ac, sensitivity Se and specificity Sp, defined as follows:

- Ac: The proportion of correct decisions given the total number of recordings.

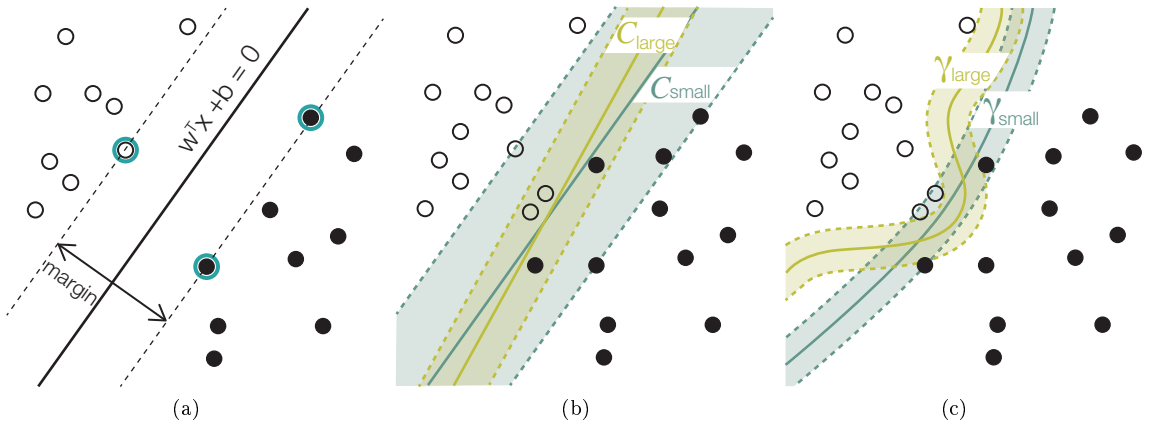


Figure 5.10: Figures (a), (b) and (c) illustrate the construction of a hyperplane (decision boundary) and its margin, and the effect of the SVM and kernel parameters on the decision boundary, given a two class-classification problem. The filled and empty circles represent the datasets of the two classes. (a) The decision boundary (solid black line) and its margin (dotted black lines), determined by the support vectors (circled blue) that are closest to the boundary, on the example of a linearly separable classification problem. (b) The effect of the penalty parameter C on the decision boundary. Whereas a large C increases the penalty error for data points close to the boundary (green boundary and margin), a small value allows for them to be ignored (blue boundary and margin). (c) The effect of the parameter γ on the decision boundary, given a fixed C value. An increased γ value increases the boundary's flexibility.

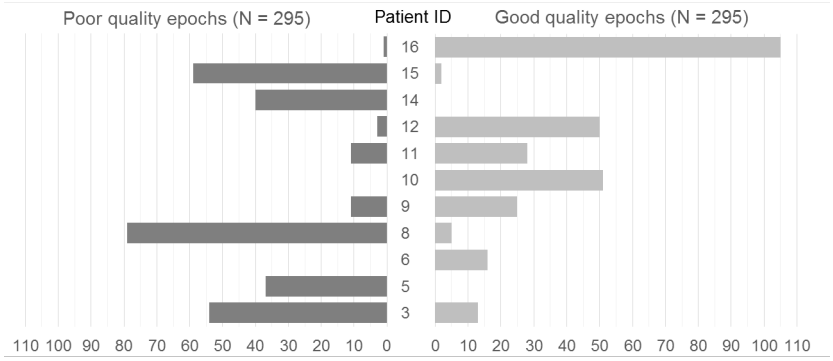
- *Se*: The proportion of true positive decisions.
- *Sp*: The proportion of true negative decisions.

5.6.3 Balancing and normalising datasets

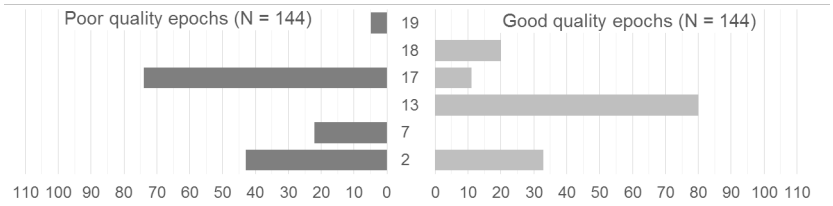
For unbiased classification the data were randomly split into a training and a test set in the approximate ratio of 2:1, stratified by patients. Each set was balanced for an equal number of 'good' and 'poor' quality epochs. Training data included 590 epochs from 11 patients, held-out test data 288 epochs from 6 patients. The contribution of data from each subject to either training or test set is illustrated in Figure 5.11.

The quality covered by epochs ranged from no beats audible in the signal to excellent with clearly identifiable beat instances. The feature matrices were normalised prior to classification. The features in the training set were normalised by subtracting the mean

of the respective feature vector and dividing by its standard deviation. For training input X, the output is therefore $(X - \mu)/\sigma$. The test set was normalised using both mean and standard deviation derived from the training data.



(a) Training set



(b) Test set

Figure 5.11: The training (a) and held out test set (b) selected randomly with an approximate split of 2:1, stratified by patients.

5.6.4 Classification performance assessment

Three experiments were conducted in order to assess the quality of both LR and SVM classifiers based on individual features and feature combinations (see Figure 5.12).

Experiment 1: Models were built for single feature input and evaluated in a 5-fold cross-validation over several iterations on the training data. The performance of individual Doppler and accelerometer based features provided insight into their classification potential and aided pre-selection of metrics to include into subsequent multi-feature classifiers.

Experiment 2: Features derived from the fetal cardiac signal were ranked using mRMR and successively added to train and assess classifiers for respective feature combinations. The step-wise inclusion of features allowed for the assessment of their contribution to

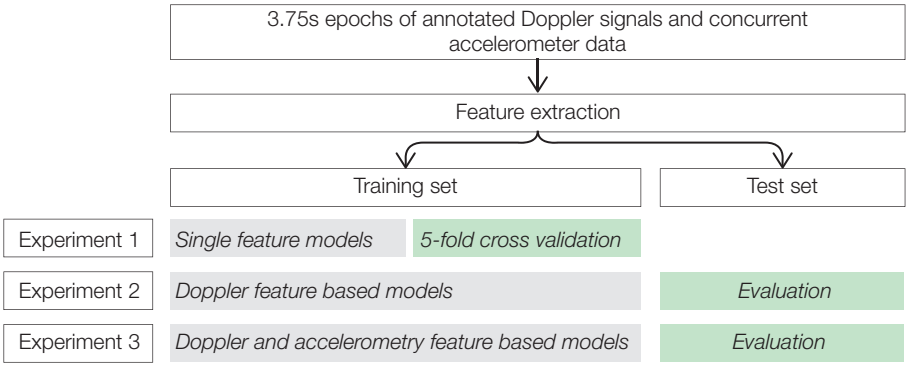


Figure 5.12: The quality of single- and multi-feature classifiers for the discrimination of good and poor fetal Doppler recordings was assessed in three separate experiments.

the classifier's performance. The multi-feature LR and SVM models were trained on the training set and evaluated on the held-out test set.

Experiment 3: Both cardiac as well as selected accelerometer features were ranked using mRMR and successively added accordingly to train and evaluate resultant LR and SVM classifiers. Experiment 3 was conducted to review the potential of using indicators of Doppler probe movement to improve classification performance.

To assess the statistical difference in accuracy estimates between models the McNemar 2-sided mid-p test was used [149]. The McNemar test is a paired non-parametric significance test appropriate for evaluating statistically significant change between two tests on the same data and was therefore considered suitable for comparing the outcomes of described classification experiments.

5.7 Results

In the first experiment all features were tested for their performance discriminating poor and good quality epochs using LR and SVM classifiers that were then compared. Accuracy was computed in a 5-fold cross-validation over 150 iterations to ensure a stable estimate. 150 iterations were concluded to be sufficient as the standard deviation of mean accuracy values had converged. The SVM penalty parameter C and RBF kernel parameter γ were optimised in the 5-fold cross-validation of the first iteration for each feature

		H_s	$E_{dc}L1$	$E_{dc}L2$	$E_{dc}L3$	$E_{dc}L4$
LR	μAc	<u>94.06</u>	88.08	86.04	88.76	72.96
	95% CI	93.73-94.38	87.79-88.37	85.71-86.37	88.32-89.21	72.61-73.31
SVM	μAc	<u>94.21</u>	88.29	85.91	88.71	74.17
	95% CI	94.05-94.37	88.01-88.56	85.30-86.52	88.24-89.17	73.61-74.73
	C	8	32	8	8	0.125
	γ	2^{-5}	2^{-5}	2^{-5}	2	2^{-3}

Table 5.3: Experiment 1a: Classification performance of single features computed from the cardiac signal, the sample entropy H_s and the energy E_{dc} present on decomposition levels 1 to 4 (assessed on training data). The mean accuracy and 95% confidence interval are reported, based on 5-fold cross validation repeated over 150 iterations. An underscore indicates the highest accuracy result for training and test data respectively.

		$\widetilde{A}_{xyz}$	A_{xyz}	D_{xyz}	P_{ax}	P_{az}	P_{dx}	P_{dz}
LR	μAc	58.86	58.47	57.59	<u>60.92</u>	56.74	59.50	58.93
	95% CI	58.31-59.40	58.06-58.89	56.91-58.27	60.36-61.49	56.14-57.33	58.93-60.06	58.44-59.42
SVM	μAc	60.91	60.99	<u>63.43</u>	61.35	57.44	59.50	59.95
	95% CI	59.57-62.25	59.75-62.22	62.40-64.47	60.50-62.21	55.76-59.12	58.76-60.25	58.57-61.34
	C	32	32	2	0.5	0.5	0.5	32
	γ	2^{-1}	2^{-1}	2	2^{-5}	2^{-1}	2^{-5}	2^{-3}

Table 5.4: Experiment 1b: The classification performance of single features computed from the accelerometer signal (assessed on training data). The mean out of sample accuracy and 95% confidence interval are reported, based on 5-fold cross validation repeated over 150 iterations. An underscore indicates the highest accuracy result for training and test data respectively.

individually. The mean accuracy with 95% confidence intervals for features derived from the cardiac and accelerometer signals are reported in Table 5.3 and Table 5.4 respectively.

The best performing single feature in both LR and SVM classification was sample entropy H_s ($m = 2$, $r = 0.1 \times SD$) achieving a mean accuracy of 94.06% with LR and a marginally higher accuracy of 94.21% using SVM.

H_s was significantly superior compared to both wavelet as well as accelerometer based features. Whereas the energy in the different frequency bands still proved of high discriminating power, stand-alone accelerometry features showed poor classification performance (Table 5.4). The percentage of acceleration in x, P_{ax}, achieved the highest LR classification accuracy with 60.92%; the cumulative distance travelled in all axes, D_{xyz}, produced the best SVM classification accuracy with 63.43%. Provided with the accuracies achieved on the held-out validation data for each individual feature, as well as the

correlation between features, only the most promising acceleration based metrics were included in further analysis ($\widetilde{A}_{xyz}$, D_{xyz} and P_{ax}).

The combination of fetal cardiac features only, successively adding features according to mRMR ranking, was assessed in the second experiment. The SVM parameters were optimised prior to fitting each model to the training data.

Using LR the highest accuracy achieved with features derived from the cardiac signal on the training data was 96.78%, leading to 93.40% accuracy on the test data. Using a SVM classifier the highest training accuracy improved to 99.15%, the same model achieving 96.18% on the test data. Comparing classifier performances it is noteworthy that the 2-feature LR model [H_s, $E_{dc}L3$] performed equivalent or superior to higher-feature combinations with models showing no statistical difference (McNemar test [149] mid-p value > 0.05). Including all wavelet features in the SVM, however, improved test results (94.10% compared to the 5-feature model accuracy 96.18%, McNemar test mid-p value $= 0.039$). The classification results of cardiac feature combinations are summarised in Table 5.5 for both training and test data.

Adding accelerometry features in the third experiment increased LR accuracy on the training data to 98.81%, achieving 95.14% on the test set. The highest accuracy using a SVM classifier was 98.98%, however, this reduced the test accuracy to 95.14% suggesting overfitting. The best test result was close to LR with 95.83% (training accuracy 98.31%) and no improvement over cardiac features only. The classification results of cardiac and accelerometer feature combinations are summarised in Table 5.6.

The overall best performance on the test set using a SVM classifier was 96.18%, based on all five cardiac input features. The best performance on the test set using a LR classifier was 95.41%, based on three cardiac as well as accelerometer features. The logistic function for this quality measure is given by:

$$logit(p) = 1.10 + 4.69x_1 - 3.18x_2 + 1.43x_3 \tag{5.17}$$

	Features	Training Ac	Se	Sp	Test Ac	Se	Sp		
LR	H_s, $E_{dc}L3$	96.61	95.59	97.73	<u>93.40</u>	86.81	100.00		
	H_s, $E_{dc}L3$, $E_{dc}L1$	<u>96.78</u>	95.93	97.72	<u>93.40</u>	86.81	100.00		
	H_s, $E_{dc}L3$, $E_{dc}L1$, $E_{dc}L4$	96.61	95.93	97.39	93.06	86.11	100.00		
	H_s, $E_{dc}L3$, $E_{dc}L1$, $E_{dc}L4$, $E_{dc}L2$	<u>96.78</u>	96.27	97.39	92.71	85.42	100.00		
	Features	Ac	Se	Sp	Ac	Se	Sp	C	γ
SVM	H_s, $E_{dc}L3$	98.64	97.29	100.00	94.10	88.19	100.00	32	2
	H_s, $E_{dc}L3$, $E_{dc}L1$	98.81	97.63	100.00	94.10	88.19	100.00	32	2
	H_s, $E_{dc}L3$, $E_{dc}L1$, $E_{dc}L4$	98.98	98.31	99.67	93.75	87.50	100.00	32	2^{-1}
	H_s, $E_{dc}L3$, $E_{dc}L1$, $E_{dc}L4$, $E_{dc}L2$	<u>99.15</u>	98.64	99.67	<u>96.18</u>	92.36	100.00	2	2

Table 5.5: Experiment 2: The classification performance of feature combinations computed from the cardiac signal only. Features were successively added according to mRMR ranking. The classification accuracy, sensitivity and specificity are reported for both the training data and the held-out test set. An underscore indicates the highest accuracy result for training and test data respectively.

where $x = (x_1, x_2, x_3)$ are the sample entropy (H_s), wavelet ($E_{dc}L3$) and accelerometer (P_{ax}) observations respectively.

5.8 Discussion

Sample entropy of the ultrasound segment was shown to be superior compared to single wavelet and acceleration features for classifying Doppler recording quality. This may be expected given that sample entropy indicates the presence or absence of reoccurring heart beats, the most fundamental characteristic of a successful recording. The percentage of energy in the frequency band 500 Hz-1 kHz (wavelet decomposition level 3), anticipated to contain both cardiac wall and valve motion, showed the highest discriminatory potential compared to remaining bands. Combining both features could raise the accuracy of LR and SVM classifiers. The combination of sample entropy with multiple wavelet features did not lead to a significant further improvement of LR. Adding all wavelet features to the SVM classifier on the other hand improved test results.

Whereas adding accelerometer features showed an increase in test accuracy for LR, this was not the case for SVM models. The class of noisy signals was highly heterogeneous with respect to causes of poor quality. Transducer movement clearly only accounted for

	Features	Training Ac	Se	Sp	Test Ac	Se	Sp		
LR	H_s, $E_{dc}L3$	96.61	95.59	97.73	93.40	86.81	100.00		
	H_s, $E_{dc}L3$, P_{ax}	97.63	97.63	97.68	<u>95.14</u>	90.28	100.00		
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$	97.46	97.29	97.69	<u>95.14</u>	90.28	100.00		
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$	<u>98.81</u>	98.64	99.00	<u>95.14</u>	90.97	99.36		
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$, $E_{dc}L2$	<u>98.81</u>	98.64	99.00	<u>95.14</u>	90.97	99.36		
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$, $E_{dc}L2$, D_{xyz}	98.47	98.31	98.67	91.32	88.89	94.38		
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$, $E_{dc}L2$, D_{xyz}, $E_{dc}L4$	98.47	98.31	98.67	92.01	90.28	94.30		
	Features	Ac	Se	Sp	Ac	Se	Sp	C	γ
SVM	H_s, $E_{dc}L3$	98.64	97.29	100.00	94.10	88.19	100.00	32	2
	H_s, $E_{dc}L3$, P_{ax}	97.97	97.29	98.68	94.44	89.58	99.37	8	2^{-1}
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$	98.31	97.63	99.01	<u>95.83</u>	91.67	100.00	32	2^{-3}
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$	<u>98.98</u>	98.31	99.67	95.14	91.67	98.72	2	2^{-3}
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$, $E_{dc}L2$	98.81	98.31	99.33	94.44	89.58	99.37	8	2^{-5}
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$, $E_{dc}L2$, D_{xyz}	<u>98.98</u>	98.31	99.67	93.06	89.58	96.86	8	2^{-5}
	H_s, $E_{dc}L3$, P_{ax}, $E_{dc}L1$, $\widetilde{A}_{xyz}$, $E_{dc}L2$, D_{xyz}, $E_{dc}L4$	<u>98.98</u>	98.31	99.67	94.10	89.58	98.74	32	2^{-5}

Table 5.6: Experiment 3: The classification performance of combinations of all selected features. Features were successively added according to mRMR ranking. The classification accuracy, sensitivity and specificity are reported for both the training data and the held-out test set. An underscore indicates the highest accuracy result for training and test data respectively.

a fraction of noise interference and therefore added additional information to only some epochs. Both experiments 2 and 3 showed eventually a stagnation of, followed by a decrease in, test accuracy with successively added features. This may be explained by the strong correlation of wavelet features and the occurrence of overfitting the models to the training data.

Specificity, the rate of correctly identifying good quality epochs, was found to be higher than sensitivity in all experiments, which again may be explained with the more homogeneous signal signature of good quality epochs compared to that from poor quality epochs. The SVM models were generally superior to the LR models using the same features and the SVM indeed could cope better with redundant features (experiment 2). Superiority however, was marginal. With the combination of best performing sample entropy H_s, wavelet $E_{dc}L3$ and accelerometer P_{ax} features, LR reached an out of sample accuracy of 95.14%, statistically not different from the overall best SVM test result (96.18%) requiring 5 features (McNemar test mid-p value > 0.05).

5.9 Conclusion

Recording Doppler signals of good quality onto a mobile phone and accurately discriminating poor from good epochs were successfully demonstrated here. Accelerometer data as a quality index are a novel parameter, not derived from the ultrasound signal directly. Its potential to improve the correct discrimination between good and poor quality data was apparent in the LR classification results. The percentage of acceleration in the x-axis showed most promise and is most likely related to the influence of hand-held recording. As the poor quality class of epochs was highly heterogenous in signal characteristics, it is not surprising that the accelerometer signals could explain poor quality only to some extent. It would be of interest to explore the use of accelerometer features further, with a view to indicating complete signal loss caused by device movement.

Results demonstrated that all three assumed characteristics for signal discrimination (differences in device movement, signal periodicity and frequency distribution) contributed to the classification task. The top three features selected by the mRMR approach included one feature from each of the three groups of quality indices (derived from the signal's sample entropy, the wavelet coefficients and the phone's accelerometer data).

Given the simplicity of the LR and the lower number of features (Equation 5.17), together with the statistical equivalence to the 5 feature SVM, the former is implemented and used in future work.

Given the limited number of annotators, the input for classification was controlled for epochs of homogeneous quality and 100% annotator agreement in order to restrict inter-observer variability. Relaxing these criteria may reduce the accuracy of the system to classify data quality, however the presented results give confidence that good accuracy can be achieved. Every attempt was made to cover a wide range of signal quality issues in the data used to develop presented SQI. Issues of poor quality include cardiac signal loss, varying levels of interference with the ultrasound beam and noise introduced by probe movement. However, provided the limited number of patients (N=17) available, not all

scenarios could possibly be represented. Continued data collection and annotation will allow for a more robust and extensive database for future exploration.

The definition of acceptable data quality might differ according to the diagnostic task and could be further investigated.

The underlying dataset to develop the presented signal quality measure did not contain any fetal cardiac arrhythmia, which may result in abnormal rhythms being identified and excluded as noisy. For the application of IUGR detection this may be considered favourable, as heart rate traces computed from pathological arrythmic signals do not adequately reflect the working of the fetal autonomic nervous system. For a general implementation of Doppler quality assessment, however, it would be valuable to train the signal quality index over different heart rhythms to ensure abnormal rhythms are not disregarded as noisy signals, resulting in a system that classifies everyone as normal.

Chapter 6

Development and evaluation of a hidden Markov model approach to fetal cardiac signal segmentation on 1D ultrasound data

6.1 Introduction

The detection of IUGR based on Doppler-derived heart rate variability metrics has shown promise in fetuses affected by early-onset growth retardation (see Chapter 3). Commercial fetal ultrasound monitors employ autocorrelation algorithms to estimate the fetus' heart rate (see Chapter 2), which is the basis for the computation of variability metrics. Both variability features (short- and long-term variability) central to the detection algorithm described in Chapter 3 were computed from a correlation-based heart rate trace. The drawback of autocorrelation methods is their averaging nature. The diagnostic potential of heart rate variability markers has been assumed to be greater with improved accuracy of the variation in the cardiac signal, as can be achieved with a fetal electrocardiograph or with magnetocardiography [89].

It can be argued that the Doppler signal is inherently unsuitable to provide true beat-to-beat measurements due to the nature of the monitoring modality. In contrast to the electrocardiogram, which records the electrical activity that originates from the de- and repolarisation of the heart muscle during each beat, Doppler-based modalities record the mechanical activity of the heart, which includes the valve or heart wall motion that constitutes each beat. The Doppler signal morphology may therefore vary depending on the angle of the fetal heart with respect to the ultrasound beam and can do so throughout the length of a recording (typically between 10 and 60 minutes). However, auscultation Doppler remains the most widespread and affordable monitoring modality in use to date. This has motivated the effort to develop a method for 1D Doppler beat-to-beat interval estimation described in this chapter, with the aim to ultimately improve on the modality's risk assessment potential.

A physiological signal closely related in its characteristics to the 1D Doppler signal is the phonocardiogram. A phonocardiogram is an acoustic recording of cardiac activity by an electronic stethoscope. Both the fetal cardiac Doppler signal and the phonocardiogram are non-stationary in nature, have dynamic spectral characteristics and are a representation of the same sequential physiological process, the cardiac heart cycle.
One of the primary concerns when processing phonocardiograms to identify clinically relevant information is the accurate identification of the first heart sound (S1 - originating from the closing of the mitral and tricuspid valves), the systolic period, the second heart sound (S2 - originating from the closing of the pulmonary valves), and the diastolic period; all four together constituting one cardiac cycle. The successful use of hidden Markov Models in the analysis of time-sequential patterns, most notably in speech processing, has led to several examples of its application in phonocardiographic heart sound segmentation [150–153]. HMMs are a probabilistic approach which are well-suited to the time-series analysis of signals whose properties successively change. They have been reported to outperform other heart sound segmentation methods, such as sound segmentation based on envelograms in combination with peak detection [154,155].

The positive results presented when applied to phonocardiograms has motivated the use of hidden Markov Models to segment the fetal 1D Doppler signal in order to accurately identify beat-to-beat intervals. This chapter is structured as follows: the data selection and pre-processing steps are laid out, followed by an introduction to hidden Markov Models and the case-specific implementation thereof. Algorithm components such as the selected features and estimated duration probabilities are then explained and the reference signal described. The chapter is finally concluded with an evaluation of the algorithm's performance in epoch heart rate and beat-to-beat interval estimation.

6.2 Database

Work in this chapter is based on the dataset (*JR RDDB*) collected at the John Radcliffe Hospital, Oxford, using the mHealth Doppler system as described in Chapter 4. The dataset contains approximately 2h of raw Doppler data from 17 healthy patients, manually annotated for both quality and heart beat instances (see database characteristics under Section 1.3). The data collection procedure and patient characteristics were summarised previously in Section 5.2.

6.3 Data selection and preprocessing

For the purpose of algorithm development and evaluation the data were segmented into epochs according to their signal quality. An epoch duration was chosen to be 3.75 seconds, the standard heart rate analysis window used in cCTG machines [108]. To assess the impact of signal quality on the detection accuracy epochs were grouped into the classes 2 ('bad'), 3 ('intermediate'), 4 ('good') and 5 ('excellent') according to manual annotations. An epoch was assigned to a class if at least two of the annotators agreed on the quality within the window and the third disagreed by no more than one quality point. Data that did not meet these criteria and data belonging to class 1, which is defined as containing

no cardiac information, were excluded. A detailed description of the manual quality annotation procedure is given in Section 5.3.

6.3.1 Spike removal

Noise spikes of selected epochs were removed by a method described by Schmidt et al. [151], whereby samples greater than a certain threshold within the analysis window are set to zero. The threshold was chosen to be three times the median of absolute values within the analysis window; the window size was selected to contain a minimum of one beat. The minimum heart rate was set to be 80 bpm, equal to a beat interval of 750 ms, which is the physiological limit based on thresholds of normality [63].

6.3.2 Frequency range of interest

To assess the frequency range of interest the frequency content of collected data was analysed and insights reviewed against findings in literature. The expected fetal cardiac information contained within the Doppler signal is composed of blood flow, cardiac wall and valve motion (see Figure 2.2 and Table 2.1 for an illustration and description of the cardiac cycle). The dominant phenomena in the waveform are expected to result from cardiac tissue motion, which has lower velocity, but which results in higher intensity signals compared to blood flow (as fluids reflect only minimal Doppler energy). Tutschek et al. [156] found that peak cardiac wall velocities, ventricular motion in particular, could be recorded from early on in gestation (12 weeks onwards) by tissue Doppler echocardiography, whereas the detection rate of valve motion increased with gestational age. Based on this observation the assumption was made that cardiac wall motion will be most likely present in the majority of data collected and will constitute the observable beat representation.

The Doppler shift was previously defined in Equation 4.1. To reiterate, f_D is the Doppler shift frequency (Hz), f_o is the frequency of the ultrasound beam (Hz), V is the tissue

velocity (m/s), and c denotes the speed of sound in soft tissue (m/s).

The frequency emitted by the handheld transducer (f_o) used to record the data described in this chapter was 3.3 MHz [157]. The speed of sound in human tissue (c) against which standard-compliant ultrasound machines are calibrated is 1540 (m/s) [158].

The axial cardiac wall velocities of the right and left ventricle during systole were reported to be $V_{RV} = 0.017x^2 - 0.5944x + 9.0522$ and $V_{LV} = 0.009x^2 - 0.2104x + 5.0742$ respectively, where x is the gestational age in weeks [156]. The expected velocities were computed for the gestational age range in the database, for both the right and left ventricle. Based on those velocities and the constants f_o and c the frequency range f_{Dmin} to f_{Dmax} was estimated given the largest and smallest angle at which the ultrasound beam could meet the direction of wall motion ($\theta = 0°$ and $\theta = 90°$). The estimated lower and upper frequencies given the gestational age of the fetus are shown in Figure 6.1. Figure 6.2 provides a comparison with the frequency content found at the point of highest intensity in each training epoch (see Figure 6.3 for an example, and section 6.3.3 for a description of the training data set). The velocities and data relevant minimum and maximum frequency cut-offs are summarised in Table 6.1.

Variables	RV	LV	RV	LV
GA [weeks]	20		40	
V [m/s]	0.04	0.04	0.12	0.11
f_{Dmin} ($\theta = 90°$) [Hz]	**0**	**0**	**0**	**0**
f_{Dmax} ($\theta = 0°$) [Hz]	**170**	**191**	**535**	**474**

Table 6.1: Expected frequency content originating from cardiac wall contraction, more specifically the right and left ventricle (RV and LV).

To remove unwanted low- and high-frequency noise all epochs were bandpass filtered with cut off values of 25 and 600Hz (zero-phase FIR equiripple bandpass filter). The frequency band was selected as the frequency content analysis of the available data and the computation of expected frequencies (see Figure 6.2 and Table 6.1) suggested that

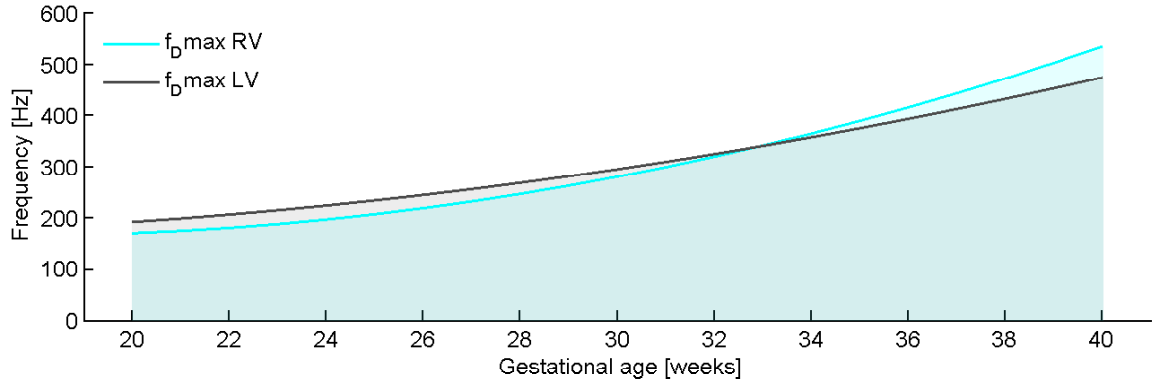


Figure 6.1: Expected frequencies in Doppler signals originating from right (RV) and left (LV) ventricular motion, dependent on gestational age; Frequencies at an angle approximating 90° ($f_D min$) are zero, $f_D max$ are frequencies expected at an angle approximating 0°, plotted here as the upper boundary of expected frequency content. The curves are computed using Equation (4.1), constants c and f_0, and tissue velocity estimates by Tutschek at al. [156].

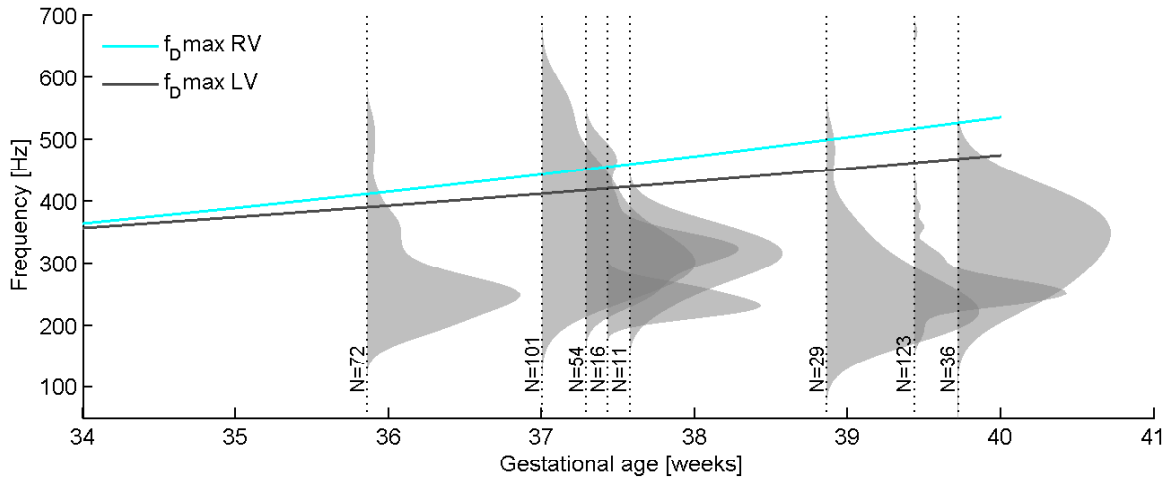


Figure 6.2: The frequency content at the highest point of intensity in each training epoch was measured, the shaded areas represent the distribution at respective gestational age at time of recording (gestational age range in training set 35+6 to 39+5 [weeks+days]). The distributions are compared for agreement to the expected frequency range ($f_D min$ = 0 to $f_D max$ for the left (grey curve) and right (turquoise curve) ventricular motion) derived from literature (visualised in Figure 6.1).

