

# Signal Processing for Newborn Survival

from Labour to Resuscitation

by

**Jarle Urdal**

Thesis submitted in fulfilment of  
the requirements for the degree of

PHILOSOPHIAE DOCTOR  
(PhD)



---

University of  
Stavanger

Faculty of Science and Technology  
Department of Electrical Engineering and Computer Science  
2020

University of Stavanger  
N-4036 Stavanger  
NORWAY  
[www.uis.no](http://www.uis.no)

© Jarle Urdal, 2020  
All rights reserved.

ISBN 978-82-7644-942-6  
ISSN 1890-1387

PhD Thesis UiS no. 535

# Preface

This thesis is submitted as partial fulfilment of the requirements for the degree of *Philosophiae Doctor* at the University of Stavanger, Norway. The research has been carried out at the Department of Electrical Engineering and Computer Science at the University of Stavanger. A period of three months, from March 2018 to June 2018, was spent at the Computer Vision and Behaviour Analysis Lab, Universitat Politècnica de València, Spain.

This thesis is based on a collection of five papers, four published and one currently under review. For increased readability, the papers have been reformatted for alignment with the format of the thesis and are included as chapters.

*Jarle Urdal, April 2020*



# Abstract

Stillbirths are a worldwide challenge, with an estimated 2.6 million stillbirths in 2015, of these 1.3 million are estimated to have died during labour and birth, i.e. fresh stillbirth. In addition to the 2.6 million, one million newborns die within their first and only day of life. Complications due to birth asphyxia are the primary cause of these deaths. The vast majority, 98%, of stillbirths and early neonatal deaths are found in low resource settings.

This thesis investigates two main challenges related to neonatal deaths, 1) fetal heart rate (FHR) and labour analysis, and 2) improving newborn resuscitation. The FHR is known to be important for effectively assessing the well-being of the fetus during labour. In high resource countries, the FHR is measured using cardiotocography for all high-risk labours. While in low income countries, assessment of the FHR is often done manually using a Pinard. With the use of continuous FHR monitoring in low income countries, abnormalities in FHR could potentially be identified at an earlier stage. In this thesis, we facilitate for further analysis of FHR signals by proposing a method to remove less trustworthy time periods of the measured signal, such as noise. And how missing data can be estimated using dictionary learning to allow for continuous analysis. The FHR signals should be interpreted in combination with the uterine activity. We therefore propose a method for detecting uterine contractions using an accelerometer mounted together with the Doppler ultrasound FHR sensor. Finally, we explore how FHR develops during labour and how this trend differs for labours with a normal and adverse outcomes.

For newborns who are unable to start breathing themselves after birth, immediate help from the healthcare workers are crucial. We therefore explore which parameters during newborn resuscitation are important for the resuscitation outcome. One of the identified parameters is the amount of stimulation, i.e. rubbing the back of the newborn. To get a greater understanding of how stimulation affects newborn resuscitation, and how it should be applied, a large number of annotated resuscitation

episodes are required. Manual annotation is both a time consuming and challenging process for the reviewer. We have proposed a complete system for automatically annotating stimulation by using ECG and accelerometer signals measured on the abdomen of the newborn.

# Acknowledgements

I want to express my greatest gratitude to my advisor Prof. Kjersti Engan. No matter the problem or the time, you have always dedicated your time and energy to help with me with my problems. I could not have asked for a better advisor! Thank you to my co-advisor Prof. Trygve Eftestøl for always seeing the bigger picture when the rest of us was stuck looking at the details. A large thank you also goes to my second co-advisor Prof. Hege Ersdal for helping us understand the clinical aspect of our work. A special thank you to Valery Naranjo at Instituto de Investigación e Innovación en Bioingeniería (I3B), Universitat Politècnica de València for welcoming me to Valencia and into her research group.

Thank you to my colleagues and co-authors at Laerdal Medical: Helge Myklebust for running the Safer Births project together with my co-advisor Hege Ersdal, and Joar Eilevstjønn, Sara Brunner and Solveig Haukås Haaland for their work on the Moyo and resuscitation data. A big thank you also goes to my co-authors in Tanzania: Benjamin Kamala, Paschal Mdoe, and Hussein Kidanto. Without all the great work you do, my work would not be possible.

A special thanks to my parents for always cheering me on during my never-ending academic career. Without your support, I would have left academia a long time ago.

Finally, I wish to thank my wife, Anne-Marthe, for all the understanding and support throughout the years as a PhD student. I would not have been where I am today were it not for you.

*Jarle Urdal, April 2020*



# List of publications

The main part of this dissertation is made up of the following published scientific papers:

- **Paper 1**

---

**Noise and contraction detection using fetal heart rate and accelerometer signals during labour**

J. Urdal, K. Engan, Trygve Eftestøl, Ladislaus Blacy Yarrot, Kidanto Hussein, Hege Ersdal

Published in the Scandinavian Health Informatics Conference 2019, SHI 2019.

- **Paper 2**

---

**Estimation of Missing data in fetal heart rate signals using shift-invariant dictionary**

F. Barzideh, J. Urdal, K. Engan, Karl Skretting, Paschal Mdoe, Benjamin Kamala, Sara Brunner, Kidanto Hussein

Published in the 26th edition of the European Signal Processing Conference, EUSIPCO 2018.

- **Paper 3**

---

**Fetal heart rate development during labour**

J. Urdal, K. Engan, Trygve Eftestøl, Solveig H. Haaland, Benjamin Kamala, Paschal Mdoe, Kidanto Hussein, Hege Ersdal

The paper is currently under review.

- **Paper 4**

---

**Signal processing and classification for identification of clinically important parameters during neonatal resuscitation**

J. Urdal, K. Engan, Trygve Eftestøl, Hussein Kidanto, Ladislaus Blacy Yarrot, Joar Eilevstjønn, Hege Ersdal

Published by the IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017.

- **Paper 5**

---

**Automatic identification of stimulation activities during newborn resuscitation using ECG and accelerometer signals**

J. Urdal, K. Engan, Trygve Eftestøl, Valery Naranjo, Ingunn Anda Haug, Anita Yeconia, Kidanto Hussein, Hege Ersdal

Published in Computer Methods and Programs in Biomedicine, Elsevier, 2020.

# Glossary

<b>Acc</b>	Accelerometer
<b>BW</b>	Birth weight
<b>ACOG</b>	American College of Obstetrics and Gynecology
<b>BMR</b>	Bag-mask resuscitator
<b>CV</b>	Cross-validation
<b>CTG</b>	Cardiotocography
<b>CWT</b>	Continuous Wavelet Transform
<b>DL</b>	Dictionary Learning
<b>DFT</b>	Discrete Fourier Transform
<b>DWT</b>	Discrete Wavelet Transform
<b>ECG</b>	Electrocardiography
<b>FHR</b>	Fetal Heart Rate
<b>FSB</b>	Fresh stillbirth
<b>FSDL</b>	Flexible Structure Dictionary Learning
<b>FIGO</b>	International Federation of Gynecology and Obstetrics
<b>GA</b>	Gestational age
<b>HLH</b>	Haydom Lutheran Hospital
<b>HR</b>	Heart rate
<b>IA</b>	Intermittent auscultation
<b>LIC</b>	Low Income Countries
<b>LMIC</b>	Low Middle Income Countries
<b>LNRM</b>	Laerdal Newborn Resuscitation Monitor
<b>LOOCV</b>	Leave-one-out cross-validation
<b>MHR</b>	Maternal Heart Rate
<b>MNH</b>	Muhimbili National Hospital

<b>MOD</b>	Method of Optimal Directions
<b>NCU</b>	Neonatal care unit
<b>NICE</b>	National Institute for Health and Care Excellence
<b>PDF</b>	Probability density function
<b>PO</b>	Pulse oximetry
<b>SI-FSDL</b>	Shift-invariant Flexible Structure Dictionary Learning
<b>SIA</b>	Shift-invariant atom
<b>SNR</b>	Signal to noise ratio
<b>STFT</b>	Short Time Fourier Transform
<b>TRRH</b>	Temeke Regional Referral Hospital
<b>UC</b>	Uterine Contraction
<b>WHO</b>	World Health Organization

# Nomenclature

$x(n)$	A signal vector
$x(n1)$	Sample $n1$ of a signal $x$
$\dot{x}$	The derivative of $x$
$\hat{x}$	The estimate of $x$
$D$	Dictionary as a $N \times K$ matrix
$d_i$	$i$ th atom of the dictionary, $i$ th column of the matrix $D$
$w(i)$	Coefficient corresponding to the atom $d(i)$
$fhr(n)$	Fetal heart rate at sample $n$
$mhr(m)$	Maternal heart rate at sample $m$
$acc(i)$	Acceleration signal at sample $i$
$\ \cdot\ _F$	Frobenius norm
$\ \cdot\ _2$	P-norm



# Contents

<b>Preface</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>List of publications</b>	<b>ix</b>
<b>Glossary</b>	<b>xi</b>
<b>Nomenclature</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Challenge 1: Improving fetal assessment . . . . .	2
1.2 Challenge 2: Improving newborn resuscitation . . . . .	3
1.3 Safer Births research and innovation project . . . . .	3
1.4 Thesis objectives . . . . .	4
1.5 Main contribution . . . . .	5
<b>2 Medical background</b>	<b>9</b>
2.1 Fetal heart rate monitoring . . . . .	9
2.2 Heart rate during newborn resuscitation . . . . .	14
<b>3 Technical background</b>	<b>17</b>
3.1 Doppler Ultrasound . . . . .	17
3.2 Dictionary learning . . . . .	18
3.3 Time-frequency analysis . . . . .	20
3.4 Classification . . . . .	23
3.5 Feature selection . . . . .	28
<b>4 Data material</b>	<b>31</b>

4.1	Safer Births . . . . .	31
4.2	Data collection devices . . . . .	32
4.3	Data material . . . . .	35
<b>5</b>	<b>Fetal heart rate and labour analysis</b>	<b>45</b>
5.1	Previous work . . . . .	45
5.2	Contribution overview . . . . .	46
5.3	Noise detection . . . . .	47
5.4	FHR ambiguity . . . . .	52
5.5	Contractions . . . . .	54
5.6	Estimation of missing data . . . . .	58
5.7	Fetal heart rate development . . . . .	63
<b>6</b>	<b>Newborn resuscitation</b>	<b>71</b>
6.1	Previous work . . . . .	71
6.2	Contribution overview . . . . .	74
6.3	Identification of clinically important parameters . . . . .	75
6.4	Automatic annotation of stimulation . . . . .	80
<b>7</b>	<b>Discussion and conclusion</b>	<b>87</b>
7.1	Fetal heart rate and labour analysis . . . . .	87
7.2	Newborn resuscitation . . . . .	91
<b>Paper 1: Noise and contraction detection using fetal heart rate and accelerometer signals during labour</b>		<b>95</b>
8.1	Introduction . . . . .	99
8.2	Data material . . . . .	100
8.3	Method . . . . .	101
8.4	Experiments and results . . . . .	105
8.5	Discussion . . . . .	108
8.6	Conclusion . . . . .	111
8.7	Acknowledgement . . . . .	111
<b>Paper 2: Estimation of Missing data in fetal heart rate signals using shift-invariant dictionary</b>		<b>113</b>
9.1	Introduction . . . . .	117
9.2	Data material . . . . .	118
9.3	The proposed method . . . . .	119
9.4	Experiments . . . . .	122
9.5	Discussion . . . . .	125

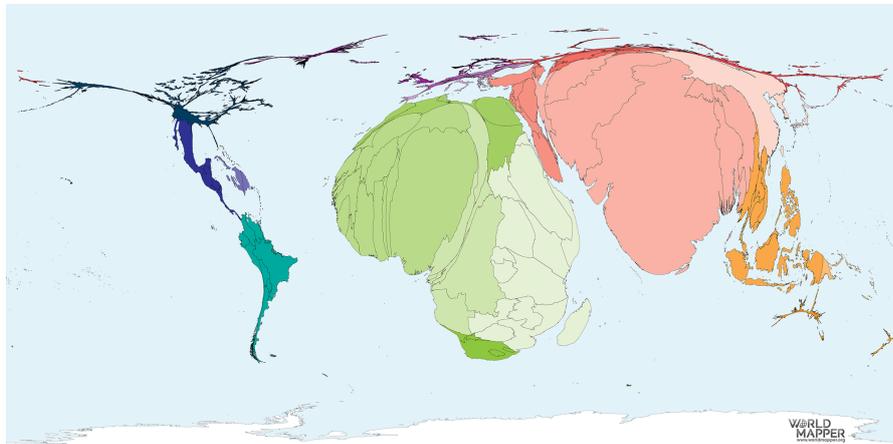
9.6	Conclusion . . . . .	127
9.7	Conflict of interest . . . . .	127
<b>Paper 3: Fetal heart rate development during labour</b>		<b>129</b>
10.1	Background . . . . .	133
10.2	Results . . . . .	134
10.3	Discussion . . . . .	139
10.4	Conclusion . . . . .	143
10.5	Methods . . . . .	144
<b>Paper 4: Signal processing and classification for identification of clinically important parameters during neonatal resuscitation</b>		<b>153</b>
11.1	Introduction . . . . .	157
11.2	Data material . . . . .	158
11.3	Proposed method . . . . .	161
11.4	Experiments and results . . . . .	167
11.5	Discussion . . . . .	169
11.6	Conclusion and future work . . . . .	172
<b>Paper 5: Automatic identification of stimulation activities during newborn resuscitation using ECG and accelerometer signals</b>		<b>173</b>
12.1	Introduction . . . . .	177
12.2	Data material . . . . .	180
12.3	Activity recognition system . . . . .	183
12.4	Experiments . . . . .	186
12.5	Results . . . . .	190
12.6	Discussion . . . . .	192
12.7	Conclusion . . . . .	195
12.8	Conflict of Interest . . . . .	196
<b>Bibliography</b>		<b>197</b>



# Chapter 1

## Introduction

Stillbirths are a worldwide challenge, with an estimated 2.6 million [uncertainty range 2.4-3.0] stillbirths in 2015 [1], of these 1.3 million are estimated to have died during labour and birth, i.e. fresh stillbirth (FSB). In addition to the 2.6 million, one million newborns die within their first and only day of life [1, 2]. Birth asphyxia is the primary cause of these very early deaths. The vast majority, 98%, of stillbirths and neonatal deaths (4 weeks) are found in low resource settings [1]. Using a world map, shown in Figure 1.1, where each country is scaled to indicate the number of newborns who die during their first 4 weeks of life, we see that most of these deaths occur in Africa and Asia. These are regions where access to healthcare may be limited.



**Figure 1.1:** Scaled world map illustration the proportion of countries where newborns died during the first 4 weeks of life. Credit: <https://worldmapper.org/maps/neonatal-deaths-2015/> No changes were made. License: CC BY-NC-SA 4.0<sup>1</sup>

<sup>1</sup><https://creativecommons.org/licenses/by-nc-sa/4.0/>

The Safer Births research and innovation project, which this PhD project has been a part of, aim to reduce fresh stillbirths and newborn mortality (i.e. perinatal mortality) through establishing new knowledge and developing new innovative products to save lives at birth.

Reducing perinatal mortality can be divided into two main challenges: 1) improving fetal assessment to reduce the number of stillbirths and to reduce the need of resuscitation, and 2) improving the resuscitation process.

## 1.1 Challenge 1: Improving fetal assessment

Fetal heart rate (FHR) monitoring is a widely used method to assess the status of the fetus during pregnancy, labour, and birth. In high resource countries, continuous monitoring of the FHR is done using cardiotocography (CTG) for labours categorized as high risk. In low- and middle-income countries (LMIC), an intermittent auscultation is the norm for all labours.

Current guidelines state that the FHR should be within 110-160 beats per minute (bpm) during labour [3, 4], and that auscultation of FHR should be conducted every 15-30 minutes during the first stage of labour, and every 5-15 minutes during the second stage of labour. Each auscultation should last for at least one minute [5]. These auscultation intervals are not possible without a nurse:patient ratio of 1:1 [6] and are a challenge to follow in LMIC where the ratio of healthcare workers to the number of labours is much lower.

A limitation of intermittent auscultations used in LMIC, independent of the device used, is that the status of the fetus is only checked at specific points in time. When the time between each auscultation increases, the possibility of detecting an abnormal FHR may be reduced. To improve assessment of the fetal well-being in LMIC, a system with continuous monitoring of FHR is highly desired. With the introduction of a continuous monitoring system, it is possible to add functionality, such as an alarm, to alert the healthcare personnel if a manual assessment is required. It may also be possible to recognize new elements of risk with the use of machine learning techniques.

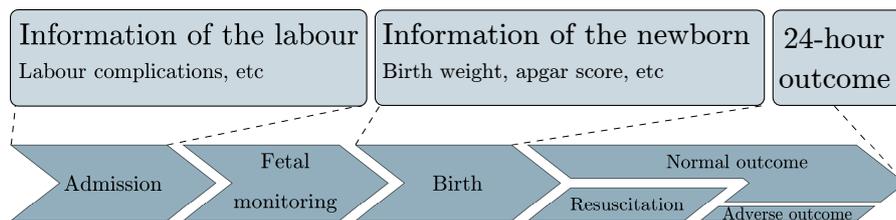
## 1.2 Challenge 2: Improving newborn resuscitation

Guidelines on newborn resuscitation are published by both the World Health Organization and others [7, 8]. The general guideline is to start resuscitation within the first minute after birth if the newborn is unable to start breathing [9]. A gap between the medical guidelines and what is performed has been observed [10].

The therapeutic activities performed during newborn resuscitation includes stimulation, such as firmly rubbing the back of the baby and drying, removal of mucus and obstructions in the airway by suction, and bag mask ventilation. While resuscitation immediately after birth is a crucial part of saving newborn lives, the full understanding of how to best apply therapeutic activities has not been reached. To reduce the perinatal mortality, it is crucial to ensure that the optimal treatment is available and provided during labour, delivery, and immediately after birth when the mortality risk is highest.

## 1.3 Safer Births research and innovation project

Safer Births is a collaborative research and innovation project between multiple Norwegian and international research institutions, as well as hospitals in Tanzania. Within the Safer Births project, observational and signal data describing labours and the newborn have been collected from the first FHR assessment on admission until 24 hours after the time of birth. An overview of the entire data collection period is shown in Figure 1.2.



**Figure 1.2:** A brief overview of a labour from admission and until 24-hours after birth.

When the mother was admitted, an initial FHR assessment was performed and additional data were collected. During labour, the FHR was

continuously measured using a FHR monitor developed by Laerdal Global Health for use in this project. If the newborn was in need of resuscitation immediately after birth, signal data of the newborns heart rate and any ventilation were collected using a resuscitation monitor developed by Laerdal Global Health. Dedicated research assistants were observing and timing important aspects of the resuscitation, such as the time from birth to start of bag-mask ventilation. Status of the newborn 24 hours after birth was set as the final end point.

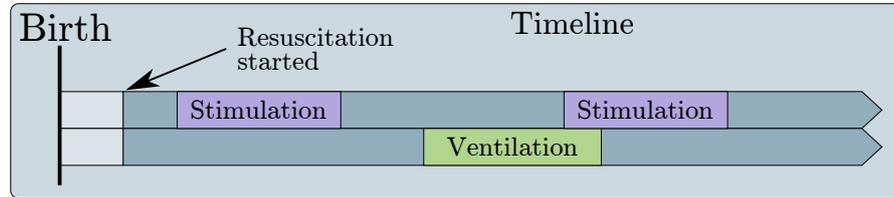
## 1.4 Thesis objectives

The main objectives of this thesis can be summarized by:

- O1: Build the foundation for assessment of the fetal well-being using FHR data collected in the Safer Births project.
- O2: Identify important parameters during newborn resuscitation and propose a complete system to automatically describe the therapeutic activities.

Related to objective 1, FHR data is collected by the Safer Births project from labours assessed as normal on admission from hospitals in a low-income country (LIC). The information contained in this data can potentially be used to improve fetal assessment. It may give us valuable information of the labour process and why labours assessed as normal on admission can get adverse outcomes. To bridge the gap between established signal analysis techniques and field collected data, several signal preprocessing techniques are investigated.

With an automatic description of the therapeutic activities performed during newborn resuscitation it will be possible to evaluate a large number of episodes to identify which resuscitation activities are improving the state of the newborn. Description of the therapeutic activities can be visualized using a timeline, shown in Figure 1.3. In the timeline example, we can see that the resuscitation was started a fixed time after birth, and that both stimulation and ventilation were performed. These timelines can also be used as a debriefing tool, and to identify if the current guidelines are followed.