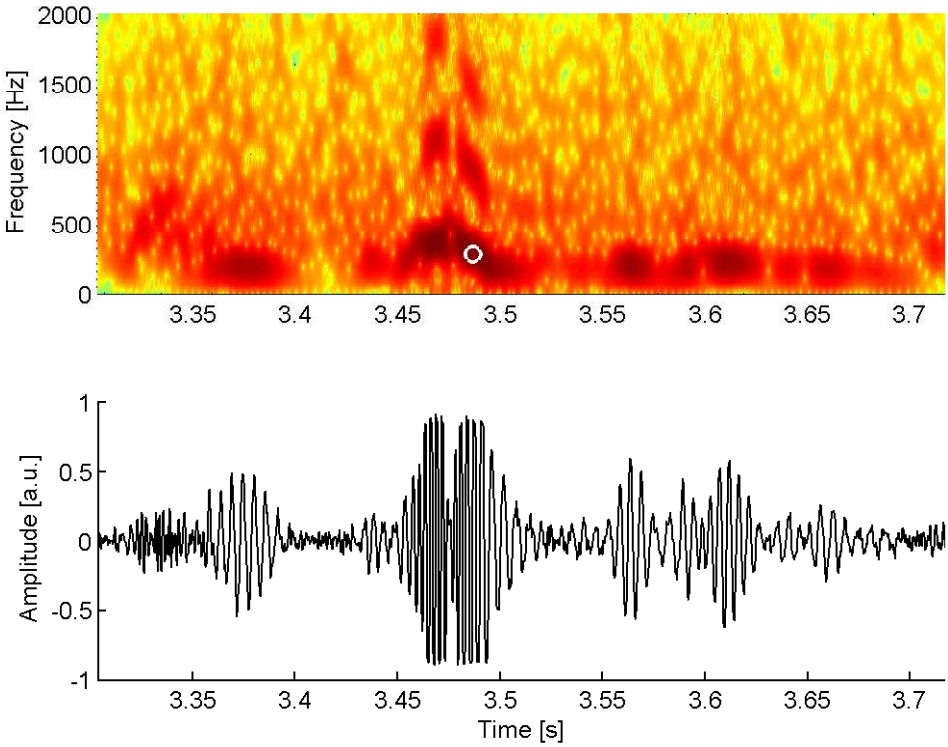


Figure 6.3: Short-time Fourier transform (STFT) spectogram of approximately one fetal cardiac cycle (upper plot) and corresponding time series (lower plot). In each training epoch the point of highest intensity within the spectogram was selected (white circle) and its frequency recorded as an example of the frequency content to be expected of cardiac tissue motion (STFT used with 10 ms window length and overlap of one sample at 4 kHz). The waveform in the time domain allowed visual verification that the selected sample coincided with a cardiac event and that the segment was of sufficient quality.

the dominant cardiac oscillations are found below 600 Hz. Removing low frequencies is thought to minimise some of the disturbance introduced by fetal movement, while high frequencies were removed to reduce the effect of blood flow.

6.3.3 Balancing the dataset

For unbiased performance evaluation the selected and pre-processed signal epochs were split into a training and test set. Training data were selected from good quality epochs only. Approximately two thirds of epochs identified by annotators as of quality 4 ('good') and 5 ('excellent') were randomly allocated to the training, the remainder to the test set, ensuring no recordings from the same patient were in both training and test sets. The test set also contained all epochs of qualities 2 ('bad') and 3 ('intermediate'). To reiterate, epochs of quality 2 are epochs that annotators found to be extremely noisy, however containing at least some cardiac events. Epochs of quality 3 are epochs found to be noisy, however, where fetal beats could be identified with some effort. Epochs of quality 1, containing no cardiac events, were excluded entirely. The individual subjects' contributions to training and test sets is visualised in Figures 6.4 and 6.5.

6.4 Hidden Markov Models

A hidden Markov model (HMM) is an extension of a discrete Markov process or chain. The output of a recorded process is no longer a set of states that result from an observable, physical event, but is a probabilistic representation of the state [159]. The model is therefore called a hidden Markov model, also described as a doubly-embedded stochastic process. In other words, using the example of fetal cardiac monitoring: the physical cardiac event, the motion of cardiac tissue and valves, cannot directly be observed, instead a representation of the cardiac activity is recorded in the form of the Doppler waveform. An HMM is built to explain recorded observations and therefore to uncover desired information, in this context the underlying cardiac states.

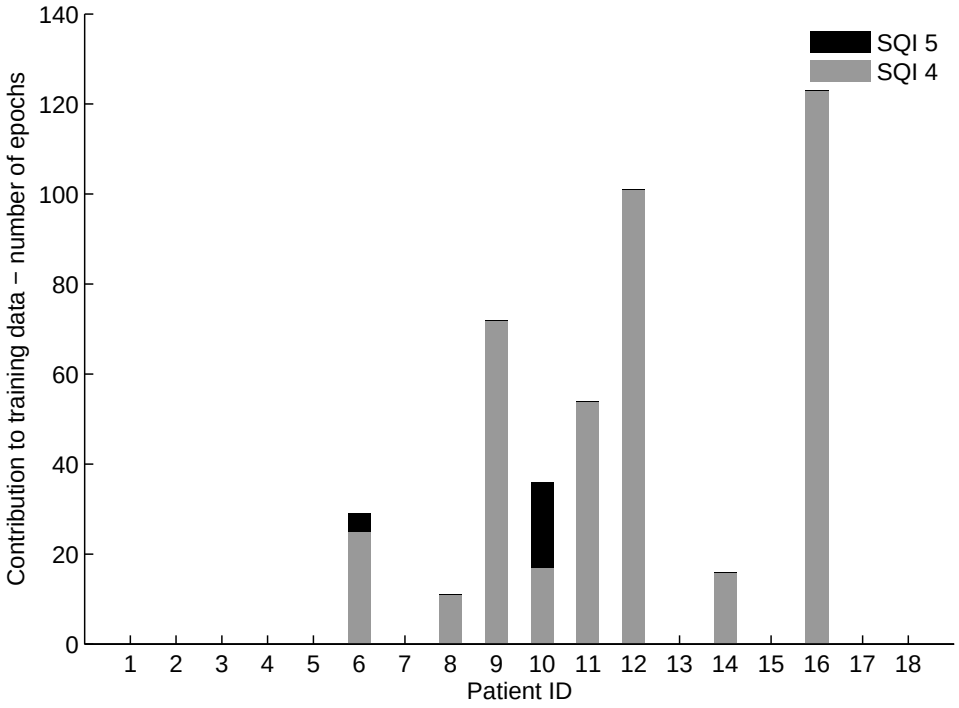


Figure 6.4: The contribution of data from each patient to the training set. Epochs were grouped according to their quality label assigned by three independent annotators. The training set contained a total of 442 3.75 second epochs (28 minutes).

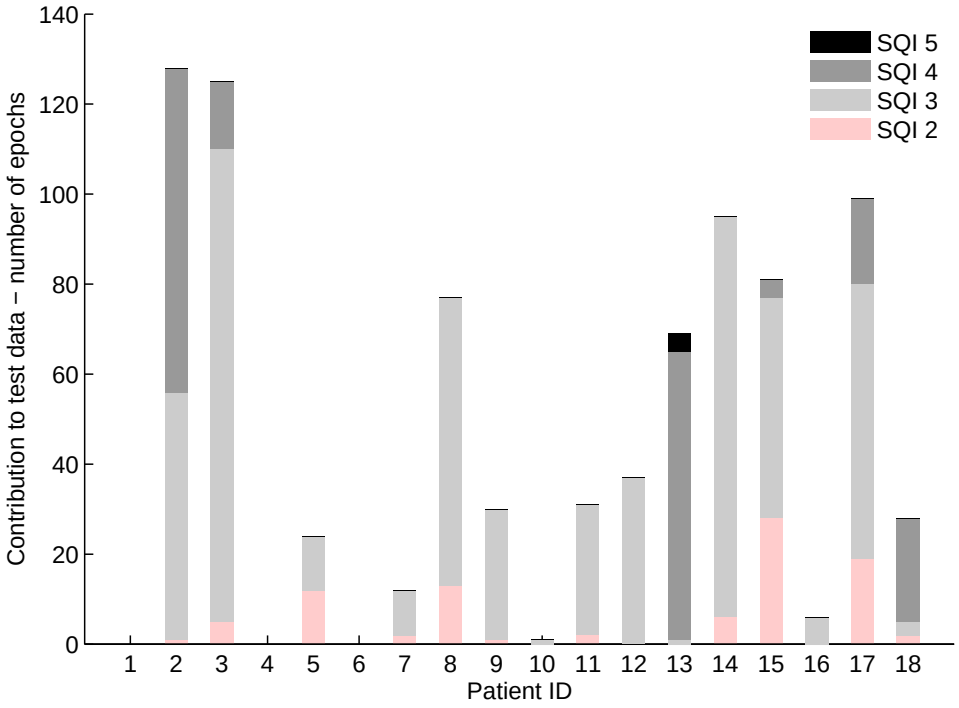


Figure 6.5: The contribution of data form each patient to the test set. Epochs were grouped according to their quality label assigned by three independent annotators. The test set contained a total of 201 epochs of good quality (labelled 4 and 5), 551 epochs of intermediate quality and 91 epochs of poor quality (a total of 53 minutes).

An HMM is governed by [159]:

- N - The number of states in the model. The number of states typically corresponds to the events to be uncovered, in this case the periodic activity of the fetal heart. States are denoted $S = S_1, S_2, ..., S_N$, the state at time t is q_t.
- M - The range of possible observation values; values are denoted $V = v_1, v_2, ..., v_M$.[1]
- A - The state transition probability distribution, an N x N state transition matrix, denoted as $A = a_{ij}$ where $a_{ij} = P[q_{t+1} = S_j | q_t = S_i]$; $1 \leq i, j \leq N$.
- B - The emission probability distribution, i.e. the probability that the observation belongs to a certain state j denoted $B = b_j(k)$, where $b_j(k) = P[v_k \text{ at } t | q_t = S_j]$, $1 \leq j \leq N$ and $1 \leq k \leq M$.
- π - The initial state distribution, $\pi = P[q_1 = S_i]$, $1 \leq i \leq N$.

In summary an HMM requires the identification of the two model parameters N and M and the three distributions A, B, and π, conventionally noted as:

$$\lambda = (A, B, \pi) \tag{6.1}$$

Given the changes in morphology throughout a Doppler signal and the uncertainty about which valve or wall motion will be present in each cycle, the number of states N was set to two states, the first to represent the dominant beat event and the second the intervals between beat events.

Multiple observations $O = O_1, O_2, ...O_t$ (feature vectors) were computed from the recorded Doppler signal, described in more detail in Section 6.5. In contrary to a discrete case of a standard HMM, observation values were continuous in this investigation and therefore

[1]The definition provided here applies to the standard discrete HMM in which the number of possible observations is finite. HMM observations may also be continuous (often from a Gaussian distribution) as true for this investigation. The implications are further described in Sections 6.4.2 and 6.4.3.

governed by Gaussian distributions.

The model parameters λ were initially unknown and had to be optimised to maximise the probability of the observations O given the model. To iteratively maximise $P(O|\lambda)$ the Baum-Welch method was employed [159,160]; the method and how it was adapted to the context of this work is described in Section 6.4.2. The purpose of this investigation was ultimately to identify the state sequence which best explains the observations O recorded. The common method to find the single optimal state sequence is by use of the Viterbi algorithm [159,161]. The Viterbi algorithm and how it has been adapted for the purpose of reliably uncovering fetal cardiac events is thus described below in Section 6.4.3.

6.4.1 Hidden Semi-Markov Model

The conventional HMM has an inherent duration probability density $p_i(d)$, which for state S_i and self-transition probability a_{ii} is written $p_i(d) = (a_{ii})^{d-1}(1 - a_{ii})$. In this model the probability of remaining in a state decreases exponentially with time. For the analysis of cardiac activity this is unsuitable. The cardiac cycle is governed by reoccurring events of durations greater than a single time-step. The standard model can be extended by introducing state duration probabilities. The probability of remaining in a state is then no longer governed by the self-transition probabilities (they are set to zero) but by the duration densities. The adapted model is called a Hidden Semi-Markov Model (HSMM). HSMMs have been successfully used in speech recognition [162] and heart sound segmentation [151]. The difference between an HMM and HSMM is illustrated in Figure 6.6 [159].

6.4.2 Training HMM parameters - The Baum-Welch algorithm

The Baum-Welch method iteratively computes the maximum likelihood for the observations O given an HMM model $\lambda = (A, B, \pi)$:

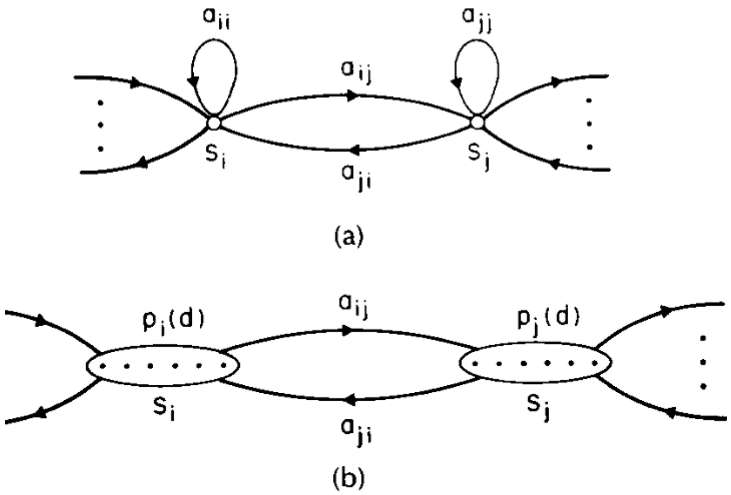


Figure 6.6: (a) The standard HMM with exponential state duration densities based on the self-transition coefficients. S_i and S_j denote the states, a_{ii} and a_{jj} are self-transition probabilities and a_{ij} and a_{ji} the probabilities to transition from one state to the other. (b) An HMM with no self-transitions but specific state duration densities, denoted by $p_i(d)$ and $p_j(d)$. Adapted from [159].

$$arg \max_{\lambda} P(O|\lambda) \tag{6.2}$$

The three primary data characteristics that had to be considered for the implementation of the Baum-Welch method were: the input of multiple observation vectors, their continuous variable values, and the input of independent training segments. Three feature vectors were derived from a Doppler recording (see Section 6.5), which meant that the algorithm had to be able to process multiple observation inputs. Furthermore, the features computed from the Doppler signal are not discrete in value, but continuous variable observations. The emission probability densities were therefore to be governed by a Gaussian multivariate density function, denoted in the remainder of this chapter as B_{pdf}. The training data consisted of several independent epochs from a number of different subjects, which required an additional modification to process individual segments. This is further described below. Based on [159] and [163] the following was implemented:

1. Initialise input parameters:

- The state transition matrix A, initialised at $A = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$, under the stochastic constraint that $\sum_{j=1}^{N} a_{ij} = 1, 1 \leq i \leq N$;
- The initial state probabilities π, initialised at $\pi = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$ in the assumption that one is equally likely to observe either state, under the stochastic constraint that $\sum_{i=1}^{N} \pi_i = 1$;
- The mean μ of each Gaussian distribution, initialised at the mean of the respective feature vector with added random noise, and respective standard deviation σ, initialised at the standard deviation of the respective feature vector for both states.

2. Set iteration parameters:
   - The maximum number of iterations, empirically set to 300;
   - The convergence tolerance, empirically set to 1.0e-8, against which the change in model log-likelihood was compared.

3. Compute the emission probabilities for each feature vector:

$$b_j(o_t) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{(o_t - \mu_j)^2}{2\sigma_j^2}} \tag{6.3}$$

   Where μ_j and σ_j denote the mean and standard deviation associated with respective state.

4. Compute forward probabilities α for each observation sequence: the probability of seeing partial observation sequence $o_1, o_2, ..., o_t$ and ending in state i at t:

$$\alpha_i(1) = \pi_i b_i(o_1) \tag{6.4}$$

$$\alpha_j(t+1) = [\sum_{i=1}^{N} \alpha_i(t) a_{ij}] b_j(o_t + 1) \tag{6.5}$$

5. Compute backward probabilities β for each observation sequence, the probability of the partial observation sequence from $t+1$ to the end (time T), given state S_i at time t and the model λ:

$$\beta_i(T) = 1; \tag{6.6}$$

$$\beta_i(t) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_j(t+1) \tag{6.7}$$

6. To enable training on data of multiple independent segments, the forward and backward algorithm is run on each training segment with α_{seg} and β_{seg} concatenated to create α and β respectively [164].

7. Compute ξ_{ij}, the probability of the joint event that the model is in state S_i at time t and S_j at time $t+1$:

$$\xi_{ij}(t) = \frac{\alpha_i(t) a_{ij} b_j(o_{t+1}) \beta_j(t+1)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij} b_j(o_{t+1}) \beta_j(t+1)} \tag{6.8}$$

8. Compute γ, the probability of being in state S_i at time t, given the observation sequence and the model:

$$\gamma_i(t) = \frac{\alpha_i(t) \beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t) \beta_j(t)} \tag{6.9}$$

9. For E observation sequences calculate the update equation for π:

$$\pi_i = \frac{\sum_{e=1}^{E} \gamma_i^e(1)}{E} \tag{6.10}$$

10. For E observation sequences calculate the update equation for A:

$$a_{ij} = \frac{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \xi_{ij}^e(t)}{\sum_{e=1}^{E} \sum_{t=1}^{T_e} \gamma_i^e(t)} \tag{6.11}$$

11. For E observation sequences calculate the update equation for each μ:

$$\mu_i = \frac{\sum_{e=1}^{E}\sum_{t=1}^{T_e}\gamma_i^e(t)o_t^e}{\sum_{e=1}^{E}\sum_{t=1}^{T_e}\gamma_i^e(t)} \tag{6.12}$$

12. For E observation sequences calculate the update equation for each σ:

$$\sigma_i = \frac{\sum_{e=1}^{E}\sum_{t=1}^{T_e}\gamma_i^e(t)(o_t^e-\mu_i)(o_t^e-\mu_i)^T}{\sum_{e=1}^{E}\sum_{t=1}^{T_e}\gamma_i(t)} \tag{6.13}$$

6.4.3 Uncovering the optimal state sequence - The Viterbi algorithm

With the model parameters estimated, the challenge remains to identify the optimal state sequence that best describes given observations, also called the inference problem in hidden Markov models [159]. The problem can be expressed as:

$$arg\max_{Q} P(Q|O,\lambda) \tag{6.14}$$

where Q is the state sequence with the maximum probability of being the state path that produced given observations. Stating the problem as finding a single most probable state sequence allows for the application of the Viterbi algorithm [161,165], a dynamic programming solution that has been successfully applied to this problem in related domains, such as speech and heart sound processing [151,159]. Dynamic programming is an approach applicable to problems that can be divided into a series of subproblems that are solved, stored and ultimately combined to an overall solution [166]. Such a method has already been used to find the forward, α, and backward, β, variables in Equations 6.5 and 6.7.

The Viterbi algorithm has been implemented here according to Rabiner [159], including

the introduction of duration probabilities and an adaptation to account for edge effects further described below.

To find the optimal path $Q = q_1, q_2, ..., q_t$ the Viterbi algorithm requires definition of the variable $\delta_t(i)$:

$$\delta_t(i) = \max_{q_1,q_2,...,q_{t-1}} P[q_1 q_2 ... q_t = i, O_1 O_2 ... O_t | \lambda] \tag{6.15}$$

which expresses the probability of the best single sequence at time t, accounting for the first t observations and ending in state S_i. By maximising according to the previous state i, δ at time step $t+1$ can be written:

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] b_j(O_{t+1}) \tag{6.16}$$

To back-track the optimal state sequence the argument which maximises 6.16 needs to be stored. In order to do so a second quantity, ψ_t, needs to be defined:

$$\psi_t(j) = arg \max_i [\delta_t(i) a_{ij}] \tag{6.17}$$

As mentioned previously, given multiple observations the emission probabilities are governed by normal distributions. The emission densities were therefore computed as k-dimensional Gaussian multivariate density function B_{pdf} in the Viterbi implementation as follows:

$$b_j(o_t) = \frac{1}{\sqrt{|\Sigma_j|(2\pi)^k}} e^{-\frac{1}{2}(o_t - \mu_j)^\top \Sigma^{-1} (o_t - \mu_j)} \tag{6.18}$$

Where μ_j and Σ_j denote the mean and covariance matrix associated with the respective state.

To avoid the problem of under- or overflow [159], the algorithm has been implemented using logarithms as follows:

1. Initialisation: The algorithm is initialised by computing the probability of the model being in state i at time $t = 1$ and generating observation O_1. $\delta_1(i)$ and $\psi_1(i)$ become:

$$\delta_1(i) = log(\pi_i) + log(b_i(O_1)), 1 \leq i \leq N \tag{6.19}$$

$$\psi_1(i) = 0 \tag{6.20}$$

2. Recursion: The self-transition probabilities are set to zero in order to introduce state duration dependency, as described earlier in Section 6.4.1. The transition matrix is therefore re-written as $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. The standard HSMM implementation works under the 'simplifying' assumption that the first state starts at time $t = 1$ and the last state ends at $t = T$ [167]. This assumption does not hold in the context of the described investigation. Instead the 'general' assumption has been implemented whereby the first state can start before or with observation o_1 and the last state may end at or after the last observation o_T [167]. The recursion step therefore takes into account a time window of the maximum state duration D_{max} [2] before and after the first and last observation respectively.

   The transition to the next state is estimated by maximising over the duration d and the previous state i:

$$\delta_t(j) = arg\max_d arg\max_i [\delta_{t-d}(i) + log(a_{ij}) + log(p_j(d) + \sum_{s=1}^{d} log(b_j(O_{t-s}))] \tag{6.21}$$

   For $2 \leq t \leq d$, where $t - d < 1$ the previous state is represented by $\delta_1(i)$; where $t - d + 1 \leq 1$ the observation probability is replaced by $b_j(O_1)$.

   For $T + 1 \leq t \leq T + d - 1$, where $t - d > T$ the previous state is represented by $\delta_T(i)$; where $t - d + 1 \geq T$ the observation probability is replaced by $b_j(O_T)$. For each t the duration D_t and the preceding state ψ_t which maximise $\delta_t(j)$ are stored:

$$D_t(j) = d \tag{6.22}$$

[2] The maximum state duration is the maximum duration of the primary oscillation or interval, whichever greater, as estimated for each input signal. The state durations are further explained in Section 6.6.

$$\psi_t(j) = i \tag{6.23}$$

3. Termination: Before the state path can be back-tracked the time T_{max} and state $q_{T_{max}}$ of maximum δ is identified in the added maximum duration window D_{max}.

$$arg \max_{T_{max}=t} arg \max_{q_{T_{max}}=i} [\delta_t(i)] \quad , T \leq t \leq T + D_{max} - 1 \tag{6.24}$$

4. Path back-tracking: The optimal state path q_t is then back-tracked from the identified time point of maximum δ given the information provided by D_t and ψ_t. The back-tracking algorithm was implemented as follows:

Assign the last state in the signal:
$t = T_{max}$
$t_{onset} = t - D_t(q_t) + 1$ Find the onset of the state given the stored state duration
$q_{t_{onset}:t} = q_t$ Assign the state for associated period to the path
$q_t = \psi_t(q_t)$ Update the state to preceding state

Back-track through the remainder of the signal:
while $t_{onset} > 1$
  $t = t_{onset} - 1$ Update end of preceding state
  $t_{onset} = t - D_t(q_t) + 1$ Update onset of preceding state given associated duration
  if $t_{onset} < 1$ or $t = 1$
    $t_{onset} = 1$ Update onset of preceding state
    $q_{t_{onset}:t} = q_t$ Assign state to path
  else
    $q_{t_{onset}:t} = q_t$ Assign state to path
    $q_t = \psi_t(q_t)$ Update state to preceding state
  end
end

6.5 Signal features

The selection of input features is critical for the performance of a system and can be considered more important than the choice of algorithm.

The cardiac Doppler signal contains important information in both the time and frequency domains. Cardiac events typically manifest themselves in higher amplitudes in the 1D Doppler signal; the signal envelope representing the signal's amplitude component can

therefore provide relevant information on event location. The homomorphic envelope is a suitable choice given its ability to smooth out jagged peaks and has been implemented with good results in heart sound analysis [151,152,168]. Analysis based solely on envelope extraction in the time domain, however, introduces the risk of missing low-energy events or erroneously identifying noise spikes as of cardiac origin. Given the non-stationary character of the signal, frequency analysis is ideally localised in time. Two additional envelope features in the time-frequency domain have been computed: a wavelet envelope and the power-spectral density envelope.

6.5.1 Homomorphic envelope

The amplitude signal was extracted by homomorphic filtering (with order 120, zero-phase, FIR low pass filter, cut-off 8 Hz) in combination with the Hilbert transform as in [169]. The Doppler signal is considered a multiplicative system, consisting of a slow-varying amplitude component and a fast varying component responsible for the signal oscillations, and can thus be written as:

$$x(t) = a(t) \times o(t) \tag{6.25}$$

where $a(t)$ represents the amplitude component and $o(t)$ the signal oscillations. Its logarithmic transformation can be written as:

$$ln|x(t)| = ln|a(t)| + ln|o(t)| \tag{6.26}$$

which can be low-pass filtered using aforementioned lowpass filter to extract the low frequency component which is then exponentiated to form the envelogram. The envelope was normalised by subtracting its mean and dividing by its standard deviation to avoid inter-epoch variability.

6.5.2 Wavelet envelope

Wavelet analysis was considered as it captures both frequency and location information in time. To compute the wavelet envelope the discrete wavelet transform (DWT) was used, as previously introduced in Section 5.5.2.

Wavelet analysis requires the selection of both the mother wavelet and the number of frequency sub-bands to assess (the decomposition level). The optimal wavelet and the decomposition level were selected experimentally. The wavelet families investigated were those suitable for discrete wavelet analysis available in the Matlab toolbox: Daubechies wavelets ('*db*N', assessed: *db*1-*db*10), Symlets ('*sym*N', assessed: *sym*2-*sym*8), Coiflets ('*coif*N', assessed: *coif*1-*coif*5), biorthogonal wavelets ('*bior*Nr.Nd', assessed: *bior*1.1-*bior*6.8), reverse biorthogonal wavelets ('*rbio*Nr.Nd', assessed: *rbio*1.1-*rbio*6.8) and the discrete approximation of the Meyer wavelet ('*dmey*'). N determines the number of vanishing moments responsible for the smoothness of the respective function (the same applies for Nr and Nd in biorthogonal wavelets, 'r' denoting 'reconstruction', 'd' denoting decomposition) [170]. Example members of all listed wavelet families are included in the previous chapter in Figure 5.7, with the exception of the biorthogonal wavelet family, an example of which is added here for completeness (see Figure 6.7).

The primary cardiac oscillation and the period to the next oscillation were annotated by the author in one good quality example beat from all patients. The sample Doppler recordings were decomposed into approximation and detail coefficients using the discrete wavelet transformation testing all wavelets listed above and decomposition levels 1-7 (given that the Doppler signal has a sampling frequency of 4 kHz, the 7th level contains frequencies between 31.25 and 62.5 Hz; no information of interest below this band is expected as illustrated in Figure 6.2). The wavelet and decomposition level that produced the highest ratio of 1) the sum of the oscillation detail coefficients to 2) the sum of the detail coefficients of periods between oscillations was selected. The ratio gives an indication of how well cardiac activity can be distinguished from periods in between those regions of interest.

The wavelet envelope was computed from the absolute value of the detail coefficients using wavelet *rbio*3.9 and decomposition level 3. The selected wavelet function with associated scaling function and high- and low-pass filters are shown in Figure 6.8)

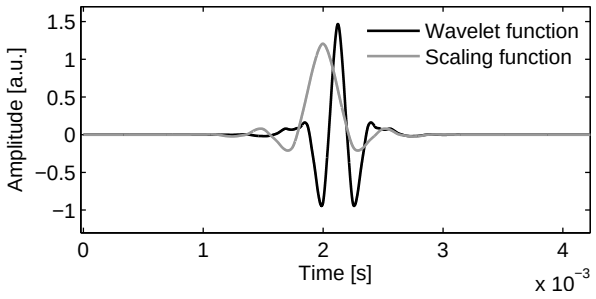


Figure 6.7: *bior*6.8, an example member of the biorthogonal wavelet family. The remaining wavelet families suitable for analysis by DWT were introduced in Figure 5.7.

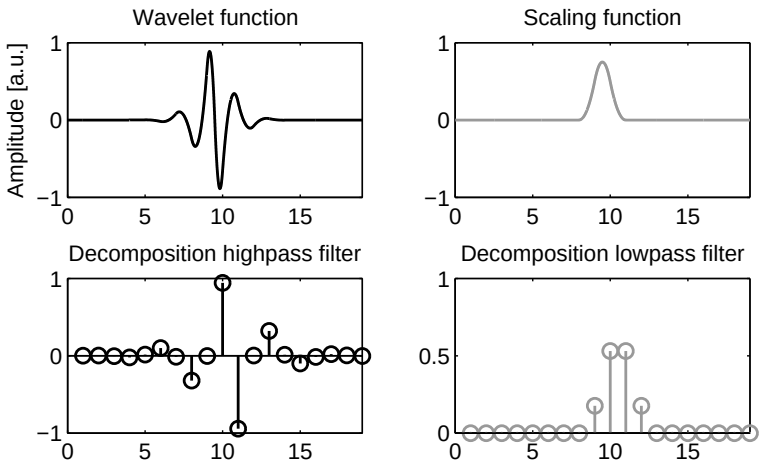


Figure 6.8: The reverse biorthogonal wavelet function *rbio*3.9, chosen to compute the wavelet envelope feature. Also shown are the scaling function and the associated high- and low-pass filters.

6.5.3 PSD envelope

Using the short-time Fourier transform the power spectral density was computed for segments defined by a window length of 30ms and an 80% overlap. The median amplitude across all frequency bands was taken as the power spectral density (PSD) value for each sample. An example epoch and derived feature envelopes are shown in Figure 6.9.

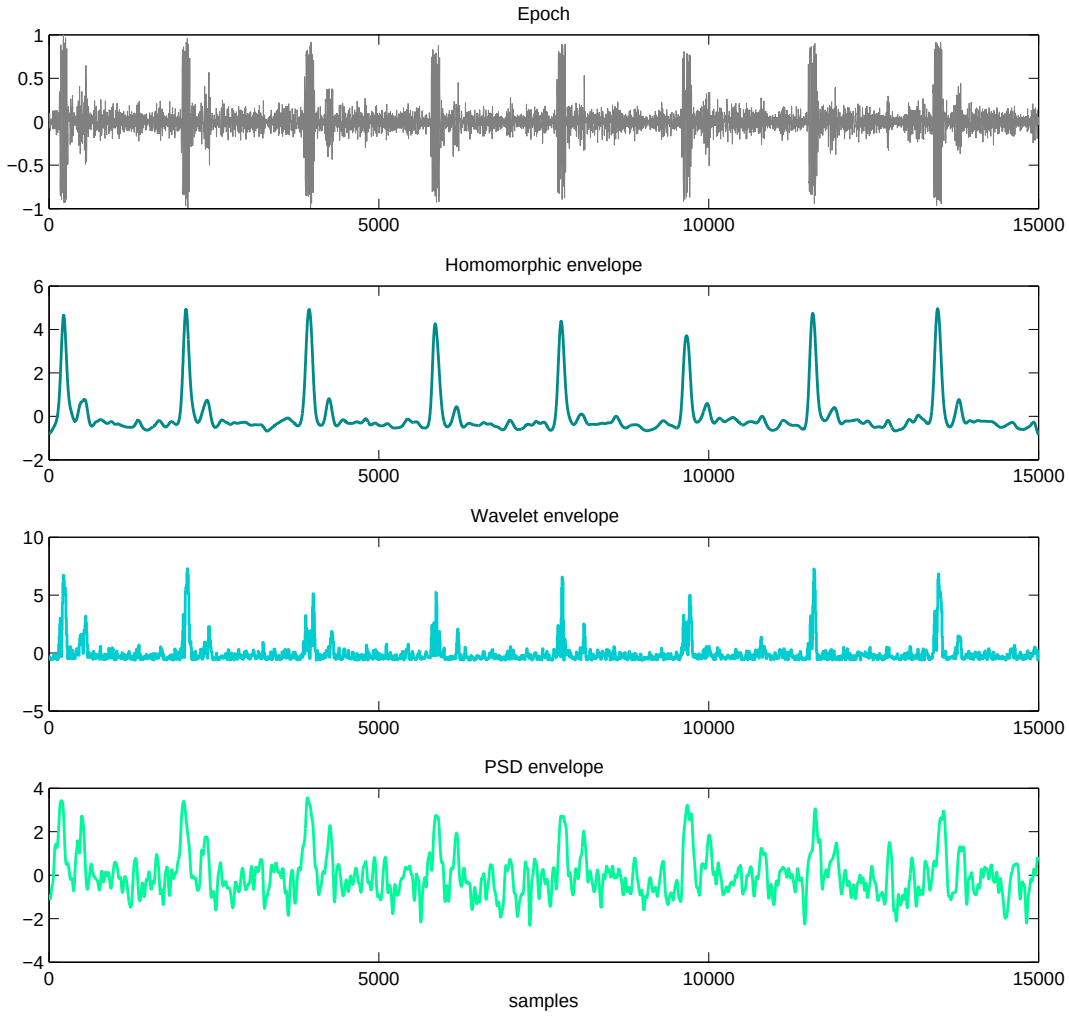


Figure 6.9: Example epoch and its derived features, the homomorphic, wavelet and power spectral density envelopes.

6.6 The distribution of Doppler state durations

Given the nature of the auscultation Doppler signal, one cannot assume a consistent heart cycle morphology between recordings nor the stationarity of the signal's morphology throughout its duration. The search was therefore simplified and the HMM parameters trained to identify a cardiac event by its primary oscillation.

In the absence of a reference signal such as a concurrent fetal electrocardiogram, the primary oscillation was defined as the oscillation of highest amplitude in the cardiac cycle, rather than by its physiological origin. The mean of the primary oscillation durations was estimated from the training data. The primary oscillations in an epoch were identified as the peaks with maximum amplitude in each cycle. The zero-crossings of the envelope either side of a primary oscillation were assumed the onset and offset of the oscillation, marking the primary oscillation duration (p_{pOsc}) (Figure 6.10). The mean oscillation duration ($\overline{d\mu}_{pOsc}$) was predetermined from the training data ($N_{osc} = 1941$) as 69.70 ms with a standard deviation of 21.50 ms.