**Figure 1.3:** One of the objectives in this thesis is to generate timelines describing when stimulation and ventilation are performed during newborn resuscitation.

## 1.5 Main contribution

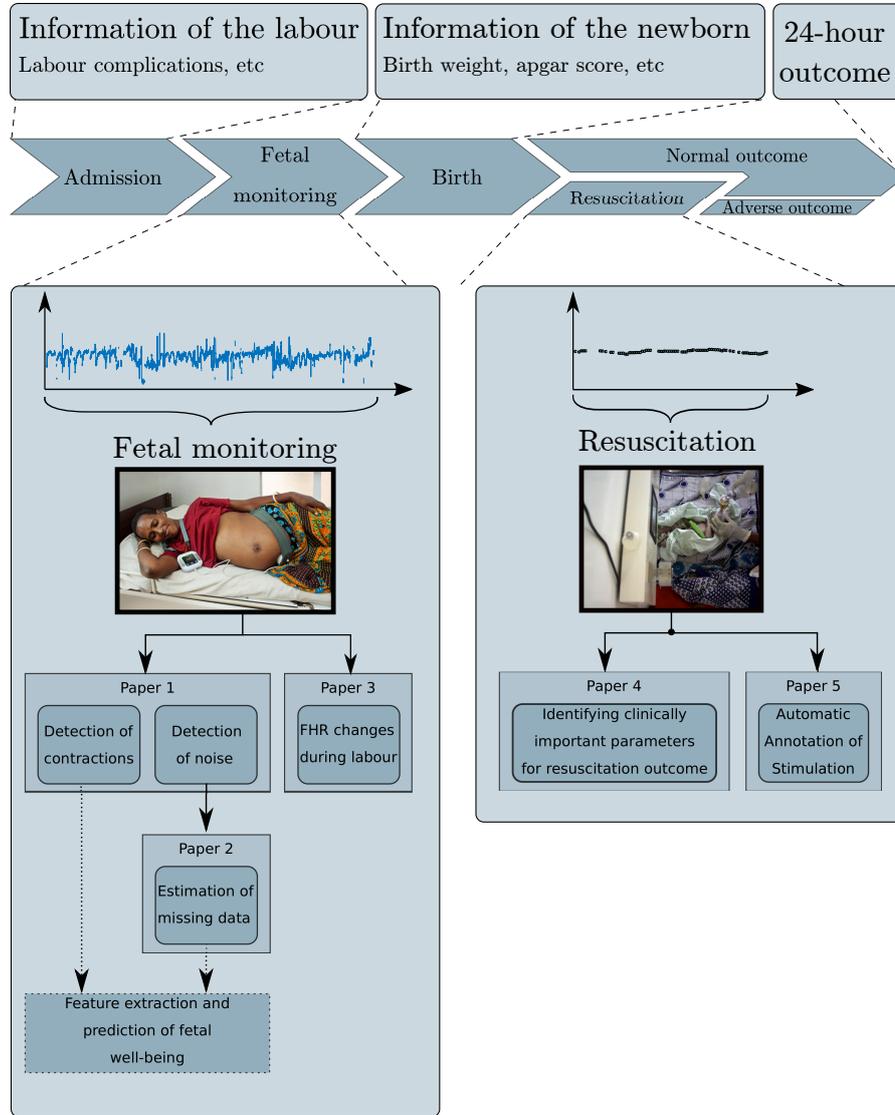
The contribution of this thesis consists of three papers on FHR monitoring and two papers on newborn resuscitation. Of the five papers, four are published and the remaining paper is currently under review. The relationship between all five papers is shown in Figure 1.4.

The left half of Figure 1.4 illustrates the work that has been proposed on fetal monitoring. In paper 1, we propose a method for indicating when contractions occur based on accelerometer signals and a method for identification of noise segments in the FHR signal. To facilitate for continuous time series analysis of the FHR signals, a method is proposed in paper 2 for filling time periods with missing data in the measured FHR signal. Finally, in paper 3, we study how the FHR changes during labour for newborns with a normal and with an adverse outcome. The dashed lines indicate a part not yet implemented but are considered an important future work in our goal for increasing the perinatal survival.

The right half of Figure 1.4 illustrates the work done on newborn resuscitation. In paper 4, we explore which parameters are important for the resuscitation outcome. One of the important parameters is the amount of *time* stimulation activities were performed during the resuscitation. To aid in the work of understanding how stimulation affects the resuscitation, we propose a system in paper 5 for automatically annotating when stimulation activities are being performed using acceleration and ECG signals. This automatic annotation can in turn be used to create useful timelines describing the resuscitation event.

### 1.5.1 Thesis outline

An introduction to the medical background for fetal monitoring and newborn resuscitation is presented in chapter 2 and the technical background in



**Figure 1.4:** An overview of the contributions and relationship between the parts of this thesis. We proposed a method to detect noise in the measured FHR signal, as well as estimating missing data to facilitate for a continuous analysis in the future. We also propose a method to indicate when uterine contractions occur using accelerometer signals, and we explore how FHR changes during labour based on the newborn outcome 24 hours after birth. For newborns that need resuscitation immediately after birth, we explore which parameters are important for the resuscitation outcome, and we propose a complete system for automatically annotating stimulation during newborn resuscitation using ECG and accelerometer signals.

## 1. INTRODUCTION

---

chapter 3. The data material and various data subsets, as well as the devices used for the data collection are presented in chapter 4. Previous work, and the contributions related to FHR monitoring is presented in chapter 5. The chapter include a method of identifying noise, and a method for identifying contractions using an accelerometer. This is followed by an approach for estimating missing data in FHR signals, and finally an observational study of how the FHR changes during labour based on the perinatal outcome. Previous work, and the contributions related to newborn resuscitation is presented in chapter 6. The chapter first includes an analysis of identifying clinically important parameters during newborn resuscitation, followed by a system for creating timelines describing the resuscitation by automatically annotating stimulation using acceleration and ECG signals. A discussion of the contribution to both topics, conclusion and further work are presented in chapter 7. Finally, all five original papers are reformatted to fit in the thesis layout and presented in the remaining chapters.



## Chapter 2

# Medical background

In this chapter, an introduction to FHR monitoring, with focus on guidelines and measurement techniques, is first presented. In the following section, the technique for measuring heart rate after birth is presented.

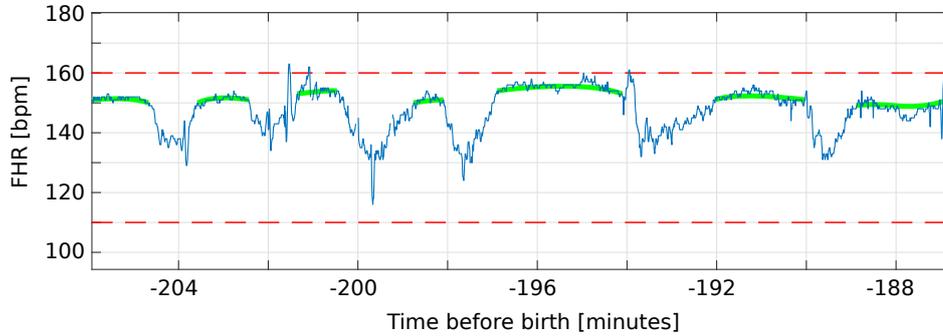
### 2.1 Fetal heart rate monitoring

Assessment of FHR is known to be an effective method to identify the fetal well-being during labour. While FHR is measured and used to assess the fetal well-being in most countries, the techniques and interpretation varies. In this section, current guidelines for interpretation of FHR, as well as measurement techniques for FHR will be presented.

#### 2.1.1 Guidelines for fetal heart rate interpretation

Guidelines for interpretation of the FHR during labour is in active use throughout the world. Three of the most used guidelines are defined by the International Federation of Gynecology and Obstetrics (FIGO) [3], the American College of Obstetrics and Gynecology (ACOG) [11], and the National Institute for Health and Care Excellence (NICE) [12]. In addition, the World Health Organization (WHO) has recommendations on intermittent [13] and continuous FHR monitoring [4].

The guidelines define how CTG features such as baseline, variability, and decelerations should be interpreted. An introduction of these features are given next.



**Figure 2.1:** Example of FHR signal. The FHR is shown in blue and the dashed lines indicate the normal region of FHR. The green lines indicate a baseline in the normal regions. The six time periods with a reduced FHR are known as decelerations.

### Baseline

The FHR baseline is the mean level in the most horizontal time periods and is estimated in 10-minute intervals. In the FIGO guidelines [3], the normal baseline is defined to be in the region of 110-160 bpm. If the heart rate stays above 160 bpm for for than 10 minutes, it is known as tachycardia, and a rate lower than 110 for 10 minutes is known as bradycardia. It is important to notice that values in the 100-110 regions can occur for all labours. In the FHR example, shown in Figure 2.1, the baseline is indicated as the green lines in the flat regions.

### Variability

The variability describes average bandwidth amplitude of the oscillations in the FHR signal, and is evaluated in 1-minute segments. The FIGO guidelines [3] defines a normal variability to be in the range of 5-25 bpm. If the variability stays below 5 bpm for at least 50 minutes, or for at least 3 minutes during a deceleration it is defined as *reduced*. If the variability is  $> 25$  bpm for at least 30 minutes, it is defined as *increased*.

### Decelerations

Decelerations are temporary decreases of  $> 15$  bpm in the FHR for at least 15 seconds [3]. They are typically seen in correlation with uterine contractions and are believed to be caused by compression of the fetal head [3].

### 2.1.2 Measuring techniques

Different approaches for measuring the FHR is used throughout the world. These differences are likely a result of the cost and power requirement of the more advanced methods. An overview of the three most used methods are described in the following subsections.

#### Pinard

The Pinard horn, shown in Figure 2.2, is a simple device used for manual auscultation of FHR. The Pinard is placed on the maternal abdomen and the healthcare worker can then listen and count the heartbeats for a defined time, i.e. 30 seconds, to obtain the FHR. As the device is simple to manufacture, and does not require any power to operate, it is still the primary instrument for midwives in low income countries. An effective use, does however, require training of the healthcare personnel. The pressure applied to the maternal abdomen could also make the auscultation uncomfortable for the mother [14]. While manual auscultation may be sufficient in determining the FHR baseline, it is challenging to determine the FHR variability.



**Figure 2.2:** Pinard fetoscope License: CC0 1.0 Universal (CC0 1.0) <sup>1</sup>

---

<sup>1</sup><https://creativecommons.org/publicdomain/zero/1.0/deed.en>



**Figure 2.3:** The Sonoline B by Baby Doppler, used for intermittent FHR assessment. No changes made. License: CC BY-SA 4.0 <sup>2</sup>

### Doppler Ultrasound

Doppler ultrasound based fetal monitoring is the most common non-invasive approach for measuring FHR. Doppler ultrasound devices commonly provide a simulated sound to allow the healthcare workers to listen to the heartbeats. The monitors utilize an ultrasound sensor attached to the maternal abdomen and the Doppler principle to detect movements in the fetus to compute the heart rate. The ultrasound frequency is usually in the 2-3MHz range. Lower frequencies are less sensitive early in the pregnancy but is often sufficient closer to term. Frequencies in the 3MHz range can be used to detect FHR earlier in the pregnancy. Even higher frequencies can be used for location of blood vessels, or other vessels. A technical introduction to Doppler ultrasound can be found in section 3.1.

---

<sup>2</sup><https://creativecommons.org/licenses/by-sa/4.0/deed.en>

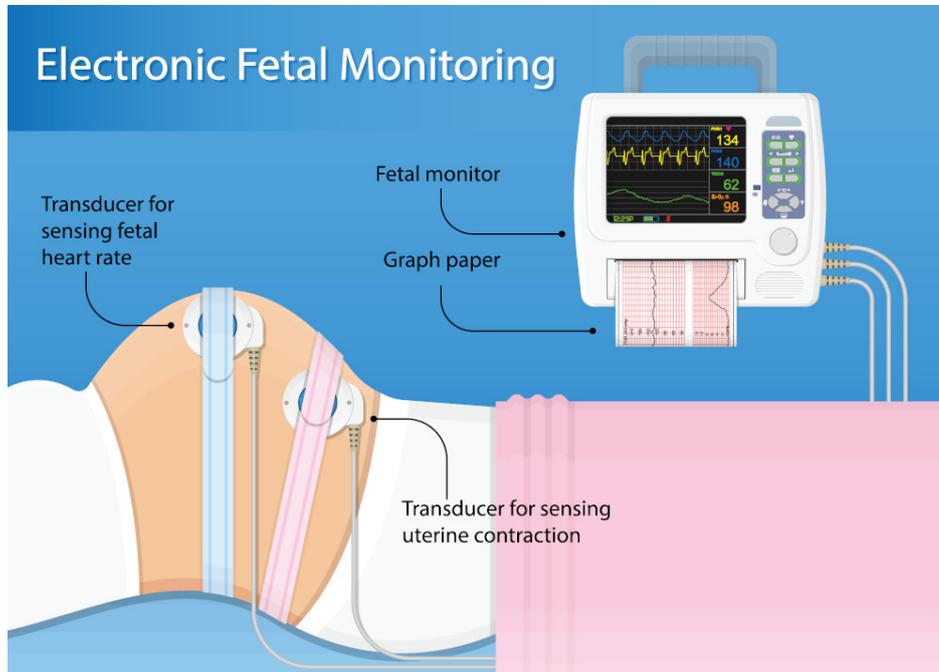
Doppler based fetal monitoring can be divided into two main categories, intermittent- and continuous monitoring, where the main difference is the duration of time the device is used. For intermittent use, a Doppler monitor device can replace manual auscultation. An example of an intermittent monitor is shown in Figure 2.3. When used as a continuous monitor, the device can be attached to the maternal abdomen during the entire labour, i.e. using elastic bands. It can then give the midwives immediate feedback of the FHR. More advanced devices also include a display where the heart rate can be shown for a defined period of time, i.e. 30 minutes. This allows midwives to take baseline and variability changes over time into account when assessing the fetal well-being.

Due to the nature of the Doppler principle, it is susceptible to both halving and doubling of the true rate. If the sensor is placed incorrectly on the maternal abdomen, the device can also incorrectly pick up the maternal heart rate instead of the FHR. If this happens while no midwives are present, it can be challenging to assess what has happened.

### **Cardiotocography**

In high income countries, cardiotocography (CTG) is used to monitor high risk labours. A CTG device can measure the uterine activity using a toco sensor in addition to the FHR measurement found in simpler devices. An example of a CTG monitor is shown in Figure 2.4. The large display provides information of the current FHR and uterine activity, as well as an overview of the history for both measurements.

FHR in CTG is normally measured using a Doppler based ultrasound probe. The probe is attached to the maternal abdomen using elastic bands, illustrated in Figure 2.1.2. If the healthcare providers are unable to get good signal quality during labour, an electrode can be attached directly on the fetus scalp to obtain better measurements. The uterine activity is measured using a strain gauge sensor, attached to the maternal abdomen. This allows us to measure both the frequency and duration, but not the strength of the uterine activity [15]. To obtain information of the strength, an intrauterine catheter (IUPC) can be placed in the uterus. While IUPC is considered the gold standard for measuring uterine activity [15, 16, 17], it requires a ruptured membrane and a cervical dilation of at least 1-2 cm [15].



**Figure 2.4:** Illustration of CTG monitoring during labour. The large display gives the healthcare personnel easy access of the current FHR and uterine activity measurements. It can also show how both measurements develop over time. Credit: rumruay/Shutterstock.com

## 2.2 Heart rate during newborn resuscitation

The state of the newborn at birth can effectively be evaluated by assessing the heart rate [8], and changes in observed heart rate may be a result of resuscitation activities.

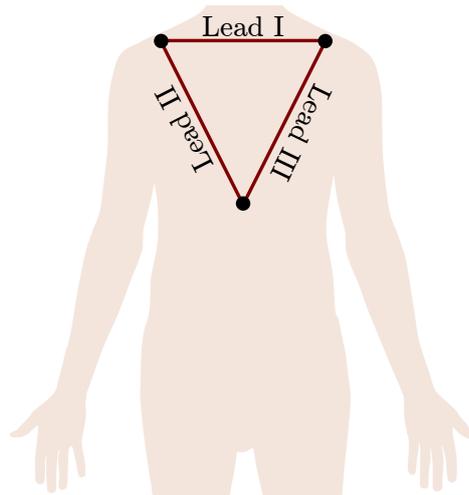
The most commonly used methods to measure the heart rate immediately after birth is by auscultation, palpation of the cord, using pulse oximetry (PO), or electrocardiogram (ECG) [8, 18, 19]. The use of ECG to measure the heart rate does, however, not reduce the need for PO, as ECG does not evaluate the oxygenation of the newborn. The use of PO is, however, shown to underestimate the newborns heart rate immediately after birth, and may therefore cause unneeded interventions [20].

The principle of ECG is to record the electric activity of the heart from multiple locations on the body. The central electrical activity in the heart is the depolarization of the heart cells. When this happens, the depolarization

## 2. MEDICAL BACKGROUND

---

propagates and create an electrical wave out in the body that can be measured. The ECG is found by computing the difference, known as lead, between these measurements. A common approach of measuring the heart rate is using 3-leads, placed in the Einthoven triangle, illustrated in Figure 2.5. A simpler technique, using only 1-lead ECG is normally used for basic heart monitoring.



**Figure 2.5:** Illustration of the sensor placement for 3-lead ECG using Einthoven's triangle. License of body illustration: Public domain

While measurement of the heart rate using ECG is the norm, the International Liaison Committee on Resuscitation (ILCOR) expect that acquiring methods to rapidly apply the ECG leads in the delivery room will take time [8]. A study found the median time from the newborn was placed on the resuscitation table, to the last sensor on a 3-lead ECG was attached was 26 seconds, while placement of pulse oximetry took 38 seconds [21]. After the sensors were attached, it took another 2 seconds for the ECG and 24 seconds for pulse oximetry to obtain a heart rate signal [21]. A study using 1-lead ECG sensor in the NeoBeat prototype, used in the Safer Birth, reduced the median time for sensor placement down to 3 seconds [22]. This reduced time to attach the sensor may be crucial in a resuscitation event.

## 2. MEDICAL BACKGROUND

---

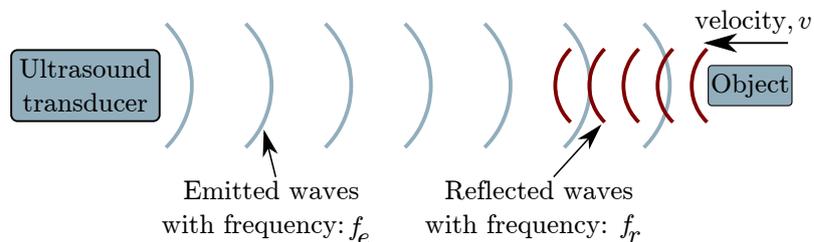
## Chapter 3

# Technical background

In this chapter, an introduction to Doppler ultrasound is first presented. In the following section, the short-time Fourier transform and wavelet transform are presented. Finally, we give an introduction to classification and feature selection.

### 3.1 Doppler Ultrasound

Doppler ultrasound is a commonly used technique to measure flow and movement in application where a non-invasive method is required. The principle is to study a frequency shift, known as Doppler shift, of ultrasound waves reflected from an object.



**Figure 3.1:** Illustration of the Doppler shift. A transducer emits ultrasound waves, the waves hit an object and are reflected. If the object moves towards the transducer, the reflected waves will have a higher frequency than the signal emitted from the transducer. The velocity of the object can then be found based using this shifted frequency.

An illustration of the Doppler shift principle is shown in Figure 3.1. Waves are emitted from the ultrasound transducer. When the waves hit a reflective object, the frequency of the reflected waves will be affected by the movement of the object, known as the Doppler shift. If the object moves towards the transducer, the frequency of the reflected waves will be

increased. The frequency will correspondingly decrease if the object moves away from the transducer.

The reflected frequency,  $f_r$ , is computed using the emitted frequency,  $f_e$ , the speed of sound in the medium,  $c$ , and the velocity,  $v$ , of the reflecting object.

$$f_r = \frac{v}{c} f_e \sin \alpha \quad (3.1)$$

An additional parameter,  $\alpha$ , is used if the object is located on an angle to the ultrasound axis.

Doppler ultrasound measurements are performed by emitting either a continuous wave, or a pulsed wave from the transducer. Both approaches have their advantages and disadvantages. A continuous-wave Doppler can record higher velocities but suffers from lack of depth information. The use of a pulsed-wave Doppler gains information of the depth, but at the cost of aliasing [23, 24]. If the object has a periodic movement, the frequency of this movement can be extracted using a time-frequency analysis. In cases where the measurement contain noise, the first harmonic may have a higher peak than the base frequency, resulting in falsely detecting a doubled frequency of the movement.

## 3.2 Dictionary learning

Sparse representation and dictionary learning are based on the idea that it is possible to represent a signal class sparsely in a domain and that a learned dictionary can represent this domain. Given a  $N \times 1$  signal,  $x$ , its approximation  $\hat{x}$  can be found using a sparse representation,  $Dw$ , of a dictionary,  $D$ .

$$\hat{x} = Dw \quad (3.2)$$

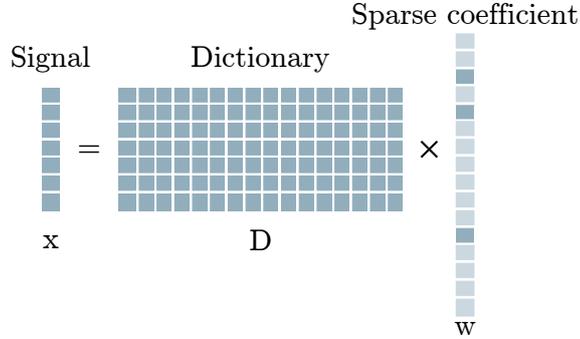
where  $D$  is a matrix of size  $N \times K$ , with the columns ( $d_i$ ) forming dictionary *atoms*. and  $w$ ,  $K \times 1$ , is the vector of *sparse* coefficients. An illustration of this estimation is shown in Figure 3.2. The darker boxes in the sparse coefficient corresponds to the non-zero elements.

The dictionary learning problem is formulated as follows:

$$W, D = \underset{W, D}{\operatorname{argmin}} \|X - DW\|_F^2 \quad s.t. \begin{cases} w_i \text{ is sparse} \\ \|d_i\|_2 = 1. \end{cases} \quad (3.3)$$

### 3. TECHNICAL BACKGROUND

---

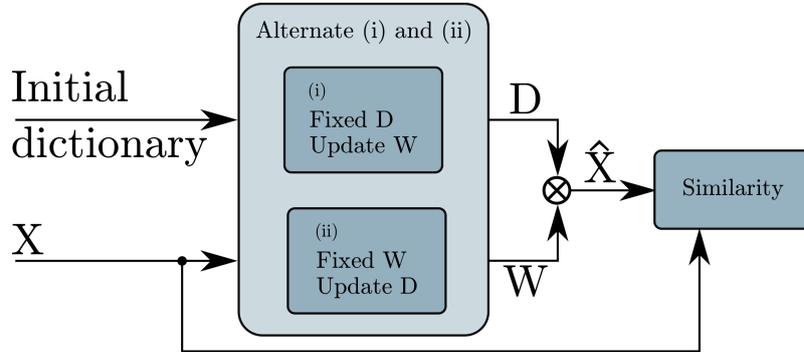


**Figure 3.2:** Illustration of signal estimation using a sparse dictionary representation. The darker boxes in the sparse coefficient corresponds to the non-zero elements.

where  $W$  and  $X$  are formed from concatenation of coefficient  $w_i$  and signal vectors  $x_i$ , respectively, and  $\|\cdot\|_F$  is the frobenius norm. Since equation 3.3 is not tractable, it is usually broken into two steps:

- (i) *sparse coding*: find a sparse  $W$  while keeping  $D$  fixed.
- (ii) *dictionary update*:  $D$  is found while keeping  $W$  constant.

The two steps are then alternated until a set criteria or similarity is achieved [25, 26]. An overview of a dictionary learning framework is shown in Figure 3.3.



**Figure 3.3:** A dictionary learning framework. The sparse coefficient vector,  $W$ , is first found while keeping the dictionary,  $D$ , fixed. The dictionary can then be updated while keeping  $W$  fixed. The two steps are alternated until a set similarity is achieved.

Dictionary learning and sparse approximation have been shown to produce state of the art results in estimation of missing data [27, 28, 29].

$$\begin{bmatrix} \star_0 & 0 & \star_0 & 0 & 0 & \alpha_0 & 0 \\ \star_1 & \star_0 & \star_1 & \star_0 & 0 & \alpha_1 & 0 \\ \star_2 & \star_1 & 0 & \star_1 & \star_0 & 0 & \alpha_0 \\ 0 & \star_2 & 0 & 0 & \star_1 & 0 & \alpha_1 \end{bmatrix}$$

**Figure 3.4:** A simple shift-invariant dictionary with 3 shift-invariant atoms (SIAs). The first two SIAs have 1 shift while the last one has 2 shifts.

A learned dictionary introduces less artefacts during processing, feature extraction, and time-frequency analysis.

When estimating large gaps in missing data, unstructured dictionaries produced by general dictionary learning methods such as Method of optimal directions (MOD) [30] or K-SVD [31] require large atom lengths. Resulting in a large number of free variables. This leads to slow training and usage, as well as the possibility of overfitting.

### 3.2.1 SI-FSDL

The shift-invariant flexible structure dictionary learning (SI-FSDL) imposes a shift-invariant structure onto a FSDL dictionary. This allows estimation of larger gaps by using larger shift-invariant atoms than a general dictionary while keeping the number of free variables fixed.

An example of a small shift-invariant dictionary with three shift-invariant atoms (SIA), is shown in Figure 3.4. This example does not have circular shifts, i.e. the shifting ends as the last non-zero element of a SIA reaches the bottom row of the dictionary matrix. SI-FSDL handles variable length and variable shift atoms as Figure 3.4 illustrates.

## 3.3 Time-frequency analysis

A time-frequency analysis is often used to characterize or manipulate a signal where the signal statistics change over time. In the following subsections, an introduction will be given to the short-time fourier transform and the wavelet transform.

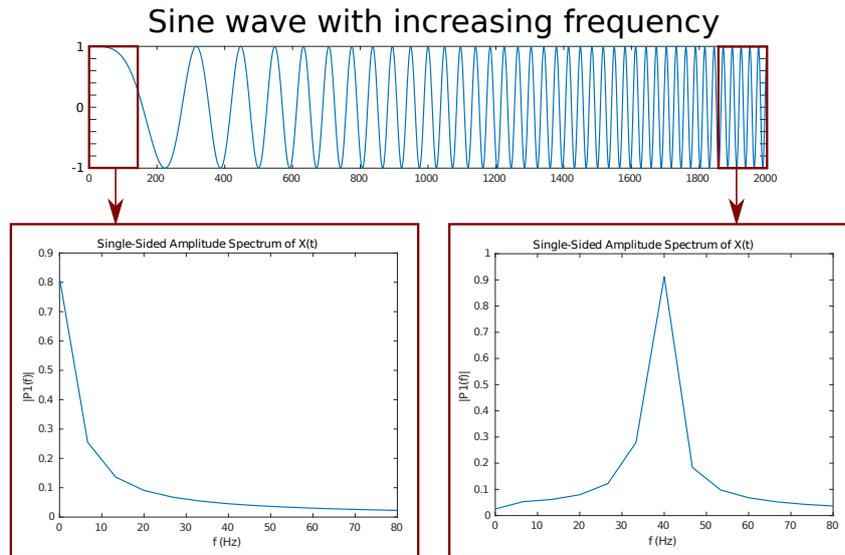
### 3.3.1 Short-time Fourier transform

The short-Time Fourier transform (STFT) is used to analyze the frequency content over time using a sliding window in non-stationary signals. The STFT of a signal,  $x$ , at time  $n$  is found by computing the discrete Fourier transform,  $\mathcal{F}$ , on a sliding window,  $g$  of length  $L$ .

$$\begin{aligned} X(w) &= \mathcal{F}(x(n)) \\ &= \sum_{n=-\infty}^{\infty} x(n)g(n-L)e^{-iwn} \end{aligned}$$

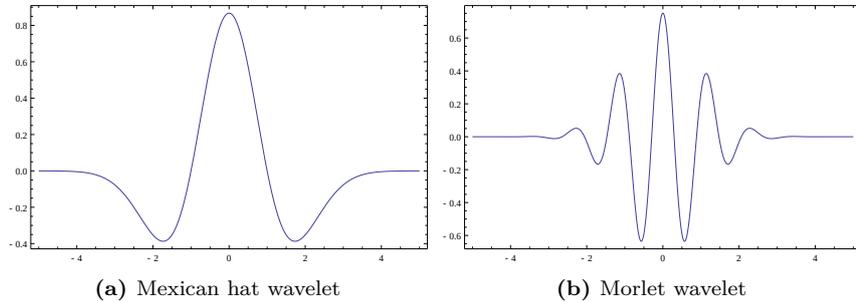
The  $i$  denotes the imaginary unit and  $w$  the frequency in radians. The window function,  $g$ , is often tapered at the edges to avoid spectral ringing.

An example signal with a sine wave of increasing frequency from 0 to 40 Hz is shown in Figure 3.5. In the example, the DFT is computed for the first and last 150 samples of the signal, illustrating how we can detect that the main frequency components changes over time.



**Figure 3.5:** Example of STFT using a sine wave with an increasing frequency from 0 to 40 Hz. The STFT is computed for the first and last 150 samples of the signal. The STFT of the first 150 samples shows a peak frequency of approximately 0 Hz, and the STFT of the last 150 samples of the signal shows a peak frequency of approximately 40 Hz.

Features extracted from the STFT have previously been used with promising results in multiple fields ranging from fingerprint enhancement [32] to



**Figure 3.6:** Example of common wavelets. License for both figures: CC BY-SA 3.0<sup>1</sup>, no changes were made.

ECG arrhythmia classification [33].

### 3.3.2 Wavelet transform

A limitation of the STFT is that the window size is predefined, and therefore equal for all frequencies. A widely used and popular alternative is the wavelet transform [34, 35, 36]. In the wavelet transform, the signal is analyzed in relation to the scale. Longer intervals are used to describe the lower frequency information and shorter intervals are used to describe the high frequency information. The principle is to decompose a signal,  $x(n)$ , using a small oscillating wave known as a wavelet. Using scaled and translated versions of this wavelet, all signals can be represented. Two examples of well-known wavelets, namely the Mexican hat and the Morlet wavelet, are shown in Figure 3.6.

Scaling and translation are performed using a scaling function,  $\phi$ ,

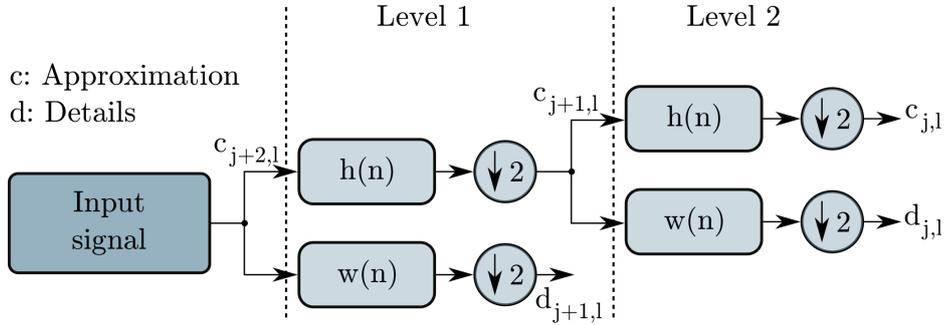
$$\phi_{j,k}(t) = 2^{\frac{j}{2}} \phi(2^j t - k) \quad (3.4)$$

The scaling parameter,  $j$ , indicates the width of the wavelet and the translation,  $k$ , gives its position.

The wavelet function,  $\psi$ , is a short oscillating wave starting and ending at zero, it has a zero mean and a square norm of 1. The wavelet function,  $\psi$ , using coefficients,  $w_k$ , is given by

$$\psi(t) = \sum_k w_k \sqrt{2} \phi(2t - k) \quad (3.5)$$

<sup>1</sup><https://creativecommons.org/licenses/by-sa/3.0/deed.en>



**Figure 3.7:** Multiresolution wavelet decomposition. At each level in decomposition, the signal is split into a low-frequency approximation part ( $c$ ) using a low-pass filter  $h(n)$  and high frequency details ( $d$ ) using a high-pass filter  $w(n)$ . The signal is downsampled by a factor of two in each level.

A popular approach when implementing the discrete wavelet transform (DWT) is using a hierarchical filter structure [37]. Figure 3.7 illustrates a 2-level hierarchical structure generating a 2-level decomposition. This allows for multi-resolution analyses, as the signal is decomposed into several sub-bands. The approximation,  $c$  is found using a low-pass filter,  $h(n)$ , such as

$$c_{j,k} = \sum_l h_{l-2k} c_{j+1,l} \quad (3.6)$$

Details in the signal are found using a high-pass filter,  $w(n)$ , such as

$$d_{j,k} = \sum_l w_{l-2k} c_{j+1,l} \quad (3.7)$$

The wavelet function can be tailored specifically for an application. An example, where it has performed well is in identifying specific segments in ECG signals [38]. The wavelet transform has been utilized in analysis of pathological pregnancies [39] and denoising applications [40, 41] with promising results.

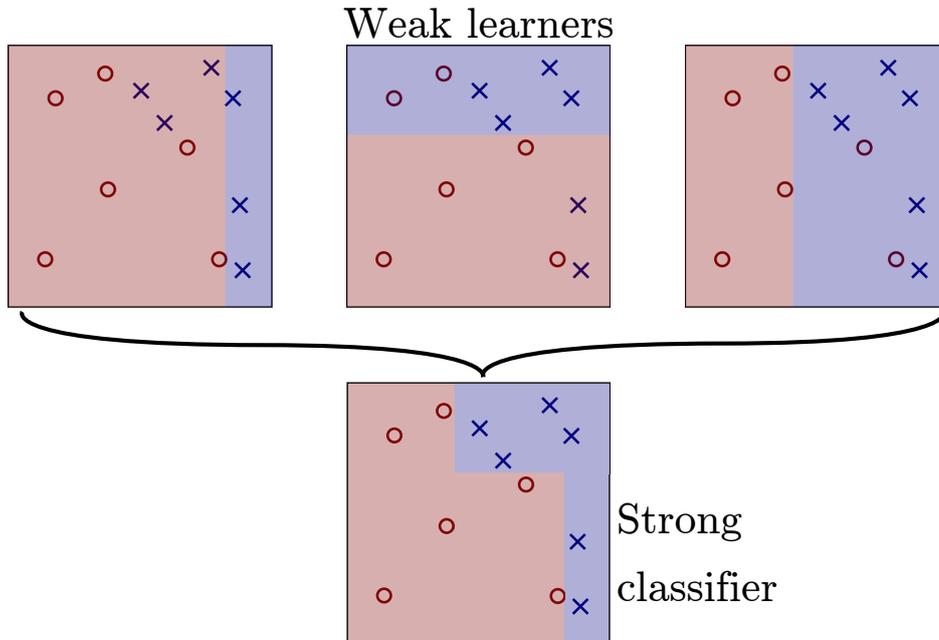
### 3.4 Classification

The principle of classification is to distinguish new observations into different sub populations based on known models. While different methods have been proposed over the years, training a classifier where the number of samples in one class greatly outnumbers another class can be challenging, known as

class imbalance. In cases with class imbalance, traditional methods tend to create models favouring the largest class [42]. Multiple algorithms have been proposed to alleviate this challenge. Some common techniques include data sampling, boosting, or a combination of the two [43].

### 3.4.1 Boosting and data sampling

Boosting [44] is the process of creating a collection of models, where all models are better than random guessing. The models can in turn be combined to create a strong classifier to distinguish the different classes. An example where we use 3 weak learners to create the strong learner is shown in Figure 3.8.



**Figure 3.8:** An example of boosting where we use 3 weak learners to create a strong learner.

By combining boosting with under- or oversampling, we can use the boosting principle on unbalanced data sets. One approach was proposed by Chawla et al. [45] with SMOTEBoost, using synthetic data sampling in combination with AdaBoost [44]. In SMOTE, new synthetic data points are created by upsampling the smaller class based on the existing points.

### 3. TECHNICAL BACKGROUND

---

A drawback of SMOTEBoost is the increased computational cost due to its oversampling technique. As an alternative, Seiffert et al. proposed RUSBoost [42], a hybrid approach using random undersampling in combination with AdaBoost to alleviate a class imbalance. The random undersampling is performed by eliminating data points in the large class until class balance is achieved. The primary drawback of using undersampling is the loss of information. The combination of undersampling and boosting overcome this, as the removed examples are likely to be used in other iterations of the boosting technique. While RUSBoost is a simpler and faster technique it performs comparably to SMOTEBoost [42].

#### 3.4.2 Model evaluation

Evaluation of classifier models are normally conducted using a set of performance metrics. For a two-class problem we can construct a confusion matrix indicating the amount of correct and incorrect classifications of both classes, illustrated in table 3.1.

		Predicted	
		Predicted condition positive	Predicted condition negative
True	Condition positive	True positive (TP)	False negative (FN)
	Condition negative	False positive (FP)	True negative (TN)

**Table 3.1:** Confusion matrix

Using the confusion matrix, accuracy for a model can easily be computed using:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.8)$$

where TP is the true positive, FP the false positive, FN the false negative, and TN the true negative. For models where the performance of each class is of interest, the true positive rate, known as sensitivity, and true negative rate, known as specificity, are often used. The sensitivity is given by:

$$Sensitivity = \frac{TP}{TP + FN} \quad (3.9)$$

And the specificity by:

$$Specificity = \frac{TN}{TN + FP} \quad (3.10)$$

These metrics are ideal when working on two-class problems but suffers from interpretation challenges when the number of classes increases. An alternative approach is therefore to utilize precision and recall. Where precision is the percentage of correct classified observations in the group of observations classified as the given class, given by:

$$Precision = \frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}} \quad (3.11)$$

And recall is the percentage of a true class which has been correctly classified.

$$Recall = \frac{\sum \text{True positive}}{\sum \text{Condition positive}} \quad (3.12)$$

Precision and recall are then computed and evaluated for each class in the model.

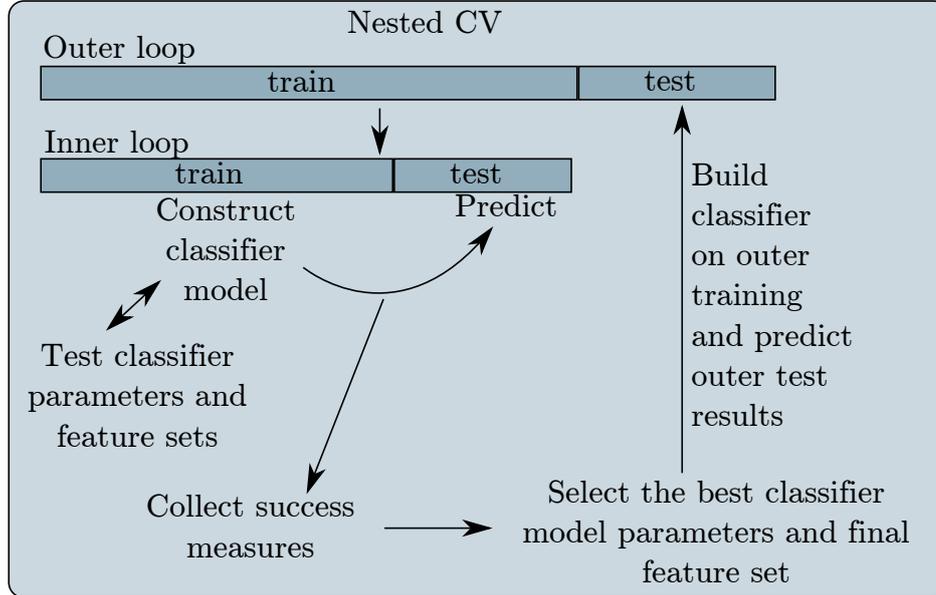
To avoid bias of the computed performance metrics, the data set should be separated into designated train, validation and test subsets. In cases where this is not feasible due to the data size, an alternative approach is the use of cross-validation, described in section 3.4.3.

### 3.4.3 Cross-validation

In cases where dividing the data set into dedicated train, validation and tests sets are not feasible due to the size of the data set, cross-validation (CV) is an often used alternative [46]. This is often the case in the medical field where data collection can be challenging. With CV, it is possible to test on the entire data set, over multiple folds, or iterations.

Two implementations of CV are typically used, Leave-one-out cross-validation (LOOCV) and K-fold CV. The main difference is the amount of data used for testing in each iteration. Given a data set of size N, LOOCV will in each iteration train using N-1 observations and test using the last observation, thus requiring a total of N iterations for the CV. K-fold on the other hand uses N/K of the data set for testing in each iteration, reducing the required amounts of iterations down to K.