The fetal heart rate is highly variable, where normal may range from 110 to 180 bpm [63]. The duration between primary oscillations was therefore estimated anew for each test epoch. The mean cardiac cycle duration of the epoch (eCC) was computed by autocorrelation of the signal envelope (see Figure 6.11) and the interval duration ($\overline{d\mu}_{oInt}$) determined as follows:

$$\overline{d\mu}_{oInt} = eCC - \overline{d\mu}_{pOsc} \tag{6.27}$$

The standard deviation of the interval showed a moderate correlation with the interval duration in the training data. The relation between the two was therefore established by regression analysis and the interval's standard deviation $\overline{d\sigma}_{oInt}$ computed as follows (an approach also taken by [151]):

$$\overline{d\sigma}_{oInt} = 0.2\overline{d\mu}_{oInt} - 0.05 \tag{6.28}$$

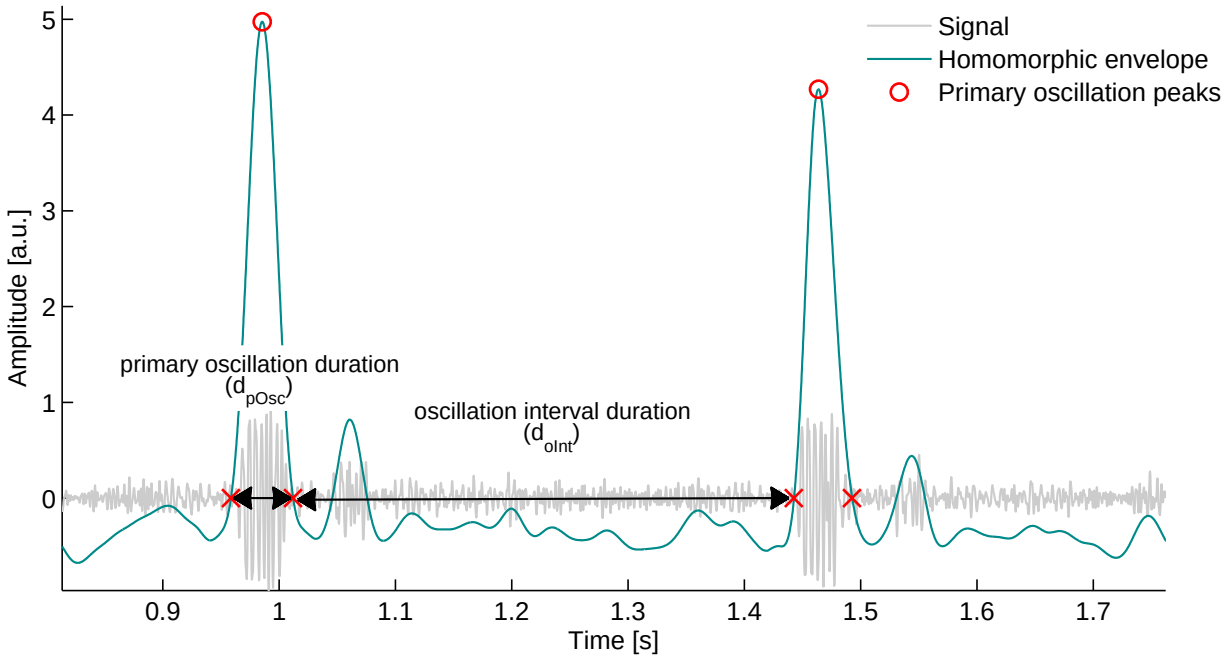


Figure 6.10: The duration of a primary oscillation was estimated from the training data. The maximum peak of each cycle in the homomorphic envelope of the epoch was identified and the zero-crossings either side adopted as onset and offset of the oscillation. The duration between oscillations was derived for each test epoch individually based on the respective heart cycle duration.

The duration probabilities for oscillations and intervals were modelled for each epoch with a Gaussian probability density function. An example is shown in Figure 6.12. The minimum oscillation duration was predetermined as 24.75 ms, the shortest measured in the training data. The maximum oscillation duration was specified as 134.27 ms, three standard deviations from the mean. The interval duration was equally bound by ± three standard deviations from the mean. If these values exceeded physiologically feasible limits, they were corrected to 300 ms (equivalent to 200 bpm) and 667 ms (equivalent to 90 bpm).

6.7 Heart rate estimation from the HSMM state path

As previously described, identified states represent the primary cardiac oscillations in a cycle (state 2) and the interval between successive oscillations (state 1). The beat in-

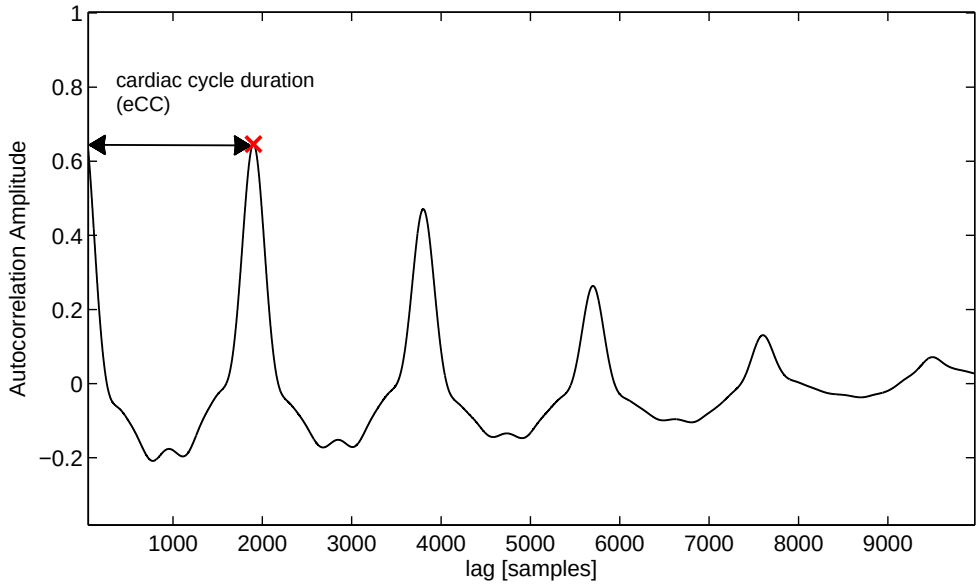


Figure 6.11: The average cycle duration within an epoch as estimated by autocorrelation of the homomorphic envelope.

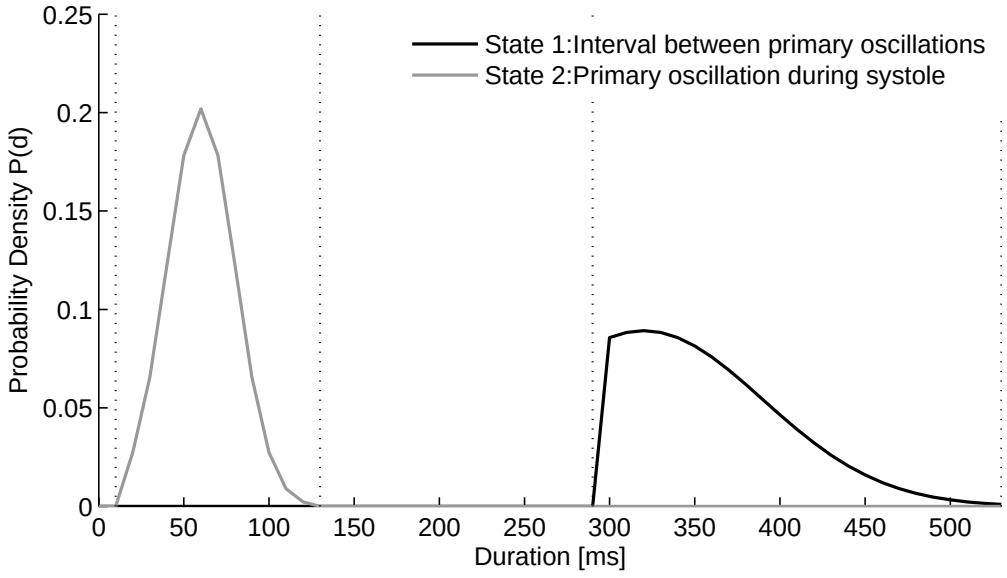


Figure 6.12: Duration probability density functions for both the primary oscillation and the oscillation interval of an example epoch. The dotted vertical lines indicate the minimum and maximum duration of those states respectively.

stances in an epoch were estimated from the duration dependent state path. The state transitions were identified and the mid point selected to represent a beat location in time. The average fetal heart rate for each epoch was computed as the number of beats per minute.

Figures 6.13 and 6.14 show example epochs of intermediate and excellent signal quality respectively, including state paths and derived beat estimates. To visualise the effect of introduced state durations which suppress secondary oscillations in the signal (Section 6.6), the state path estimated by implementation of the standard Viterbi algorithm, and the duration-dependent variant of the algorithm, are both shown.

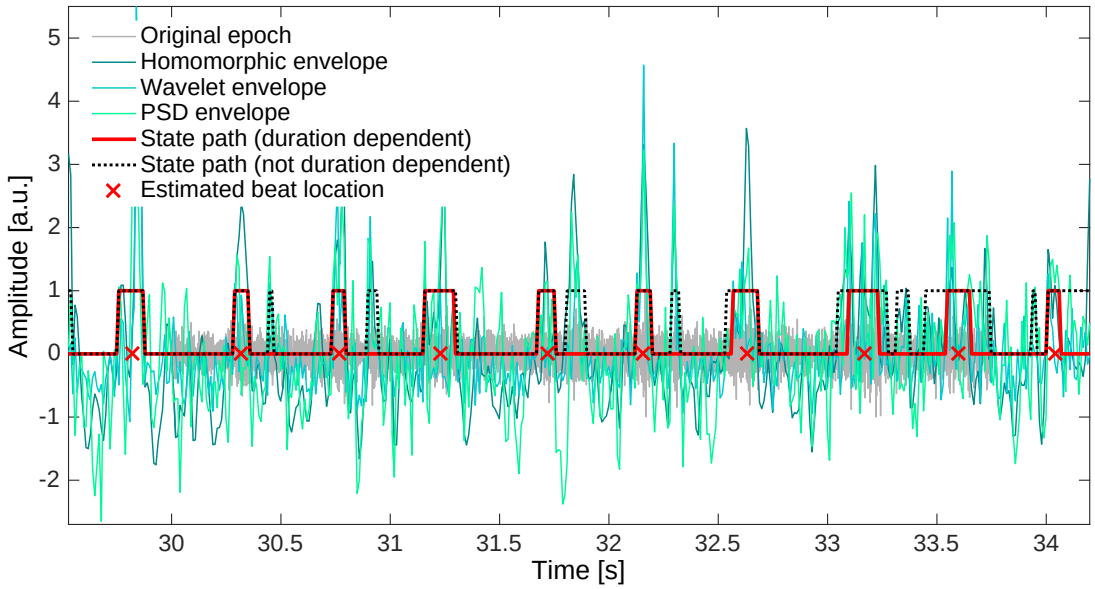


Figure 6.13: An example epoch of intermediate signal quality, showing the original epoch signal, its features, the computed state path and resultant beat estimates. To visualise the effect of duration probabilities both the duration dependent and standard path estimates are shown.

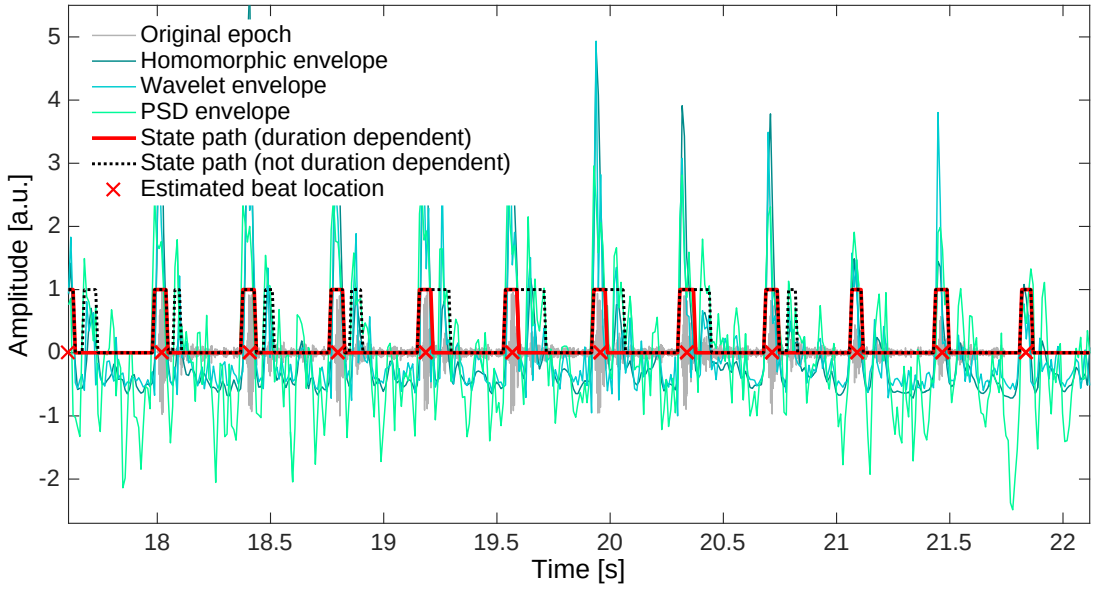


Figure 6.14: An example epoch of excellent signal quality (epoch SQI 5), showing the original epoch signal, its features, the computed state path and resultant beat estimates. To visualise the effect of duration probabilities both the duration dependent and standard path estimates are shown.

6.8 Reference signal

6.8.1 Manual beat annotation

Three independent annotators were asked to review recorded Doppler signals as described in Section 5.3. As part of the signal quality assessment, annotators were asked to listen to each recording indicating all audible beats by mouse click. In the absence of a synchronously recorded gold standard signal, such as fetal electrocardiograms, the resultant beat annotations provided the basis for the reference signal used to evaluate the developed segmentation algorithm.

6.8.2 Human reaction time

Heart beat annotation by audio assessment introduced a non-stationary reaction delay dependent on the fetal heart rate, the heart rate variability over time and noise levels. Annotations showed, for example, a greater delay at the beginning of a recording, which typically became more steady as annotators became familiar with the rhythm and again

became less regular with deterioration in signal quality. There was no clear intra- and inter-annotator delay pattern. The manual annotations were therefore processed as follows.

The three annotation sequences for each epoch were individually examined for missed beats based on a cardiac interval of 90 bpm, empirically chosen to be lower than the accepted threshold of normality [62] to account for physiological outliers and annotator delay. The annotator's beat markers were subsequently grouped as belonging to respective beats according to the minimum distance between them, and duplicate beat annotations were disregarded (see Figure 6.15 for an example). The annotations' median time was taken to indicate a beat location. Given the dependency of the annotations on the human reaction time as described above, only the number of beats in the analysis window and therefore the average beat interval was considered to be true. The beat location was adjusted in two steps. First the highest amplitude peak in the signal's envelope closest to the last median annotation was identified (within a search window starting one interval prior to the last annotation and ending 1/4 interval thereafter (both values empirically chosen)). Preceding beat locations were identified by searching for the maximum peak within an interval's distance ± 15 bpm preceding the last beat. Beat-to-beat variability in the healthy fetus ranges from 5 to 15 bpm, which determined the choice of search window [63]. Figure 6.16 illustrates the process of beat location adjustment.

6.9 Results of HSMM Doppler segmentation

6.9.1 Performance assessment of heart rate estimation

The American National Standard on cardiac monitors, heart rate metres, and alarms determines the acceptable read out error for heart rate estimation to be no greater than ± 10 percent of the input rate or ± 5 bpm, whichever is greater [171].

Both heart rate estimation by autocorrelation of the epoch's envelope, standard in CTG machines, and by HMM state detection were assessed and compared to the processed

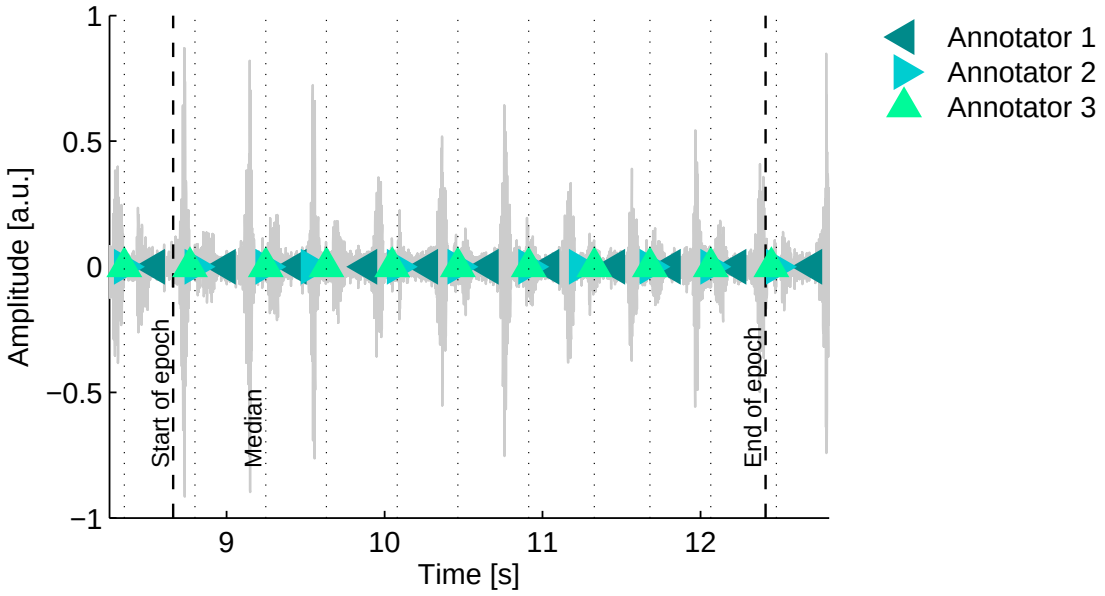


Figure 6.15: The manual annotation sequences were aligned according to their proximity to each other. Additional beats were disregarded. The median value was computed indicating a beat location.

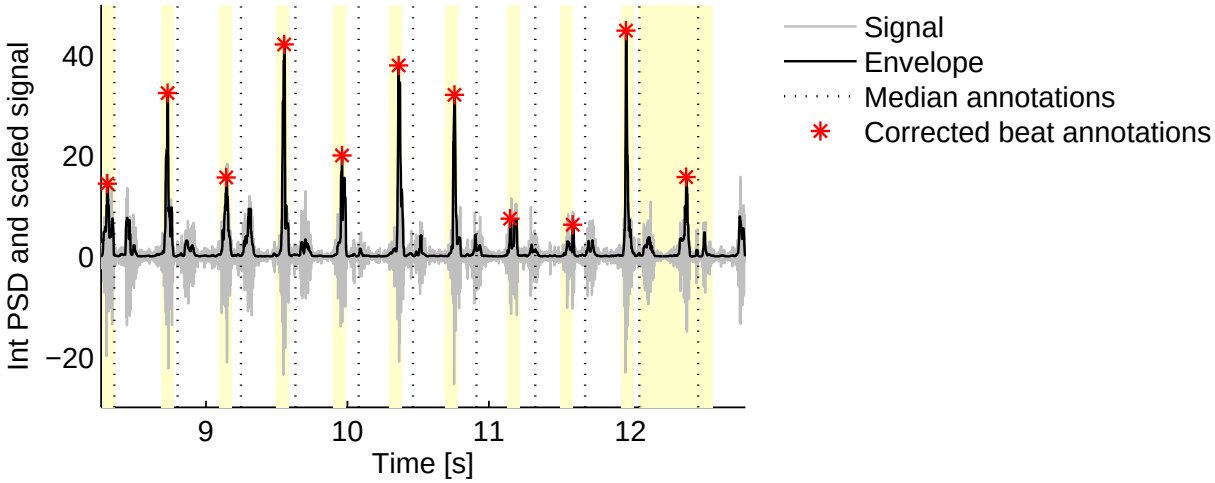


Figure 6.16: The median annotations were adjusted to the maximum peak within the specified search window. The shaded areas in the plot indicate respective search onset and offset. The last beat in the epoch was selected as the maximum peak identified closest to the last median annotation within a window governed by the average interval of the manual annotations. Preceding peaks were identified as the maximum within an interval distance +/- 15 bpm, which is the maximum expected beat-to-beat variation.

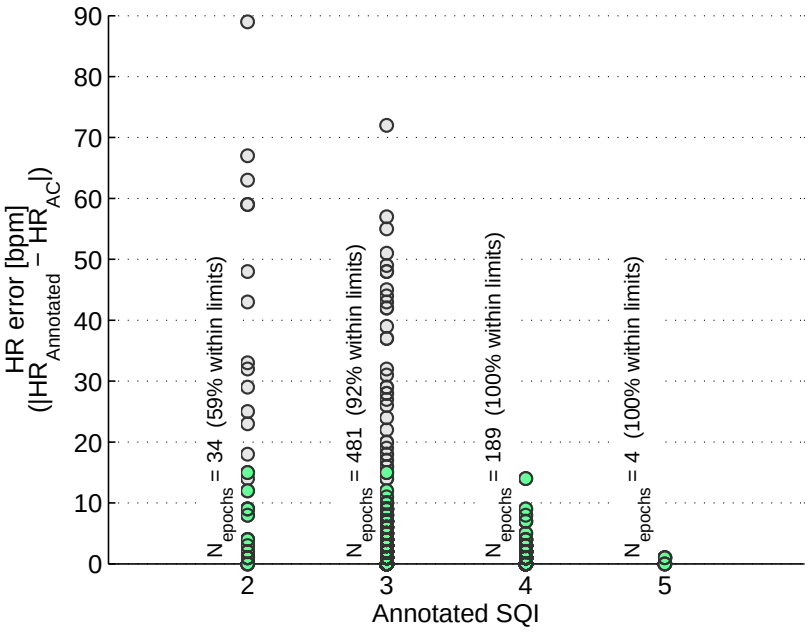


Figure 6.17: The error in heart rate estimation by autocorrelation compared to manual annotation is plotted against the signal quality assigned to respective epochs by the annotators. The number of epochs analysed per signal quality class and the percentage of estimate values within acceptable bounds ($\pm 10\%$ of the input heart rate) are shown alongside.

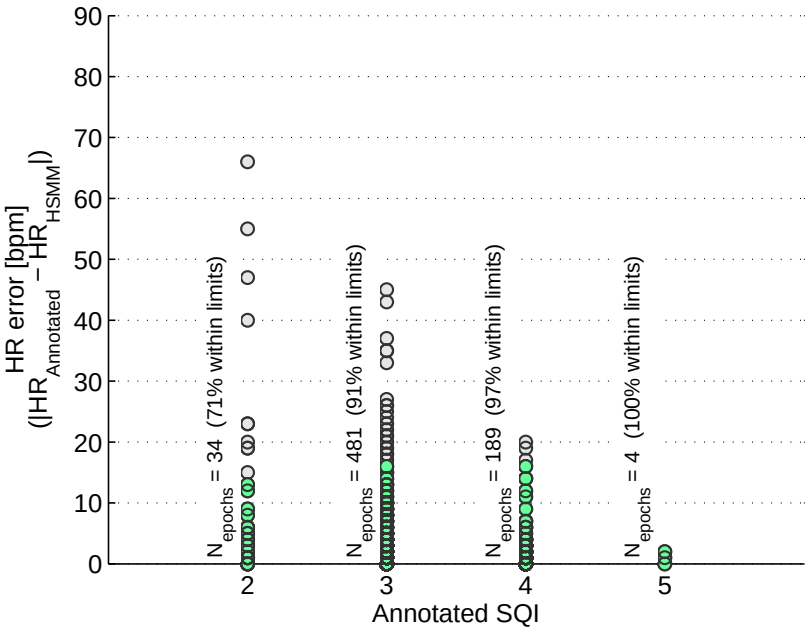


Figure 6.18: The error in heart rate estimation base on HSMM segmentation compared to manual annotation is plotted against the signal quality assigned to respective epochs by the annotators. The number of epochs analysed per signal quality class and the percentage of estimate values within acceptable bounds ($\pm 10\%$ of the input heart rate) are shown alongside.

manual annotation reference. To illustrate the dependence on signal quality the heart rate error was plotted against the manually assigned SQI in Figures 6.17 and 6.18. The error in heart rate estimation decreased for both methods with increasing signal quality. Both the autocorrelation based heart rate estimation and the segmentation approach performed well for intermediate to excellent signal qualities. Autocorrelation estimates were within the 10% tolerance limit for all test epochs of SQI 4 and 5 and 92% of SQI 3, while the heart rate estimates by HMM segmentation when within the 10% tolerance in 97% and 91% respectively (see Figures 6.19 and 6.20 showing results for good quality epochs (SQI 4 and 5)). Provided with poor quality data the autocorrelation algorithm only estimated 59% of the epochs correctly, substantially lower than the 71% achieved by HSMM. It is important to note that the manual annotation accuracy may equally be affected by signal quality as Figure 6.21 suggests.

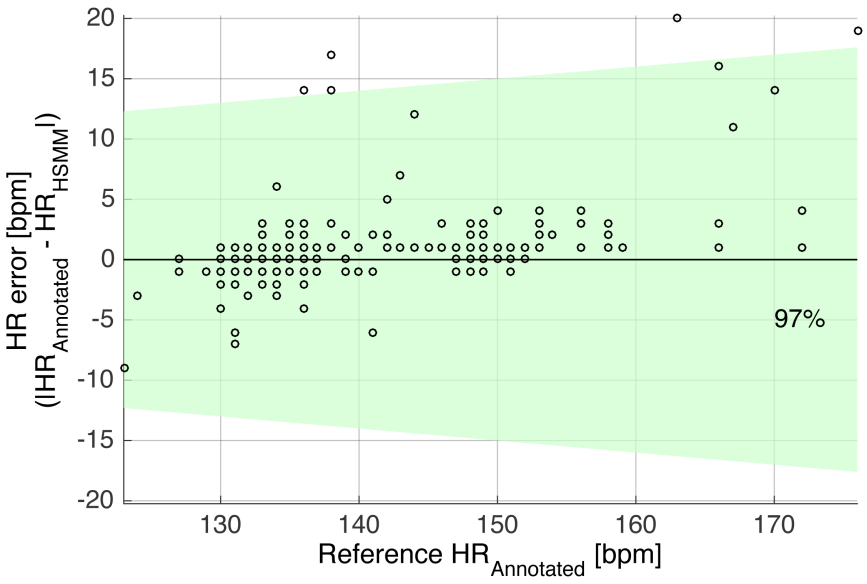


Figure 6.19: The error between HMM state derived HR estimation and the manual reference plotted against the manual gold standard (N = 201) for good quality data (SQI 4 and 5). The shaded area indicates the 10% accuracy tolerance according to the applicable American National Standard [171].

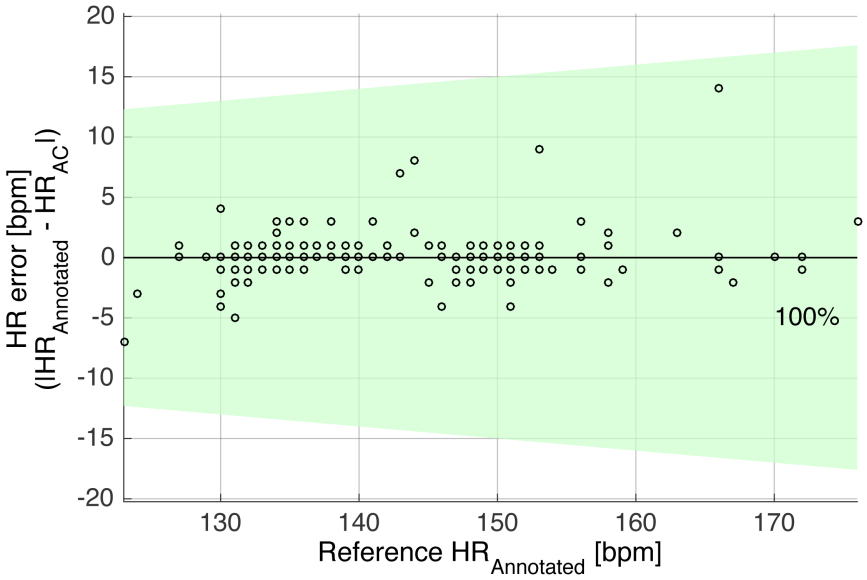


Figure 6.20: The error between autocorrelation derived HR estimation and the manual reference plotted against the manual gold standard (N = 185) for good quality data (SQI 4 and 5). The shaded area indicates the 10% accuracy tolerance according to the applicable American National Standard [171].

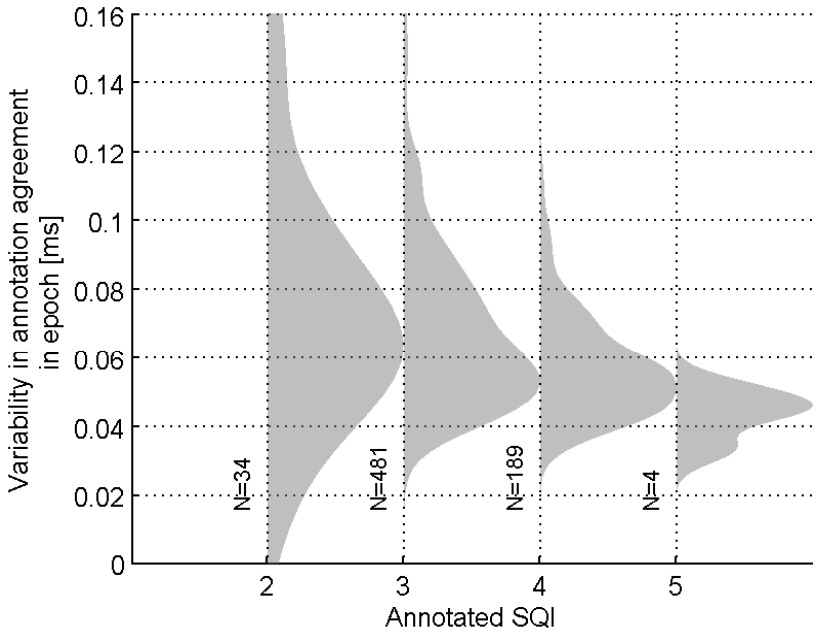


Figure 6.21: The mean standard deviation of the difference in beat annotations between the three annotators was computed as a measure of annotator agreement. To visualise the effect of signal quality on the manual annotations, the annotation variability was plotted against the epoch's assigned SQI.

6.9.2 Performance assessment of beat-to-beat interval estimation

For the performance assessment of cardiac rhythm algorithms, the Association for the Advancement of Medical Instrumentation (AAMI) suggests a temporal window of 150ms within which a detected beat may be assumed correct [172]. As this standard has been developed for an adult population, the window was adjusted to 50ms, in order to reflect the typically much higher fetal heart rate. This is a threshold previously used in fetal ECG studies as reported in the literature [173,174]. The adjusted manual reference and the HSMM derived heart rate signal may not necessarily identify the same fiducial points in the cardiac event. To assess the performance of the HSMM segmentation approach the duration accuracy of successive intervals was therefore compared to the reference instead of the beat locations in time. Considering the goal of improved heart rate variability estimation the change in interval duration is the primary measure of interest.

The sensitivity (Se), positive predictive value (PPV), and the F1 score as a measure of the algorithm's accuracy were computed :

$$Se = \frac{TP}{TP + FN} \tag{6.29}$$

$$PPV = \frac{TP}{TP + FP} \tag{6.30}$$

$$F1 = 2 \cdot \frac{PPVSe}{PPV + Se} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \tag{6.31}$$

Where TP denotes the true positives, FP the false positives and FN the false negatives, based on the adapted threshold of 50ms.

Given the observation made in the previous section on the influence of signal quality on the heart rate estimation performance and the impact on annotator agreement, results were computed for signal qualities 3, 4 and 5 only. Results per patient and as an average over all intervals are reported in Tables 6.2, 6.3, 6.4, and 6.5.

The accuracy of estimated beat-to-beat intervals of good and excellent quality, computed for each patient individually, ranged from 94.2% to 100%. This averaged to an overall accuracy of 97.6% (Table 6.2). The performance results per patient allow to review their individual impact on the gross performance computed over all intervals, which proved higher at 98.4% (Table 6.4). The accuracy results for intermediate quality data per patient were spread much more widely from 77.6% to 100%, averaging to an overall accuracy of 93.3%. The gross accuracy computed over all intermediate quality intervals was slightly lower at 91.8%.

Test patient ID	N	Se [%]	PPV [%]	F1 [%]
Patient 2	520	98.3	99.6	98.9
Patient 3	111	94.6	98.1	96.3
Patient 13	558	99.5	99.8	99.6
Patient 15	32	100.0	100.0	100.0
Patient 17	150	92	96.5	94.2
Patient 18	127	95.5	98.4	96.4
Average	1498/6	96.5	98.7	97.6

Table 6.2: The performance of HMM beat-to-beat interval estimation per subject on data with quality labels SQI 4 and SQI 5. N indicates the number of heart beat intervals. The average is the mean of the statistics calculated individually for each subject in the test data set.