**Figure 3.10:** An example of nested cross-validation. By testing different models, parameters and features sets, an optimal classification model can be found using the inner cross-validation. The performance is then evaluated by training and testing the found model using outer loop. As the trained model may change for each validation in the outer loop, the use of nested CV will only give an estimation of the performance that is possible to achieve using a feature subset.

### 3.5 Feature selection

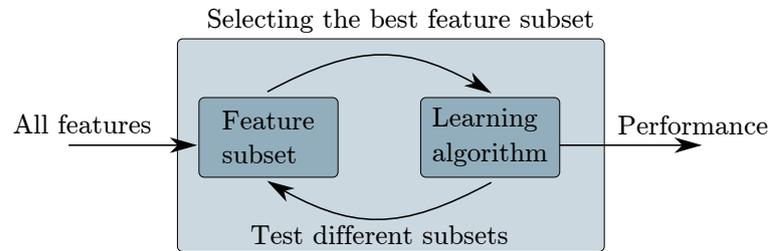
Feature selection identifies the optimal features, which allows us to remove any redundant and irrelevant features as they do not contribute to distinguishing between the classes [48].

In a wrapper-based feature selection [48], shown in Figure 3.11, feature subsets are generated and tested to identify the best set available. As the number of features increase, the number of possible subsets increase and thus the computational complexity. To overcome this, a forward or backward greedy selection is often used. In a greedy forward selection, only the best feature is found in the first iteration. In the next iteration, the best feature complementing the first feature is found. When the order of all features is found, we can identify the number, and the features, which gives the best performance. In a backward feature selection, all features are used in the first iteration. The least important feature is then identified,

### 3. TECHNICAL BACKGROUND

---

and removed in each iteration.



**Figure 3.11:** Wrapper based feature selection. Feature subsets is generated and used in combination with a learning algorithm to identify the best feature subset. The system performance can then be computed.

### 3. TECHNICAL BACKGROUND

---

## Chapter 4

# Data material

The data used throughout this thesis was collected as a part of the large collaborative research project Safer Births. In this chapter, an introduction to the project is given, followed by a description of the devices used for data collection. Finally, an overview of the data sets used in the work is presented.

### 4.1 Safer Births

Safer Births<sup>1</sup> is a large and collaborative research and innovation project between multiple Norwegian and international research institutions, as well as hospitals in Tanzania. The overall goal of Safer Births is to establish new knowledge, and to develop new innovative products to save lives at birth.

The data collection has been conducted at three hospitals in Tanzania, all partners in the Safer Births project. Haydom Lutheran Hospital (HLH) is in the Manyara region, a rural part of Tanzania. The two other hospitals, Muhimbili National Hospital (MNH) and Temeke Regional Referral Hospital (TRRH), are both located in the city of Dar-es-Salaam.

The Safer Births project was approved by the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest), and National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434). Parental verbal consent was obtained for monitoring of both fetal and resuscitated episodes at HLH. Parenteral written consent was obtained for all fetal monitoring episodes at MNH and TRRH. Within the Safer Births project, different subprojects have been subject to randomized trials. However, for the work presented in this thesis, the data collection has been part of an observational study, not an intervention study.

---

<sup>1</sup><http://www.saferbirths.com/>



**Figure 4.1:** Illustration of the Moyo fetal heart rate monitor, Laerdal Global Health AS, Norway. The monitor consist of a sensor device, which can be attached to the maternal abdomen using an elastic band, and a display unit. Illustration reproduced with permission [49]

## 4.2 Data collection devices

The data has been collected using two devices. The Moyo fetal heart rate monitor (Moyo) has been used to collect data during labour, and the Laerdal newborn resuscitation monitor (LNRM), has been used to collect data during newborn resuscitation. The following subsections describe the two devices.

### 4.2.1 Moyo fetal heart rate monitor

The Moyo fetal heart rate monitor, illustrated in Figure 4.1, measures FHR using a 9-crystal pulsed wave Doppler ultrasound sensor. The sensor operates at a frequency of 1MHz, with an intensity of less than  $5\text{mW}/\text{cm}^2$ . The heart rate is computed and logged twice per second, equivalent of a sampling rate of 2Hz. The device is equipped with an accelerometer, sampled at 50Hz, and a temperature sensor, both mounted in proximity of

#### 4. DATA MATERIAL

---

**Table 4.1:** Comparison between Moyo fetal heart rate monitor and conventional CTG

	Moyo	CTG	Comments
FHR	Doppler	Doppler	
Contractions	No*	Yes	*We propose a method to estimate contractions using an accelerometer
Accelerometer	Yes*	No	
Moveable	Yes	No	
Maternal HR	Yes	No	

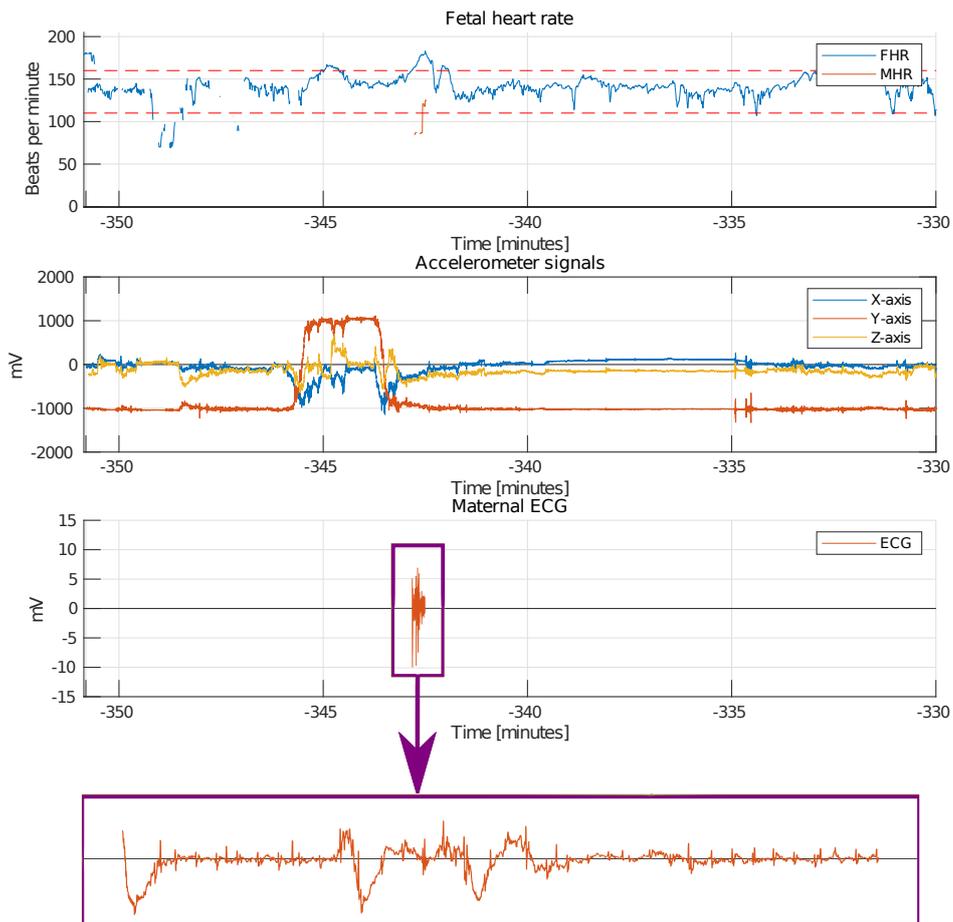
the Doppler ultrasound sensor. Dry-electrode ECG sensors for intermittent measurement of the maternal heart rate (MHR) is mounted on the display unit of the device. The ECG sensor, used to measure the MHR, require the mother to keep one finger from each hand on the monitor. It is therefore only suitable to intermittently assess the MHR and can be used to determine if the Doppler measurement captures the true FHR or if it falsely detects the MHR.

An example of a segment from the signals collected using Moyo during labour is shown in Figure 4.2. The top plot shows FHR and MHR, in blue and orange correspondingly, in relation to the time of birth. The normal region for FHR, i.e. 110 – 160 bpm, is indicated by red dashed lines. The second subplot shows movement captured by the three axis accelerometer. The MHR, in the top plot, is computed using the ECG signal shown in the third subplot.

The use of the Moyo FHR monitor is similar to conventional CTG using external Doppler for measurement of the FHR. The most noticeable difference is the lack of uterine activity measurements. An overview of the main similarities and differences between Moyo fetal heart rate monitor and conventional CTG are shown in Table 4.1. One of the contributions in this thesis, is a method to estimate the time of uterine contractions during labour using accelerometer signals.

#### 4.2.2 Laerdal Newborn Resuscitation Monitor

The LNRM, shown in Figure 4.3, was used to collect data during newborn resuscitation. The resuscitation monitor consists of a main processing unit with a display to show the measured heart rate, as well as a heart rate sensor and a bag-mask resuscitator (BMR). The green heart rate sensor is



**Figure 4.2:** Example signals from the Moyo fetal heart rate monitor. The top plot includes the FHR, shown in blue, and MHR, shown in orange. The red dashed lines indicate the normal region for FHR. The MHR, in the top plot, is computed based on the measured ECG, shown in the bottom plot. The second plot shows the movement captured by the three-axis accelerometer.

## 4. DATA MATERIAL

---



**Figure 4.3:** Laerdal newborn resuscitation monitor with the various sensors indicated. The measured heart rate is shown on the display to give immediate feedback to the healthcare personnel. The green buckle with accelerometer and dry-electrode ECG is a prototype of the NeoBeat. The bag-mask resuscitator includes sensors for measuring pressure, flow, and  $CO_2$ .

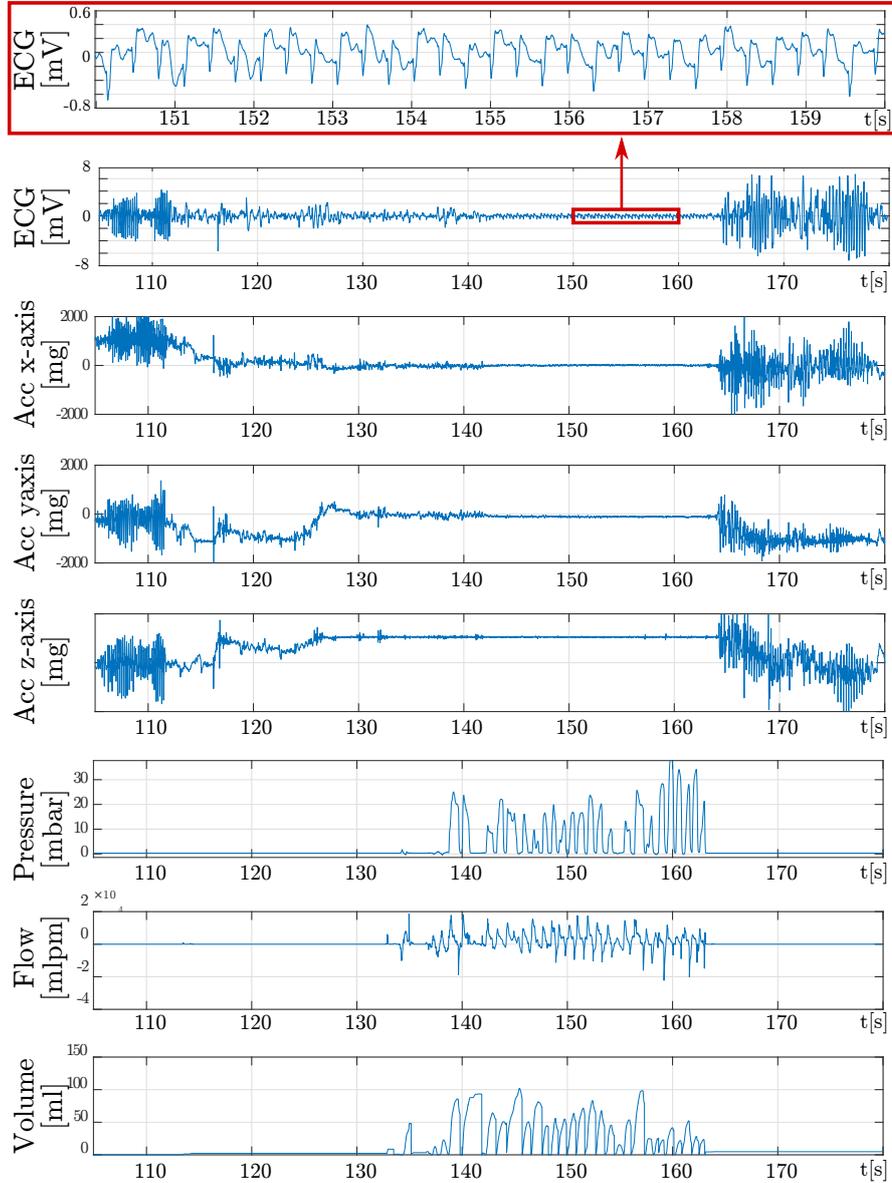
a prototype of the now available product, Laerdal NeoBeat<sup>2</sup>, designed as a part of the Safer Births project. The prototype contains dry-electrode ECG, sampled at 500Hz, and a three-axis accelerometer to monitor movement of the newborn, sampled at 100Hz. The heart rate sensor is designed to be placed over the abdomen of the newborn. This design allows the healthcare personnel to quickly attach the ECG sensor to the newborn and monitor the heart rate. The healthcare personnel can therefore focus on giving the best treatment possible without struggling with gel and placement of traditional ECG sensors. The BMR include sensors for measurement of the pressure and flow, sampled at 100Hz, as well as a sensor for measurement of the  $CO_2$ , sampled at 20Hz. An example of ECG, accelerometer and ventilator signals measured using the NeoBeat prototype and the BMR is shown in Figure 4.4. Due to the combination of dry-electrode ECG sensors and an environment with a lot of movement, the ECG signal contains more noise than what is seen when using traditional ECG sensors in settings with less movement.

### 4.3 Data material

The signal data material was collected between October 2013 and June 2018 by the Safer Births project. All data were pseudonymized using a key before transfer to researchers. The key is kept at the hospitals in Tanzania.

The healthcare workers involved in clinical care using the equipment (i.e. Moyo and LNRM) were trained to follow the existing Helping Babies

<sup>2</sup><https://laerdalglobalhealth.com/products/moyo-fetal-heart-rate-monitor/>



**Figure 4.4:** Example signals of ECG (50 Hz filtered), acceleration in three axis, ventilation, pressure, flow, and volume (integrated from flow). A magnified section of the ECG is included to illustrate the dynamic range of measurements with little noise. An intervention from the healthcare workers are seen during the first 15 seconds, followed by a movement of the newborn. A ventilation sequence is seen from 135 to 165 seconds, followed immediately by a new intervention from the healthcare workers.

Breathe (HBB) guideline<sup>3</sup> for newborn resuscitation. The HBB guideline states what should be assessed and what action to perform if a newborn is asphyxiated and need help to start breathing. A poster with the guideline was posted on the wall above each resuscitation table. A limitation of the guideline is that it only define which activities to perform, and not the amount, or how long the activities should be performed.

Additional clinical data related to the labour and resuscitation process was observed and recorded continuously by designated research assistants present at the labour ward. This data includes information from:

- Admission, such as gestational age and maternal age.
- Labour, such as presence of abnormal FHR and any interventions.
- Birth, such as weight of the newborn, sex, and 1- and 5-minute Apgar score.
- Outcome, such as newborn status 30 minutes and 24 hours after birth.

A full overview of the additional collected data in HLH can be found in Mdoe et al- [50], and the additional data collected at MNH and TRRH in Kamala et al. [51].

The prospective data collection was ongoing during the time period of this PhD work, and as a result of this, the data sets were updated between the various publications. Details of the data collection and differences of data sets between the papers are described in the following subsections.

### 4.3.1 Fetal heart rate data set, Moyo

The FHR data consists of labour episodes all measured using Moyo, presented in section 4.1. In addition, data is collected at admission and birth, including maternal age, duration of the stages of labour, any interventions, 30-minute and 24-hour outcome.

Data from 1087 labours were collected at TRRH, from 669 labours at MNH, and from 1617 labours at HLH. Only labours assessed as normal on admission were included in the study. An overview of the collected data and used data sets are shown in Figure 4.5.

---

<sup>3</sup><https://shop.aap.org/helping-babies-breathe-2nd-ed-action-plan-wall-poster-paperback/>

The data collection has been an ongoing process, and as a result of this, the amount of data available for analysis was expanded throughout this work. In Paper 2, 1399 episodes in  $F_{D2}$  were available. This data set consisted of all data from TRRH (n=1087), 227 episodes from MNH and 85 episodes from HLH. In Paper 1 all data (n=3807),  $F_{D1}$ , were included. For Paper 3, all collected data was used. However, data from 96 episodes were excluded as the measured data could not be matched to a specific labour. The data set used in paper 3,  $F_{D3}$ , is further divided into the following four outcome groups based on the newborn status 24 hours after birth:

- Normal
- Newborn care unit (NCU)
- Early neonatal death (END)
- Fresh stillbirth (FSB)

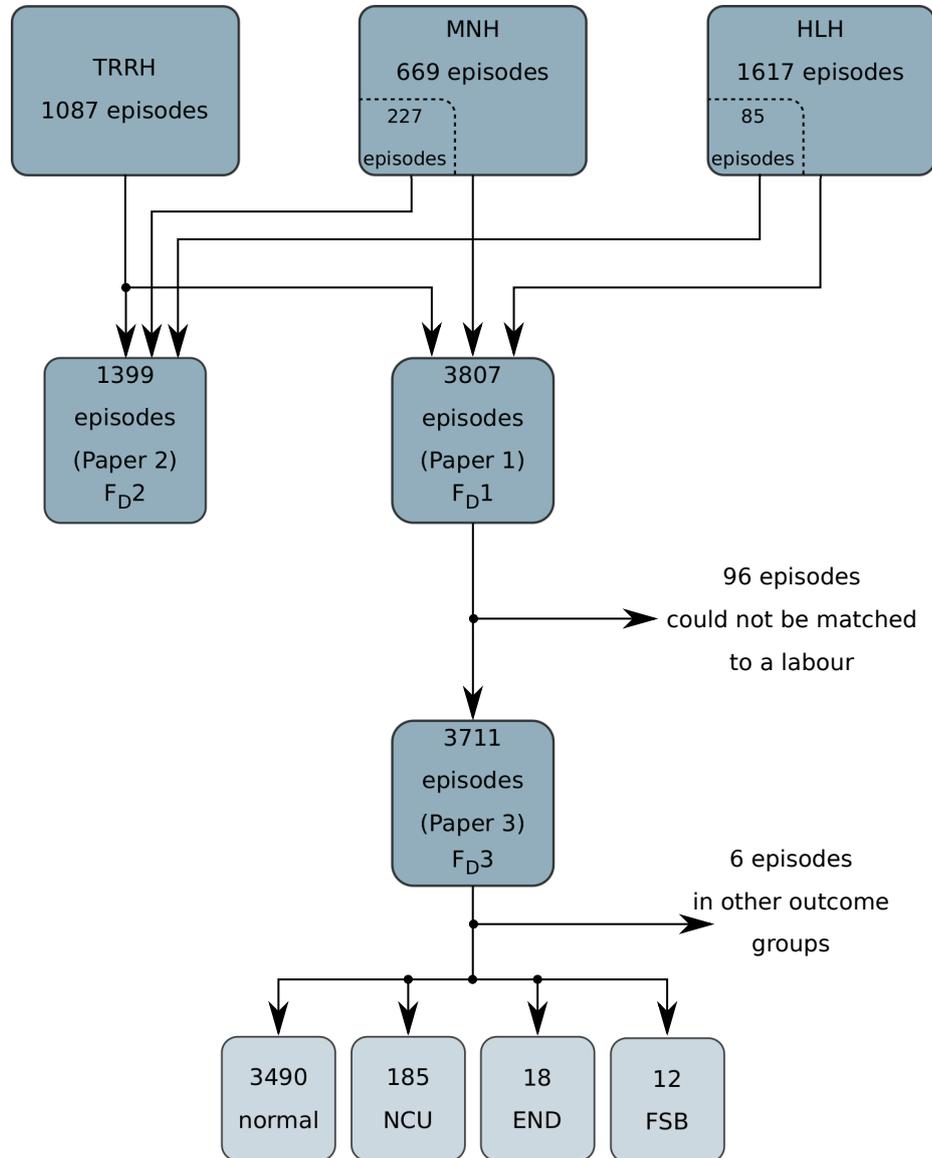
6 episodes were excluded as they were annotated as seizures, referred or the outcome was missing. An overview of the data sets, and number of episodes are shown in table 4.2.

#### 4.3.2 Resuscitation data set

The resuscitation data consists of resuscitation episodes all measured at HLH using the LNRM, presented in section 4.3. This data consists of synchronized measurements of the ECG, heart rate, pressure, flow, CO<sub>2</sub>, and accelerometer. In addition, data was collected at admission and at birth, including information of the maternal age, duration of the stages of labour, any interventions, as well as 30-minute and 24-hour outcomes. Videos overlooking the resuscitation table were also collected to facilitate for manual annotation of the resuscitation episodes at a later time. The resuscitations have been grouped based on the newborn status 24 hours after birth:

- Normal
- Still in neonatal care unit (NCU)
- Death

#### 4. DATA MATERIAL



**Figure 4.5:** Overview of the Moyo data. A total of 3807 episodes were collected from the three hospitals, used in Paper 1. 96 episodes were excluded as they could not be matched to a specific labour. In the remaining 3711 episodes, used in Paper 3, 6 episodes were excluded as they did not belong in the four defined outcome groups: normal, neonatal care unit (NCU), early neonatal death (END), and fresh stillbirth (FSB). A smaller subset consisting of 1399 episodes were used in Paper 2 as this was the data available at the given time.

**Table 4.2:** Summary of data sets

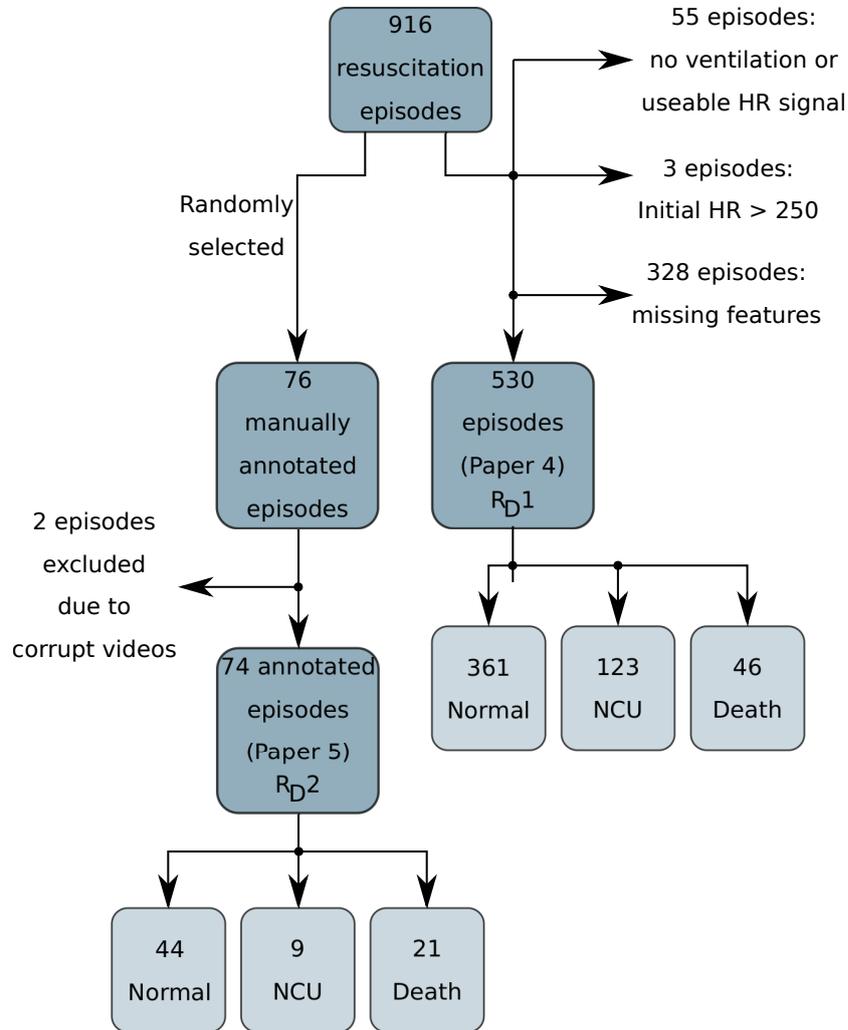
	Name	Episodes	Comment
FHR	$F_D1$	3807	Paper 1
	$F_D2$	1399	Paper 2
	$F_D3$	3711	Paper 3
	$F_D3s_1$	3490	Newborns from $F_D3$ with a normal outcome
	$F_D3s_2$	185	Newborns from $F_D3$ admitted to NCU at 24 hours
	$F_D3s_3$	30	Newborns from $F_D3$ which died during the first 24 hours
Resuscitation	$R_D1$	530	Paper 4
	$R_D2$	74	Paper 5
	$R_D2_{blk}$		Block subset of $R_D2$
	$R_D2_{win}$		Sliding window subset of $R_D2$

A total of 916 resuscitation episodes were recorded between October 2013 and August 2016 at HLH. Outcome at 24 hours included 617 labelled as normal, 194 still admitted to NCU, 48 deaths, and 57 episodes classified as fresh stillborn. Heart rate is however observed on 27 episodes (mis)classified as fresh stillborn by the midwife. Resuscitation was attempted in these episodes and are therefore reclassified and included in the study as *death*, thus the *death* group includes  $(48 + 27 =)75$  episodes. An overview of the data collection and used data sets are shown in Figure 4.6.

### Manually annotated data

A subset of 76 episodes were selected and annotated by carefully watching the corresponding videos and making a timeline synchronized with all the collected signals for the episode. The subset consists of all resuscitation episodes ending in death at the time of annotation, as well as randomly selected cases from episodes where the newborn survived.

The resuscitation episodes were annotated by two independent reviewers watching the videos; one neonatologist and one human factor engineer. If the resuscitation lasted longer than seven minutes, only the first seven



**Figure 4.6:** Overview of the resuscitation data. A total of 916 resuscitation episodes were collected. 55 episodes were excluded as they did not include any ventilation or had a usable heart rate signal, 3 episodes were excluded as the initial heart rate was above 250 beats per minute, and 328 episodes were excluded as the defined features could not be computed. Of the collected 916 resuscitation episodes, 76 episodes were randomly selected and manually annotated. 2 of these episodes were excluded due to corrupt video files, resulting in a total of 74 episodes with manually annotated timeline data.

minutes were annotated. In cases with agreement score less than 80%, the two reviewers sat together and obtained consensus. The following activities and/or categories were annotated: 1) *stimulation*, 2) *suction*, 3) *uncovered*, 4) *other*, 5) *obscured view*, and 6) *start/stop of resuscitation*. Stimulation and suction are considered two of the three primary treatment events performed during resuscitation in addition to ventilation of the newborn. *Uncovered* describes how much of the newborn that is covered by a blanket. This is considered important information, as covering the newborn will result in a lower heat loss. The fourth category includes other activities that are considered relevant for the treatment, for example clamping of the umbilical cord and injections.

The heart rate sensor was observed to be detached and later reattached during a resuscitation episode. As this will contribute to errors and missing data in the data set, attachment of the heart rate sensor was annotated. When classifying stimulation, only time periods where the heart rate sensor is fully attached to the newborn were used.

#### **Data set to identify clinically important parameters**

A data subset,  $R_D1$ , was created to indicate which parameters were clinically important for the resuscitation outcome. Features describing the initial condition of the newborn, the given treatment, and early response were defined. More details of the defined features can be found in section 6.3.

When determining which of the features are important, we only included episodes where all features could be extracted. 55 of the recorded episodes were excluded, as they did not contain any ventilation or a usable heart rate signal. 3 episodes were excluded as they contain an initial heart rate above 250, and 328 episodes were excluded as not all features can be extracted. An overview of the exclusions are shown in table 4.3. Of the 328 excluded episodes, 83 episodes were missing 1 feature, 46 episodes were missing two features, and 199 episodes were missing 3 or more features. The largest contributors were heart rate 30 and 60 seconds after the first ventilation which were missing in 238 of the 328 excluded episodes.

An overview of the newborn outcomes for the 530 episodes in  $R_D1$  is shown in Figure 4.6.

#### 4. DATA MATERIAL

---

**Table 4.3:** Exclusion of resuscitation episodes in creation of the data subset  $R_{D1}$ .

	# episodes
Original data set	916
Episodes without useful heart rate signal or ventilation	55
Episodes with first heart rate $> 250$	3
Episodes with missing features	328
Final data set	530

#### Annotated data set

A smaller data set,  $R_{D2}$ , was defined based on the manually annotated data. This data set was used to train and validate a system for automatic annotation of resuscitation episodes. The trained system can in turn be used to annotate new resuscitation episodes. Two of the manually annotated episodes were excluded as the video files were corrupt and could therefore not be used to annotate when the sensor was attached. The data set used in this work therefore consists of 74 episodes of newborn resuscitation.

Newborn resuscitation can involve the effort of multiple healthcare providers, resulting in more than one therapeutic activity being performed at the same point in time. While most of the resuscitations were observed to be performed by a single healthcare provider, time periods with multiple healthcare providers will introduce challenges in distinguishing between the therapeutic activities. A data subset is therefore created with all time periods where some activity is being performed, and only stimulation or no therapeutic activity is manually annotated. Time periods with an activity is found using VuDetector [52], further details of this method is presented in section 6.1.1. Using a sliding window of 1 second, and 900ms overlap, the subset,  $R_{D2_{win}}$ , consists of 15,958 time points of stimulation and 3,653 time points of non-therapeutic activities. Grouping these based on the manual annotations, we obtained  $R_{D2_{blk}}$  with 464 regions with stimulation, and 357 regions with non-therapeutic activities. An overview of the data sets is shown in table 4.2.



## Chapter 5

# Fetal heart rate and labour analysis

In the following chapter, previous work on FHR and labour analysis is first presented, followed by our contribution to the topic.

### 5.1 Previous work

Assessment of the FHR is known to be an effective method to determine the fetal well-being during labour, but a visual interpretation of the FHR obtained in CTG is shown to be insufficient [53]. CTG is the usual method of monitoring labours at risk in high income countries, and the method is associated with a reduction in the number of neonatal seizures and an increased number of caesarean sections. But there is no clear association with improved mortality rate [54]. Any potential intervention based on the assessed FHR should also consider the maternal condition [55]. A study in LMIC using intermittent auscultation found no evidence that higher detection rates of abnormal FHR having any impact on the perinatal outcome [56].

While CTG cannot be considered a therapeutic device, it can be used as a screening tool to identify risk [57]. Several systems for continuous analysis of the CTG are under development, including Sisporto [58] and OXSYS [59, 60]. With the use of computerized systems for assisting in the analysis of CTG from fetuses at risk, two main approaches have surfaced: 1) Mimicking what the clinicians are doing, and 2) data-driven interpretation [57].

As measurements of FHR using Doppler ultrasound is susceptible to signal dropouts, described in section 2.1.2, any missing data should be

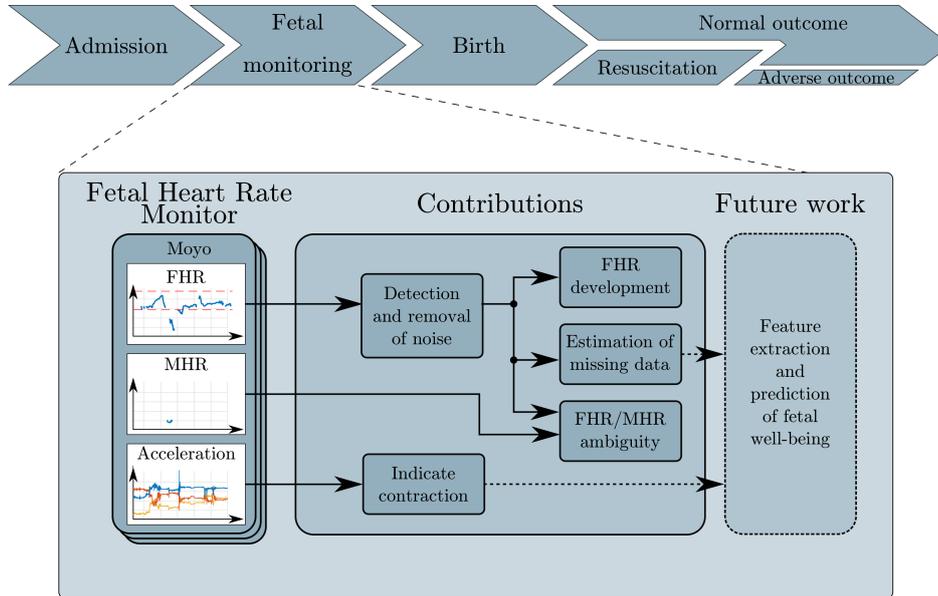
estimated prior to any automated analyses. Simple methods such as linear interpolation[61] and cubic Hermite spline interpolation[62], and more complex methods such as Gaussian processes [63] and K-SVD [29] have previously been used to estimate the missing samples on FHR recorded through CTG. However, depending on the length of the gaps, these methods can affect computation of the traditional heart rate *features* such as variability. It is also important to ensure that the MHR is not mistaken for the FHR during the labour [64].

Deep learning has emerged in the later years with promising results using ECG signals for detecting myocardial infarction [65] and arrhythmias [66]. Using data from 35,429 deliveries, Petrozziello et al. [67] showed that a convolutional neural network (CNN) could also be used on CTG data. The system outperformed long short-term memory (LSTM), clinical practice and Oxsys 1.5 when distinguishing between normal and compromised labours with a sensitivity of 42%. An artificial neural network consisting of extracted signal features and clinical features has also shown promising results to identify the outcome based on pH [68].

A challenge in many of the current studies, is the lack of adverse outcomes, such as stillbirths, brain damage or deaths [57]. This creates a challenge in how the endpoint of intrapartum monitoring should be defined for these analyses. In the literature, single cord gas parameters, such as pH or base deficit, or hypoxic-ischemic encephalopathy (HIE) is often used as the endpoint. But a low pH or high base deficit do not always indicate that the newborn would require any special care [69, 70].

## 5.2 Contribution overview

In the following sections, the contribution of the methods illustrated in each of the blocks in Figure 5.1 will be presented. Noise detection and removal in FHR signals are presented in section 5.3. Followed by an analysis of the MHR and MHR/FHR ambiguity in section 5.4, estimation of when contractions occur based on analysis of the acceleration signal in section 5.5, and estimation of missing data in section 5.6. Finally, an analysis of how the FHR develops during labour for newborns with normal and adverse outcome is presented in section 5.7.

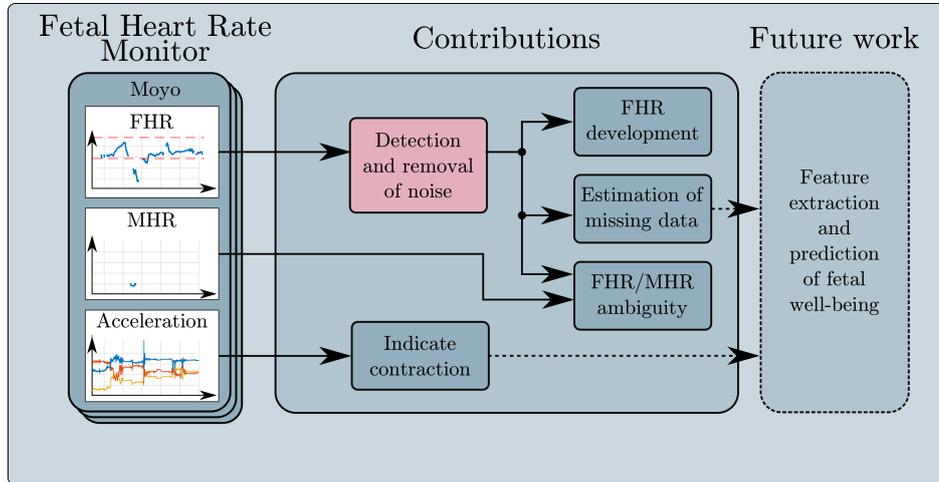


**Figure 5.1:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work.

### 5.3 Noise detection (Paper 1)

In this section, the contribution in the pink *Detection and removal of noise* block in Figure 5.2 is presented. Full details of this method can be found in paper 1.

The use of Moyo, section 4.2.1, at several hospitals in Tanzania, provides the opportunity to study the FHR changes and patterns without relying on human interventions to conduct periodic measurements. Well-known problems with continuous Doppler ultrasound devices, section 2.1.2, like traditional CTG, are that they are susceptible to noise and missing signal data. While the noise may be the result of not detecting the fetus correctly, it may also be caused by a doubling and halving of the FHR signal due to the Doppler principle. Missing data can be caused by sensor movement and suboptimal placement of the sensor. Any artefacts due to noise may affect the interpretability and should be removed for both visual interpretation and further digital analysis. Methods for classification and suppression of noise [71], and removal of the MHR [72] have previously been used on fetal ECG. In the following section, we propose a method to identify this



**Figure 5.2:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work. This is a repetition of Figure 5.1 with the part presented in this section, detection and removal of noise, marked in pink

noise using only the sampled heart rate. This allows for use on low-cost continuous FHR monitoring devices, where the ECG is not available.

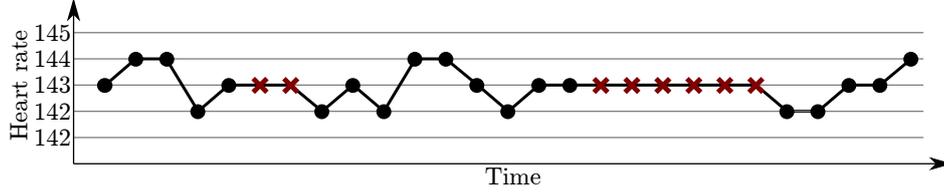
### 5.3.1 Method

The FHR collected using Moyo does not contain information of which samples can be regarded as noise, or which samples are affected by doubling or halving. Due to this lack of truth data, a conservative approach is used to only correct or remove measurements which cannot be explained from a physiological perspective.

To identify the time periods, hereafter called segments, where variations in the FHR cannot be explained from a physiological perspective, we first fill missing data in the FHR using forward replication, given by

$$fhr_{rep}(n+1) = fhr(n) : fhr(n+1) = 0 \quad \forall n \quad (5.1)$$

In practice, any missing samples is filled using the last known sample before the gap, also known as zero-order-hold. An illustration of forward replication is shown in Figure 5.3. The black dots indicate measurements of the heart rate, and the red crosses indicate samples inserted using forward replication.



**Figure 5.3:** Example of forward replication, also known as zero-order-hold. Black dots indicate measurements of the heart rate, and red crosses indicate missing samples filled in using forward replication.

Let  $s$  be a pair of time indexes  $(t_s, k)$  representing the start time and length of a segment. Let  $A$  be a set of  $s$ ,

$$A = \{s : |f\dot{h}r_{rep}(t_s)| > c \cap |f\dot{h}r_{rep}(t_s + k)| > c \cap k < T_k\} \quad (5.2)$$

As the measured FHR is a result of a biological process, physiological limitations exist for how fast the heart rate can change. The threshold  $c$  is therefore set to 30 beats per minute. The threshold  $T_k$  is set to 25 seconds based on Barzideh et al. [73].

The segments are checked in order from the shortest to the largest to see if the large signal variation is a doubling or halving caused by a Doppler shift error. Let  $fhr_h(n)$  and  $fhr_d(n)$  denote the intersample variation for halving and doubling corresponding to:

$$fhr_h(n) = |2 \cdot fhr(n) - fhr(n-1)| \quad (5.3)$$

$$fhr_d(n) = |0.5 \cdot fhr(n) - fhr(n-1)| \quad (5.4)$$

And the cleaned signal,  $fhr_c$ , be defined as:

$$fhr_c(n) = fhr(n) \quad (5.5)$$

The shift errors are identified by comparing the intersample variation to a threshold  $T_D$ , allowing for some intersample variability. The shift errors are then corrected using:

$$fhr_c(n) = 2 \cdot fhr(n) : fhr_h(n) < T_D \quad (5.6)$$

$$fhr_c(n) = 0.5 \cdot fhr(n) : fhr_d(n) < T_D \quad (5.7)$$

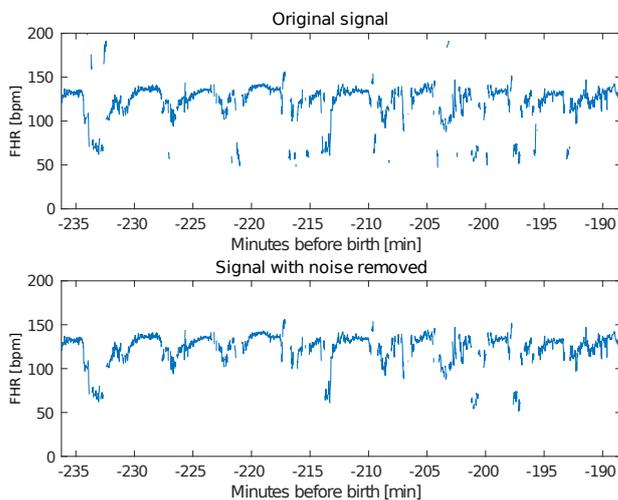
The threshold  $T_D$  is set to 5 based on empirical observation. If the sharp variations do not correspond to doubling or halving, the segment is considered as noise and removed:

$$fhr_c(i) = 0 : |f\dot{hr}(i)| > T_D \quad (5.8)$$

When all segments of length  $< T_K$  are checked, the process is repeated using backward replication. This is done as some segments can be  $> T_K$  due to replication of missing data in the end of the segment. A cleaned FHR signal is finally returned. An overview of the method is shown in Algorithm 1: Noise detection.