6.10 Discussion

The performance of the HSMM heart rate estimator is only marginally inferior to the autocorrelation algorithm when compared to manual annotations. Both are well within bounds of the requirements set by the applicable industry standard for good quality recordings. Results highlighted the susceptibility of autocorrelation algorithms to high noise levels, which was apparent in the poor heart rate estimation performance in low quality signals (SQI 2) of 59% accuracy. In addition to the autocorrelation's averaging nature this is another drawback, particularly given challenging use environments. The

Test patient ID	N	Se [%]	PPV [%]	F1 [%]
Patient 2	575	93.4	95.4	94.4
Patient 3	767	89.7	93.1	91.4
Patient 5	79	74.7	80.8	77.6
Patient 7	83	92.8	97.5	95.1
Patient 8	512	81.5	88.2	84.7
Patient 9	264	83.3	95.7	89.1
Patient 10	9	100	100	100
Patient 11	231	93.5	96.4	94.9
Patient 13	316	91.1	95.0	93.1
Patient 14	42	100	100	100
Patient 15	226	96.9	97.8	97.3
Patient 17	229	94.3	96.4	95.4
Patient 18	170	92.9	94.6	93.8
Patient 19	7	100	100	100
Average	3510/14	91.7	95.1	93.3

Table 6.3: The performance of HMM beat-to-beat interval estimation per subject on data of intermediate quality (quality label SQI 3). N indicates the number of heart beat intervals. The average statistics are the mean of the statistics calculated individually for each subject in the test data set.

	Total intervals	Se [%]	PPV [%]	F1 [%]
Average all intervals	1498	97.5	99.2	98.4

Table 6.4: The gross performance statistics of HMM beat-to-beat interval estimation on good quality data (quality labels SQI 4 and SQI 5), reported over all intervals in the test dataset.

	Total intervals	Se [%]	PPV [%]	F1 [%]
Average all intervals	3510	89.8	93.9	91.8

Table 6.5: The gross performance statistics of HMM beat-to-beat interval estimation on intermediate quality data (quality label SQI 3), reported over all intervals in the test dataset.

accuracy achieved by HSMM was distinctly superior at 71%, however, both results may be considered unsatisfactory.

The primary aim of this investigation was to assess whether a HSMM segmentation approach may be suitable for beat-to-beat interval estimation. Beat-to-beat resolution is recognised as an important quality of a heart rate trace for the accurate detection of fetal

compromise such as IUGR.

Assessing the beat-to-beat interval accuracy of the HSMM estimator compared to manual annotations produced promising results, confirming the potential of this approach.

A deterioration in accuracy on both heartrate estimation and interval segmentation could be observed with a reduction in signal quality. The same however applies to the manual annotations, where an increase in annotator variability lowers the confidence in the derived reference signal. The manual annotation process as described has several limitations, particularly with reduced signal quality. It is therefore strongly recommended to test the beat-to-beat accuracy of the HSMM segmentation against a more robust measure such as fetal ECG.

Furthermore, it is important to evaluate the described approach on a larger dataset, especially including gestational ages lower than 35 weeks. This is of particular interest with a view to employ such beat-to-beat segmentation for the risk analysis of early onset IUGR (as described in Chapter 3).

6.11 Conclusion

The HSMM approach proved a viable approach to fetal heart rate estimation and most importantly to beat-to-beat interval segmentation. Manual annotations however showed to be problematic, especially when the signal quality reduced. Even though a deterioration in confidence with decreasing signal quality may persist for alternative reference signals, further algorithm development would benefit significantly from a more robust alternative. A robust reference could also make the estimation procedure (Baum-Welch parameter estimation, described in Section 6.4.2) redundant, as HMM model parameters can be estimated from training data provided there are reliable annotations. The fetal ECG as measured by scalp electrode may be considered the gold standard, however it cannot be recorded antenatally. An alternative is the abdominal fetal ECG (fECG), introduced in Section 1.1.2. In order to both increase the database and address this issue,

data collection of synchronous Doppler and fECG has already commenced in collaboration with the Technical University in Dresden and the University Hospital in Leipzig, Germany. This is discussed further in Section 8.1.1.1.

Chapter 7

Data collection and validation of the mHealth fetal monitoring system in rural Guatemala

7.1 Introduction

To test the feasibility and ultimately the potential impact of the mHealth intervention proposed in this thesis, the early involvement of target user communities is essential. The non-profit organisation Wuqu' Kawoq, introduced in Chapter 1, Section 1.1.4, works closely with traditional birth attendants (TBAs) as part of their programme delivering primary healthcare in an under-served rural and predominantly indigenous region of Guatemala. The development of the mHealth application and overall healthcare intervention was designed as an iterative process and is being executed in collaboration with Wuqu' Kawoq's medical staff, their extensive network of 'comadronas' (traditional birth attendants) and their patients.

The development process including all stages completed or in progress is summarised in Figure 7.1. Chapters 1 and 4 outlined the need and motivated the mHealth approach to improved fetal and maternal health monitoring. The first prototype (See Chapter 4)

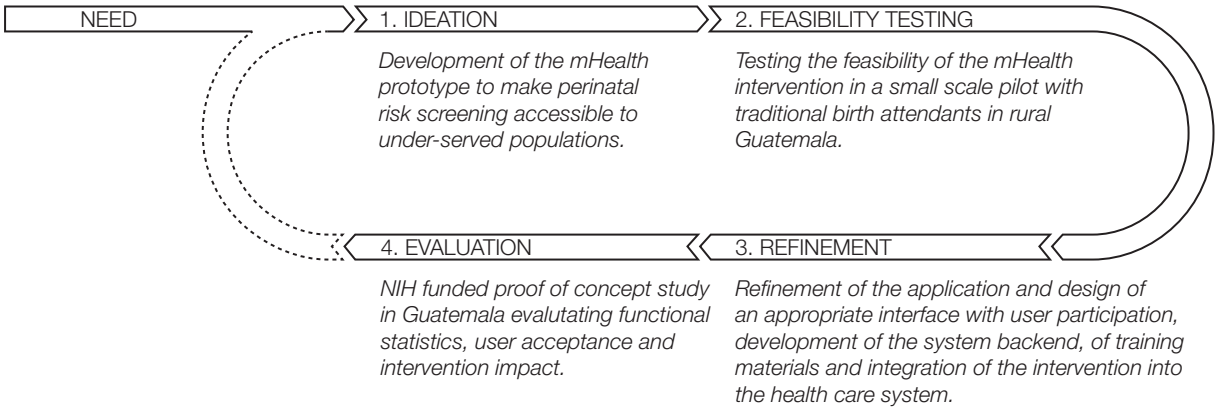


Figure 7.1: The iterative design process followed in the development of the mHealth perinatal monitoring intervention. The need and development of the mHealth approach were introduced in Chapters 1 and 4 respectively. Outcomes of the feasibility pilot and the implementation of system improvements are discussed in this chapter. The project is currently in its evaluation stage.

was piloted with comadronas in order to assess the acceptability of such intervention by both healthcare worker and patient, and to evaluate the usability and technical feasibility of such system. Based on collected insights and by closely working with Wuqu' Kawoq's team and TBAs, the mobile application was extended in its functionality and an appropriate user interface developed. Both feasibility testing and system refinement are discussed in this chapter. The project is currently entering its evaluation phase. A proof of concept study has been designed and is supported by a National Institutes of Health (NIH) grant to assess the impact of the mobile application and overall healthcare intervention. An overview of this study concludes this chapter.

7.2 Feasibility testing

The feasibility assessment was designed to evaluate whether frontline healthcare workers could operate the study equipment (1D fetal Doppler, pulse oximeter and recording application, as described in Chapter 4) and record signals successfully using a smart phone. The feasibility study also set out to gauge user need and to assess the acceptability by both healthcare provider and patient to inform subsequent system refinement and the proof-of-concept pilot, as well as to determine training requirements. The study proto-

col titled “Preliminary evaluation of a mobile health perinatal monitoring system” was approved by the Wuqu’ Kawq Institutional Review Board (WK-2014-003).

7.2.1 Data collection

Recording staff included nine healthcare workers drawn from the Wuqu’ Kawoq network of perinatal care providers: four traditional birth attendants (TBAs), two physicians (PH), one nurse (NU), one healthcare worker (HCW) specialising in women’s health and one health promoter (HP) (Participant information is summarised in Table 7.1). They were instructed to attach the pulse oximeter to the mother’s wrist and index finger, start the recording application, locate the fetal heart beat with the Doppler (using ultrasound gel for better contact with the skin) and to record the cardiac activity for a minimum of ten minutes. They were then asked to save the recording, switch off and remove all devices (See Figure 7.2). An alternative set-up was trialled with staff that had some degree of experience in the use of fetal Doppler. Those staff were asked to position the Doppler probe on the abdomen with an elastic belt and start recording thereafter. This was to assess the option of continuous monitoring for advanced cardiac signal analysis, such as screening for IUGR risk. If necessary, researchers provided guidance during the recording. Participants were given headphones or speakers to listen to the fetal cardiac activity while recording. No analysis of the fetal cardiac signal was provided at this stage of the investigation.

A total of 22 pregnant women were enrolled with an average age of 26 years (Range: 16 to 37 years), carrying a fetus with an average gestational age of 27 weeks 3 days (Range: 6 weeks 5 days to 39 weeks 6 days). All pregnancies were singleton pregnancies with no known fetal complications. Recordings were collected during five home visits and in four different community health centres, where 25 unique staff-patient interactions were observed (three women were recorded twice by different staff) and 28 recording attempts. The average recording length was 11 minutes and 10 seconds (±3m49s).

Participants	PH_{01}	PH_{02}	NU_{01}	HCW_{01}	TBA_{01}	TBA_{02}	TBA_{03}	TBA_{04}	HP_{01}
Education	University degree			No formal education					
Experience (yrs)	2	3	1	15	15	25	<1	25	<1
NRecordings	6	5	5	2	1	2	1	2	1
NRec no support	2	3	2	1	0	0	0	0	0
NRec min support	4	2	3	1	0	0	0	0	1
NRec ext support	0	0	0	0	1	2	1	2	0

Table 7.1: Participants were Wuqu' Kawoq staff and healthcare workers recruited from Wuqu' Kawoq's network with varying levels of education and experience working in perinatal care. The total number of patients recorded by each participant was dependent on their current patient load. Recordings were observed and classified into instances requiring no support, minimum support (verbal) or extensive support (hands-on) support by study staff.

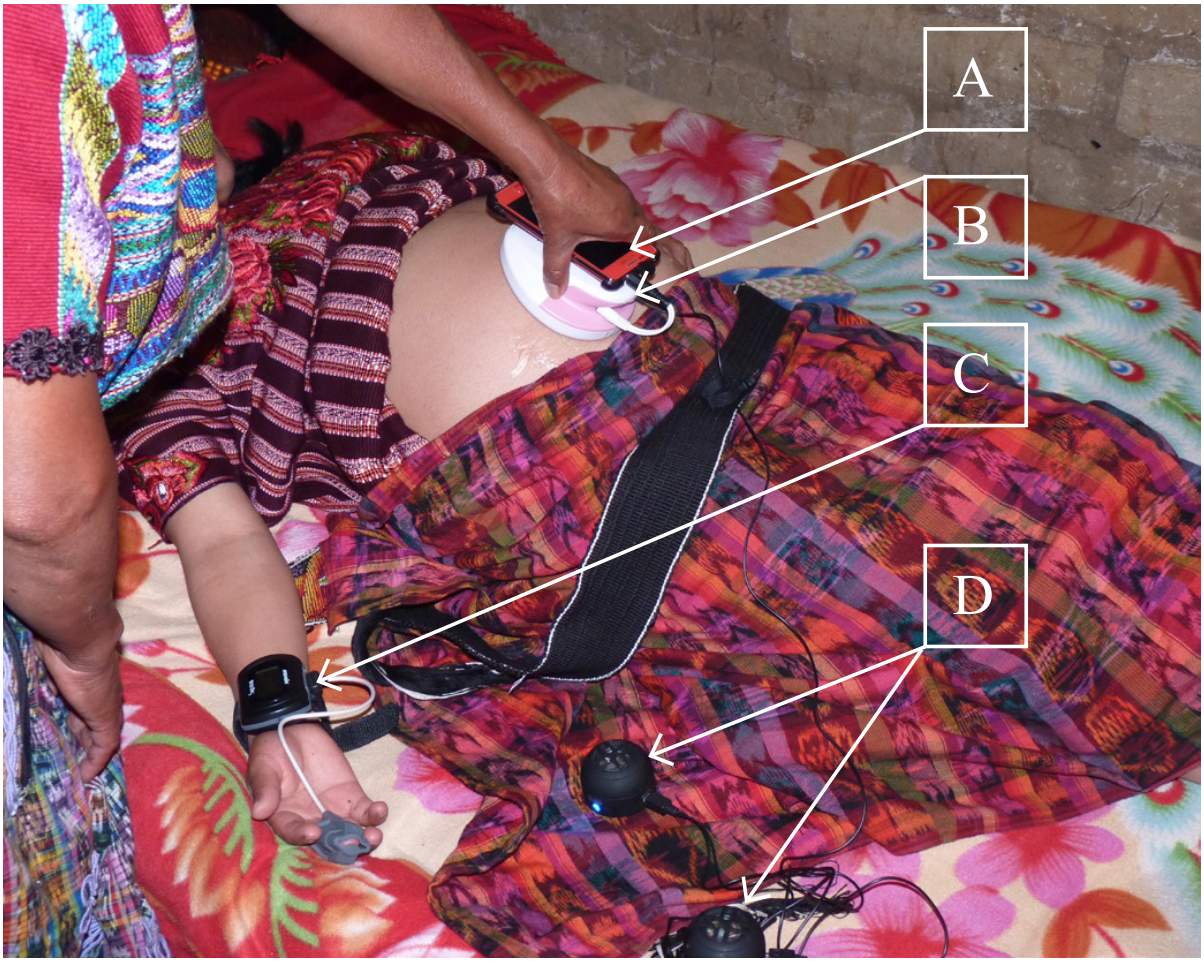


Figure 7.2: TBA with prototype equipment for recording fetal heart activity in a remote village in the Guatemalan highlands. The separate components: A: Smart phone, B: Low-cost 1D Doppler, C: Pulse oximeter D: Speakers to enable both healthcare worker and patient to listen to the fetal cardiac signal.

7.2.2 Evaluation tools

To evaluate whether Wuqu' Kawoq's antenatal care providers could successfully handle the study equipment for physiological signal capture, both qualitative and quantitative research tools were employed, assessing data quality and implementation realities. Participating staff were invited to provide feedback rating the potential utility of such a system in their practice and the ease-of-use of the individual system components on a Likert scale. Each visit and recording procedure was also observed in detail by an accompanying researcher to gain insights into the work environment of the target users, their routine consultations and in particular their interaction with the study equipment. Further quantitative metrics collected were the success rate of recording all signals onto the phone concurrently and the success rate of staff identifying fetal heart sounds.
The quality of Doppler recordings was furthermore reviewed using developed signal quality metrics, as described in Chapter 5.

7.2.3 Feasibility testing outcomes

7.2.3.1 Device use

Five out of the nine staff required no or minimal support (occasional oral instructions or reminders) when setting up the prototype system (attaching pulse oximeter, switching on Doppler probe and starting the recording application). In only 21% of the recording instances (consultations) additional instructions had to be provided when starting the Doppler probe, and in 24% when attaching and starting the pulse oximeter. Navigating the smart phone application to start recording proved more difficult, required guidance in 40% of all cases. In total, 25 recordings were successfully started. In 16 cases staff located fetal cardiac activity with ease, in seven with some difficulty, while in two recordings no heart sounds could be found (this was however anticipated, as the gestational ages were 6 weeks 5 days and 10 weeks respectively). In 48% of all cases the fetal cardiac signal could be well maintained over the duration of the recording. A number of participants were

not aware whether they had recorded cardiac activity from the placenta or the fetal heart directly. Similarly two participants seemed to not notice maternal interference and did not attempt to find a clearer signal. Several participants appeared not to be conscious of the differences in signal quality and the clarity in heart sound that could be achieved.
Seven out of the nine staff could end the procedure with ease. There was no difference in the number of sessions requiring support in navigating the app (saving the signals) and those requiring help switching off and removing the devices. In summary, eight recordings required no intervention, eleven minimal guidance (i.e. oral instructions), whereas six demanded hands-on support (i.e. entering the patient ID into the smart phone). Required levels of support are summarised in Table 7.1. Staff participants with experience in handling similar devices were generally more at ease. This was particularly true for operating the smart phone but also applied to the use of the Doppler probe. The five out of nine participants who were familiar with operating smart phones, were all staff with a higher educational level or of younger age.
In total, participating staff collected 24 successful recordings (all signals recorded for desired length of time), in three instances a second recording attempt was required as the pulse oximeter communication had timed out (the application was started before the pulse oximeter), one recording failed as it was accidentally saved and the application exited too early. All staff managed to identify and record fetal heart activity, only two recordings did not contain any heart sounds, however this was expected due to the gestational age of the fetuses ($<$12 weeks). Four out of nine staff participants appeared to be capable of handling all devices in the system with the very basic training provided by the researchers.

7.2.3.2 Staff and patient response

All pregnant women approached for participation agreed to be enrolled. Neither staff nor patients voiced concerns over the use of the study equipment prior, during or after recording. All recording staff appeared comfortable using the study devices on their patients, and, even if unfamiliar with their use, applied them with confidence and no noticeable

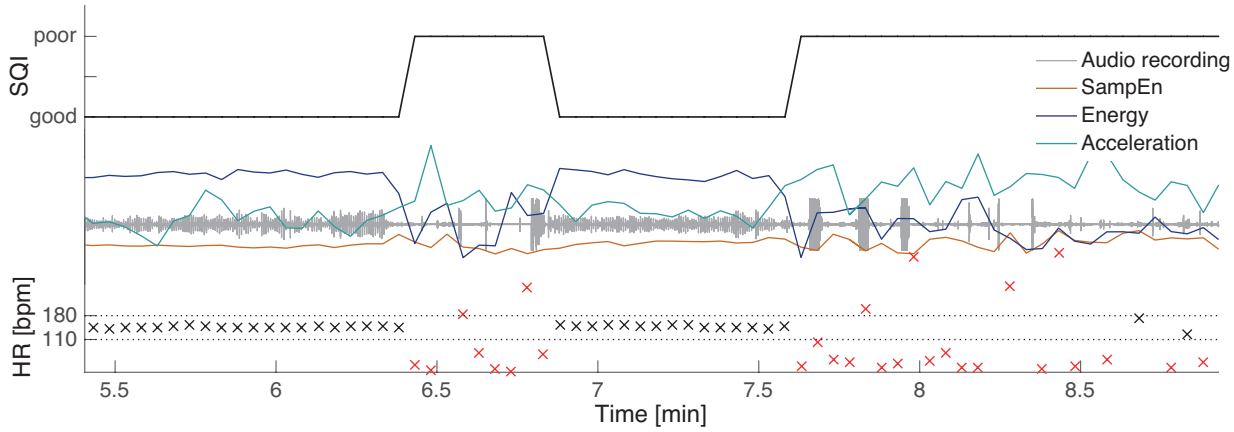


Figure 7.3: An example showing an original recording segment, the computed quality features (SampEn, Energy, Acceleration, defined in Chapter 5), the associated heart rate, and assigned quality score.

hesitation. The six staff who provided written feedback found all system components easy to operate and independently agreed on the potential benefit of such system. Participating healthcare workers (TBAs and health promoters unfamiliar with the use of Doppler and pulse oximetry devices) demonstrated great eagerness to complement their skills, as well as enthusiasm given the prospect of being able to provide their patients with additional health information and the ability to share the experience of listening to the fetal heart beat.

7.2.3.3 Doppler signal quality assessment

The signal quality of the fetal cardiac recordings was assessed using the logistic regression based SQI classifier (described in Chapter 5, Equation 5.17). In addition to the signal's quality index (SQI) assigned by the algorithm, signal segments were regarded as poor if the heart rate computed by autocorrelation was out of a rhythm's normal range (110-180 bpm [63]). SQI and heart rate were computed for 3.75 epochs at 0.33 Hz sampling frequency. An example recording including assigned quality score is shown in Figure 7.3. Of the 4 hours and 16 minutes recorded (probe in contact with abdomen) 41% were found to be of good and 59% to be of poor (including intermediate) quality. The number of years working in antenatal care showed not to be positively correlated with an increase in the

quality of data collected. On the contrary, quality appeared to decrease for longer serving staff, which may be attributed to their age and lack of exposure to technology of similar kind (Figure 7.4). The percentage of high quality data was greater for staff informally trained in their profession compared to university educated participants. Doppler signals collected by lay midwives and healthcare workers (N recordings = 9) contained 47% good and 53% poor quality, signals recorded by the nurse and clinicians (N recordings = 13) 37% and 63% respectively. Figure 7.5 however illustrates how gestational age influences the ease with which good quality data can be collected. The proportion of good quality data in the signal increases substantially towards the third trimester of gestation. The difference in the proportion of high quality recordings between informally and formally educated participants can likely be attributed to the greater number of early gestation patients recorded by the latter group.

The percentages of signal quality are comparable to those of the data collected at the John Radcliffe Hospital by healthcare professionals using the same prototype (see Section 5.2 for details on data collection). Applying the SQI algorithm (Equation 5.17), 55% of the total dataset was classified as poor, 45% as of good quality. The algorithm may be considered conservative as signal segments labelled 'poor' can contain data of intermediate quality, which may explain the rather low percentage of good quality signal. What is important to note however is that there is no significant discrepancy between data collected in the hospital environment compared to data recorded in the field.

7.2.4 Discussion

Three key areas for successful fetal cardiac signal recording have emerged, the ability to operate the equipment provided, the clinical knowledge to identify and to target the best possible signal, and the degree of willingness to employ such technology.

The technology-related difficulties participants experienced were largely related to the participants' educational level (including a very high illiteracy rate amongst TBAs) and experience in handling similar equipment. As the instructions provided by the study team

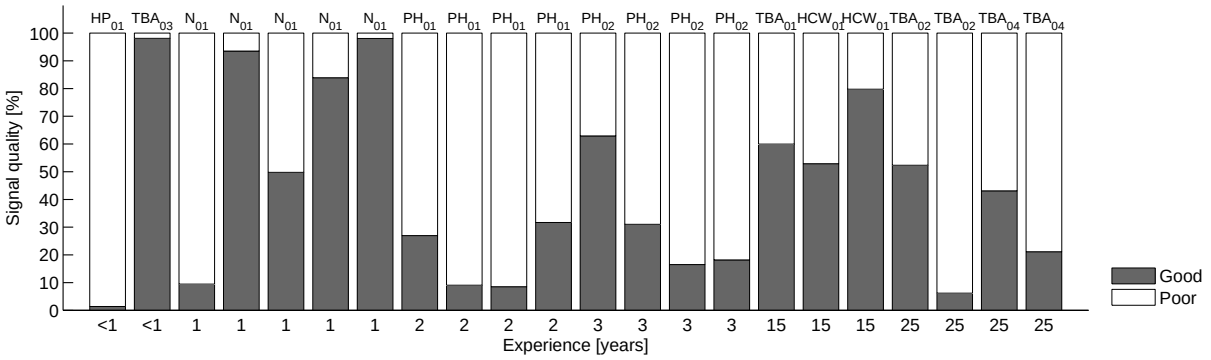


Figure 7.4: The proportion of good and poor quality in recordings sorted according to the staff's experience in antenatal care.

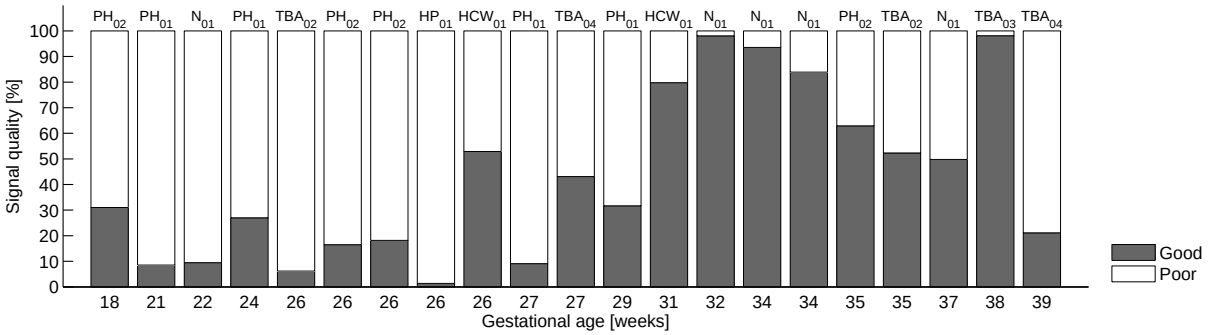


Figure 7.5: The proportion of good and poor quality in recordings sorted according to gestational age.

prior to the recording sessions were minimal and did not consider the various skill levels of the participants, several of the observed challenges could to some extent be attributed to a lack of appropriate training and guidance material.

It is noteworthy that despite only rudimentary instructions, all participants were able to successfully locate the fetal cardiac signal and complete a recording.

Participants with formal clinical education (physicians and nurse) who had used 1D Doppler at least occasionally in their practice had the ability not only to operate the device but also to understand the cardiac signal obtained, its origin, its susceptibility to maternal interference and the varying degrees of quality. Participants not trained in the use of similar devices, however, quickly developed a grasp of the technology and were able to successfully handle most recording steps with the limited instructions provided. Even though differences between formal and informally trained healthcare providers in the ease

of handling devices were apparent, the educational level did not appear to impact the quality of data collected.

Observation confirmed that 1D Doppler can be operated with minimal training only, suggesting that wider dissemination is possible. To record good quality data early on in gestation appeared difficult to both groups and will require additional training in locating, identifying and maintaining an adequate signal.

Participating staff, independent of education or experience level, were generally comfortable using the study equipment on their patients and demonstrated a high degree of confidence. Similarly all patients agreed to participate and appeared at ease throughout the recordings. These observations allow for the assumption of a well-functioning doctor/healthcare worker-patient relationship, considered essential for high quality healthcare delivery.

7.2.5 Conclusion and system implications

All participants were able to record and to identify fetal heart sounds with only six recordings (out of 25) requiring hands-on support of study staff. The high rate of successful recordings, with only minimal instructions provided prior to each patient encounter, in combination with the positive participant feedback, were encouraging indicators that the introduction and adoption of such equipment in the communities visited is feasible.

Several areas for improvement were identified during the feasibility testing. The familiarity with the technology and the clinical understanding of the physiological signals and their characteristics, the system's usability and ultimately signal quality, could all be improved by a combination of educational measures such as training and training manuals, an appropriate application interface and built-in guidance on the phone.

The implementation of signal analysis capabilities, such as the discrimination of maternal from fetal cardiac signal by comparing the pulse oximeter to the ultrasound measure, would provide real time quality improvement and reduce false readings. In addition, user feedback throughout the intervention, such as the acknowledgement of successful record-

ings, may prevent unnecessary loss of data. In the next section, improvements and their implementation are discussed.

7.3 System refinement and intervention design

The perinatal monitoring system is anticipated to be used primarily by front-line TBAs who are typically informally educated in their profession, often illiterate and currently do not use monitoring technology in their practice. By bringing together engineers[1], physicians[2], public health specialists and medical anthropologists[3], and working with the local network of healthcare workers, the aim was to create a system suited to the existing healthcare infrastructure and appropriate to the needs of the local population. The system was designed with the ultimate goal of becoming an intuitive tool for data collection, improving the diagnostic capabilities of practicing midwives and assisting with timely and accurate referral. Towards this goal three areas of system refinement and system extension were identified for the next stage of development:

1. Mobile application

    - Continuum of care: Extension of the application from prenatal assessment only, to the inclusion of intrapartum and postnatal health checks.
    - Comprehensive health checks: Inclusion of maternal blood pressure measurements and checks for pregnancy-related complications into the application work flow.
    - User interface (UI): The design of a user-centred mobile application interface, to be inclusive of an illiterate user group.
    - Signal analysis: Implementation of heart rate analysis including discrimination between the maternal and fetal rhythm.

[1] Department of Engineering Science, University of Oxford, Oxford, UK and Department of Biomedical Informatics, Emory University, Atlanta, US
[2] Maya Health Alliance Wuqu' Kawoq, Guatemala and Harvard Medical School Boston, US
[3] Department of Anthropology and Public Health, Agnes Scott College, Atlanta, US

- Emergency response: Implementation of logic to trigger calls to higher level care staff in case of detected emergencies.

2. Integration into medical records

   - Data consolidation: Automated upload of recorded signals, patient and health information to Wuqu' Kawoq's medical record system, openMRS, a globally adopted open source electronic medical record (EMR) system.

3. User training and education

   - User manuals: The development of on-phone user manuals, inclusive of an illiterate user group.
   - User training: The development of training sessions, which will teach participants in the application of all devices and the appropriate clinical knowledge to enable the recording of good quality physiological signals and the assessment of complications. (The development of the training material will not be covered in this thesis.)

7.3.1 Mobile application design

As outlined above, the mobile application was significantly extended in its functionality. This included comprehensive health checks antenatally, during birth and postnatally, on-phone signal analysis and emergency communication. A flow diagram illustrating the application's additional functionality is shown in Figure 7.6.
The Android software developed for the study is stand-alone and does not depend on existing systems such as Sana Mobile[4].

[4]Sana Mobile is a customisable mobile telehealth platform. Website URL: http://sana.mit.edu

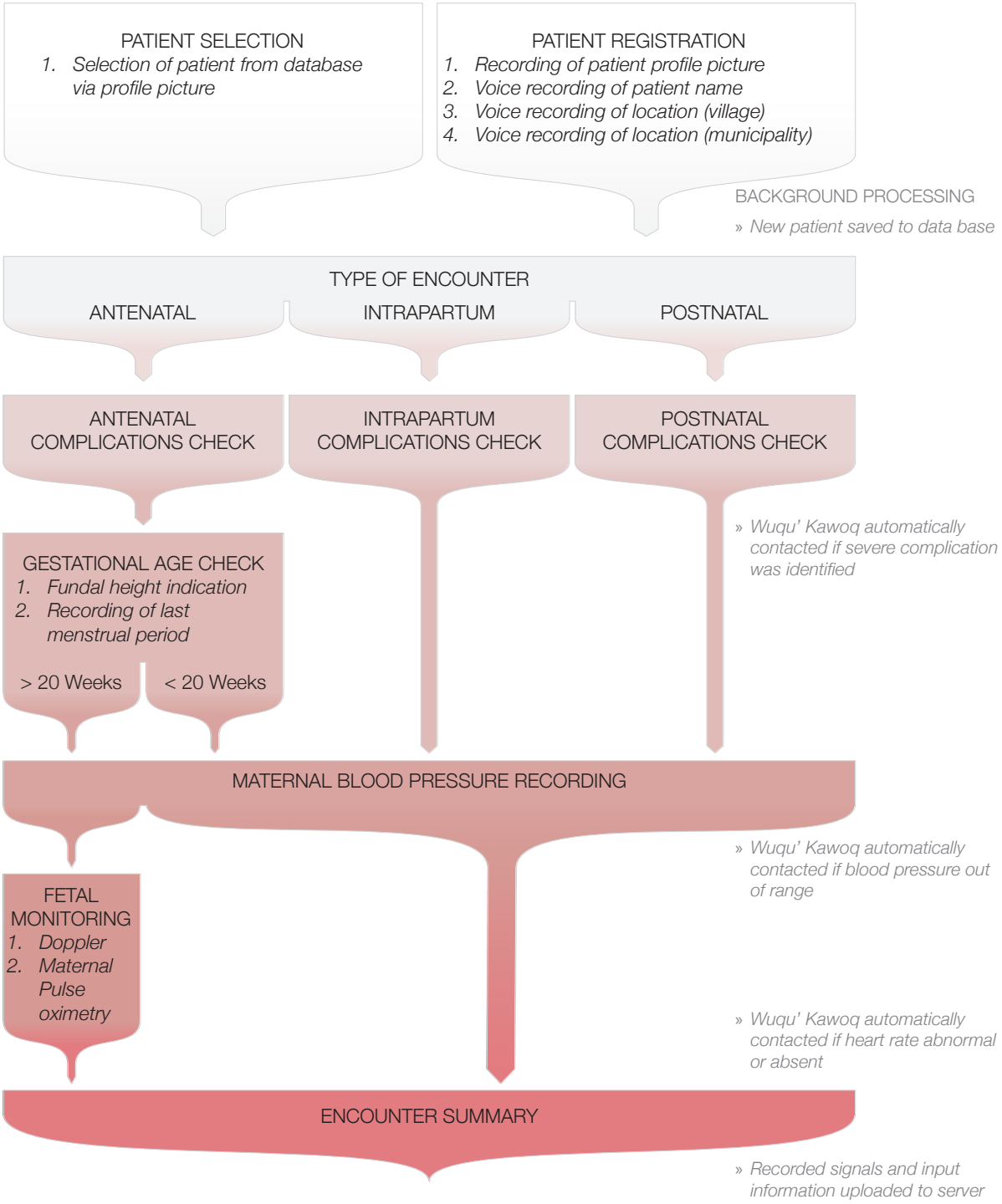


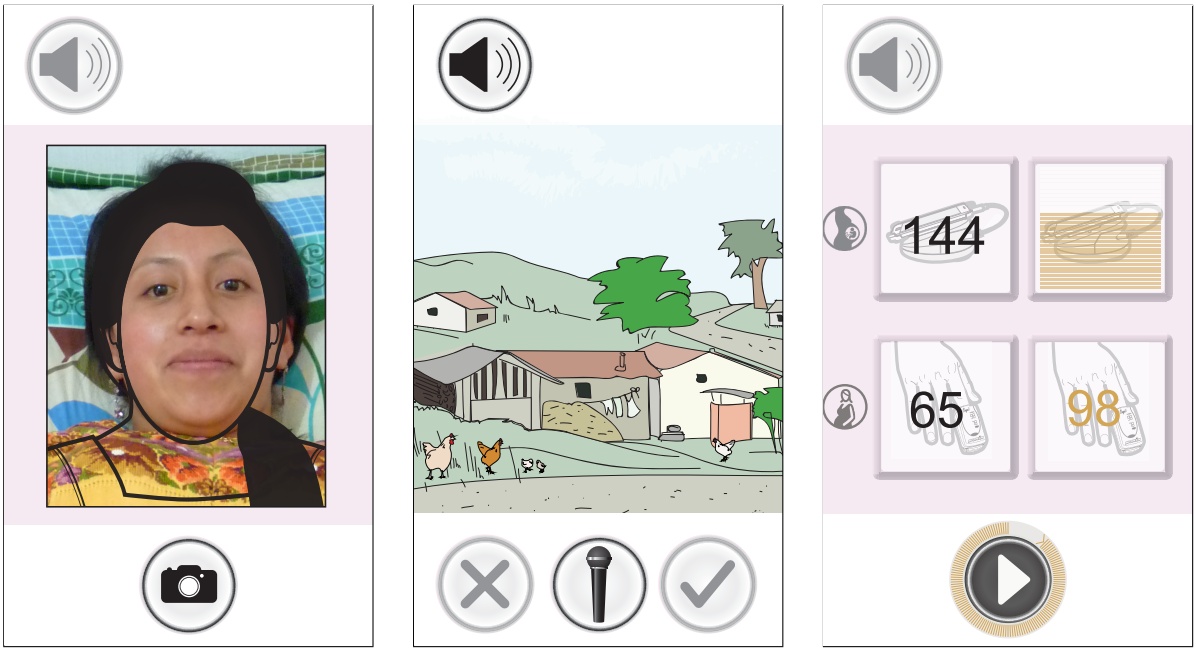
Figure 7.6: Flow diagram summarising the application steps the user is guided through during a patient visit. The application combines patient registration, guidance on and recording of pregnancy related complications and the recording of physiological signals to assess maternal and fetal wellbeing. The encounter summary and recorded signals are uploaded to Wuqu' Kawoq's central patient records.