### 5.3.2 Results

An example FHR signal with artefacts, and the resulting signal after the noisy regions are removed is shown in Figure 5.4. The method successfully identifies many of the outliers as noise, while some segments in the 75bpm region are kept. As it is difficult to determine with certainty which part of the measured FHR signal is noise, only time periods where the signal is very unlikely to contain information of the fetal status are removed. This conservative approach may result in some time periods containing noise being kept.



**Figure 5.4:** Example of noise detection and removal. Original signal on top, with some artefacts. Filtered signal on the bottom.

**Algorithm 1:** Noise detection

---

**Input:** fetal heart rate,  $fhr$   
Variation threshold,  $c$   
Maximum length of segment,  $T_K$   
Doubling/halving variation threshold,  $T_D$

**Output:** cleaned fetal heart rate,  $fhr_c$   
 $fhr_c = fhr$

**for**  $direction \in \{forward, backward\}$  **do**  
     $fhr_{rep}(n) = fillGaps(fhr_c(n), direction)$   
    /\* Note:  $fhr$  denotes the derivative of  $fhr$  \*/  
     $A = \{s : |fhr_{rep}(t_s)| > c \cap |fhr_{rep}(t_s + k)| > c \cap k < T_k\}$   
    **for all**  $s \in A$  **sorted from smallest**  $k$  **do**  
        **for all**  $i \in \{t_s, t_s + k\}$  **do**  
             $fhr_d(n) = |2 \cdot fhr(n) - fhr(n - 1)|$   
             $fhr_d(n) = |0.5 \cdot fhr(n) - fhr(n - 1)|$   
             $fhr_c(n) = 2 \cdot fhr(n) : fhr_h(n) < T_D$   
             $fhr_c(n) = 0.5 \cdot fhr(n) : fhr_d(n) < T_D$   
             $fhr_c(i) = 0 : |fhr(i)| > T_D$   
             $fhr_{rep}(n) = fillGaps(fhr_c(n), direction)$

return  $fhr_c$

**def**  $fillGaps(fhr_{rep}, direction)$ :  
    **if**  $direction = forward$  **then**  
        |  $fhr_{rep}(n + 1) = fhr(n) : fhr(n + 1) = 0 \quad \forall n$   
    **else**  
        |  $fhr_{rep}(n - 1) = fhr(n) : fhr(n - 1) = 0 \quad \forall n$   
    return  $fhr_{gaps}$

---

Doppler shift were detected and corrected in 0.22 percent of the samples in the data set  $F_D1$ . Overall, 2.73 percent of all samples were identified as noise and removed, as shown in table 5.1.

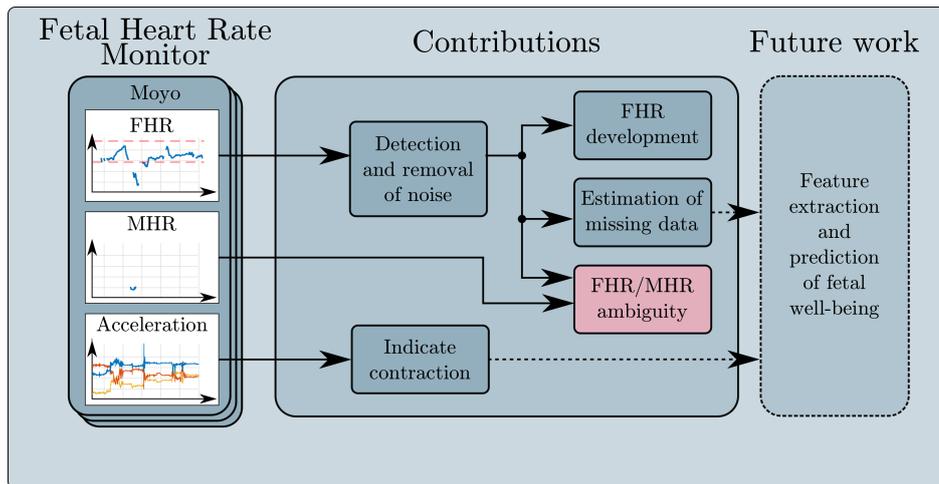
The method identifies many small sections of the FHR signal as noise, as seen in Figure 5.4. By removing these sections, a cleaner version of the FHR signal, better suited for further analysis can be obtained. This results in improved visual interpretation as well as it opens up for automated signal analysis and feature extraction in future work.

**Table 5.1:** Overview of the detected noise in  $F_{D1}$ .

Number of episodes	3807
Total duration of all episodes	14201 hours
Percentage of all samples with detected, and corrected, Doppler shift error	0.22
Percentage of all samples removed	2.73

## 5.4 FHR ambiguity (Paper 3)

In this section, the contribution in the pink *FHR/MHR ambiguity* block in Figure 5.5 is presented. Full details of this method can be found in paper 3.



**Figure 5.5:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work. This is a repetition of Figure 5.1 with the part presented in this section, FHR/MHR ambiguity, marked in pink.

Doppler based FHR measurements, presented in section 2.1.2, are susceptible of incorrectly picking up the MHR due to sub-optimal sensor placement [74]. The MHR can also mimic an expected FHR, making it challenging to distinguish true MHR from true FHR signals [75]. If the measured FHR is within 5 bpm of the measured MHR, it can be classified as an MHR/FHR ambiguity [76]. The amount of MHR/FHR ambiguity in Doppler CTG has been found to be  $1.22 \pm 1.9$  percent during the first stage

of labour and  $6.2 \pm 9.0$  percent during the second stage of labour [76]. As measurements of MHR using Moyo is intended to be intermittent and not continuous, the amount of MHR varies from labour to labour. Thus, the possibility of verifying whether the measured heart rate from the Doppler ultrasound is maternal or fetal at a given time point is therefore limited.

#### 5.4.1 Method

Computation of MHR/FHR ambiguity on Moyo data is performed using all time points where both the cleaned FHR signal,  $fhr(n)$ , and MHR signal,  $mhr(m)$ , are present. Let  $h_t$  be heart rate pairs of the measured FHR and MHR, sampled at the same time point,  $t$ , but with different sampling rate.

$$h_t = [fhr(n_t), mhr(m_t)] \quad (5.9)$$

Let  $H$  be the set of all such matching heart rate pairs,  $h_t$

$$H = \{h_t : fhr(n_t) > 0 \cap mhr(m_t) > 0\} \quad (5.10)$$

The ambiguity for each of the pairs,  $h_t$ , can be defined using the indicator function  $\mathbf{I}(h_t)$

$$\mathbf{I}(h_t) = \begin{cases} 1; & \text{if } |h_t(1) - h_t(2)| \leq T_{mhr} \\ 0; & \text{if } |h_t(1) - h_t(2)| > T_{mhr} \end{cases} \quad (5.11)$$

where  $T_{mhr}$  is a threshold to allow some inequalities due to the different measurement techniques. The similarity threshold,  $T_{mhr}$ , is set to 5 when computing the ambiguity, according to the study of Reinhard et al. [76].

The MHR/FHR ambiguity,  $mhr_{amb}$ , in an episode is then calculated as a fraction of the time where both signals are present, defined as

$$mhr_{amb} = \frac{1}{\#H} \sum_{h_t \in H} \mathbf{I}(h_t) \quad (5.12)$$

where  $\#H$  is the number of heart rate pairs.

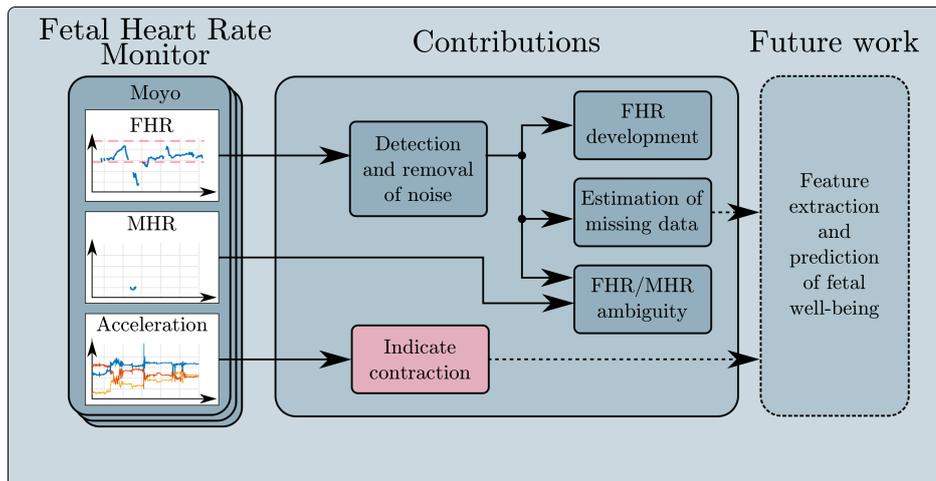
#### 5.4.2 Result

The dry-electrode ECG sensor for MHR is used in 30.54% of the episodes in the data set,  $F_D3$ . In these episodes, the MHR is measured in  $0.412 \pm 0.542$  percent of the episode duration.

The average MHR/FHR ambiguity in time points where both heart rates are measured is  $3.29 \pm 8.95$  percent.

## 5.5 Contractions (Paper 1)

In this section, the contribution in the pink *indicate contractions* block in Figure 5.6 is presented. Full details of this method can be found in paper 1.



**Figure 5.6:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work. This is a repetition of Figure 5.1 with the part presented in this section, indicate contractions, marked in pink.

In CTG, presented in section 2.1.2, both the FHR (cardio) and uterine activity (toco) is measured. Interpretation of the measured FHR signal is normally conducted in relation to the corresponding uterine contraction (UC). Simple Doppler based monitors typically lack dedicated sensors for measuring UC, but accelerometers may be available, i.e. in the Moyo. Accelerometers have previously been used to monitor muscle contraction of the quadriceps muscle, a large muscle on the front of the thigh [77]. Accelerometers in mechanomyogram have also been used to monitor muscular behaviour of the biceps under fatiguing exercises [78]. In a step towards removing movement-induced artefacts in electrohysterogram, the signal from an accelerometer mounted on the maternal abdomen was shown to correlate with the uterine movement [79]. Encouraged by the findings using accelerometer for muscle contraction detection [77, 78], and correlation to uterine movement [79], we explored the possibility of using the accelerometer in the Moyo for detecting uterine contractions.

### 5.5.1 Method

The FHR signal can often be used to identify the location of UC based on observations of the decelerations. However, as the FHR signal quality acquired using Doppler ultrasound often contains a lot of noise and missing samples, a more robust approach is desired. With the use of accelerometer signals for indicating UC, the missing data challenge is eliminated. As the data set,  $F_{D1}$ , is collected using Moyo, truth data for when contractions occur is not available. We have therefore chosen to indicate the time for when UC occur based on the accelerometer signal, and then verify using decelerations in the FHR signals.

The accelerometer captures small movements in the abdomen muscles as well as larger movements due to the mother changing positions. The amplitude of the acceleration signal for these movements is, however, typically vastly different. As the sensor location and orientation may be different between each labour, a trend describing the movement is computed using the acceleration energy,  $Acc_E(n)$ , given by:

$$Acc_E(n) = \sqrt{Acc_x^2(n) + Acc_y^2(n) + Acc_z^2(n)} \quad (5.13)$$

As the acceleration energy signal contains high frequency components, an upper envelope is computed to obtain the movement trend. The envelope of the acceleration energy,  $Acc_{env}(n)$ , is computed using a 20 second window. A set of positions,  $C$ , indicating contractions at time points,  $t_c$ , are found as local peaks of the envelope, given by

$$C = \{t_c : \dot{Acc}_{env}(t_c) = 0 \cap T_1 < Acc_{env}(t_c) < T_2\} \quad (5.14)$$

The thresholds  $T_1$  and  $T_2$  are set to  $10^{-2}$  and  $10^{-1}$  standard gravity,  $g_0$ , correspondingly. The thresholds are chosen to avoid detecting small movements, as well as larger movements due to the mother changing position, as contractions. The FIGO guidelines for intrapartum fetal monitoring [3] states that  $< 5$  contractions per 10-minute window averaged over 30 minutes is considered normal. We therefore require the onset of two consecutive indicated contractions to occur at least 2 minutes apart. The indicated contractions are hereafter called detected contractions. A pseudocode of the proposed contraction detection is depicted in Algorithm

2.

**Algorithm 2:** contractions**Input:** Acceleration signals,  $Acc_x, Acc_y, Acc_z$ **Output:** Set positions for detected contractions,  $C$ 

$$Acc_E(n) = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2}$$

$$Acc_{env}(n) = envelope(Acc_E(n))$$

$$C = \{t_c : \dot{Acc}_{env}(t_c) = 0 \cap T_1 < Acc_{env}(t_c) < T_2\}$$

**5.5.2 Result**

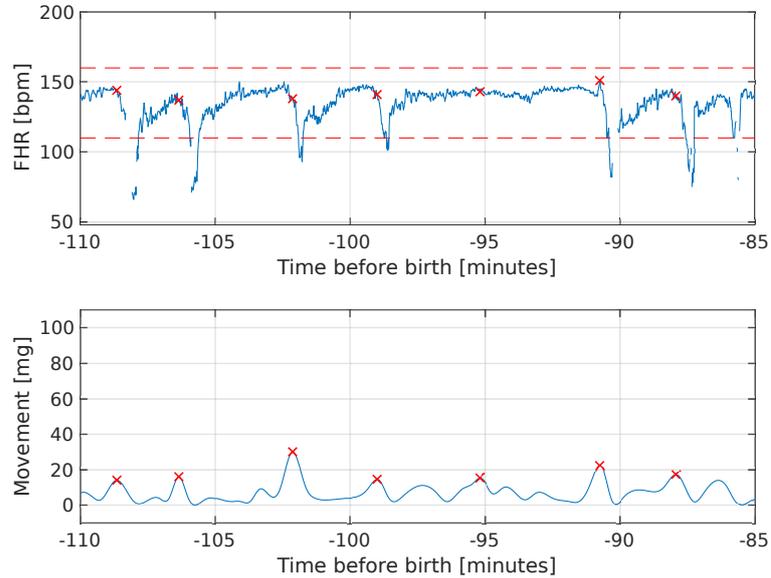
The proposed method to detect contractions using an accelerometer attached in proximity of the Doppler ultrasound sensor was run on the entire data set,  $F_D1$ . An overview of the amount of detected contractions are shown in table 5.2. The proposed algorithm was also run on signals with low-, medium-, and high energy in the acceleration signal to determine the performance in the different scenarios.

**Table 5.2:** Overview of the detected uterine contractions in  $F_D1$ . Values are given as median [q1, q3].

Episodes with detected contractions	3753
Episodes without detected contractions	54
Median number of detected contractions per episode	29 [14, 51]
Median length of episode [minutes]	171 [90, 304]
Mean time between contractions [minutes]	6.27

**1) Contractions on signals with low energy in the acceleration signal**

An example of a recording with low amount of energy in the acceleration signal was chosen from  $F_D1$ . The original signal and detected contractions are shown in Figure 5.7. Decelerations, which typically occur as a fetal response to a contraction are clearly visible in the FHR. Time points of detected contractions are shown using red markers. Contractions corresponding to the 6 largest decelerations are detected. The contraction associated to

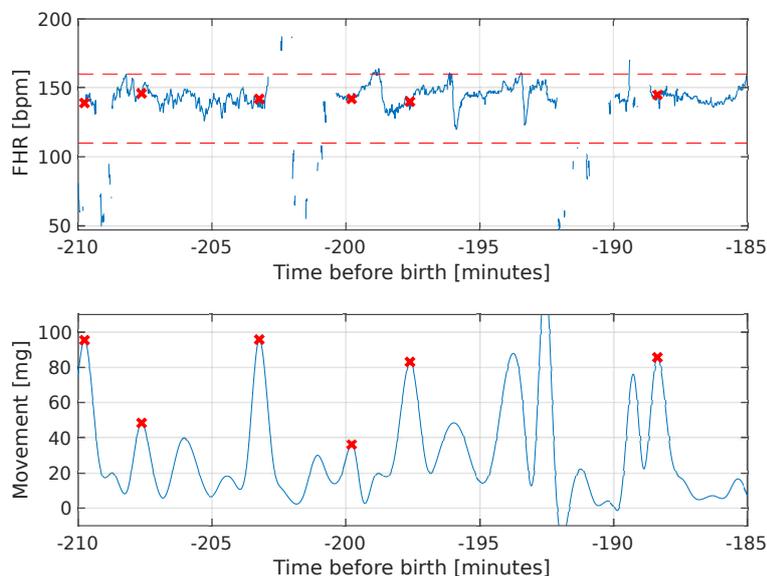


**Figure 5.7:** Detected contractions on a signal with low energy in the acceleration signal. The red dashed lines indicate the normal range of the FHR. The red crosses indicate the detected contractions.

the deceleration with a smaller drop in heart rate, at approximately 86 minutes before birth, is not considered to be caused by a contraction as it is too close to the previous detected contraction. An additional uterine contraction is detected at approximately 95 minutes before birth, without a corresponding deceleration in the FHR.

## 2) Contractions on signals with medium energy in the acceleration signal

An example of a recording with medium amount of energy in the acceleration signal was chosen from  $F_D1$ . The original signal and the detected contractions are shown in Figure 5.8. Contractions are detected periodically in the first half of the signal, while only one contraction are detected in the second half. Due to the quality of the FHR signal, it is challenging to assess if these are actual uterine contractions.



**Figure 5.8:** Detected contractions on a signal with a medium energy in the acceleration signal. The red dashed lines indicate the normal range of the FHR. The red crosses indicate the detected contractions.

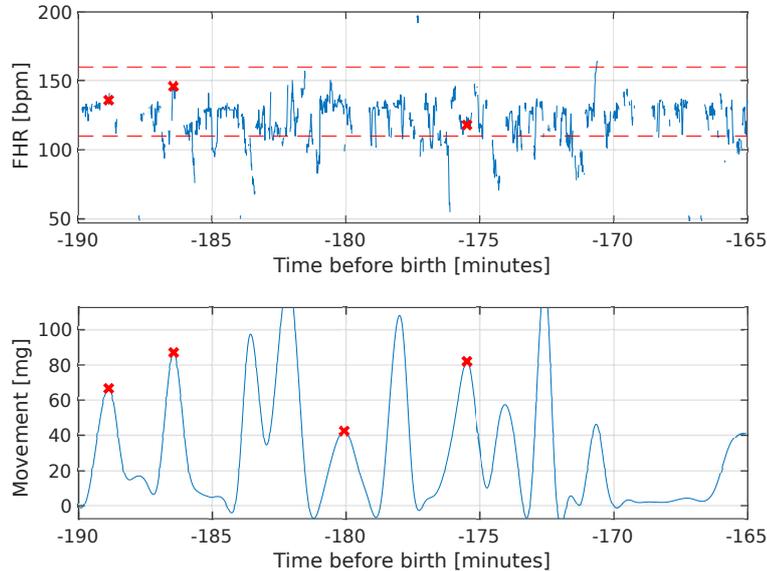
### 3) Contractions on signals with high energy in the acceleration signal

An example of a recording with high amount of energy in the acceleration signal was chosen from  $F_D1$ . The original signal, and the detected contractions are shown in Figure 5.9. Four uterine contractions are detected in the 25-minute window, but it is challenging to assess if these are actual contractions due to the poor FHR signal quality.

## 5.6 Estimation of missing data (Paper 2)

In this section, the contribution in the pink *Estimation of missing data* block in Figure 5.10 is presented. Full details of this method can be found in paper 2.

A well-known problem of measuring FHR using Doppler ultrasound is signal dropouts due to both fetal and maternal movements in addition to sensor displacement, see section 2.1.2. The use of wearable devices for monitoring FHR allows the mother to move freely, but it can increase



**Figure 5.9:** Detected contractions on a signal with a high energy in the acceleration signal. The red dashed lines indicate the normal range of the FHR. The red crosses indicate the detected contractions.

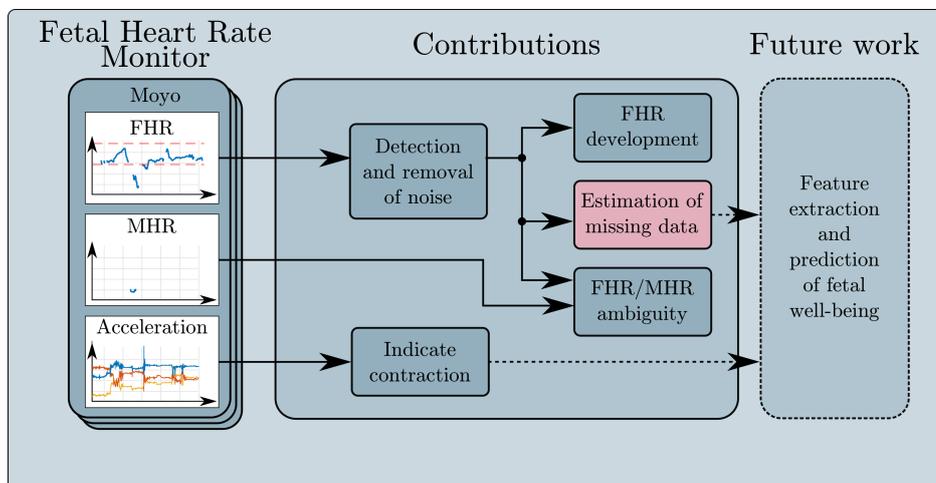
both the number of signal dropouts as well as their length. These signal dropouts are a challenge when determining traditional features to assess the fetal well-being, such as the short- and long-time variability of the FHR. It will also create challenges when doing time-frequency analysis on the FHR signal.

Dictionary learning and sparse approximation have been shown to produce state of the art results in estimation of missing data [27, 28, 29]. An important advantage of using dictionary learning over methods such as linear or spline interpolation is that it introduces less artefacts during processing, feature extraction, and time-frequency analysis.

We propose to use shift-invariant dictionary by utilizing SI-FSDL, section 3.2.1, a dictionary learning method for shift invariant dictionaries recently proposed by our group [80] to estimate the missing data in FHR signals.

### 5.6.1 Method

The measured FHR signal, and the location of missing samples differ from patient to patient. Looking at these analyses in retrospect, we wish to



**Figure 5.10:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work. This is a repetition of Figure 5.1 with the part presented in this section, estimation of missing data, marked in pink.

learn a dictionary on the signals with missing data. The dictionary can then be tailored to the patient, before performing the reconstruction. For this reason, the information about the missing samples is built into a mask matrix. One possible mask matrix is an identity matrix with its rows removed on locations where the data is missing. This means that applying it to a vector or a matrix, removes the rows corresponding to missing samples.

To reconstruct the FHR signal, a shift-invariant dictionary SI-FSDL [80] is used. One of the benefits of using this dictionary is that larger gaps can be addressed, by using larger shift-invariant atoms, than a general dictionary while keeping the number of free variables fixed.

To incorporate the mask information, the approach presented by Oikonomou et al. [29] is used, which only requires a change to the sparse coding step. The altered steps can be summarized by

- (i) Keep the dictionary,  $D$ , fixed. Remove the rows of the dictionary and the signal vectors corresponding to the missing samples for each signal vectors.
- (ii) Normalize the dictionary columns to 1.
- (iii) Find the coefficients,  $w$ .

- (iv) Find the signal estimate,  $\hat{x} = Dw$ .
- (v) Reconstruct the signal,  $x$ , by replacing the missing samples with the estimates,  $\hat{x}$ .
- (vi) Update the dictionary,  $D$ , and normalize the columns to 1.

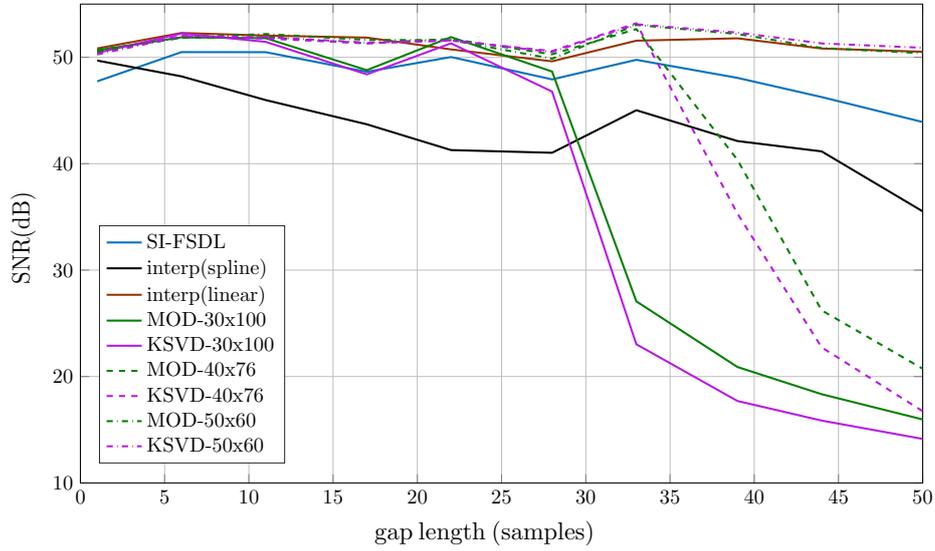
### 5.6.2 Result

Examples of signal segments *without* missing samples were chosen from the data set  $F_{D2}$ . Parts of these signals were randomly removed to make up a set of test signals. To measure the performance, SI-FSDL was compared with the traditional dictionary learning techniques MOD [30] and K-SVD [31], as well as spline and linear interpolation.

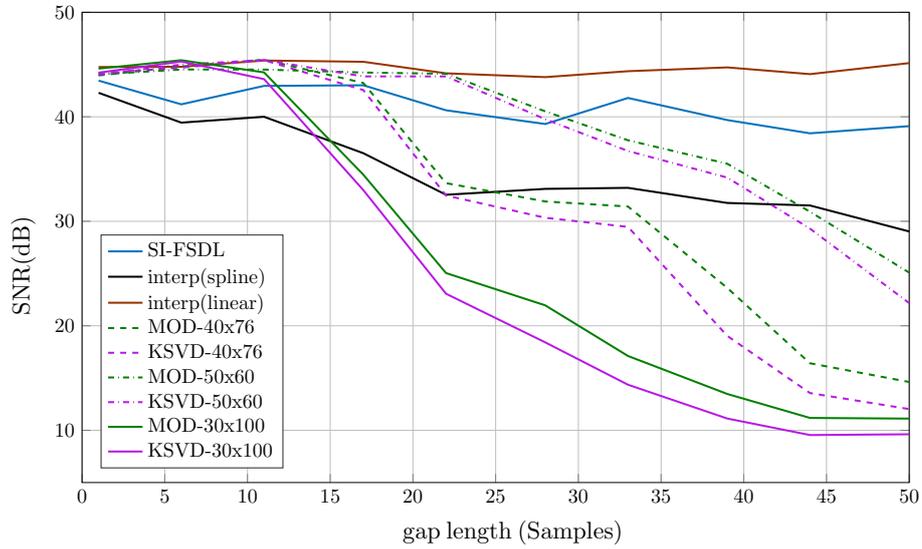
During the experiments, the number of free variables in the dictionaries are kept constant at approximately 3000. The ratio of non-zeros coefficients to the number of elements in the signal is 0.1. Signal blocks are chosen in an overlapping fashion for the dictionary learning methods. Block lengths,  $N$ , of 30, 40 and 50 are used for MOD and K-SVD, and 500 for SI-FSDL.

Estimation of missing data was done by varying the gap length of the missing segments, while keeping the amount of missing data fixed at 10%, Figure 5.11, and 30%, Figure 5.12. Performance of the tested methods for 30% missing data are shown in Figure 5.12. All methods achieve similar performance for short interval lengths. When the amount of missing data increase to 30%, the performance for MOD and K-SVD for all segment lengths decrease faster than the case of 10%. The performance of linear interpolation and SI-FSDL, remains almost the same regardless of the missing sample length.

Both SI-FSDL and linear interpolation achieve a high signal to noise ratio (SNR) for the case where 30% of the data is missing. A visual assessment of the signal reconstruction was performed using continuous wavelet transform on the reconstructed signals to see how similar the time-frequency distribution of the reconstructed signal is to the original signal. The time-frequency response for a short section of FHR where 3 time periods of the signal has been removed are shown in Figure 5.13. In close-ups of the signal around each mask, the original signal is shown in blue, estimations using linear interpolation in dotted red, and SI-FSDL in dashed black. By visual inspection of the three blocks with estimated signals in Figure 5.13, it is apparent that the SI-FSDL reconstruction

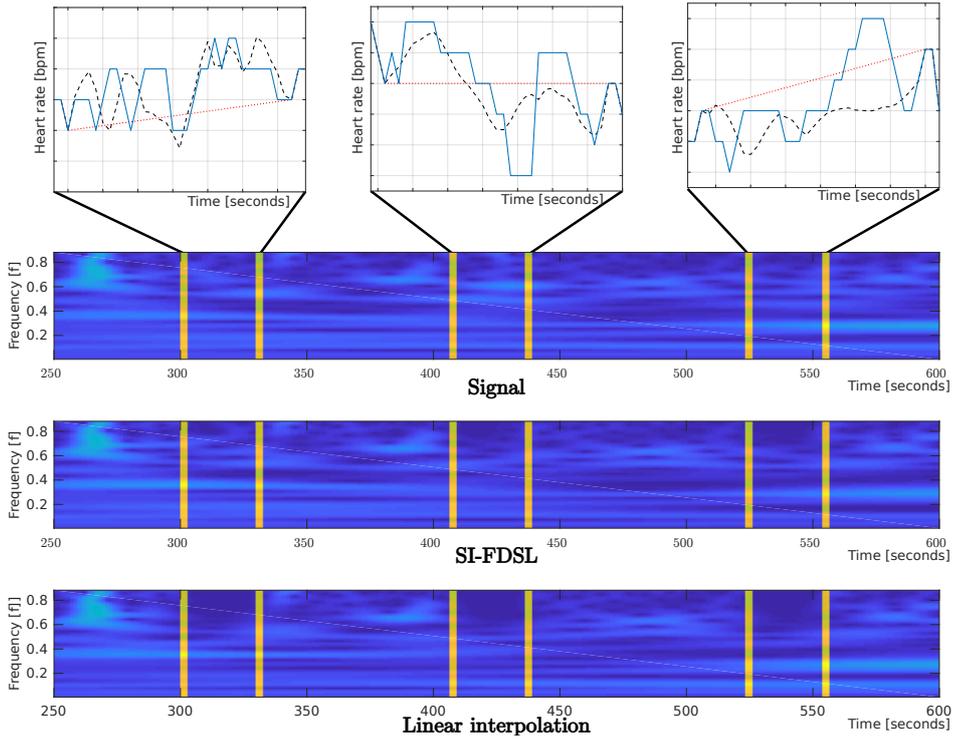


**Figure 5.11:** Recovery performance for different methods when the missing intervals are increased from 1 to 50 samples, but the total missing percentage is kept fixed at 10%



**Figure 5.12:** Recovery performance for different methods when the missing intervals are increased from 1 to 50 samples, but the total missing percentage is kept fixed at 30%

restores the time-frequency properties better than the linear interpolation, even though the latter has a higher SNR. Based on this, the computed SNR is not considered to be a good measure to describe the reconstructed signal quality.

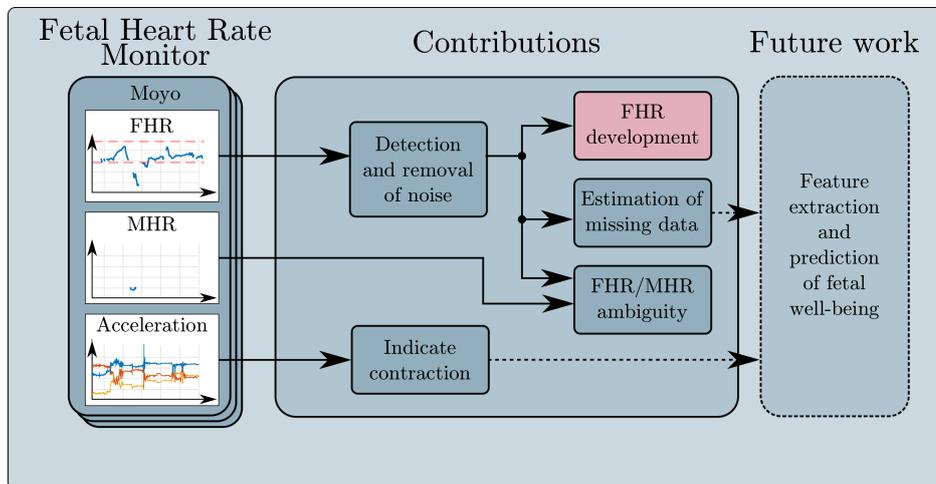


**Figure 5.13:** Continuous wavelet transform over a short signal with 3 time periods of the signal removed. The original signal is shown in blue, linear interpolation shown as a dotted red line, and SI-FSDL shown as dashed black line.

## 5.7 Fetal heart rate development (Paper 3)

In this section, the contribution in the pink *FHR development* block in Figure 5.14 is presented. Full details of this method can be found in paper 3.

Continuous FHR measurement of labours with different outcomes can give valuable information on both normal and abnormal FHR patterns. In high income settings, continuous FHR monitoring is primarily used for



**Figure 5.14:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work. This is a repetition of Figure 5.1 with the part presented in this section, FHR development, marked in pink.

high risk labours. This study, does however, only include labours assessed as low risk with normal FHR on admission. This inclusion criteria gives us the possibility to study the FHR development for labours with a normal outcome.

The heart rate of newborns is reported to increase shortly after birth [22]. But to the authors knowledge, a corresponding trend in how the FHR changes just before birth has not been reported. With the use of continuous FHR monitoring in many labours, it is possible to determine how the FHR typically develops during labour, and to make a model of a normal development vs a development with adverse outcome. These models can in turn be useful to determine progress in future labours.

### 5.7.1 Method

The time period where the FHR is measured varies from labour to labour. To compare different episodes, they are compared relative to time of birth,  $t_{ob}$ . The sample index,  $n$ , is defined in the measured FHR signal based on the measured FHR with sample rate, 2 Hz, the elapsed time,  $t$ , and a defined start point before birth,  $t_0$ , so that  $t_{ob} - t_0$  is constant, such as

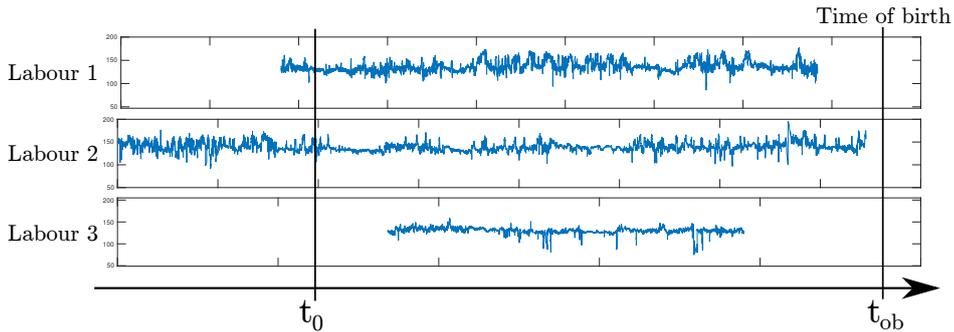
$$n = 2(t + t_0) \quad (5.15)$$

## 5. FETAL HEART RATE AND LABOUR ANALYSIS

**Table 5.3:** Overview of the data subsets in  $F_{D3}$ , and the removed noise from FHR signal in each labour outcome group.

Outcome	#episodes	Subset	Missing data [%]	Removed data [%]
All	3705	$F_{D3}$		
Normal	3490	$F_{D3s_1}$	$27.83 \pm 19.87$	$1.79 \pm 1.35$
NCU	185	$F_{D3s_2}$	$31.28 \pm 20.38$	$1.82 \pm 1.31$
END	18	$F_{D3s_4}$	$29.22 \pm 24.34$	$1.31 \pm 0.83$
FSB	12	$F_{D3s_5}$	$40.50 \pm 28.60$	$1.92 \pm 1.61$
Perinatal mortality	30	$F_{D3s_3}$		

An example of the varying start and end point of 3 episodes, as well as  $t_{ob}$  and  $t_0$  is shown in Figure 5.15. The time of birth is known for all episodes and is therefore used as a reference point between the episodes.



**Figure 5.15:** Illustration of the how the start and end point of FHR measurements varies from labour to labour. The end point in this analysis is therefore set to the time of birth,  $t_{ob}$ , and the starting point  $t_0$  a fixed amount of time prior to the time of birth.

The subset is defined based on the neonatal status 24 hours after birth. The subset  $s_1$  includes all labours where the newborn was assessed as normal. Subset  $s_2$  includes all labours where the newborn was still admitted to neonatal care unit (NCU) at 24 hours. Subset  $s_3$  includes all episodes where the newborn died during the 24-hour period or died during labour (FSB). FSB and END are grouped together in subset  $s_3$  due to the low quantity of episodes within each outcome. An overview of the used subsets, and the amount of missing and removed data in each subset is shown in table 5.3.

### Fetal Heart Rate Development

The measured FHR within a defined time interval,  $\Delta$ , in an episode,  $i$ , from the subset,  $s_k$ , is extracted from the start time,  $t$ , and throughout the duration,  $t + \Delta$ . Let the trend,  $mFHR_s(p)$ , be defined as the median of all measured heart rate in the interval, of all episodes in the subset,  $s$

$$mFHR_s(p) = \text{median} \left( \left( fhr_{1,s}^t(n), \dots, fhr_{L_s,s}^t(n) \right) \right) \quad n \in \{t, t + \Delta\} \quad (5.16)$$

Where  $L_s$  is the number of episodes in the subset  $s$ , and the sampling index  $p$  is given by

$$p = \frac{1}{\Delta}(t + t_0) \quad (5.17)$$

To describe the spread at each time interval, the 1st and 3rd quartiles,  $q_1, q_3$ , called  $HRq_1(t)$  and  $HRq_3(t)$ , are computed using the concatenated vector of all FHR measurements in the time interval,  $\left( fhr_{1,s}^t(n), \dots, fhr_{L_s,s}^t(n) \right) \forall n \in \{t, t + \Delta\}$ .

### Fetal Heart Rate Distribution

A normalized histogram is used to estimate the changes over time in a probability density function (pdf). The histogram is created using an interval defined by the starting point,  $t$  and end point  $t + \Delta$  for all episodes in a subset. When computing two or more distributions, these can be used to identify how the distribution changes over time. Let  $h_i^t(l)$  be the histogram of the measured FHR in episode  $i$ , in the interval with start point  $t$  and end point  $t + \Delta$ ,

$$h_s^t(l) = \sum_{i \in s} h_i^t(l) \quad \forall \quad l = \{50, 51, \dots, 200\} \quad (5.18)$$

The normalized histogram,  $\bar{h}_s(l)$ , is then given by

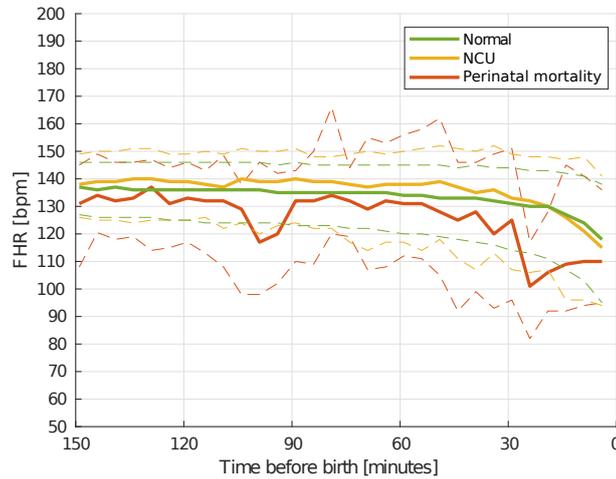
$$\bar{h}_s^t(l) = \frac{1}{\#N} h_s^t(l) \quad (5.19)$$

Where  $\#N$  is the total number of  $\sum_l h_s^t(l)$

## 5.7.2 Result

### Fetal Heart Rate Development

The FHR development is computed as the median FHR,  $mFHR(t)$ , using non-overlapping intervals of fixed size. Intervals of 5- and 10-minute duration were used to obtain multiple resolutions of the heart rate trend in the last 150 minutes before birth. An overview of  $mFHR(t)$  using 5- and 10-minute intervals are shown in Figures 5.16 and 5.17, respectively. The solid lines indicate  $mFHR(t)$ , and dashed lines the  $HR_{q_1}(t)$  and  $HR_{q_3}(t)$ . The green lines show the trend for newborns with a normal outcome,  $F_D3s_1$ , yellow newborns still admitted to NCU,  $F_D3s_2$ , and red newborns either defined as FSB or dead during the first 24 hours,  $F_D3s_3$ .