7.3.1.1 User interface

The feasibility study revealed the limited familiarity with smart phones and confirmed the high illiteracy rate especially amongst TBAs of older age. To be maximally inclusive the application interface was therefore adapted using a combination of audio and visual instructions only, replacing any written content. Likewise the mechanism to record patient information manually was changed to voice and image capture, eliminating the need for written input. As a result, each screen presents an illustration serving as a visual clue to prompt a certain action. The illustration is accompanied by an audio playback button which provides additional audio instructions to the midwife, if required. Illustrations were drawn based on the specific use context and care was taken to guarantee their cultural relevance. All audio instructions were recorded by Wuqu' Kawoq staff in collaboration with TBAs in the local Mayan dialect Kaqchikel. To ensure the quality of the information being collected, midwives are asked to confirm each input, presented back to them visually or aurally (via a tick mark button), before progressing through the application. Three example application screens are shown in Figure 7.7. The entire application including all screens can be found in Appendix A.

7.3.1.2 Emergency response

To provide support to TBAs and to initiate immediate higher level care intervention if required, the application alerts a clinician on duty should a complication be detected. A single number is programmed into the application which uses a dynamic call forwarding system to ensure the on-call doctor is always reached. Complication detection occurs in a twofold manner. Firstly the application guides the birth attendant through a number of pregnancy related complications. If abnormalities are present, those will be confirmed on the phone by the TBA. As with all other instructions, guidance through symptom assessment is audio and pictogram based. A print version (produced for the control group of the evaluation study, described in Section 7.4) can be found in Appendix B. All visit-



(a) Screen prompting to take a patient profile photo for registration.

(b) Screen prompting to record the patient's home village.

(c) Screen providing feedback on the physiological signals being recorded.

Figure 7.7: Example screens of the mobile perinatal monitoring application. (a) Patient registration: In order to register a new patient the TBA is prompted to take a patient profile picture, guided by a positioning frame. (b) Location information: The image provides a visual clue about the information to be collected, the loudspeaker button plays audio instructions if required, and the microphone button voice records the name of the patient's home village provided by the midwife or patient. (c) The physiological measurements are displayed real time to indicate successful recording. Signals are recorded for 30 minutes, the time elapsed is visualised by a progress wheel. See Appendix A for the full set of images. Interface illustrations were produced for the specific use context, ©Lisa Stroux 2015, released under the creative commons license attribution 4.0 international (CC BY 4.0)

specific complications and the alerts they trigger are summarised in Table 7.2.

Secondly, the physiological signals recorded (1D fetal Doppler, maternal pulse, maternal oxygen saturation and blood pressure) may be indicative of an underlying fetal or maternal pathology. In addition to fetal heart rate abnormalities, indicative of fetal distress (See Chapter 2, Section 2.2.2), maternal blood pressure and oxygenation levels are reviewed with the primary aim of detecting pre-eclampsia [175–177]. The logic triggering an alert given the requisite heart rate, oxygen saturation and blood pressure is outlined in Table 7.3.

Prenatal	Delivery	Postnatal
Seizures	Membrane rupture without labour	Seizures
Haemorrhage	Prolonged labour	Postpartum haemorrhage
Membrane rupture	Multiple pregnancy	Difficult breathing
Abdominal pain	Inadequate fetal position	Abdominal pain
Difficult breathing	Placenta praevia	Fever
Blurred vision	Umbilical cord around neck	Blurred vision
Fever	Haemorrhage	Headache
Edema	Placental retention	Difficult breathing (new born)
Headache	Previous C-section	Fever (new born)
	Difficult breathing	Cold baby
	Abdominal pain	Very little baby
	Blurred vision	Difficult breastfeeding
	Fever	Umbilical infection
	Headache	Jaundice
	Edema	

Table 7.2: The list of complications the TBA is guided through dependent on the type of visit. The identification of severe complications (in red) automatically triggers an emergency phone call to the clinician on duty. If there is no signal it alerts the user to it and queues an emergency SMS which can often get through as soon as even a weak or transient signal is detected. All complications are summarised at the end of the consultation and are sent via SMS to the Wuqu' Kawoq medical record (See Figure 7.8 for an illustration of the application's data transfer backend.)

7.3.2 Back end infrastructure

Patient information is transmitted via voice, short message service (SMS), mobile data service (GPRS) and wireless network access (WIFI). A diagram summarising the back

Measurement	Logic	Detection trigger	Indication
Heart rate [bpm]	Maternal HR ≤ 60	SMS alert	Maternal bradycardia
	Maternal HR ≥ 100	SMS alert	Maternal tachycardia
	Fetal HR ≤ 120	Call	Fetal bradycardia [62]
	Fetal HR ≥ 160	Call	Fetal tachycardia [62]
Oxygen saturation [%]	Oxygen ≤ 90	Call	Maternal hypoxemia [178]
Blood pressure [mmHg]	Systolic ≤ 70	Current: Addressed at data review Future: Call	Low blood pressure
	Systolic ≥ 140, diastolic ≥ 90	Current: Addressed at data review Future: SMS alert	Mild hypertension [179]
	Systolic ≥ 155, diastolic ≥ 105	Current: Addressed at data review Future: Call	Severe hypertension [179]

Table 7.3: The logic implemented to alert clinicians if maternal and fetal heart rate, maternal oxygen saturation and maternal blood pressure readings are abnormal. Readings that require urgent attention trigger an automated call at the end of the encounter, all other measurements are included in the encounter summary SMS to inform medical staff. Blood pressure readings are currently transmitted as photo and are therefore addressed when reviewed by medical staff after data upload; optical character recognition is being implemented for the next iteration to convert the image into numeric values automatically triggering the appropriate alert. There are currently no Bluetooth enabled blood pressure monitors approved for use in pregnancy (for the detection of pre-eclampsia).

end infrastructure is shown in Figure 7.8.

To ensure an immediate higher level healthcare response, the phone automatically connects to the clinician on call for remote consultation if severe complications are identified (see Section 7.3.1.2 and Tables 7.2 and 7.3).

Once a patient assessment is completed, the application sends a text message via the SMS management software FrontlineSMS[5] to the OpenMRS[6] electronic medical record (EMR). The SMS contains coded patient information providing a summary of the patient encounter.

The system interfaces with the NGO's OpenMRS through synchronisation of a local SQLite database on the phone with the backend OpenMRS database. This occurs whenever the phones have a strong GPRS or WiFi connection and a subset of data, such as blood pressure readings or gestational age, are uploaded. This is a two-way process to ensure

[5] FrontlineSMS is an open source desktop software that enables two-way communication to mobile handsets. Website URL: http://www.frontlinesms.com

[6] OpenMRS is an open source software platform, which enables the design of customised medical record systems. Website URL: http://openmrs.org

synchronised patient data on phone and server.

The complete set of collected patient data is uploaded once a week at Wuqu Kawog's offices via WIFI to the data storage service Amazon S3[7]. This includes all large raw signal files such as Doppler and pulse oximetry recordings, which can have a size of several megabytes. Wuqu' Kawoq medical staff can review data on Amazon S3, transcribe information from Amazon S3 to the central EMR, check for missing data between data bases and review and complement patient information in the EMR.

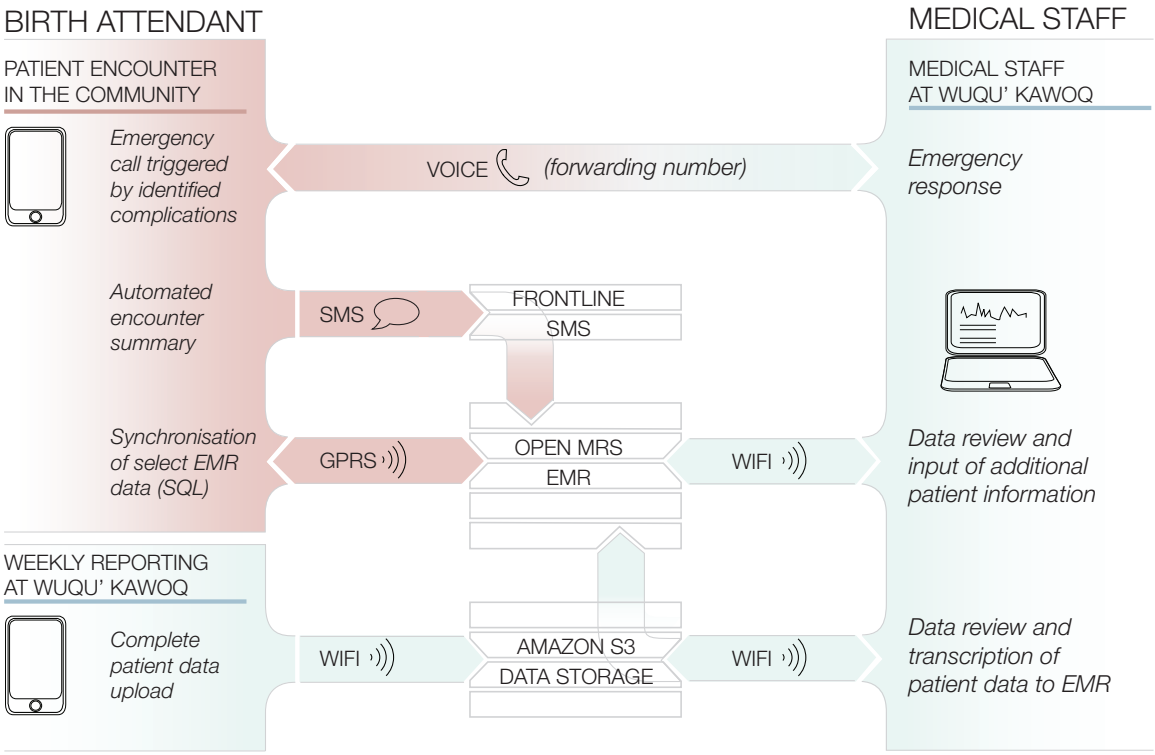


Figure 7.8: The back end infrastructure of the perinatal mobile application. Patient information is communicated via voice, text and data services dependent on urgency and connectivity. All patient information is ultimately collated in Wuqu' Kawoq's central electronic medical record system.

[7] Part of the Amazon Web Services (AWS), Amazon Simple Storage Service (S3) is a web-based data storage service. Website URL: https://aws.amazon.com/s3

7.3.3 User manuals

To train midwives in the use of the equipment and to allow for a review of how to apply all devices during a patient encounter, pictogram based manuals were developed and implemented on the phone. All pictograms were drawn based on photos depicting each action required when taking blood pressure, recording the fetal cardiac Doppler signal and the maternal pulse and oxygen levels. Photos were staged with the help of Wuqu' Kawoq staff to ensure they were familiar and culturally appropriate. Before each recording action requiring the use of a sensor, the midwife has the opportunity to remind herself of the correct procedure in a step-by-step manner. The training manuals are included in Appendix D.

7.4 System evaluation

The mHealth system described in this chapter is now being evaluated as part of the NIH funded study titled "Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala" (Research Project Grant (R01) Number: 1R21HD084114-01). The detailed study protocol can be found in Appendix C. The study was approved by the Wuqu' Kawoq and Emory University institutional review boards (Wuqu' Kawoq IRB approval number: WK-2015-001, Emory University protocol record: IRB00076231) to ensure compliance with ethical standards.

The study is a randomised control trial (RCT) involving 50 traditional birth attendants (TBAs) in the department of Chimaltenango, and an overall enrollment of approximately 500 pregnant women and their newborns over the duration of 12 months. Initially, 25 midwives have been randomly assigned to receive the study equipment (mobile phone, 1D Doppler, pulse oximeter and automated blood pressure cuff) at the start of the project[8].

[8] Randomisation was constrained to balance both travel time to Wuqu' Kawoq headquarters and age of TBAs. Participants were randomised by village to reduce interaction between arms and potential envy.

The remaining 25 will receive laminated decision support cards which mimic the logic of the phone application and use the same pictograms as are included in the mobile phone app. They will receive the same conceptual training for using the cards as the 25 midwives that receive training on the mHealth system. Equal training time has been alloted to both arms of the study. In order to establish a study baseline (such as referral rates) the remaining 25 will receive the equipment with a delay of 6 months at which point both arms of the study will be trained (or retained) in the use of the equipment. Recruitment is currently being performed by dedicated Wuqu' Kawoq staff. TBA training will commence in January 2016 and patient encounters are anticipated from February 2016 onwards.

Table 7.4 provides a summary of the data collection and processing features deployed for testing to date, and of the plans for future implementation work, which will continue in parallel to the study.

Algorithms / collected parameters	On-phone	Cloud-based	Implementation status
Patient enrollment	X		Implemented
Gestational age	X		Implemented
Ante-, intra- and postpartum complications	X		Implemented
Blood pressure readings	X		Implemented
Blood pressure logic (Table 7.2)		X	Future plan: point-of-care assessment
Maternal O_2 reading and logic (Table 7.2)	X		Implemented
Maternal HR reading and logic (Table 7.2)	X		Implemented
Doppler SQI (Chapter 5)	X		Implementation work in progress
Fetal HR by autocorrelation (Section 2.3.1.1)	X		Implemented
Fetal HR by HSMM approach (Chapter 6)		X	Future plan: Beat-to-beat HR for IUGR detection
IUGR detection		X	Future plan: remote IUGR screening

Table 7.4: The status of Android implemented study parameter collection and algorithm deployment. The table also provides an indication whether data capture and processing occurs on the Android device at point-of-care or cloud-based post patient visit. Future plans include the on-phone assessment of blood pressure readings, which in the absence of pregnancy-approved Bluetooth devices is anticipated in the form of OCR (optical character recognition) processing, the further development and cloud implementation of the beat-to-beat HR estimation algorithm presented in Chapter 6 and post-visit remote IUGR risk assessment (Chapter 3).

7.4.1 Expected study outcomes

Two primary outcomes are anticipated in this proof-of-concept study: an increased appropriate antenatal referral rate and an increase in timely evaluation of neonates, conducted

by TBAs at delivery and by higher level care staff where indicated after initial assessment. The data being collected through the mHealth-EMR system will furthermore allow for the analysis of the system's technical efficacy and process evaluation. Clinical measures such as maternal, fetal and neonatal mortality and morbidity will also be collected, however, the study is not powered sufficiently to translate those into authoritative indicators (see Section 1.1.4 for the context specific health statistics). Participating TBAs will be tested following training provided to assess the amount of clinical knowledge and tech-literacy gained. Post-intervention interviews with the 50 TBAs will be conducted to further understand their attitudes toward the mHealth-EMR system, perceptions of self-efficacy while using the system, and input for further implementation. Direct observation will also give additional insight into the ease-of-use of the system.

The key process outcomes and health indicators that will be extracted from the EMR, from interviews and observation are summarised in Table 7.5.

Study indicators and outcomes
• Success and accuracy of data transmission; • Intelligibility/ease of use of decision support algorithms, phone and peripheral devices; • Direct observation of correct device use by study staff; • Device failure (breaking, faults, damage).
• Medical knowledge/background of TBAs following training; • Proportion of pregnancies in which device was used in routine antenatal care; • Rate of prenatal referral to higher level of care based on TBA knowledge/assessment alone, with use of device/decision support; • Proportion of time postnatal team aware that a birth is imminent, birth has occurred; • Hours from time-of-birth to evaluation of neonate by postnatal team; • Proportion of referrals generated through TBA consultations; • Proportion of referrals assigned in timely manner (<30 days); • Proportion of referrals handled via 'Straight-Through-Processing' from electronic system, TBA-directed referrals only; • Proportion of follow-ups generated through TBA assessments using project associated communications, additional communication channels.
• Proportion of adverse maternal outcomes; • Neonatal mortality rate.

Table 7.5: List of study indicators collected and anticipated outcomes analysed during the mHealth system evaluation period.

7.5 Conclusion

The initial user testing provided a positive indication that introducing monitoring technology into the practice of traditional birth attendants is indeed feasible. This led to the system's extension and refinement as described in this chapter, working closely with Wuqu' Kawoq staff and representative local TBAs. The approach of including frontline healthcare workers in the design process shows promise to not only create a system able to address healthcare shortages prevalent in the region, but a system that relates to the cultural context, maximising the chances of successful implementation and ultimately adoption.

The proof-of-concept evaluation study under way will test those assumptions and is expected to uncover additional improvements required, with respect to technology, process and usability.

The mHealth intervention presented in this thesis takes the approach of integrating TBAs into formal care delivery. This acknowledges TBAs as a valuable part of the community, as opposed to side-stepping their services in order to achieve higher levels of facility-based care as suggested by some [52]. The study will show whether this alternative approach will in fact lead to improvements in early risk detection and adequate referrals as anticipated.

At present, no reliable data covering the region of Chimaltenango is available on causes of maternal and fetal compromise, the latter of which is particularly scarce. Frequent and meaningful monitoring reaching families in remote areas, together with centralised record keeping as made possible with the system described, is expected to create a richer knowledge base enabling the design of studies adequately powered to investigate the health impact of such intervention. Similarly there is no information available on TBA-initiated referral rates to higher level care facilities. The study will establish a baseline of the birth attendants' abilities to adequately refer with training on the assessment of complications only, and allow for comparison with referral rates by TBAs equipped with the monitoring

application and devices. The physiological signals together with other health metrics collected during the study period will support further algorithm development towards the reliable detection of pre-eclampsia, fetal distress and IUGR.

Great care has been taken to tailor the presented application interface and mHealth intervention to the specific social and cultural realities and infrastructural challenges of Guatemala and the Chimaltenango region in particular. The core components of the system including its back end and decision support are however directly transferable, and the approach to UI design and intervention implementation easily replicable, suggesting that use in other low-resource contexts is viable.

Chapter 8

Summary and conclusion

8.1 Summary of contributions

A number of contributions towards improved perinatal health assessment in scarcely resourced environments has been presented in this thesis.

1. The fetal cardiac Doppler signal has been used in clinical practice as an indicator for fetal wellbeing for decades, however development has largely been dominated by few industry players leading to black-box devices and raw data being inaccessible for development efforts. The Doppler signal is a low-cost and relatively easy-to-obtain information source and hence suited for use in settings with often only informally trained healthcare workers available. The raw data collected during the course of this thesis formed the basis for the investigation into novel approaches to fetal Doppler processing. A substantial amount of work has gone into laying the foundations for continued data collection to significantly grow the unique dataset of raw 1D Doppler signals over the next 12 months to enable future development.

2. A first mHealth prototype was built for data capture (fetal Doppler and maternal pulse oximetry) and to test the feasibility of introducing such technology in rural Guatemalan communities not familiar with its use or equivalent devices (Chapter 4).

Building on initial insights the prototype was extended to become a screening platform tailored to the needs of practicing traditional birth attendants, empowering traditional birth attendants to improve perinatal care whilst remaining a midwife-led intervention (Chapter 7). The system extension was achieved in collaboration with engineers[1] at the Department of Biomedical Informatics, Emory University, Atlanta, medical staff and traditional birth attendants[2] at the Maya Health Alliance Wuqu' Kawoq in Guatemala, and medical anthropologists[3] at Agnes Scott College, Atlanta. The system is currently being trialled in a proof-of-concept study in rural Guatemala.

3. Given the challenging use scenarios anticipated, such as limited medical training and a difficult work environment (often poorly lit and with high noise levels), an indication of a signal's quality is crucial. A novel 1D Doppler quality measure was therefore developed and tested, as described in Chapter 5.

4. Of particular concern is the reliable detection of intrauterine growth restriction as its prevalence in developing countries is high and preparedness is crucial to enable appropriate postnatal care. The timely referral to higher level care could improve outcomes substantially, particularly in rural communities. Based on an extensive database, made available by the John Radcliffe Hospital in Oxford, the feasibility of solely using metrics computed from the Doppler heart rate trace to identify IUGR risk was assessed and a classification algorithm presented (Chapter 3).

5. IUGR risk metrics based on the fetal heart rate signal may be improved by beat-to-beat heart rate estimation. A segmentation algorithm has been introduced in Chapter 6 towards this goal.

[1]Lead in the Android and back end programming.
[2]Lead in the design of the healthcare intervention, responsible for the trial coordination and providing contextual feedback.
[3]Lead in designing the study training materials and evaluation.

Contributions are further specified in the following sections and their potential for future development outlined.

8.1.1 1D Doppler database

To date there is no raw fetal cardiac Doppler database readily available, which severely limits research in this area, in particular without commercial ties. During the course of this thesis approximately two hours of fetal cardiac Doppler signal have been collected at the John Radcliffe (JR) Hospital in Oxford and over four hours in rural Guatemala. All recordings are accompanied by a concurrent maternal heart rate signal, maternal oxygenation and the phone's accelerometer data. The data collected at the JR Hospital were also annotated by three independent annotators for signal quality and beat instances.

The current database will grow significantly over the next 12 months with a minimum of 375 antenatal monitoring encounters envisaged as part of the ongoing proof-of-concept study in Guatemala (see Chapter 7 and Appendix C for details on the study subjects and protocol). Each recording (synchronous 1D fetal Doppler, phone accelerometry and maternal pulse and oxygen levels) will last a minimum of 10 and up to 30 minutes increasing the database significantly to approximately 150 hours.

The data collected to date contains signals of healthy fetuses only. To further develop the work presented in this thesis it will be important to complement the dataset with pathological cases, in particular signals containing fetal cardiac arrhythmia (for the continued improvement of robust quality and segmentation algorithms), and signals of growth-restricted babies (for the implementation of the IUGR risk detection algorithm).

Given the substantial sample size of the ongoing study in Guatemala it is highly likely that several cases of growth restriction will be recorded as prevalence rates are typically 11% [51] within the population. The data collected may also include fetuses with irregular cardiac rhythms, however numbers are expected to be much lower as the prevalence rate for this condition is estimated at only 2% [63].

The study in Guatemala was set-up to record birth outcomes where possible, providing

important additional metrics such as gestational age at birth and fetal birth weight to retrospectively confirm growth-retardation.

To further develop and evaluate presented signal quality and heart rate estimation algorithms, quality and beat annotations of the data already collected and about to be recorded in Guatemala would be extremely valuable.

8.1.1.1 Fetal Doppler and fECG dataset

In addition to new data being recorded as part of the described study in Guatemala a collaboration was initiated in the past year with the Technical University of Dresden[4] and the University Hospital in Leipzig[5], Germany, with the specific aim of enlarging the Doppler database and building a dataset of pathological cases including a synchronous fetal abdominal reference signal. The study was designed by the collaborators with the primary objective of analysing fetal heart rate variability measures, derived from the abdominal fetal ECG, and their potential of detecting fetal compromise (a summary of study parameters is provided in Table 8.1). In addition to maternal and fetal ECG, data acquisition also includes maternal respiration, continuous blood pressure and the maternal plethysmogram. The controlled variation of breathing and blood pressure is part of the study protocol to assess their impact on the fetal-maternal cardiac regulation. The hand-held 1D Doppler was integrated into the study set-up for synchronous recording as shown in Figure 8.1. The procedure of data acquisition during a patient encounter is illustrated by the timeline in Figure 8.2.

The first phase of testing and refining the study setup has been completed in Leipzig, data acquisition is now in progress and expected to be ongoing over the next 12 months. An example of a good quality set of recorded signals is shown in Figure 8.3. The Oxford-

[4] PIs: Dr.-Ing. Sebastian Zaunseder, Prof. Dr.-Ing. Hagen Malberg, student: Fernando Andreotti, Institut für Biomedizinische Technik TU Dresden, Georg-Schumann-Straße 9, 01187 Dresden.

[5] PIs: Dr. med. Alexander Jank, Prof. Dr. med. Holger Stepan, student: Julia Kage, Abteilung für Geburtsmedizin Universitätsklinikum Leipzig, Liebigstraße 20a, 04103 Leipzig.

Study parameters	
Study period	October 2014 to October 2016
Target subject numbers	Controls: 100
	IUGR: 20
	Fetal arrhythmias: 20
	PROM: 20
	Fetal cardiac abnormalities: 20
Gestational age	20 - 28 weeks
Recording duration	30 minutes (max. 25 minutes Doppler)
Signal acquisition	Abdominal fECG (7 channels)
	mECG (1 channel)
	Maternal continuous blood pressure
	Maternal plethysmogram
	Maternal respiration
	1D fetal Doppler

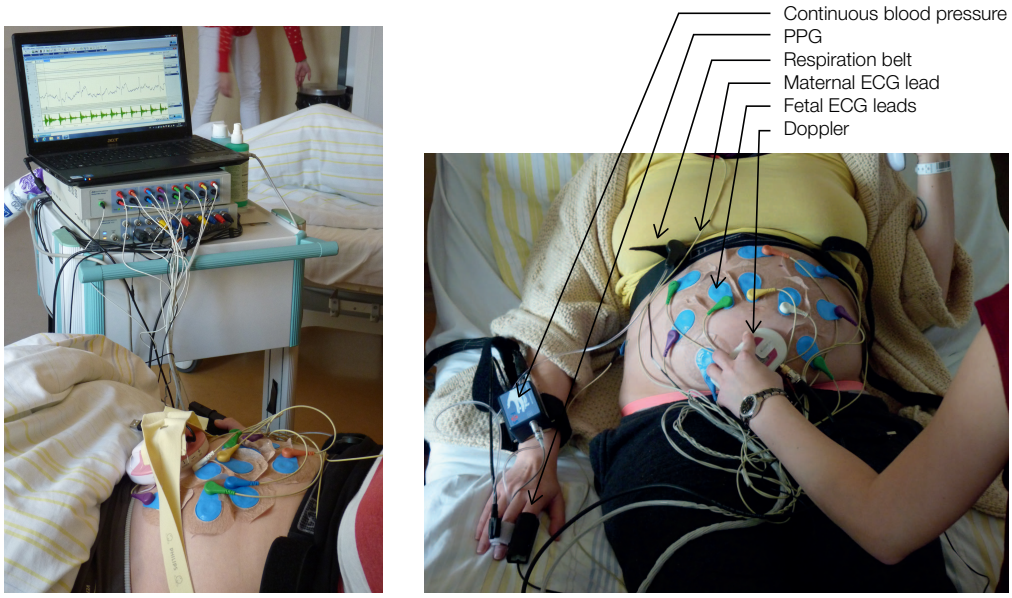
Table 8.1: The primary study parameters. The collaboration was initiated with a view to complement data already collected with relevant pathological cases (highlighted in red) for further algorithm development and evaluation. The Oxford-arm of the study will focus on the analysis of the Doppler signal utilising the fetal and maternal ECG as reference (highlighted in red).

arm of the study will focus on the analysis of the Doppler signal in conjunction with the gold-standard fECG reference.

8.1.2 SQI

A 1D Doppler signal quality index was developed in Chapter 5 with the objective to provide real-time feedback on the signal quality to the operator when monitoring the fetal cardiac activity. The low-complexity logistic regression model (Equation 5.17) selected for phone implementation combines three features based on the signal's wavelet coefficients, sample entropy and the mobile phone's acceleration data. The classifier could distinguish poor from good quality epochs with 95% accuracy.

To ensure the classifier's robustness, a database more substantial in size and diverse in pathology would be of great benefit. To significantly increase the current dataset, the Guatemalan data already collected, as well as the new data anticipated to be added over the next 12 months from Guatemala and Germany, will require expert annotation



(a) The data acquisition setup (b) Monitoring devices

Figure 8.1: The study setup for data acquisition.(a) All signals are recorded synchronously onto a study laptop using the software LabChart7. (b) Five monitoring devices are applied in the supine position.

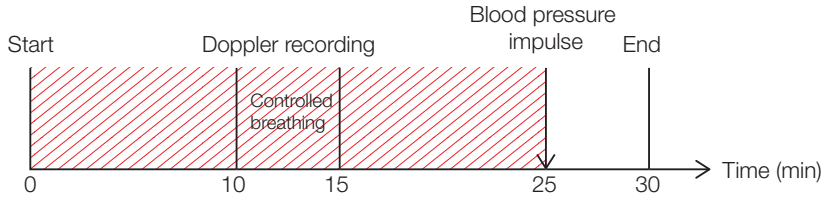


Figure 8.2: During each patient encounter signals are recorded for 30 minutes. An episode of controlled breathing is initiated at 10 minutes and a blood pressure impulse (sitting up) at 25 minutes, the Doppler being removed at that point.

of signal quality for algorithm training and testing. A web-based implementation of the annotation interface that was programmed as part of this thesis (presented in Chapter 5, Figure 5.1) is suggested to facilitate this process. As mentioned previously, some cardiac rhythm abnormalities may be expected in the data which are going to be recorded over the course of the study in Guatemala. The relatively challenging circumstances of data collection and healthcare provision, however, may make diagnosis difficult. The data arriving from Leipzig on the other hand are recorded in a hospital context and will contain cases of identified pathologies, presenting a valuable addition to the database.

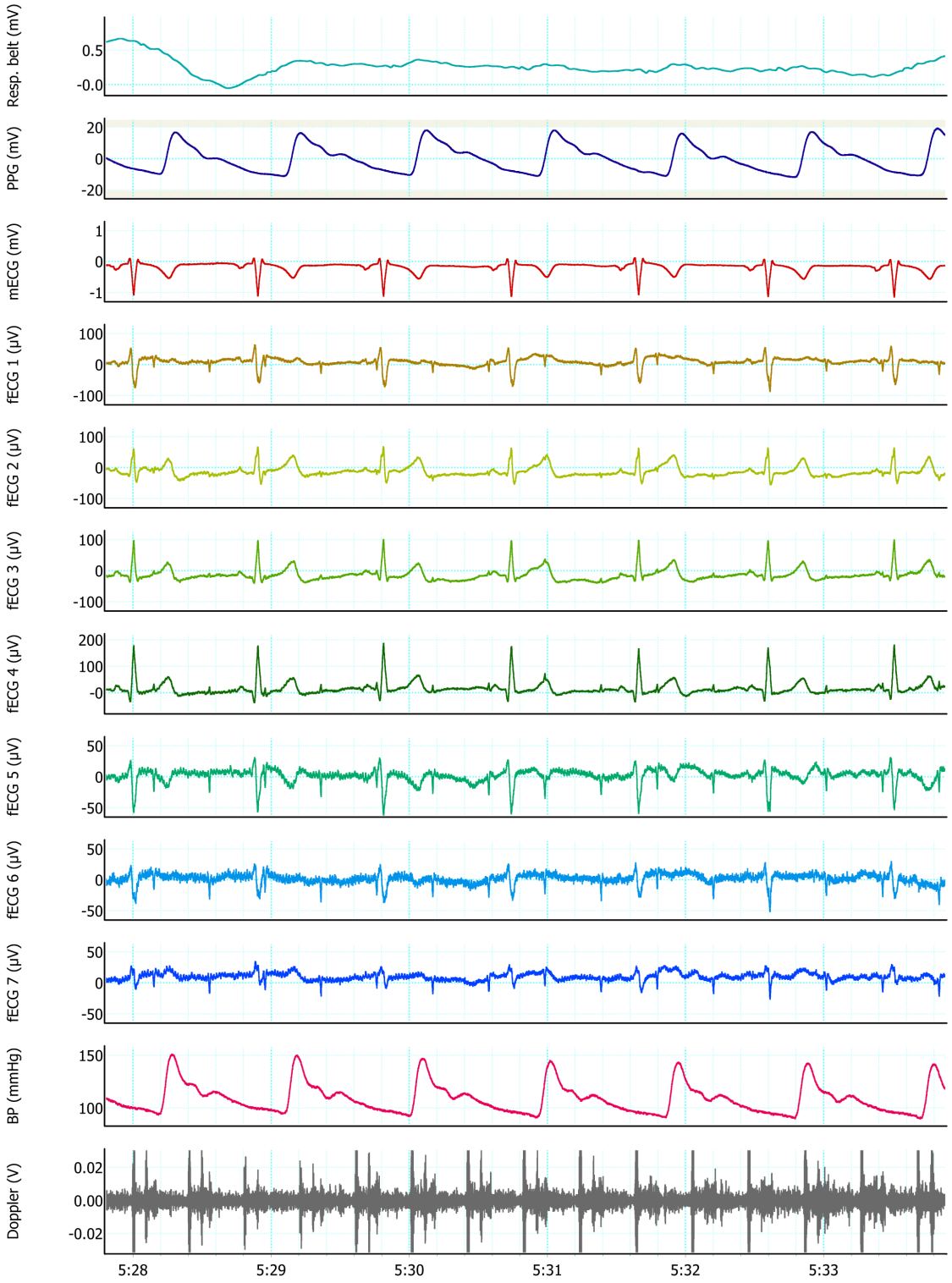


Figure 8.3: An example segment of good quality synchronous signals collected from a healthy control. All signals are recorded in parallel onto a study laptop using the software LabChart7.

8.1.3 IUGR risk detection

Chapter 3 investigated the utility of heart rate variability markers computed from the heart rate trace for the detection of IUGR, which could be confirmed for fetuses below 35 weeks of gestation; the multivariate LR model including gestational age, short- and long-term variability in high variation episodes, the average duration in active sleep and the number of active episodes in the trace resulted in an AUC of 76% on held-out test data. Results included two important insights; firstly they showed that early onset IUGR can indeed be identified, whereas discrimination in late onset IUGR was poor, and secondly they confirmed that detection improved when comparing markers computed during episodes of active sleep.

For the future implementation of the risk detection algorithm it is therefore crucial to take gestational age into account. It is furthermore recommended to also implement automated sleep state analysis as applied by the Dawes/Redman system, described in Section 3.4.2, prior to the computation of heart rate variability (HRV) markers from the heart rate trace. The risk detection algorithm will require further evaluation on raw IUGR data currently collected in Germany.

For the analysis of growth restriction risk selected variability markers will need to be computed from the heart rate trace. The sleep state analysis, the computation of variability metrics and ultimately the IUGR risk assessment are not required real-time and on-phone, but can be executed remotely.

Research into features able to detect late-onset IUGR (from 35 weeks of gestation onwards) remains an important challenge to be addressed in future investigations.

To further augment the IUGR risk assessment algorithm or to prioritise screening for IUGR amongst patients, additional indicators could be taken into account in future development work. As outlined in Section 2.2.3.1 the risk for fetal growth retardation increases with maternal and fetal complications, as well as due to socio-economic factors [65,180].

Indicators that may be recorded during midwife-patient encounters include: constitutionally small mother or low pre-pregnancy weight, extremes of maternal age (<16 years, >35 years), maternal hypertension, poor maternal weight gain and malnutrition, low socio-economic status, smoking, alcohol and drug abuse.

8.1.4 HMM heart rate trace

The hidden Markov model approach to fetal heart rate estimation was introduced in Chapter 6 with the objective to develop an alternative to autocorrelation techniques that would approximate beat-to-beat accuracy and ultimately improve the predictive utility of variability metrics for IUGR detection. The developed algorithm showed promising results compared to heart rate estimation by autocorrelation and in accuracy of interval estimation tested against manual annotations.

The chapter highlights the limitations of manual beat annotations that were available for evaluation purposes. Fetal ECG is therefore suggested as a more robust measure against which to test the beat-to-beat segmentation algorithm. The data being recorded in Leipzig are collected towards this aim; refer to Figure 8.4 for an example recording of a Doppler signal with one synchronous fetal ECG channel including identified R-peaks. The Leipzig data will also provide a set of IUGR cases required to confirm the assumption that variability measures computed from beat-to-beat heart rate traces are indeed superior in their predictive power.

8.1.5 Design of a low-cost perinatal monitoring platform

The hardware selected for the fetal monitoring system, the hand-held Doppler, mobile phone and pulse oximeter, were assembled in a robust way both regarding safe handling of the system when used on patients and with regard to reliable data capture. A bespoke audio recording cable was designed and an Android application programmed to reliably record the Doppler's audio signal (cable connection), the maternal pulseoximetry (Blue-

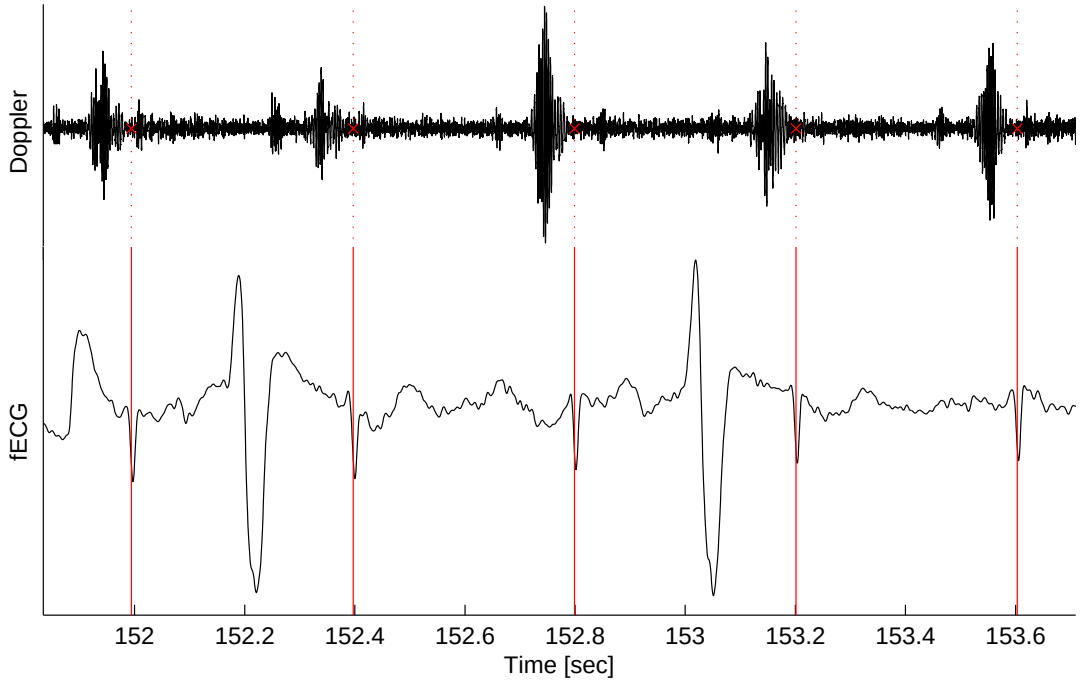


Figure 8.4: An example of a Doppler signal recorded in Leipzig, plotted alongside one of the fetal ECG reference signals with identified R-peaks marking the beat instances.

tooth) and the phone's internal accelerometer data (Chapter 4). The initial user testing in Guatemala (Section 7.2) informed the further development of the mobile Android application, implemented predominantly by engineers of the Department of Biomedical Informatics at Emory University, and the TBA training requirements for the proof-of-concept study, developed and implemented by the project anthropologist at Agnes Scott College and staff at Wuqu' Kawoq. For the extended mobile application a context specific user interface was designed (Appendix A), taking into account the characteristics of the target user group, such as high illiteracy rates and limited familiarity with smart phones. Similarly pictogram based user manuals were created to guide TBAs through the application of all monitoring devices (Appendix D). The ongoing proof-of-concept study will generate additional insights iteratively improving the mobile application and healthcare intervention over the coming months.

The Doppler data analysis implemented to date is limited to heart rate estimation by autocorrelation and logic alerting the user to an abnormal fetal heart rate or a heart

rate close or equivalent to the maternal reference (obtained by pulse oximetry); the latter suggesting that the transducer is processing a signal from the maternal aorta instead of the fetal myocardium and needs to be repositioned. Further analysis capabilities are anticipated as follows. The on-phone implementation of the SQI developed in Chapter 5 is planned in the near future to be field-tested in Guatemala. One of the primary objectives this thesis has worked towards is to enable IUGR risk screening; the detection algorithm developed in Chapter 3 was a first step towards this goal. Its implementation will require the identification of quiet and active sleep as described in Section 3.4.2 and the computation of short- and long-term heart rate variability according to Dawes and Redman [108]. Provided further investigations will confirm the accuracy of the developed HMM heart rate estimator and the superiority of the beat-to-beat heart rate trace when computing variability metrics for IUGR detection, the HMM approach should be chosen over autocorrelation based methods.

The translation of additional Dawes/Redman criteria [108] for on-phone use, such as episodes of heart rate acceleration and deceleration, could be considered to complement the fetal wellbeing assessment both antenatally as well as in labour. As mentioned above, IUGR risk assessment may occur remotely.

8.1.5.1 Scaling up

The results of the ongoing clinical trial will provide detailed information on whether proposed mHealth approach changes practices and outcomes when implemented in low resource communities. If results prove positive, the challenge of how to scale up such system across a fractured political and healthcare infrastructure needs to be tackled. Without a national healthcare service to mandate and pay for the use of the equipment, the likelihood of adoption will rest on the shoulders of non-governmental organisations (NGOs), such as Wuqu' Kawoq, or private enterprise.

If the benefit of the system developed in this thesis can be demonstrated to other NGOs, they may well adopt it and integrate it into their own workflows (adapted to their target

population concerning for example language and pictograms used). If scaling relies on private enterprise it will require the establishment of a company to supply and maintain the hardware (amortised on a per use basis down to an affordable charge per birth). This appears feasible given the low cost of the individual hardware components.

Although many of the technical, cultural and medical practice barriers have been addressed, the high level political issues remain, including the lack of EMR integration across healthcare systems. However, since the system can be tailored to each NGO's or hospital's EMR (if they have one) then this may present less of a barrier than first thought. Moreover, the system can function without an EMR component with several of the algorithms run on the phone and a call and/or text function to alert a medical professional when further advice, treatment or intervention is required.

Regardless of the route to scale-up the comprehensive coverage of medical care for those identified by the mHealth system as at risk needs to be ensured.

8.2 Conclusion

The feedback received in Guatemala, observation of study participants as well as signal quality analysis performed on data collected in Guatemala suggested that it is a feasible proposition to introduce an mHealth approach to antenatal care into an extremely low-tech environment. The acceptability of target users and their ability to handle the monitoring equipment was convincingly high. The co-development approach taken to designing an appropriate user interface and learning material demonstrated that a system inclusive of illiterate and untrained healthcare workers can be created.

Experiments demonstrated that it is possible to distinguish poor from good quality Doppler data with high accuracy. This is essential in order to provide real-time user feedback on the phone. The detection of IUGR risk was also found to be possible based on heart rate variability metrics, even though their utility is limited to the assessment of early onset growth restriction. Finally, a promising alternative to heart rate estima-

tion was introduced with the potential to improve the discriminative power of variability metrics in their prediction of IUGR.

This thesis has provided many key answers towards practical solutions for low-cost fetal wellbeing assessment in resource-constrained regions, particularly with illiterate populations with little exposure to technology. Although the results from the RCT are over a year away, it is hoped that this proves to be an impulse for continued investigation into addressing some of the challenges faced by midwives and mothers in such environments.

Appendix A

Mobile Application

The following application screens show the example of an antenatal visits as implemented on the mobile phone. Screens are laid out in chronological order and from left to right. The active buttons on the interface are labelled in black, explanatory labels are shown in grey. All illustrations were produced for the specific use context, (c) Lisa Stroux 2015, released under the creative commons license attribution 4.0 international (CC BY 4.0). Complication pictograms were adapted from [181].

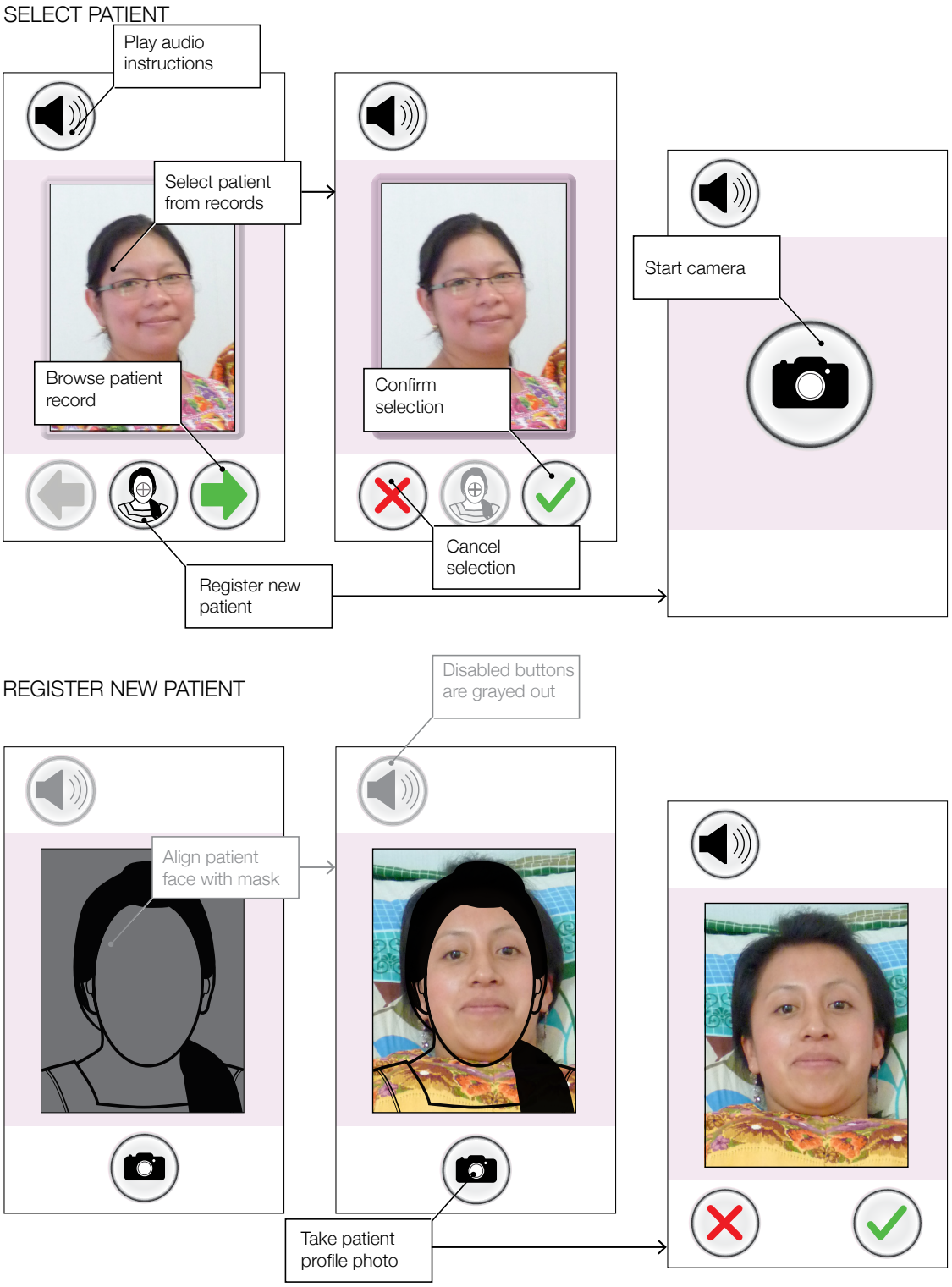


Figure A.1: To start the assessment the patient can be selected by photo from the records if she has been registered during a previous visit. Alternatively a new patient can be registered by taking a profile picture which is added to the database. Throughout the application audio instructions are provided to the TBA if desired (active speaker button).

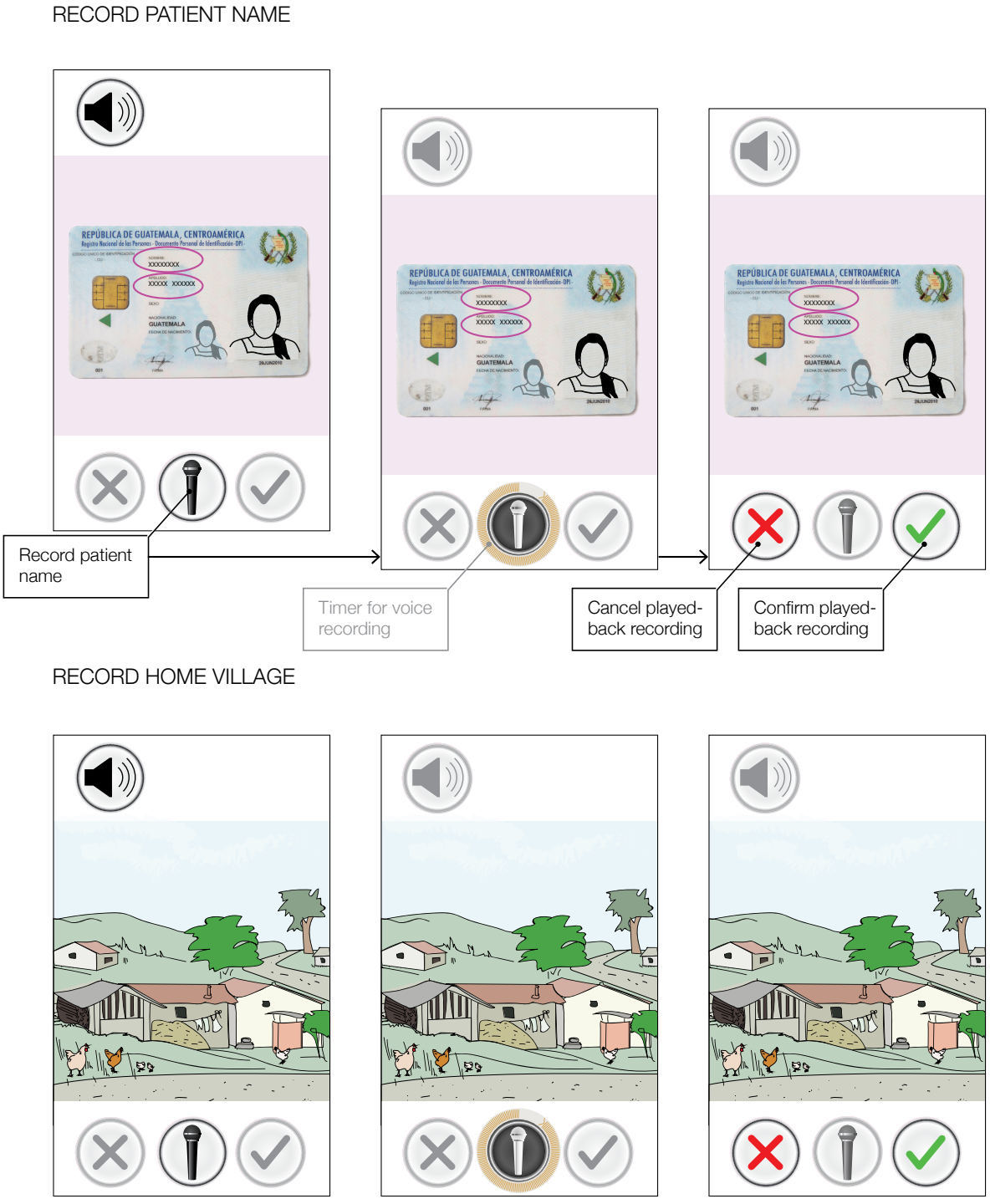


Figure A.2: Patient information such as name and home village is voice recorded by the TBA and automatically played-back for confirmation.

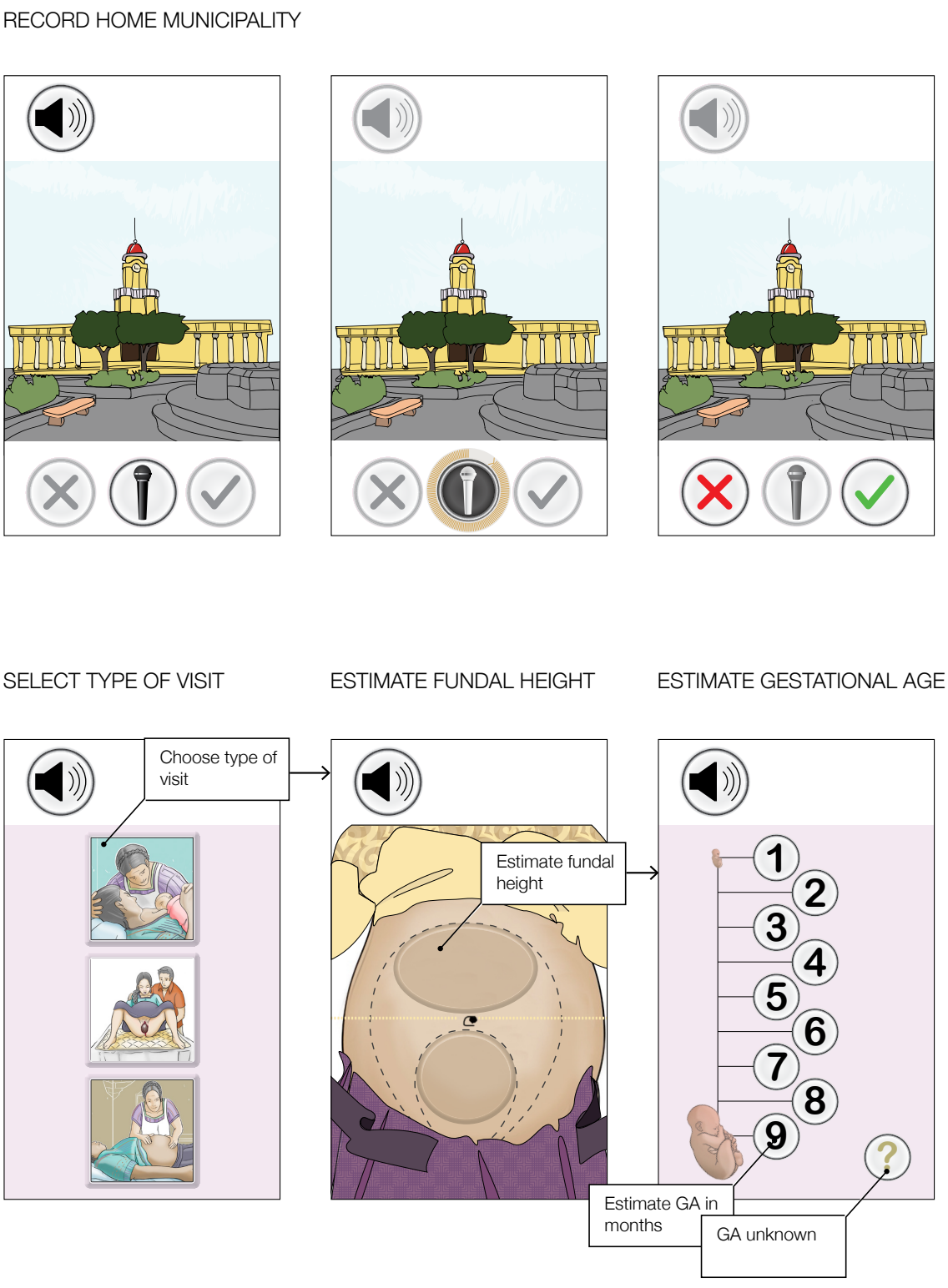


Figure A.3: In addition to the home village the municipality is recorded in the same manner. Once the patient's personal information is stored, the TBA selects the type of visit (antenatal, labour, postnatal). The antenatal examination includes the estimation of fundal height and gestational age. Gestation under 20 weeks will disable the pulse oximeter and Doppler recording function as identifying the fetal heart signal can be challenging leading to high rates of false alarms. Fundal height is estimated as above or below the umbilicus, the approximate 20 week mark, gestational age is estimated as months pregnant.

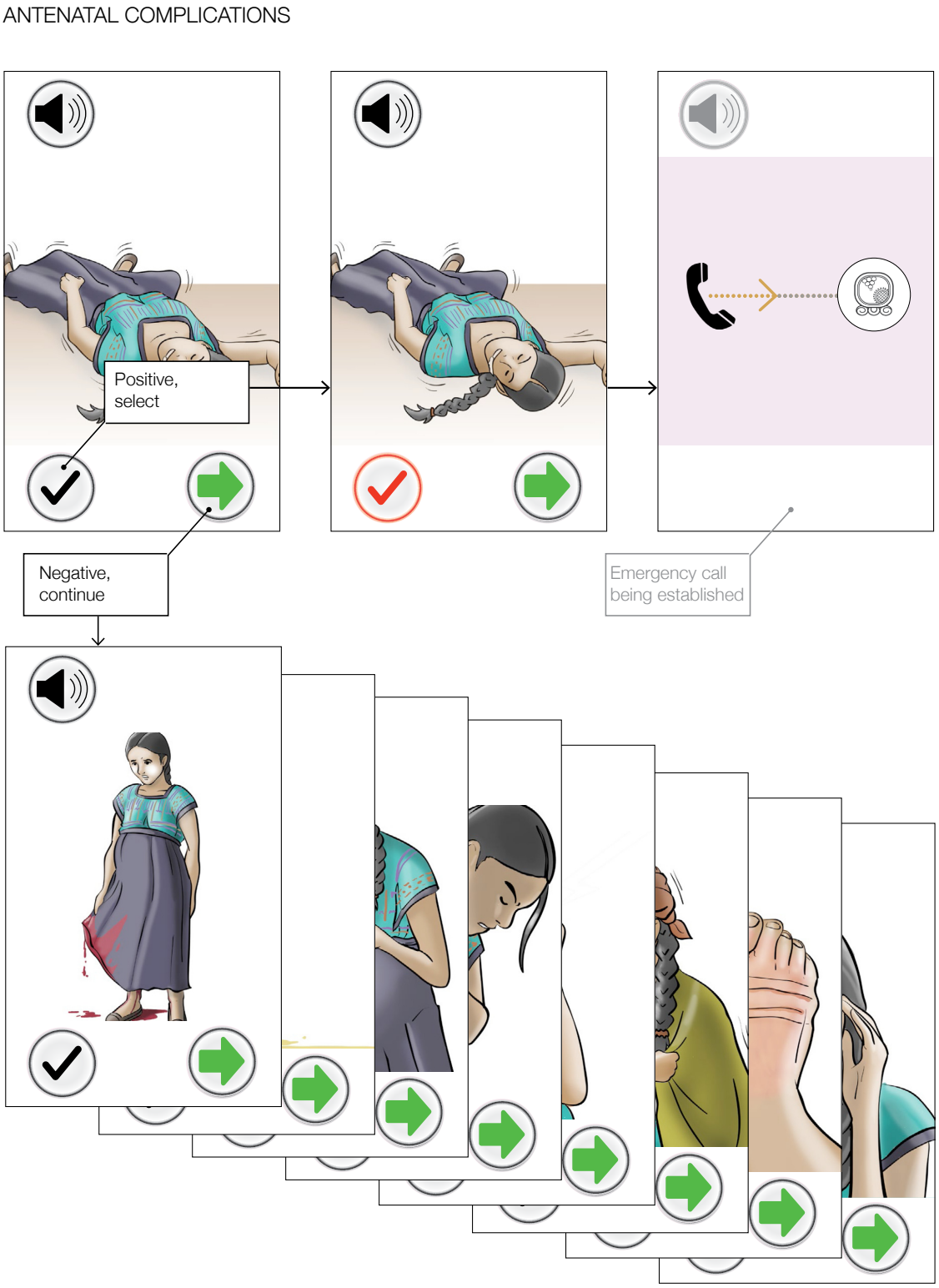


Figure A.4: The TBA is guided through applicable complications to assess. Severe complications automatically trigger an emergency phone call to the clinician on duty.

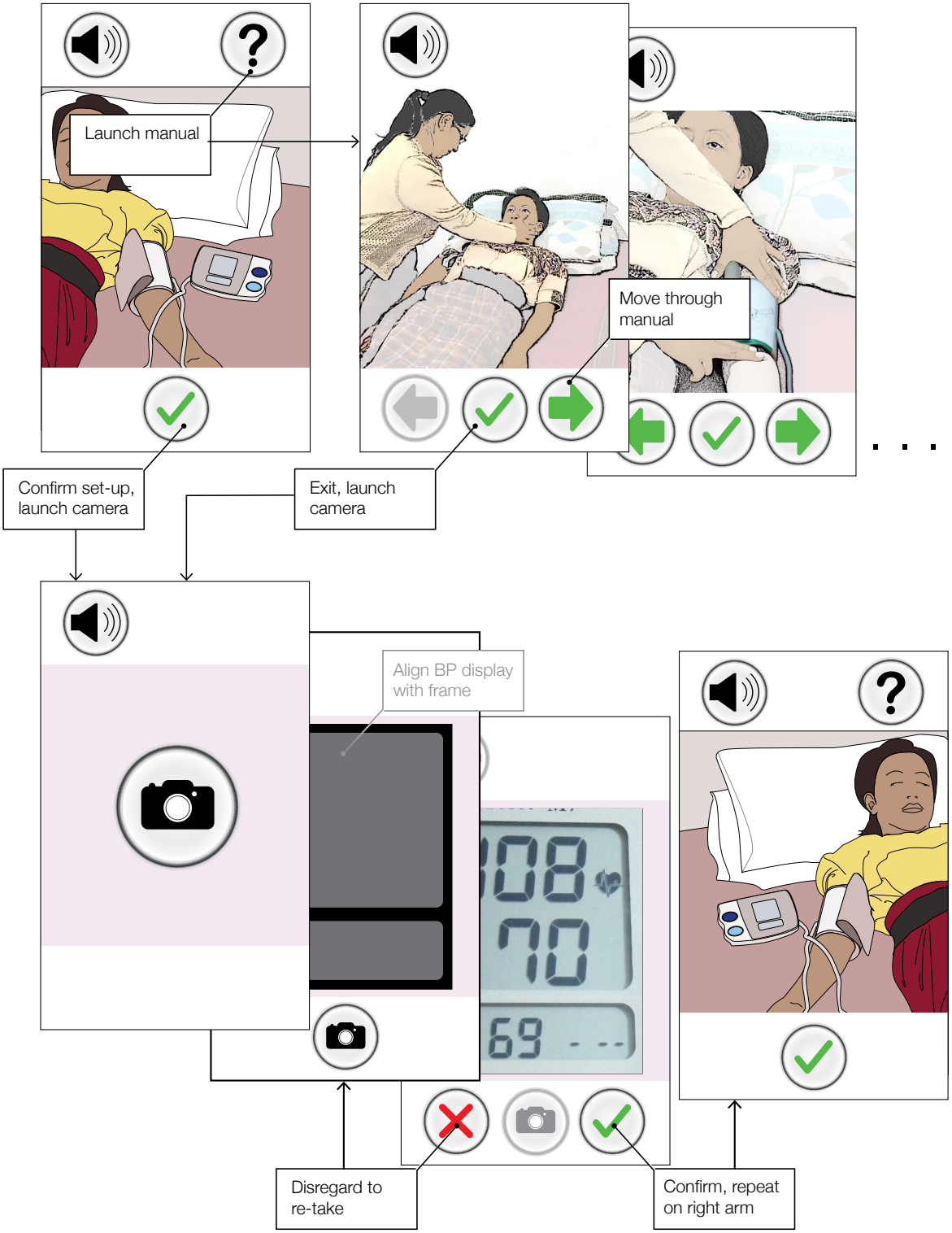


Figure A.5: The maternal blood pressure is recorded using a pregnancy-approved monitor. If required a pictogram based manual is available to the TBA to guide her through the process. As no Bluetooth enabled monitor is available for use in pregnancy to date, the reading is transferred to the phone by photo in the same manner a patient profile picture is taken. The measurement is repeated on the other arm.

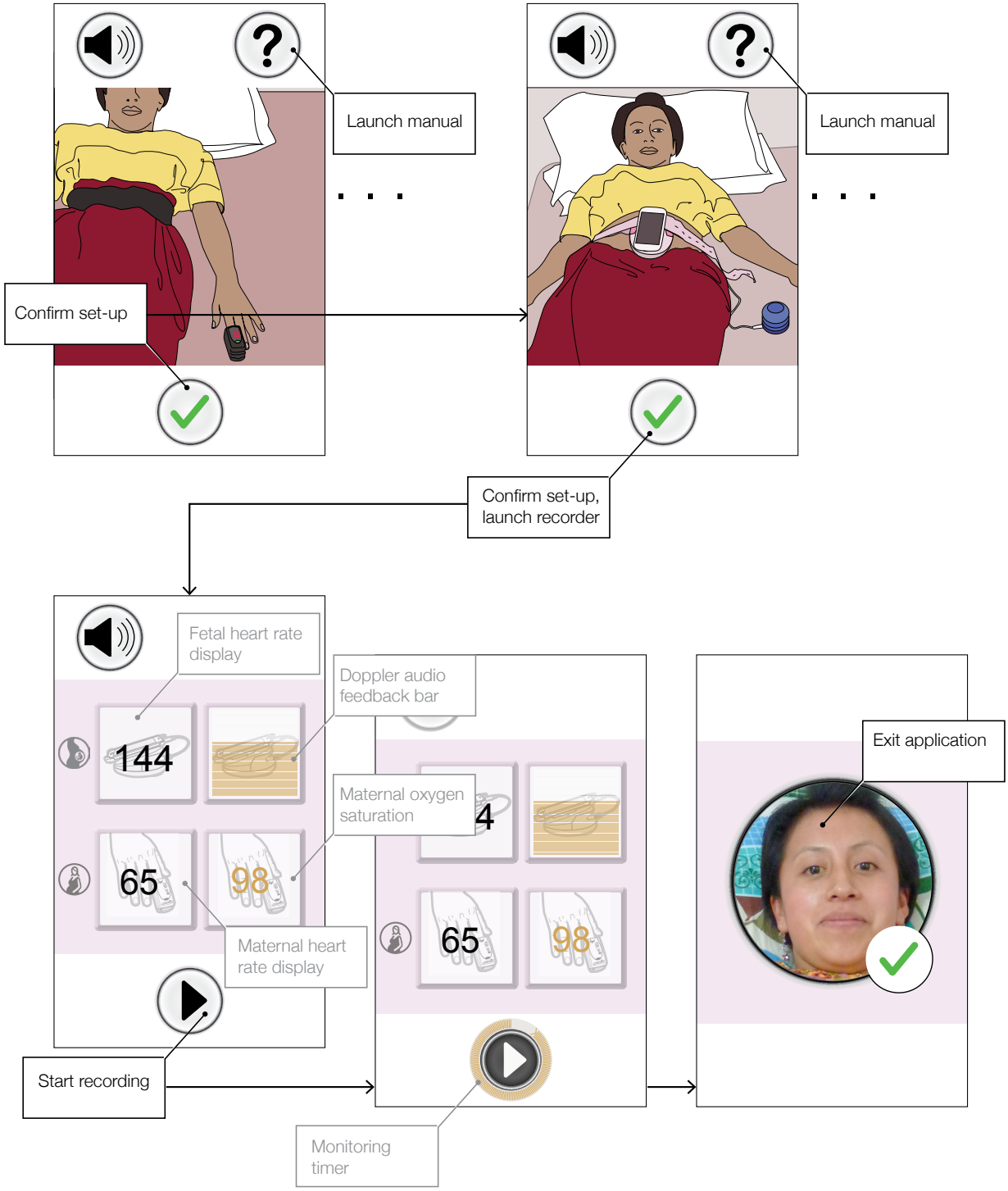


Figure A.6: The TBA is asked to attach the pulse oximeter followed by the Doppler. Step-by-step manuals are available for both procedures if required. Once the recorder is started the screen provides feedback on the fetal and maternal heart rate and maternal oxygen levels. The signals are recorded for a set time and saved to the phone. A summary SMS is sent to the Wuqu' Kawoq server and the TBA can close the application.

Appendix B

Complications Guide

The study control group will receive the same training as the intervention group. They will also receive a printed and laminated guide to antenatal, birth and postnatal complications to take with them to patient visits. The printed guide follows the same logic as the implemented triggers generating emergency calls and SMS summaries on the phone. Complications framed in red indicate complications requiring an immediate call to the clinician on duty, whereas complications framed in light red should be noted and discussed with a clinician after the encounter. The guide cards are shown in Figures B.1, B.2, B.3, B.4 and B.5.



Figure B.1: The antenatal complications guide (laminated card) for the study control group. 'xxxx-xxxx' is a placeholder for the emergency number to call when indicated. Complication pictograms are adapted from [181].

.



Figure B.2: First page of the birth complications guide (laminated card) for the study control group. 'xxxx-xxxx' is a placeholder for the emergency number to call when indicated. Complication pictograms are adapted from [181].



Figure B.3: Second page of the birth complications guide (laminated card) for the study control group. 'xxxx-xxxx' is a placeholder for the emergency number to call when indicated. Complication pictograms are adapted from [181].



Figure B.4: First page of the postnatal complications guide (laminated card) for the study control group. 'xxxx-xxxx' is a placeholder for the emergency number to call when indicated. Complication pictograms are adapted from [181].



Figure B.5: Second page of the postnatal complications guide (laminated card) for the study control group. 'xxxx-xxxx' is a placeholder for the emergency number to call when indicated. Complication pictograms are adapted from [182].

Appendix C

Scientific Protocol for Ethical Approval of Study in Guatemala

Human Subjects Involvement, Characteristics, and Design

The study will enroll participating lay midwives from the department of Chimaltenango in central Guatemala. We anticipate enrolling 50 midwives in the project. There will be no restrictions in terms of age of the enrolled participants. Participants will be required to have basic Spanish language literacy in order to enroll. Some more elderly midwives will lack the required literacy and therefore will not be eligible to enroll. However, since the project involves use of a Spanish language interface on a cellular phone, this minimum requirement is essential to program success. Additionally, in our experience, most elderly, illiterate midwives in the communities where the project will be conducted are semi-retired, and most practicing midwives are literate. Therefore, we do not anticipate that a significant portion of the eligible pool of midwives will be excluded from participation based on the literacy requirement. Recruitment identification will be by word of mouth through the local site partners well-established contacts with midwives in the community.

In addition to enrollment of midwives, who will be the primary subjects of the project, the project will also collect demographic and health data on consenting pregnant patients of each practicing midwife, as well as neonates that are born to consenting pregnant patients during the project. No data will be collected on other patient populations other than these pregnant women and their neonates. Since this is a study of the use of mHealth technology to improve pregnancy and neonatal outcomes, it is important to collect data both on the pregnant women during the pregnancy, as well as data on the neonate in the first month of life.

Of the 50 recruited midwives, 25 will be randomly assigned to receive access to the mHealth technology immediately, whereas the remaining 25 will receive the technology after 6 months. Randomization of the midwives will be performed by entering their IDs into a list and running a simple program to generate an equal number of 1s and zeros, uniquely assigned to each midwife ID. For the second half of the 12 month implementa-

tion phase, all 50 midwives will have access to the technology. This randomized stepped wedge design is critical to the evaluation of the project, since baseline data (for example, on referral rates to local hospitals) is not readily available. Therefore, randomizing 50% of the enrolled midwives to a delayed intervention phase will allow us to collect high quality internal baseline data that can be used to make comparisons to the early intervention arm (e.g., to compare rates of referral). At the 6 month time-point, all midwives who did not initially receive the technology will do so, which will allow all participating midwives to benefit from the training and technology by the end of the project.

In the proposed work site, midwives attend on average 10-20 births per year. This means that, over 12 months of implementation, we will have the opportunity to collect process and outcomes data on a minimum of 500 pregnant women and neonates. We will restrict enrollment of pregnant women to those who are older than 18 years of age. This is the legal age of consent in Guatemala, and this is a necessary arrangement to ensure that only consenting adults are participating in the project, as opposed to minors who are pregnant. The average age of first pregnancy in the region is 22 years of age, so the number of pregnant women who are excluded from the project due to being legal minors will be minimal. Selection of patients will be on an ad hoc basis, based on the willingness and availability of patients.

Data collection will be coordinated by field staff from Wuqu' Kawoq, the collaborating local partner. In the early intervention arm, all data will be doubly entered onto paper forms and into the mHealth/EMR interface directly by collaborating midwives. In the later intervention arm, data will initially be entered only on paper forms by midwives and then, after receipt of the technology, will also be entered into the mHealth interface. Wuqu' Kawoq field staff will collect the paper forms and check them against the EMR for accuracy. After coding into the EMR, paper forms will be destroyed.

Sources of Materials

The research material that will be collected from human subjects will be of two forms. First, in the case of participating midwives, tests of knowledge and competence with the technology and mHealth interface will be administered by project staff. Performance data on these tests will be tabulated in an electronic spreadsheet, where each midwife will be identified by a unique identification number. A separate master sheet linking these unique numbers to each participating midwife will be maintained, accessible only by the principal investigators. At the end of the data entry and data cleaning phases, this master sheet will be destroyed. All data will be stored on a secure, HIPAA-compliant server.

Second, demographic and health outcomes data will be entered on each participating pregnant woman and resulting neonate into both paper forms and the mHealth interface. Paper forms will be collected by local field staff, checked against the EMR for accuracy, and then immediately destroyed. Each patient entered into the EMR in this way will have a unique electronic medical record number, and the project's principal investigators will maintain a list of these medical record numbers on a secure, HIPAA compliant server. Only the PIs will have access to this list. At the end of the project, this list will be used to extract the relevant outcomes data from the EMR, at which point the data will be deidentified and the master list of medical record numbers will be destroyed.

No identifiable data will be removed from the Wuqu' Kawoq servers, and will be de-identified before being downloaded to Emory University servers in the United States.

Potential Risks to Participants

The potential risks to participants from this project are minimal. For participating midwives, there is the potential risk that early randomization to the technology will create resentment among midwives who receive late randomization to the technology. To mini-

mize this risk, all participants will receive an extensive orientation to the project design to the need to collect baseline data from half of the participants. Participants will also be reassured that all will receive equal training and access by the end of the project.

The risks to participating pregnant women and neonates are also minimal. The technologies that will be used in the project are well validated and noninvasive (Doppler ultrasound, pulse oximetry, and automated blood pressure) and routinely used worldwide in standard of care prenatal health. Therefore, no new or invasive technologies that might pose a risk to pregnancy outcomes will be employed. Furthermore, the algorithms that the mHealth interface will employ are highly conservative and designed to promote referral to physicians and to higher levels of medical care. In other words, the algorithms are designed never to 'downgrade' an individual midwife's level of clinical concern and the decision support component would never suggest against medical referral. This means that the risk that the project would decrease health care access or falsely reassure midwives against a referral is minimal.

There is also risk of psychological stress or stigma to participating pregnant women who, for cultural or family reasons may choose to conceal the fact of their pregnancy to all but the consulting midwife. Participating in this project may contribute to inadvertent disclosure of pregnancy status to community or family members. In order to minimize this risk, project staff will limit direct interactions with enrolled pregnant women during the course of the project after obtaining consent. In other words, confidentiality and intimacy in the patient-midwife relationship will be maintained to the greatest degree possible. Project staff will interact only with the midwife and will collect relevant data only from the midwife, off-site and not within the community.

One final risk is that incorrect use of the technology may lead to a reading of an absent, excessively low or excessively high heart rate, and that the relating of this information to

the patient would lead to unnecessary emotional stress. Although our device is specifically designed to avoid such scenarios, we will provide specific training to the midwives at the WK main site. We will follow standard training patterns, familiar to the midwives, for assembling midwives in small groups in a training room, which simulates the rural environment. The research team will provide training to the consented midwives using volunteer pregnant women. Training will be given on:

- How to interact with the consenter to allow non-coercive consent prior to any use of the study equipment or discussion of its utility.
- How to explain what they are doing during the data collection.
- How to answer questions about data use (after consent).
- How to use the device to obtain a good signal, and recognize when the signal is poor.
- How to ensure data is saved and later uploaded.
- How to label the data back to the patient in a de-identified manner if required.
- How to react to scenarios that may lead to stress on the patient such as:
  - Finding no heart beat
  - Heartbeat is too fast/slow
  - Device does not work
  - Device feels 'uncomfortable'

(In the first two cases the midwife would not reveal a problem initially to the patient, but will be trained to talk to a back-up Wuqu' Kawoq physician directly prior to disclosure).

In order to participate in the study, the midwife will have to be able to use all parts of the device correctly, and react correctly in the above scenarios. The study coordinator

will assess each of the midwives for passing this test and certify that they are ready to participate. The study coordinator will also visit midwives at random in the study environment and identify deviations from the protocol requiring retaining and recertification.

Adequacy of Protection Against Risks

Recruitment and Informed Consent

Written or verbal informed consent will be obtained from all participating midwives and also from all pregnant patients of each participating midwife. There are two distinct settings and population for consenting study participants. The first group of participants is the midwives, who will be recruited and consented by the study coordinator at Wuqu' Kawoq. Midwives in Wuqu' Kawoq's catchment area will be contacted by phone by the study coordinator to assess interest in the project. If a midwife expresses interest in participating they will be invited to a preliminary training session. During this training session, project staff (and in particular the study coordinator) will describe the full details of the study to potential midwife participants, including the nature of the subject's involvement in the study, the possible risks and benefits of participation, and the individual capacity to withdraw from the study at any time without consequence. The midwives will have the opportunity to ask questions about the study, after which consent to proceed will be obtained. The study coordinator will receive training in human subjects protection (e.g., CITI or the NIH module) prior to interacting with and obtaining consent with midwives.

The second group of participants that requires consenting is the pregnant mothers. All pregnant patients of each recruited midwife will be eligible for participation. Participating midwives will be encouraged to mention the study to all of their patients and to relay any potential interest in participating to the study coordinator. When a midwife identifies a pregnant subject for potential participation, a single study staff member (a trained nurse working under the supervision of the study coordinator) will discreetly

visit the patient (and the biological father, if appropriate), at a location of their choosing (most likely to be the domicile of the pregnant patient) and again describe the details of the study, the risks and benefits, and the capacity to withdraw at any time without consequence, at which point verbal consent will be obtained for collection of data both during the pregnancy and in the neonatal period.

Both midwives and patients will all be bilingual speakers of Spanish and Kaqchikel Maya. Spanish is the language that they use for technical conversations, such as medical conversation. Kaqchikel is an oral, non-written language that is used in more colloquial settings. Therefore, consent (and consent documents) will be obtained in Spanish, although research staff will be able to speak both languages. We anticipate that some informal conversation and question answering will occur in Kaqchikel, after the formal consent has been read, as this is typical of medical consultations and interactions in the region: technical conversation in Spanish followed by informal clarification in Kaqchikel. In our previous experiences, we have been able to obtained written informed consent successfully, as long as attention is paid to providing clear orientation and allowing time for informal clarification and question answering. Wuqu' Kawoq's field staff all are native Spanish and Kaqchikel Maya speakers.

Protection Against Risk

Confidentiality: To protect confidentiality, the following steps will be taken: All paper research forms will be kept in locked file cabinets and will be available only to research staff directly involved in this project. Immediately after checking paper forms against the mHealth/EMR interface, they will be destroyed. The EMR will be maintained on a secure, HIPAA-compliant cloud server, and only the principal investigators will have access to the EMR directly. Master lists of participating midwives and patients will be keyed to unique identifying numbers, and only the project principal investigators will have access to these keys, which will also be maintained separately on a secure, HIPAA-

compliant cloud server. Once data extraction and cleaning has been completed, these master lists will be destroyed, and data in the analysis phase will be completely deidentified. Laptop computers or tablets used for data entry will be routinely backed-up and will be password-protected and encrypted.

Additional Protections for Pregnant Women, Fetuses and Neonates: This research involves the collection of demographic and health outcomes data on pregnant women and neonates. Importantly, the interventional component of the study involves an mHealth (EMR) integration step and does not introduce novel or experimental technologies. All technologies that will be used in the mHealth arm are standard-of-care technologies (Doppler, blood pressure monitoring, pulse oximetry). Half of the enrolled participants will initially not receive any intervention. However, they will receive the current standard of care by lay midwives, with no interference in this usual process. Furthermore, for pregnant women and neonates who access the technology (either in the early access or later access arms), the expected result of the technology will be earlier and more timely referral to a higher level of medical care. At no point will the technology discourage accessing a higher level of care. Since the project holds out the potential for direct benefit to the pregnant woman and/or her fetus, consent will be obtained directly from the pregnant women to participate. The proposal does not include any inducements to early termination of a pregnancy, and it does not explicitly involve research on neonates of uncertain or non-viability, although in accordance with the proposal's plan to increase early access to medical care for all neonates, rates of neonatal mortality will be tracked.

Additional Protections for Children: This study involves collection of neonatal mortality rates and rates of early evaluation of neonates by a postnatal health care delivery team. As such, no direct interventions on children will be performed, and the study involves minimal risk to children. As outlined above, consent will be obtained from both biological parents at the time of enrollment of the pregnant patient in the study to permit

collection of this data.

The study will not enroll pregnant patients prior to the legal age of informed consent in Guatemala (18 years).

Adverse Events: In the event that midwives either in the early technology access or later access arms detect the need for a referral to a higher level of medical care or an adverse event, they will be instructed to proceed with arranging the referral through the usual local mechanisms. However, the local site partner, Wuqu' Kawoq, will also designate a medical officer to assist with the project, and this physician will be available to all midwives throughout the duration of the project to provide further logistical and decision support for arranging referrals and management of adverse events.

Potential Benefits to Study Participants

Participants are likely to benefit directly from this project. Midwives will benefit by gaining access to additional medical knowledge and technology. Pregnant women and neonates will benefit by earlier detection of the need for medical referral and improved care coordination. The risk of the project are minimal, and the potential benefits of the project, which include the demonstration that mHealth technology can be used to enhance the care provided by lay midwives in a developing country setting where perinatal outcomes and care coordination are generally poor outweigh these risks.

Importance of the Knowledge to Be Gained

The information gained through this project will contribute important, practical insights into the implementation of a mHealth solution to low rates of medical referral and care coordination for pregnant women and neonates. These are perennial public health problems in rural Guatemala which have until now proven to be largely intractable. In addition to suggesting new avenues for enhancing care in rural Guatemala, the results will also

be applicable to other rural populations in other developing countries which experience similar breaks in the primary care referral chain and the continuum of care.

Institutional Review Board Coverage

The project will be supervised by the IRB of Emory University. In addition, the study will also be reviewed and approved by the IRB of the local implementing partner, Wuqu' Kawoq. Both of these institutions have OHRP approved IRBs with FWA. All field staff participating in the project will approve an on-line course on the protection of human subjects participating in research.

Inclusion of Women and Minorities

Study participants will include 50 lay midwives from the department of Chimaltenango in rural Guatemala. Midwives in rural Guatemala are all female for cultural reasons. All participants will likely be of Maya ethnicity. Although there will be no restrictions in terms of ethnicity for enrolment, the area in which the study is taking place is composed almost exclusively of ethnic Maya speaking the Kaqchikel language. Indeed, the purpose of the study is to explore ways to improve maternal and child outcomes in this population, which experiences wide disparities in health care access and outcomes when compared to other inhabitants of Guatemala who are not ethnically Maya and who live in other parts of the country. See Targeted/Planned Enrollment Table for projected target sizes of subgroups based on gender and race/ethnicity.

Inclusion of Children

Because of the stigma associated with teenage pregnancy, and our desire to involve any elevated risk of stigma by inadvertently identifying teenage pregnancies against their will, only pregnant woman above the age of legal consent in Guatemala (18 years) will be included in the project.

Neonates who are born to consenting parents during the project, and data on neonatal referrals and mortality will be collected from the electronic medical record. No neonates born to consenting parents will be excluded from this data collection.

Targeted/Planned Enrollment Table

The table shows projected target sizes of subgroups based on gender and race/ethnicity.

Study Title: Mobile Health Intervention to Improve Perinatal Continuum of Care in Guatemala

TARGETED/PLANNED ENROLLMENT: Number of Subjects			
Ethnic Category	Females	Males	Total
Hispanic or Latino	800	250	1050
Not Hispanic or Latino	0	0	0
Ethnic Category: Total of All Subjects *	800	250	1050
Racial Categories			
American Indian/Alaska Native	800	250	1050
Asian	0	0	0
Native Hawaiian or Other Pacific Islander	0	0	0
Black or African American	0	0	0
White	0	0	0
Racial Categories: Total of All Subjects *	800	250	1050

* The *Ethnic Category: Total of All Subjects* must be equal to the *Racial Categories: Total of All Subjects*. ** Any children born in the study are considered part of the enrollment. With 500 anticipated births, 50% of these are anticipated to be male

Table C.1: Targeted enrollment of study participants including children born.

Appendix D

Digital Help Manuals for Mobile Application

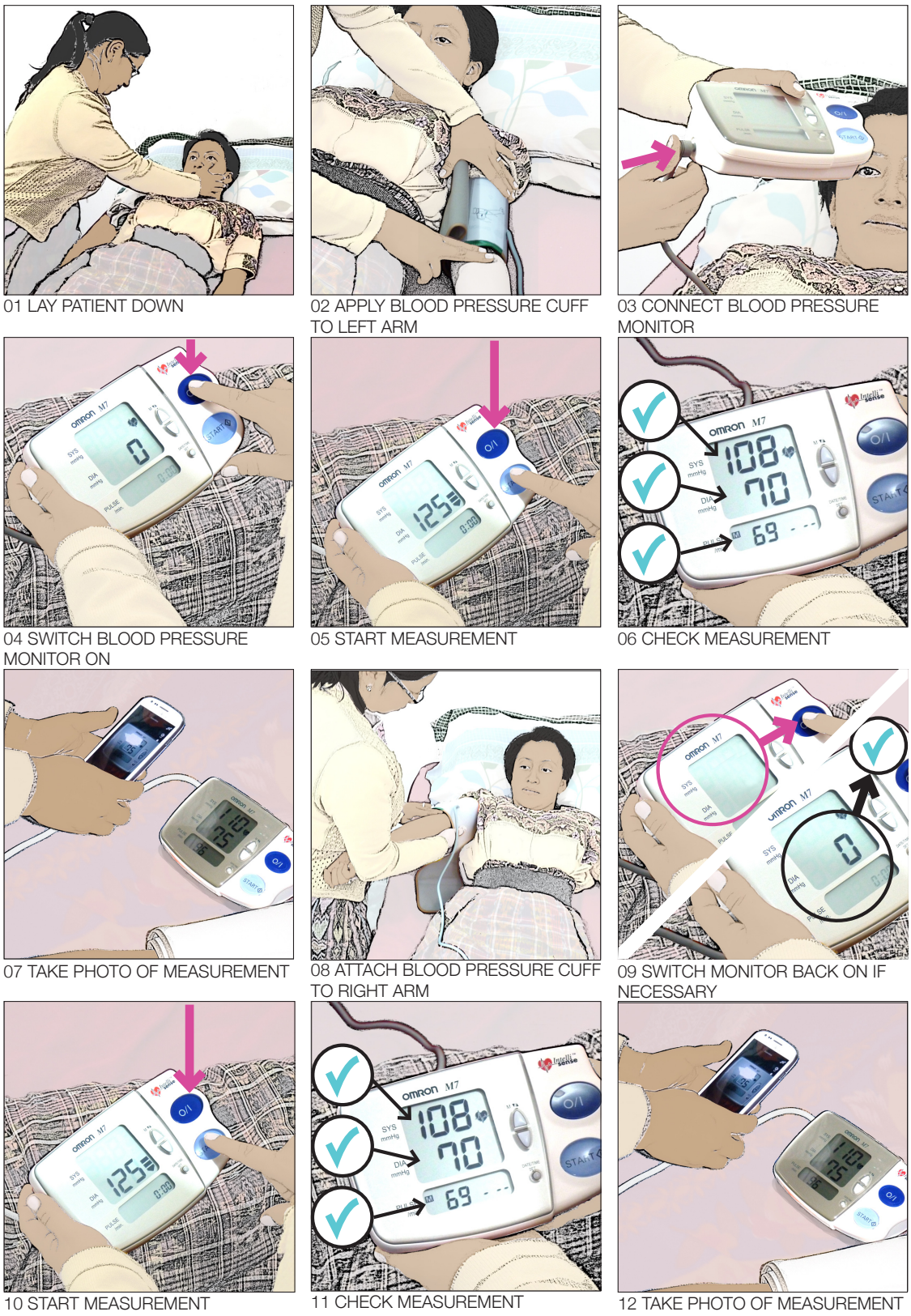


Figure D.1: Manual for use of the blood pressure monitor. Note that the instructions under each image are replaced by an audio recording in Kaqchikel on the mobile phone. (c) Lisa Stroux 2015, released under the creative commons license attribution 4.0 international (CC BY 4.0).

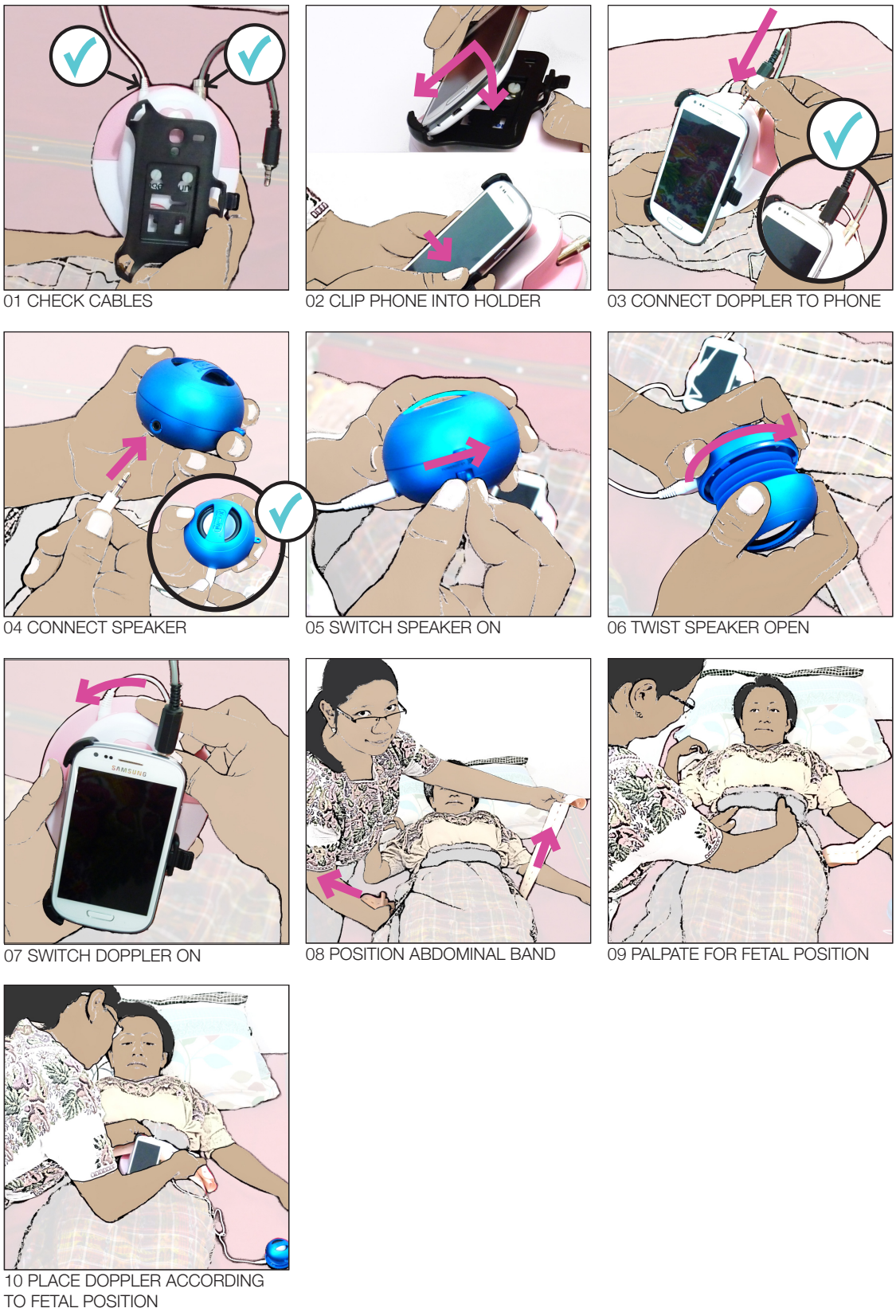


Figure D.2: Manual for use of the Doppler. Note that the instructions under each image are replaced by an audio recording in Kaqchikel on the mobile phone. (c) Lisa Stroux 2015, released under the creative commons license attribution 4.0 international (CC BY 4.0).

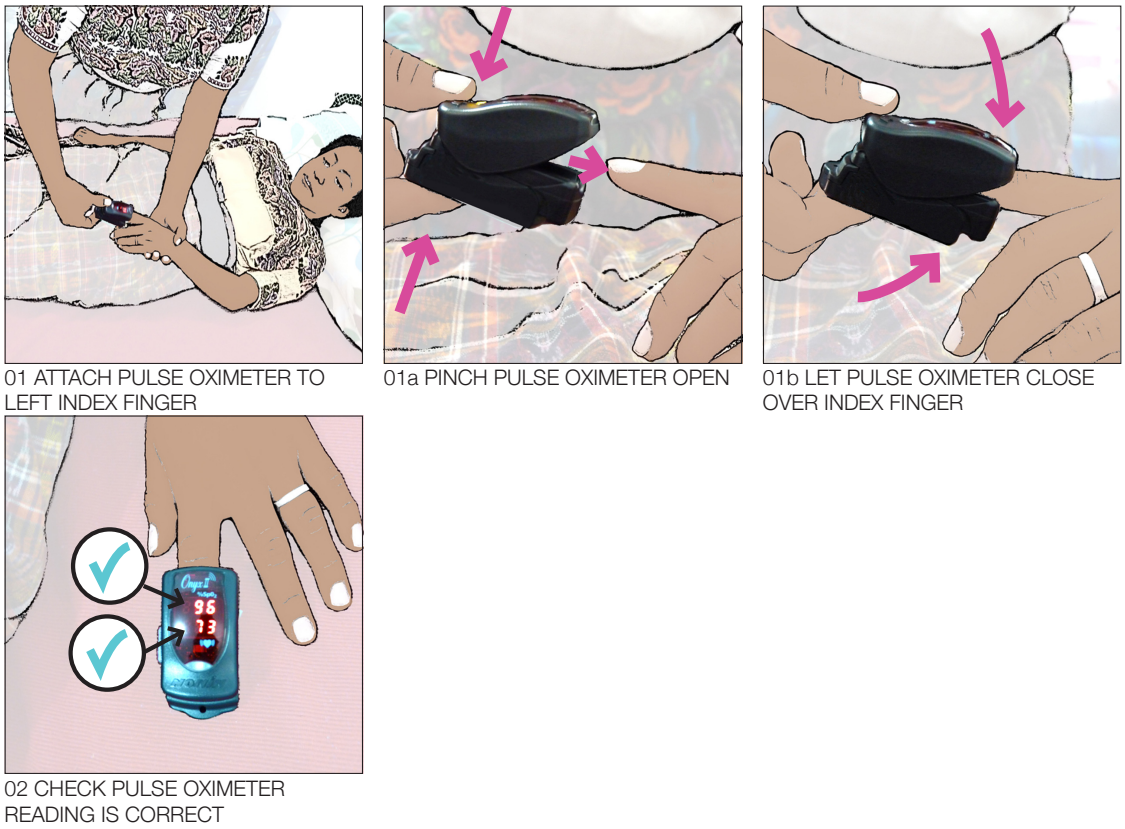


Figure D.3: Manual for use of the pulse oximeter. Note that the instructions under each image are replaced by an audio recording in Kaqchikel on the mobile phone. (c) Lisa Stroux 2015, released under the creative commons license attribution 4.0 international (CC BY 4.0).

Bibliography

[1] Maternal and perinatal health. World Health Organisation. [Online]. Available: http://www.who.int/maternal_child_adolescent/topics/maternal/maternal_perinatal/en/

[2] Milennium Development Goals. United Nations. [Online]. Available: http://www.un.org/millenniumgoals/

[3] The Millennium Development Goals report 2015. United Nations. [Online]. Available: http://www.un.org/millenniumgoals

[4] J. E. Lawn, H. Blencowe, S. Oza, D. You, A. C. C. Lee, P. Waiswa, M. Lalli, Z. Bhutta, A. J. D. Barros, P. Christian, C. Mathers, and S. N. Cousens, "Every newborn: progress, priorities, and potential beyond survival." *Lancet*, vol. 384, no. 9938, pp. 189–205, Jul. 2014.

[5] "Levels & trends in child mortality," UNICEF, World Health Organization, the World Bank Group and the United Nations, Tech. Rep., 2015.

[6] R. Woods, "Long-term trends in fetal mortality: implications for developing countries." *Bull World Health Organ*, vol. 86, no. 6, pp. 460–466, Jun. 2008.

[7] R. L. Goldenberg and E. M. McClure, "Maternal, fetal and neonatal mortality: lessons learned from historical changes in high income countries and their potential application to low-income countries." *Maternal Health, Neonatology and Perinatology*, vol. 1, no. 1, pp. 1–10, 2015.

[8] R. E. Black, L. H. Allen, Z. A. Bhutta, L. E. Caulfield, M. De Onis, M. Ezzati, C. Mathers, and J. Rivera, "Maternal and child undernutrition: global and regional exposures and health consequences," *The lancet*, vol. 371, no. 9608, pp. 243–260, 2008.

[9] M. de Onis, M. Blössner, and J. Villar, "Levels and patterns of intrauterine growth retardation in developing countries." *Eur J Clin Nutr*, vol. 52 Suppl 1, pp. S5–15, Jan. 1998.

[10] Transforming our world: the 2030 Agenda for Sustainable Development. United Nations. [Online]. Available: https://sustainabledevelopment.un.org/post2015/transformingourworld

[11] M. Peters, J. Crowe, J. F. Piéri, H. Quartero, B. Hayes-Gill, D. James, J. Stinstra, and S. Shakespeare, "Monitoring the fetal heart non-invasively: a review of methods." *J Perinat Med*, vol. 29, no. 5, pp. 408–416, 2001.

[12] D. Ayres-de Campos and J. Bernardes, "Twenty-five years after the FIGO guidelines for the use of fetal monitoring: Time for a simplified approach?" *International Journal of Gynecology & Obstetrics*, vol. 110, no. 1, pp. 1–6, 2010.

[13] R. G. Kennedy, "Electronic fetal heart rate monitoring: retrospective reflections on a twentieth-century technology." *J R Soc Med*, vol. 91, no. 5, pp. 244–250, May 1998.

[14] C. Signore, R. K. Freeman, and C. Y. Spong, "Antenatal testing - a reevaluation: executive summary of a Eunice Kennedy Shriver National Institute of Child Health and Human Development workshop." *Obstet Gynecol*, vol. 113, no. 3, pp. 687–701, Mar. 2009.

[15] R. M. Grivell, Z. Alfirevic, G. M. Gyte, and D. Devane, "Antenatal cardiotocography for fetal assessment." *Cochrane Database Syst Rev*, no. 1, 2010, art. No.: CD007863.

[16] R. Sameni and G. D. Clifford, "A review of fetal ECG signal processing; issues and promising directions." *Open Pacing Electrophysiol Ther J*, vol. 3, pp. 4–20, Jan. 2010.

[17] J. P. Neilson, "Fetal electrocardiogram (ECG) for fetal monitoring during labour." *Cochrane Database of Systematic Reviews*, vol. 4, 2012, art. No.: CD000116.

[18] S. A. Shakespeare, J. A. Crowe, B. R. Hayes-Gill, K. Bhogal, and D. K. James, "The information content of Doppler ultrasound signals from the fetal heart." *Med Biol Eng Comput*, vol. 39, no. 6, pp. 619–626, Nov. 2001.

[19] P. J. Steer, "Has electronic fetal heart rate monitoring made a difference?" *Semin Fetal Neonatal Med*, vol. 13, no. 1, pp. 2–7, Feb. 2008.

[20] D. Devane, J. G. Lalor, S. Daly, W. McGuire, and V. Smith, "Cardiotocography versus intermittent auscultation of fetal heart on admission to labour ward for assessment of fetal wellbeing." *Cochrane Database Syst Rev*, vol. 2, 2012.

[21] F. Kovács, C. Horváth, A. T. Balogh, and G. Hosszú, "Fetal phonocardiography–past and future possibilities." *Comput Methods Programs Biomed*, vol. 104, no. 1, pp. 19–25, Oct. 2011.

[22] V. Serra, J. Bellver, M. Moulden, and C. W. G. Redman, "Computerized analysis of normal fetal heart rate pattern throughout gestation." *Ultrasound Obstet Gynecol*, vol. 34, no. 1, pp. 74–79, Jul. 2009.

[23] S. Turan, J. Miller, and A. A. Baschat, "Integrated testing and management in fetal growth restriction." *Semin Perinatol*, vol. 32, no. 3, pp. 194–200, Jun. 2008.

[24] B. O. Okusanya, "Antenatal cardiotocography for fetal assessment: RHL commentary," *The WHO Reproductive Health Library*, 2010.

[25] P. Várady, L. Wildt, Z. Benyó, and A. Hein, "An advanced method in fetal phonocardiography." *Comput Methods Programs Biomed*, vol. 71, no. 3, pp. 283–296, Jul. 2003.

[26] S. Vaisman, S. Y. Salem, G. Holcberg, and A. B. Geva, "Passive fetal monitoring by adaptive wavelet denoising method." *Comput Biol Med*, vol. 42, no. 2, pp. 171–179, Feb. 2012.

[27] *Medical devices: managing the mismatch: an outcome of the priority medical devices project.* WHO Press, World Health Organization, 2010, produced by the Department of Essential Health Technologies under the direction of Josée Hansen.

[28] Moyo fetal heart rate monitor. Laerdal Global Health. [Online]. Available: http://www.laerdalglobalhealth.com/doc/2555/Moyo-Fetal-Heart-Rate-Monitor

[29] Wind-up fetal heart rate monitor. Philips Healthcare. [Online]. Available: http://www.healthcare.philips.com/pwc_hc/main/clinicalspecialities/womenshealthcare/foa/powerfree-fetal-heart-rate.html

[30] J. Boeing, T. Gatlin, E. Hendler, and E. Hyde, "Fetal heart rate monitor for resource-limited settings." University of Michigan, Tech. Rep., 2012. [Online]. Available: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/96179/me450f12project2_report.pdf?sequence=1

[31] G. D. Clifford and D. Clifton, "Wireless technology in disease management and medicine." *Annual Review of Medicine*, vol. 63, pp. 479–92, 2012.

[32] WHO, "Global observatory for ehealth series - volume 3 mhealth - new horizons for health through mobile technologies," World Health Organization, Tech. Rep., 2011.

[33] P. Yu, M. X. Wu, H. Yu, and G. Xiao, "The challenges for the adoption of m-health," in *Service Operations and Logistics, and Informatics, 2006. SOLI '06. IEEE International Conference on*, Jun. 2006, pp. 181–186.

[34] A. Labrique, L. Vasudevan, L. W. Chang, and G. Mehl, "H_pe for mhealth: More "y" or "o" on the horizon?" *International Journal of Medical Informatics*, vol. 82, no. 5, pp. 467–469, 2013.

[35] A. B. Labrique, L. Vasudevan, E. Kochi, R. Fabricant, and G. Mehl, "mhealth innovations as health system strengthening tools: 12 common applications and a visual framework." *Global Health : Science and Practice*, vol. 1, no. 2, pp. 160–171, 2013.

[36] Key ICT indicators 2015. ITU Wolrd Telecommunication. [Online]. Available: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

[37] A. Kearney, "Improving the evidence for mobile health." GSMA, Tech. Rep., 2011.

[38] J. Lemaire, "Scaling up mobile health – elements necessary for the successful scale up of mhealth in developing countries." Actevis Consulting Group for Advanced Development For Africa, Tech. Rep., 2011.

[39] C. S. Hall, E. Fottrell, S. Wilkinson, and P. Byass, "Assessing the impact of mhealth interventions in low- and middle-income countries–what has been shown to work?" *Glob Health Action*, vol. 7, p. 25606, 2014.

[40] N. Al-Shorbaji and A. Geissbuhler. (2012) Bulletin of the World Health Organization - Establishing an evidence base for e-health: the proof is in the pudding. World Health Organization. [Online]. Available: http://www.who.int/bulletin/volumes/90/5/12-106146/en/

[41] M. Tomlinson, M. J. Rotheram-Borus, L. Swartz, and A. C. Tsai, "Scaling up mhealth: where is the evidence?" *PLoS Med*, vol. 10, no. 2, 2013, e1001382.

[42] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Patel, and A. Haines, "The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis." *PLoS Med*, vol. 10, no. 1, p. e1001363, 2013.

[43] U. B. Nurmatov, S. H. Lee, B. I. Nwaru, M. Mukherjee, L. Grant, and C. Pagliari, "The effectiveness of mhealth interventions for maternal, newborn and child health in low- and middle-income countries: Protocol for a systematic review and meta-analysis." *J Glob Health*, vol. 4, no. 1, Jun. 2014, 010407.

[44] "ehealth and innovation in women's and children's health: A baseline review," Online, World Health Organization (WHO) and the International Telecommunication Union (ITU), Tech. Rep., 2014.

[45] T. Tamrat and S. Kachnowski, "Special delivery: An analysis of mhealth in maternal and newborn health programs and their outcomes around the world," *Maternal and Child Health Journal*, vol. 16, no. 5, pp. 1092–1101, 2012.

[46] A. C. Noordam, B. M. Kuepper, J. Stekelenburg, and A. Milen, "Improvement of maternal health services through the use of mobile phones." *Trop Med Int Health*, vol. 16, no. 5, pp. 622–626, May 2011.

[47] A. P. Kazantsev, A. A. Senin, J. N. Ponomareva, and M. N. Mochalova, "An mhealth approach to remote fetal monitoring." in *Healthcare Innovation Conference (HIC), 2014 IEEE*, 2014, pp. 239–242.

[48] C. S. Lee, M. Masek, C. P. Lam, and K. T. Tan, "Advances in fetal heart rate monitoring using smart phones," in *Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on*, 2009, pp. 735–740.

[49] M. A. Soltani, "Integrating a hand held computer and stethoscope into a fetal monitor," *Medical Education Online*, vol. 14, no. 3, 2009.

[50] J. L. Watterson, J. Walsh, and I. Madeka, "Using mhealth to improve usage of antenatal care, postnatal care, and immunization: A systematic review of the literature," *BioMed Research International*, 2015.

[51] UNICEF, "State of the world's children 2015 country statistical information," 2015. [Online]. Available: http://www.unicef.org/infobycountry/guatemala_statistics.html

[52] A. Garces, E. M. Mcclure, K. Hambidge, N. F. Krebs, L. Figueroa, M. Aguilar, J. L. Moore, and R. L. Goldenberg, "Trends in perinatal deaths from 2010 to 2013 in the Guatemalan Western Highlands." *Reprod Health*, vol. 12 Suppl 2, p. S14, Jun. 2015.

[53] Caracterización departamental: Chimaltenango 2013. Instituto Nacional de Estadística (INE). [Online]. Available: http://www.ine.gob.gt/sistema/uploads/2015/07/20/uYeKswXf9BHOJHZZZ7f7eR8CaMWVXnzg.pdf

[54] L. V. Walsh, "Beliefs and rituals in traditional birth attendant practice in guatemala." *J Transcult Nurs*, vol. 17, no. 2, pp. 148–154, Apr. 2006.

[55] E. Kestler, D. Walker, A. Bonvecchio, S. Sáenz de Tejada, and A. Donner, "A matched pair cluster randomized implementation trail to measure the effectiveness of an intervention package aiming to decrease perinatal mortality and increase institution-based obstetric care among indigenous women in Guatemala: study protocol," *BMC Pregnancy and Childbirth*, vol. 13, no. 1, 2013.

[56] "Sonicaid fetacare antepartum analysis - clinical application guide," Huntleigh Healthcare Ltd, Tech. Rep., 2008. [Online]. Available: http://www.arjohuntleigh.net/diagnostics/Admin/files/20081027135828.pdf

[57] D. Stanger, "An overview on fetal monitoring," Huntleigh Healthcare Ltd., Tech. Rep., April 2011.

[58] N. Archer and N. Manning, "OSH Fetal Cardiology," Oxford, UK, July 2009. [Online]. Available: //oxfordmedicine.com/10.1093/med/9780199230709.001.0001/med-9780199230709

[59] (2015, May) Anatomy & physiology. OpenStax CNX. [Online]. Available: http://cnx.org/contents/14fb4ad7-39a1-4eee-ab6e-3ef2482e3e22@7.28

[60] L. S. Baker, M. C. Beaves, and W. E. Trickey, D. J., *Fetal Surveillance, A practical guide*. Fetal Surveillance Education Program (FSEP), Royal Australian and New Zealand College of Obstetricians and Gynaecologists.

[61] S. Blackburn, *Maternal, fetal, & neonatal physiology : a clinical perspective*, 4th ed. Blackburn, Susan Tucker, 2013.

[62] S. Pildner von Steinburg, A. Boulesteix, C. Lederer, S. Grunow, S. Schiermeier, W. Hatzmann, K. M. Schneider, and M. Daumer, "What is the normal fetal heart rate?" *PeerJ*, vol. 1, pp. e82–, May 2013.

[63] L. K. Hornberger and D. J. Sahn, "Rhythm abnormalities of the fetus," *Heart*, vol. 93, no. 10, pp. 1294–1300, Oct. 2007.

[64] R. Polin, W. W. Fox, and S. Abman, *Fetal and neonatal physiology*, 3rd ed. Elsevier Saunders, 2004.

[65] A. Lausman, F. P. McCarthy, M. Walker *et al.*, "Screening, diagnosis, and management of intrauterine growth restriction," *Journal of Obstetrics and Gynaecology Canada*, vol. 34, no. 1, pp. 17–28, 2012.

[66] E. H. Hon, "The electronic evaluation of the fetal heart rate. preliminary report. 1958." *Am J Obstet Gynecol*, vol. 175, no. 3 Pt 1, pp. 747–748, Sep. 1996.

[67] D. L. Tuck, "Improvement in Doppler ultrasound human foetal heart rate records by signal correlation." *Med Biol Eng Comput*, vol. 20, no. 3, pp. 357–600, May 1982.

[68] G. Lawson, R. Belcher, G. Dawes, and C. Redman, "A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems," *American Journal of Obstetrics & Gynecology*, vol. 147, no. 6, pp. 721–722, 1983.

[69] M. Y. Divon, F. P. Torres, S. Y. Yeh, and R. H. Paul, "Autocorrelation techniques in fetal monitoring." *Am J Obstet Gynecol*, vol. 151, no. 1, pp. 2–6, Jan. 1985.

[70] C. H. L. Peters, E. D. M. ten Broeke, P. Andriessen, B. Vermeulen, R. C. M. Berendsen, P. F. F. Wijn, and S. G. Oei, "Beat-to-beat detection of fetal heart rate: Doppler ultrasound cardiotocography compared to direct ECG cardiotocography in time and frequency domain." *Physiol Meas*, vol. 25, no. 2, pp. 585–593, Apr. 2004.

[71] J. D. Cryer and K.-S. Chan, *Time Series Analysis With Applications in R*, 2nd ed. Springer-Verlag New York, 2008.

[72] X. Hua, L. Kaiqing, and Z. Zhenxi, "A new algorithm for detecting fetal heart rate using ultrasound Doppler signals," *Ultrasonics*, vol. 43, no. 6, pp. 399–403, 2005.

[73] I. Voicu, J.-M. Girault, C. Roussel, A. Decock, and D. Kouame, "Robust estimation of fetal heart rate from US Doppler signals," *Physics Procedia*, vol. 3, no. 1, pp. 691–699, 2010, international Congress on Ultrasonics, Santiago de Chile, January 2009.

[74] A. Khandoker, Y. Kimura, T. Ito, and M. Palaniswami, "Non-invasive determination of electromechanical time intervals of cardiac cycle using abdominal ECG and Doppler ultrasound signals from fetal hearts," in *Computers in Cardiology, 2007*, 2007, pp. 657–660.

[75] T. Kret and K. Kaluzynski, "The fetal heart rate estimation based on continuous ultrasonic Doppler data," *Biocybernetics and Biomedical Engineering*, vol. 26, no. 3, p. 49â56, 2006.

[76] I. Voicu, S. MÃ©nigot, D. KouamÃ©, and J.-M. Girault, "New estimators and guidelines for better use of fetal heart rate estimators with Doppler ultrasound devices." *Comput Math Methods Med*, vol. 2014, p. 784862, 2014.

[77] A. de CheveignÃ© and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, May 2002.

[78] J. Jezewski, T. Kupka, and K. Horoba, "Extraction of fetal heart-rate signal as the time event series from evenly sampled data acquired using Doppler ultrasound technique," vol. 55, no. 2, pp. 805–810, 2008.

[79] J. Jezewski, D. Roj, J. Wrobel, and K. Horoba, "A novel technique for fetal heart rate estimation from Doppler ultrasound signal." *Biomed Eng Online*, vol. 10, p. 92, 2011.

[80] D. Roj, J. WrÃ³*bel*, *K. Horoba*, *T. Przybyla*, *andT. Kupka*, "*Improvingtheperiodicitymeasureme*
−26, 2010.

[81] V. ChudÃ¡Äek, J. Spilka, M. BurÅ¡a, P. JankÅ¯, L. Hruban, M. Huptych, and L. LhotskÃ¡, "Open access intrapartum ctg database." *BMC Pregnancy Childbirth*, vol. 14, p. 16, 2014.

[82] E. L. Nyboe, "An algorithm based on the dawes/redman criteria for automated fetal heart rate analysis," Master's thesis, Chalmers University Of Technology, Gothenburg, Sweden, 2011.

[83] U. Schneider, A. Fiedler, M. Liehr, C. Kähler, and E. Schleussner, "Fetal heart rate variability in growth restricted fetuses," *Biomedizinische Technik*, vol. 51, no. 4, pp. 248–250, 2006.

[84] M. Ferrario, M. G. Signorini, and G. Magenes, "Comparison between fetal heart rate standard parameters and complexity indexes for the identification of severe intrauterine growth restriction." *Methods of information in medicine*, vol. 46, no. 2, p. 186, 2007.

[85] G. Buscicchio, L. Gentilucci, and A. L. Tranquilli, "Computerized analysis of fetal heart rate in pregnancies complicated by gestational diabetes mellitus, gestational hypertension, intrauterine growth restriction and premature rupture of membranes," *Journal of Maternal-Fetal and Neonatal Medicine*, vol. 23, no. 4, pp. 335–337, 2010.

[86] E. Huhn, S. Lobmaier, T. Fischer, R. Schneider, A. Bauer, K. Schneider, and G. Schmidt, "New computerized fetal heart rate analysis for surveillance of intrauterine growth restriction," *Prenatal Diagnosis*, vol. 31, no. 5, pp. 509–514, 2011.

[87] S. M. Lobmaier, E. A. Huhn, S. Pildner von Steinburg, A. Müller, T. Schuster, J. U. Ortiz, G. Schmidt, and K. T. Schneider, "Phase-rectified signal averaging as a new method for surveillance of growth restricted fetuses," *Journal of Maternal-Fetal and Neonatal Medicine*, vol. 25, no. 12, pp. 2523–2528, 2012.

[88] E. M. Graatsma, E. J. H. Mulder, B. Vasak, S. M. Lobmaier, S. Pildner von Steinburg, K. T. M. Schneider, G. Schmidt, and G. H. A. Visser, "Average acceleration and deceleration capacity of fetal heart rate in normal pregnancy and in pregnancies complicated by fetal growth restriction." *Journal of Maternal-Fetal and Neonatal Medicine*, vol. 25, no. 12, pp. 2517–2522, Dec. 2012.

[89] D. Hoyer, F. Tetschke, S. Jaekel, S. Nowack, O. W. Witte, E. Schleußner, and U. Schneider, "Fetal functional brain age assessed from universal developmental indices obtained from neuro-vegetative activity patterns," *PloS one*, vol. 8, no. 9, p. e74431, 2013.

[90] A. Fanelli, G. Magenes, M. Campanile, and M. Signorini, "Quantitative assessment of fetal well-being through CTG recordings: a new parameter based on phase rectified signal average." *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 5, pp. 959–966, Sep. 2013.

[91] G. S. Dawes, M. Moulden, and C. W. G. Redman, "Short-term fetal heart rate variation, decelerations, and umbilical flow velocity waveforms before labor," *Obstetrics & Gynecology*, vol. 80, no. 4, pp. 673–678, 1992.

[92] K. Hecher, C. M. Bilardo, R. H. Stigter, Y. Ville, B. J. Hackelöer, H. J. Kok, M. V. Senat, and G. H. A. Visser, "Monitoring of fetuses with intrauterine growth restriction: a longitudinal study," *Ultrasound in Obstetrics and Gynecology*, vol. 18, no. 6, pp. 564–570, 2001.

[93] V. Serra, M. Moulden, J. Bellver, and C. W. G. Redman, "The value of the short-term fetal heart rate variation for timing the delivery of growth-retarded fetuses," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 115, no. 9, pp. 1101–1107, 2008.

[94] M. Ferrario, G. Magenes, M. Campanile, I. F. Carbone, A. Di Lieto, and M. G. Signorini, "Multiparameter analysis of heart rate variability signal for the investigation of high risk fetuses," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.* IEEE, 2009, pp. 4662–4665.

[95] M. Ferrario, M. G. Signorini, and G. Magenes, "Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses," *Medical & Biological Engineering & Computing*, vol. 47, no. 9, pp. 911–919, 2009.

[96] P. Anastasiadis, A. Kotini, P. Anninos, A. Adamopoulos, J. Sigalas, and N. Koutlaki, "Chaotic and periodic analysis of fetal magnetocardiogram recordings in growth restriction," *Prenatal Diagnosis*, vol. 23, no. 5, pp. 405–409, 2003.

[97] F. Ferrario, M. G. Signorini, and G. Magenes, "New indexes from the fetal heart rate analysis for the identification of severe intra uterine growth restricted fetuses," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE.* IEEE, 2006, pp. 1458–1461.

[98] D. Hoyer, H. Friedrich, B. Frank, B. Pompe, R. Baranowski, J. J. Żebrowski, and H. Schmidt, "Autonomic information flow improves prognostic impact of task force HRV monitoring," *Computer methods and programs in biomedicine*, vol. 81, no. 3, pp. 246–255, 2006.

[99] A. Kikuchi, T. Shimizu, A. Hayashi, T. Horikoshi, N. Unno, S. Kozuma, and Y. Taketani, "Nonlinear analyses of heart rate variability in normal and growth-restricted fetuses," *Early Human Development*, vol. 82, no. 4, pp. 217–226, 2006.

[100] A. Kikuchi, N. Unno, S. Kozuma, and Y. Taketani, "Detrended fluctuation analysis of heart rate variability in normal and growth-restricted fetuses." *Gynecologic and Obstetric Investigation*, vol. 65, pp. 116–122, 2008.

[101] H. Gonçalves, J. Bernardes, and D. Ayres-de Campos, "Gender-specific heart rate dynamics in severe intrauterine growth-restricted fetuses," *Early Human Development*, 2013.

[102] R. N. Pollack and M. Y. Divon, "Intrauterine growth retardation: definition, classification, and etiology," *Clinical obstetrics and gynecology*, vol. 35, no. 1, pp. 99–107, 1992.

[103] R. Resnik, "Intrauterine growth restriction," *Obstetrics & Gynecology*, vol. 99, no. 3, pp. 490–496, 2002.

[104] *Neonatal and perinatal mortality : country, regional and global estimates.* Geneva : World Health Organization, 2006.

[105] A. Griffiths, A. Pinto, and L. Margarit, "A survey of methods used to measure symphysis fundal height." *J Obstet Gynaecol*, vol. 28, no. 7, pp. 692–694, Oct. 2008.

[106] A. Bauer, J. W. Kantelhardt, A. Bunde, P. Barthel, R. Schneider, M. Malik, and G. Schmidt, "Phase-rectified signal averaging detects quasi-periodicities in non-stationary data," *Physica A: Statistical Mechanics and its Applications*, vol. 364, no. 0, pp. 423–434, 2006.

[107] P. L. Yudkin, M. Aboualfa, J. A. Eyre, C. W. Redman, and A. R. Wilkinson, "New birthweight and head circumference centiles for gestational ages 24 to 42 weeks." *Early Hum Dev*, vol. 15, no. 1, pp. 45–52, Jan. 1987.

[108] J. Pardey, M. Moulden, and C. W. G. Redman, "A computer system for the numerical analysis of nonstress tests." *American Journal of Obstetrics & Gynecology*, vol. 186, no. 5, pp. 1095–1103, May 2002.

[109] C. W. G. Redman, *Sonicaid FetalCare Antepartum Analysis - Clinical Application Guide*, Oxford Instruments Medical Ltd., 2003.

[110] G. S. Dawes, M. Moulden, and C. W. Redman, "Criteria for the design of fetal heart rate analysis systems." *Int J Biomed Comput*, vol. 25, no. 4, pp. 287–294, May 1990.

[111] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, pp. 65–70, 1979.

[112] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[113] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.

[114] S. J. Haberman, "Maximum likelihood estimates in exponential response models," *The Annals of Statistics*, pp. 815–841, 1977.

[115] M. Alberry and P. Soothill, "Management of fetal growth restriction," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 92, no. 1, pp. F62–F67, 2007.

[116] P. Peirano, C. AlgarÌn, and R. Uauy, "Sleep-wake states and their regulatory mechanisms throughout early human development," *The Journal of Pediatrics*, vol. 143, no. 4, Supplement, pp. 70–79, 2003.

[117] T. Okai, S. Kozuma, N. Shinozuka, Y. Kuwabara, and M. Mizuno, "A study on the development of sleep-wakefulness cycle in the human fetus." *Early Hum Dev*, vol. 29, no. 1-3, pp. 391–396, 1992.

[118] M. Mirmiran, Y. G. H. Maas, and R. L. Ariagno, "Development of fetal and neonatal sleep and circadian rhythms." *Sleep Med Rev*, vol. 7, no. 4, pp. 321–334, Aug. 2003.

[119] J. A. DiPietro, D. M. Hodgson, K. A. Costigan, S. C. Hilton, and T. R. B. Johnson, "Fetal neurobehavioral development." *Child Development*, vol. 67, no. 0009-3920 (Linking), pp. 2553–2567, 1996.

[120] J. Brändle, H. Preissl, R. Draganova, E. Ortiz, K. O. Kagan, H. Abele, S. Y. Brucker, and I. Kiefer-Schmidt, "Heart rate variability parameters and fetal movement complement fetal behavioral states detection via magnetography to monitor neurovegetative development." *Front Hum Neurosci*, vol. 9, p. 147, 2015.

[121] C. Romanini and G. Rizzo, "Fetal behaviour in normal and compromised fetuses. an overview." *Early Hum Dev*, vol. 43, no. 2, pp. 117–131, Oct. 1995.

[122] M. A. van Vliet, C. Martin, Jr, J. G. Nijhuis, and H. F. Prechtl, "Behavioural states in growth-retarded human fetuses." *Early Hum Dev*, vol. 12, no. 2, pp. 183–197, Nov. 1985.

[123] B. Sriram, M. A. Mencer, S. McKelvey, E. R. Siegel, S. Vairavan, J. D. Wilson, H. Preissl, H. Eswaran, and R. B. Govindan, "Differences in the sleep states of IUGR and low-risk fetuses: An MCG study." *Early Hum Dev*, vol. 89, no. 10, pp. 815–819, Oct. 2013.

[124] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. Wiley, 2000.

[125] A. A. Baschat, "Fetal growth restriction - from observation to intervention." *J Perinat Med*, vol. 38, no. 3, pp. 239–246, May 2010.

[126] C. R. Hill, J. C. Bamber, and G. R. ter Haar, Eds., *Physical Principles of Medical Ultrasonics*, 2nd ed. John Wiley and Sons, Ltd., 2004.

[127] G. S. Dawes, G. H. A. Visser, J. D. S. Goodman, and C. W. G. Redman, "Numerical analysis of the human fetal heart rate: the quality of ultrasound records," *Am J Obstet Gynecol*, vol. 141, no. 1, pp. 43–52, 1981.

[128] S. P. Raymond and C. R. Whitfield, "Systolic time intervals of the fetal cardiac cycle." *Baillieres Clin Obstet Gynaecol*, vol. 1, no. 1, pp. 185–201, Mar. 1987.

[129] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[130] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy." *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.

[131] R. Alcaraz and J. J. Rieta, "A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms," *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 1–14, 2010.

[132] S. M. Pincus, "Assessing serial irregularity and its implications for health." *Ann N Y Acad Sci*, vol. 954, pp. 245–267, Dec. 2001.

[133] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability." *Am J Physiol Regul Integr Comp Physiol*, vol. 283, no. 3, pp. R789–R797, Sep. 2002.

[134] O. Rioul and M. Vetterli, "Wavelets and signal processing," *Signal Processing Magazine, IEEE*, vol. 8, no. 4, pp. 14–38, 1991.

[135] C. Valens, "A really friendly guide to wavelets," 1999.

[136] R. J. E. Merry, M. Steinbuch, and M. J. G. van de Molengraft, "Wavelet theory and applications," *Literature Study, Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group*, 2005.

[137] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 674–693, 1989.

[138] V. S. Chourasia and A. K. Mittra, "Selection of mother wavelet and denoising algorithm for analysis of foetal phonocardiographic signals," *Journal of Medical Engineering & Technology*, vol. 33, no. 6, pp. 442–448, 2009.

[139] Y. Murata and C. B. J. Martin, "Systolic time intervals of the fetal cardiac cycle," *Obstetrics & Gynecology*, vol. 44, no. 2, pp. 224–232, 1974.

[140] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

[141] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[142] D. A. Salazar, J. I. Véléz, and J. C. Salazar, "Comparison between SVM and Logistic Regression: Which One is Better to Discriminate?" *Revista Colombiana de EstadÃstica*, vol. 35, pp. 223–237, Jun. 2012.

[143] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[144] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[145] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[146] E. D. Ãbeyli, "Doppler ultrasound signals analysis using multiclass support vector machines with error correcting output codes," *Expert Systems with Applications*, vol. 33, no. 3, pp. 725–733, 2007.

[147] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," in *Data mining techniques for the life sciences*. Springer, 2010, pp. 223–239.

[148] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.

[149] M. W. Fagerland, S. Lydersen, and P. Laake, "The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional." *BMC Med Res Methodol*, vol. 13, p. 91, 2013.

[150] A. Ricke, R. Povinelli, and M. Johnson, "Automatic segmentation of heart sound signals using hidden Markov models," in *Computers in Cardiology, 2005*, Sept 2005, pp. 953–956.

[151] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model." *Physiol Meas*, vol. 31, no. 4, pp. 513–529, Apr. 2010.

[152] D. Gill, N. Intrator, and N. Gavriely, "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model," in *18th Biennal International EURASIP Conference Biosignal*, 2006, pp. 375–383.

[153] L. Gamero and R. Watrous, "Detection of the first and second heart sound using probabilistic models," in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 3, 2003, pp. 2877–2880.

[154] H. Liang, L. Sakari, and H. Iiro, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction," in *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, vol. 4, Oct 1997, pp. 1630–1633 vol.4.

[155] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Computers in Cardiology 1997*, Sep 1997, pp. 105–108.

[156] B. Tutschek, T. Zimmermann, T. Buck, and H. G. Bender, "Fetal tissue Doppler echocardiography: detection rates of cardiac structures and quantitative assessment of the fetal heart," *Ultrasound in Obstetrics and Gynecology*, vol. 21, no. 1, pp. 26–32, 2003.

[157] *Fetal Heart Detector JPD-100S AngelSounds Instruction Manual*, Jumper Medical Equipment Co., Ltd, 2005.

[158] S. H. Eik-Nes, "Chapter 1 - physics and instrumentation," in *Ultrasound in Obstetrics and Gynaecology*, J. W. Wladimiroff and S. H. Eik-Nes, Eds. Edinburgh: Elsevier, 2009, pp. 1–20. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780444518293000015

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

[160] A. P. Dempster and D. B. Laird, N. M. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1–38, 1977.

[161] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," vol. 13, no. 2, pp. 260–269, 1967.

[162] S. V. Vaseghi, "State duration modelling in hidden Markov models," *Signal Processing*, vol. 41, no. 1, pp. 31–41, 1995.

[163] J. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," International Computer Science Institute Berkeley CA, 94704 and Computer Science Division Department of Electrical Engineering and Computer Science U.C. Berkeley, Tech. Rep., 1998.

[164] A. M. Fraser, *Hidden Markov Models and Dynamical Systems*. Society for Industrial and Applied Mathematics, 2008.

[165] G. D. J. Forney, "The viterbi algorithm," *Proc.IEEE*, vol. 61, 1973.

[166] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. The MIT Press, 2001.

[167] S.-Z. Yu, "Hidden semi-Markov models," *Artificial Intelligence*, vol. 174, no. 2, pp. 215–243, 2010, special Review Issue.

[168] K. Hassani, K. Bajelani, M. Navidbakhsh, D. Doyle, and F. Taherian, "Heart sound segmentation based on homomorphic filtering." *Perfusion*, vol. 29, no. 4, pp. 351–359, Feb. 2014.

[169] I. Rezek and S. Roberts, "Envelope extraction via complex homomorphic filtering," Department of Electrical and Electronic Engineering; Imperial College of Science, Technology and Medicine, Tech. Rep., 1998.

[170] J. Tian and R. O. W. Jr., "Vanishing moments and biorthogonal wavelet systems," in *Mathematics in Signal Processing IV. Institute of Mathematics and its Applications Conference Series*. University Press, 1998.

[171] *Cardiac monitors, heart rate meters, and alarms - American National Standard (ANSI/AAMI EC13:2002)*, Association for the Advancement of Medical Instrumentation Std., 2002.

[172] *Testing and reporting performance results of cardiac rhythm and STsegment measurement algorithms - American National Standard (ANSI/AAMI EC57:2008)*, Association for the Advancement of Medical Instrumentation Std., 2008.

[173] J. Behar, A. Johnson, G. D. Clifford, and J. Oster, "A comparison of single channel fetal ECG extraction methods." *Ann Biomed Eng*, vol. 42, no. 6, pp. 1340–1353, Jun. 2014.

[174] S. Zaunseder, F. Andreotti, M. Cruz, H. Stepan, C. Schmieder, H. Malberg, A. Jank, and N. Wessel, "Fetal QRS Detection by means of Kalman Filtering and using the Event Synchronous Canceller," vol. 15, no. 1, pp. 83–89, 2013.

[175] B. Sibai, G. Dekker, and M. Kupferminc, "Pre-eclampsia," *The Lancet*, vol. 365, no. 9461, pp. 785–799, 2005.

[176] E. A. Steegers, P. von Dadelszen, J. J. Duvekot, and R. Pijnenborg, "Pre-eclampsia," *The Lancet*, vol. 376, no. 9741, pp. 631–644, 2010.

[177] A. L. Millman, B. Payne, Z. Qu, M. J. Douglas, J. A. Hutcheon, T. Lee, L. A. Magee, K. R. Walley, and P. von Dadelszen, "Oxygen saturation as a predictor of adverse maternal outcomes in women with preeclampsia." *J Obstet Gynaecol Can*, vol. 33, no. 7, pp. 705–714, Jul. 2011.

[178] F. McAuliffe, N. Kametas, E. Krampl, J. Ernsting, and K. Nicolaides, "Blood gases in pregnancy at sea level and at high altitude." *BJOG*, vol. 108, no. 9, pp. 980–985, Sep. 2001.

[179] NICE, "Hypertension in pregnancy: diagnosis and management. clinical guideline." Online, August 2010. [Online]. Available: http://www.nice.org.uk/guidance/cg107

[180] J. E. A. K. Bamfo and A. O. Odibo, "Diagnosis and management of fetal growth restriction." *J Pregnancy*, 2011.

[181] *Más que una sanadora - Manual práctico de comadronas tradicionales para una maternidad sana*, Farmacéuticos Mundi, 2014. [Online]. Available: http://www.farmaceuticosmundi.org/farmamundi/descargas/publicaciones/MATERNIDAD%20SANA%20manual%20comadrona%20Guate.pdf

[182] *Que debo hacer para tener una maternidad saludable?*, MSPAS y GTI/IEC, Guatemala, 2003.