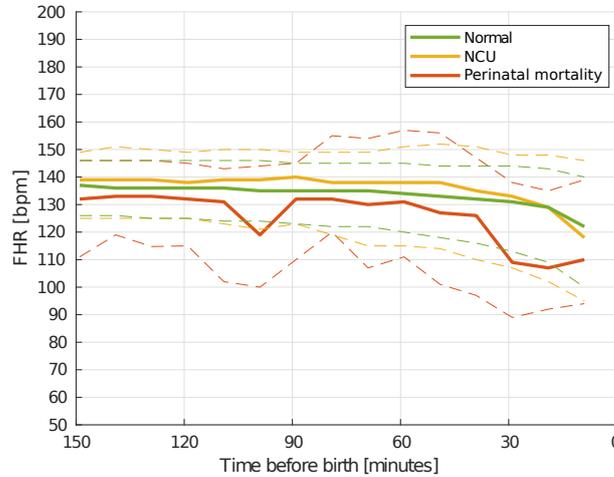


**Figure 5.16:** Trend of the FHR using 5-minute non-overlapping intervals the last 150 minutes before birth. Solid lines indicate the median heart rate, and the matching dashed lines the 25 and 75 percentiles.

The computed trend indicates a reduction in  $mFHR$  the last 30 minutes before birth. The reduction in  $mFHR$  for the perinatal mortality group,  $F_D3s_3$ , group occurs at a longer time before birth than for  $F_D3s_1$  and  $F_D3s_2$ .

### Fetal Heart Rate Distribution

An estimate of the pdf for all FHR over all episodes in each subgroup was found for the last 30 minutes before birth, and the two 30-minute intervals before. The pdfs were computed to study how the estimated pdf of FHR



**Figure 5.17:** Trend of the FHR using 10-minute non-overlapping intervals the last 150 minutes before birth. Solid lines indicate the median heart rate, and the matching dashed lines the 25 and 75 percentiles.

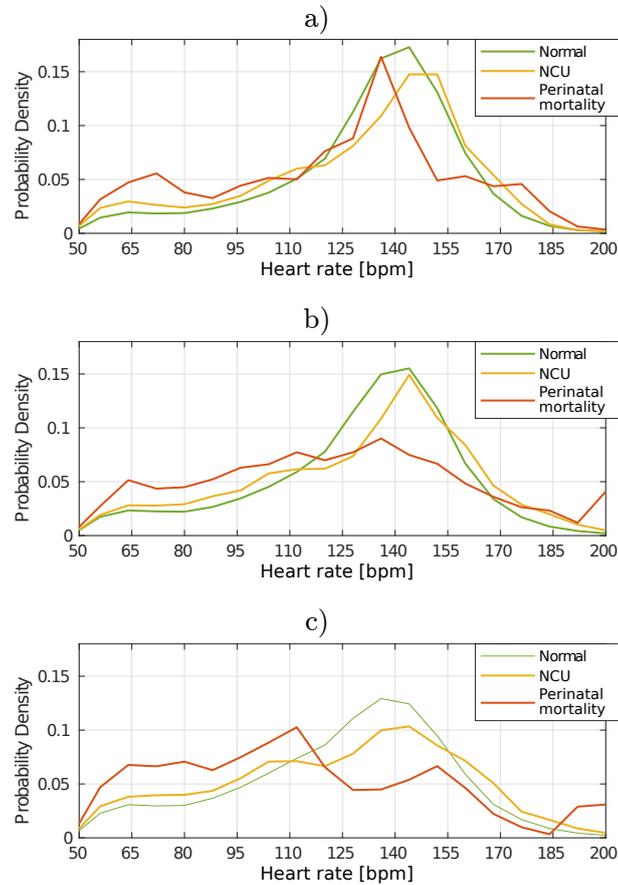
changes before and after the drop in the mFHR observed in section 5.7.2. The estimated pdf for the last 90 to 60 minutes before birth is shown in Figure 5.18a, the last 60 to 30 minutes before birth in Figure 5.18b, and the final 30 minutes before birth in Figure 5.18c. The green line indicates the estimated pdf for normal outcome, yellow line is the NCU group and red for the perinatal mortality group.

In the interval 90 to 60 minutes before birth, Figure 5.18a, all outcome groups have a peak in the 135-145 bpm region. In the interval 60 to 30 minutes before birth, a similar peak as in the previous interval is found for the normal and NCU groups. The perinatal mortality group still has a peak at 135 bpm, although not as distinct. The variance of the perinatal mortality group is also increased. In the last 30 minutes before birth, the variance increases for both the normal group and the NCU group. But the peaks stay within the same 135-145 bpm region. For the perinatal mortality group, the variance is high, and the peak has now shifted down to 110 bpm.

By combining multiple normalized histograms using continuous non-overlapping intervals, both the change in trend and spread of the FHR can be visualized, shown in figure 5.19. The red line illustrates the number of episodes with data at the corresponding time point. A peak in the computed histograms will result in a visible ridge in the 3D-visualization.

## 5. FETAL HEART RATE AND LABOUR ANALYSIS

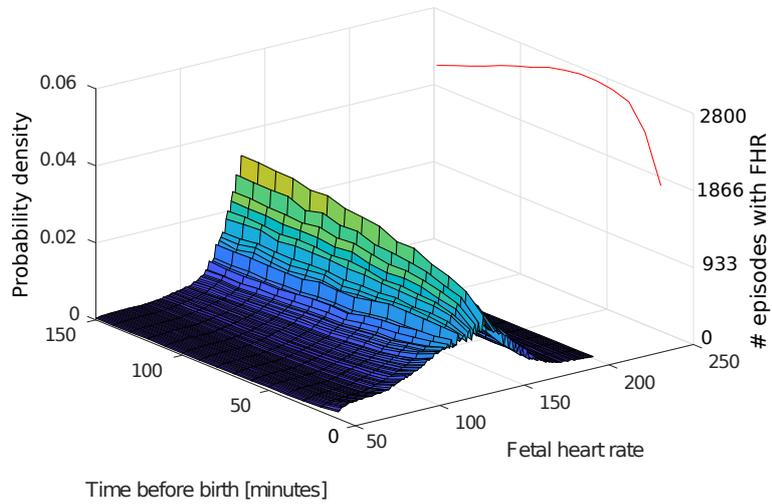
---



**Figure 5.18:** Estimated probability density function (pdf) of the measured heart rate the last 90 minutes before birth, divided in 30-minute windows. **a)** The estimated pdf for the last 90 to 60 minutes before birth. **b)** The estimated pdf for the last 60 to 30 minutes before birth. **c)** The estimated pdf for the last 30 minutes before birth.

The major findings can be summarized by the following four points:

- 1a: A reduction in mFHR is observed for all subgroups close to birth.
- 1b: The reduction in mFHR for the perinatal mortality group is larger and occurs longer time before birth.
- 2a: The variance of the estimated pdf increases for all subgroups close to birth.



**Figure 5.19:** Estimated probability density function the last 150 minutes before birth for newborns defined as normal 24 hours after birth. The red line illustrates the number of episodes with data at the corresponding time point.

2b: A larger increase in the variance, as well as shift in the peak is observed for the perinatal mortality group.

## Chapter 6

# Newborn resuscitation

In the following chapter, previous work on newborn resuscitation is first presented, followed by our contribution to the topic.

### 6.1 Previous work

Detection and recognition of activities using accelerometer data have previously been explored on healthy adults. Static and dynamic activities such as sitting positions, walking versus running has been found using accelerometers mounted on the back [81] and on the waist [82]. With the rise of wearable technology in everyday life, such as sport watches and cell phones, accelerometers are now available for activity recognition using commercially available devices [83]. To the authors knowledge, there are no reported works utilizing accelerometer and ECG signals to automatically classify therapeutic activities, except from this research group at this time.

#### 6.1.1 Activity detection

A system for detection of time periods during resuscitation where activities are likely to be performed on the newborn, from now on called VuDetector [52], has previously been proposed by our research group. The method detects these time periods based on the short time energy (STE) of the acceleration energy signal. The acceleration energy,  $Acc(n)$  is given by

$$Acc(n) = \sqrt{Acc_x^2(n) + Acc_y^2(n) + Acc_z^2(n)} \quad (6.1)$$

Where  $Acc_d$  is a low pass filtered version of the measured acceleration in axis  $d \in \{x, y, z\}$ , and  $n$  is the index in the acceleration signal. The STE,

$E_{Acc}(i)$ , is then found using a window function,  $w$

$$E_{Acc}(i) = \sum_{n=i-N+1}^i (Acc(n) \cdot w(i-n))^2 \quad (6.2)$$

The STE at index  $i$  is computed using samples from the window of length  $N$ . The computed STE signal is then thresholded to determine if an activity occurs at the current window.

$$Activity(i) = \begin{cases} 1 & \text{if } E_{Acc}(i) \geq Thr \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

The method achieves a sensitivity of 90% and a specificity of 80%, both with a standard deviation of 6%. Further details of VuDetector can be found in Vu et al. [52].

### 6.1.2 Detection of Ventilation

Ventilation is known to be a vital part of the resuscitation process. A method to detect and parameterize the manual bag-mask ventilation has previously been proposed by our research group with an accuracy of 95%, from now on called VuVentilation [84].

The method detects ventilation events ( $v_l$ ) by thresholding the pressure signals after baseline wander removal. Expired volume in the ventilation is found by extracting start of inspiration and expiration periods from the flow signal and integrating the flow signal over the relevant period. Further details of can be found in Vu et al. [84].

### 6.1.3 Activity classifier

A system to automatically identify therapeutic strategies during newborn resuscitation, from now on called VuClassifier, has previously been proposed by our research group [85]. VuClassifier is designed to identify the following categories:

- (i) Chest compression
- (ii) Stimulation
- (iii) Other activities

The classification is performed using 46 features extracted from accelerometer- and ECG-signals from the NeoBeat prototype in the LNRM. These features include time domain features, such as energy, RMS, entropy, and frequency domain features, such as wavelet for both the acceleration and ECG signals. The wavelet features were extracted using a 6-level decomposition using the Daubechies mother wavelet. More information on these features can be found in Vu et al. [85]. The classifier achieves an overall accuracy of 78.7% when distinguishing between the three classes. By combining chest compression and stimulation into one general stimulation activity, the overall accuracy is increased to 79.8%.

While the system achieves a promising performance, some limitations should be mentioned. The system needs full-episode statistics for the classification, making real-time implementations challenging. The small data set of 30 episodes used to verify the method can also be a limitation. In total, these episodes consisted of 21 sequences of *chest compression*, 250 sequences of *stimulation*, and 175 sequences of *other*.

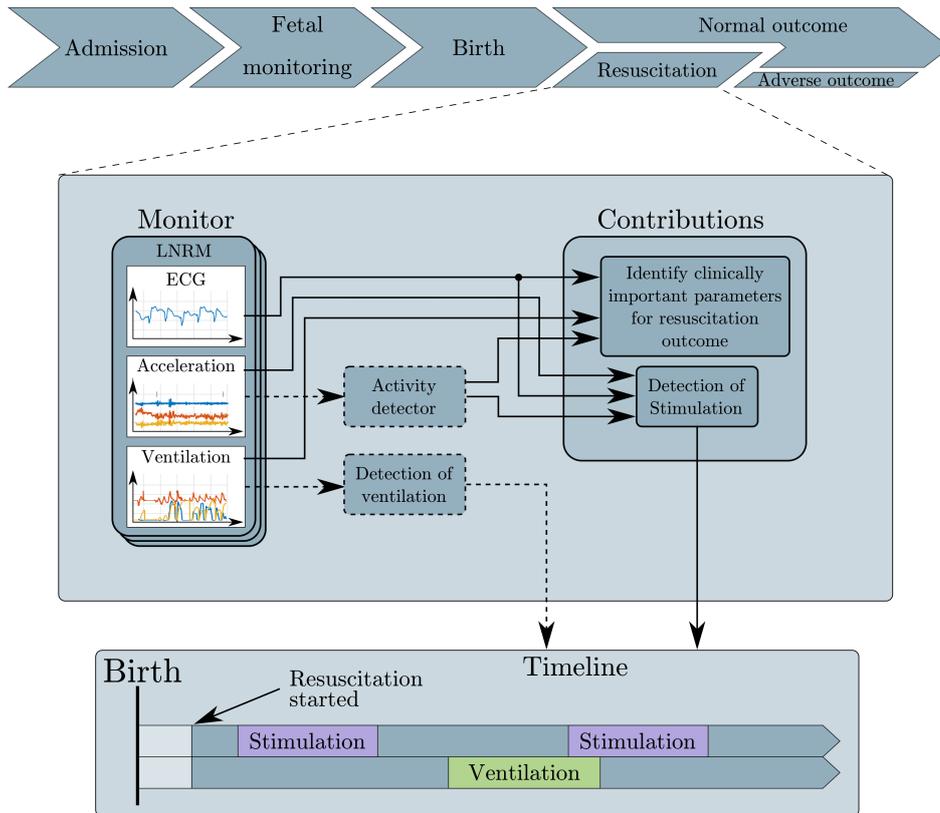
#### 6.1.4 Activity recognition using deep learning on video signals

An alternative approach for automatically annotating newborn resuscitation episodes is by using deep learning on videos overlooking the resuscitation table. One such system, ORAA-net [86], has previously been proposed by our research team. The system consists of four main steps: 1) Object detection, 2) Region proposal, 3) Activity recognition, and 4) generation of Activity timelines. The first step is based on a deep learning system to detect objects such as bag-mask ventilator, suction devices, and health care hands [86]. Regions around these objects are proposed and used in a new network to recognize stimulation, ventilation, suction, and if the newborn is covered or uncovered [87]. The recognized activities are finally used to create timelines describing the resuscitation episode.

However, video of a resuscitation may not always be available. The use of NeoBeat in combination with information from different sensors could provide timelines based on what is available: ventilation signals, NeoBeat (ECG, accelerometer) signals, videos or a combination of these sources.

## 6.2 Contribution overview

In the following sections, the contribution of each block in Figure 6.1 will be presented. Identification of clinically important parameters during newborn resuscitation is presented in section 6.3. A validation of VuClassifier and a new proposed method for annotating stimulation is presented in section 6.4.



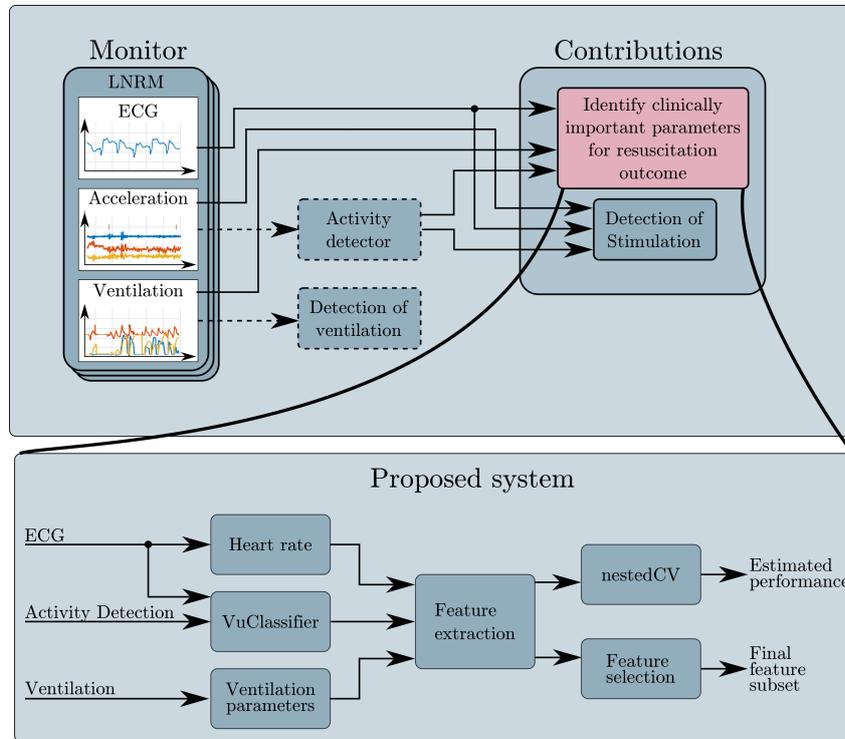
**Figure 6.1:** Overview of the contributions in newborn resuscitation. The dotted lines indicate modules previously proposed by our research group. With the use of a system to detect stimulation and ventilation during resuscitation, timelines can be created to describe the resuscitation event.

The dotted lines in Figure 6.1 indicate methods previously proposed by our research group. These methods are included as they are a vital part in a system for describing the therapeutic activities performed during newborn resuscitation. In the generated timeline, seen on the bottom of Figure 6.1, we can see that the resuscitation is started shortly after birth.

Both stimulation and ventilation are performed during the resuscitation, and both activities can be performed at the same time.

### 6.3 Identification of clinically important parameters (Paper 4)

In this section, the contribution in the pink *Identify clinically important parameters for resuscitation outcome* block in Figure 6.2 is presented. The proposed system is shown on the bottom of the figure. Full details of this method can be found in paper 4.



**Figure 6.2:** Overview of the identification of clinically important parameters during newborn resuscitation. The dotted line and boxes illustrate methods previously proposed by our research group. Features are extracted from LNRM data as well as manually logged values during labour. The features are used in a nested cross-validation to estimate the performance for distinguishing the resuscitation outcomes.

By identifying treatment factors determinant for the 24-hour neonatal outcome, better treatment, and feedback solutions to guide the therapy can

be sought. Factors for prediction of the need of neonatal resuscitation [88], and the relationship between ventilation performance and responses of newborns in term of Apgar score [89] have been explored. There is, however, still an uncertainty of which therapeutic strategies are determinant for outcome in neonatal resuscitation - what characterize a good resuscitation?

### 6.3.1 Method

Features are extracted from LNRM data as well as manually logged values during labour.

Parametrization of the resuscitation episode is distinguished in three main categories: initial conditions, treatment, and early response parameters. Treatment parameters are further divided into two subgroups, those reflecting therapeutic strategies, like stimulation vs ventilation, and those reflecting the quality of the ventilations.

#### Initial condition parameters

The following parameters were used to describe the initial condition of the newborn prior to bag-mask ventilation:

- Birth weight (BW) in grams.
- Gestational Age (GA) in weeks, estimated at birth.
- Time from birth to start of bag-mask ventilation ( $t_{BMW}$ ). Logged by research assistants present at the labour.
- Initial heart rate ( $hr_{V_i}$ ) is the heart rate when the first ventilation sequence starts ( $V_{1,start}$ ).

$$hr_{V_1} = HR(V_{1,start}) \quad (6.4)$$

#### Treatment parameters

The following parameters are extracted from the sensor data:

- Total duration of the resuscitation episode ( $T$ ). Defined as the time period from the resuscitation was started, until the end of the last stimulation or ventilation sequence.

- Stimulation time in percent ( $ST_{PRC}$ ) is the ratio of the time of all stimulation sequences ( $S_j$ ) over  $T$ .

$$ST_{PRC} = \frac{1}{T} \sum_i S_j \cdot 100\% \quad (6.5)$$

- Ventilation time in percent ( $VT_{PRC}$ ) is the ratio of the time of all ventilation sequences ( $V_i$ ) over  $T$ .

$$VT_{PRC} = \frac{1}{T} \sum_i V_i \cdot 100\% \quad (6.6)$$

- Hands-off time in percent ( $HOT_{PRC}$ ) is the ratio of the time of all hands-off sequences  $HO_k$  over  $T$ .

$$HOT_{PRC} = \frac{1}{T} \sum_i HO_k \cdot 100\% \quad (6.7)$$

- The total number of ventilations ( $n_V$ ) is the sum of all ventilation events ( $v_l$ ) in all  $V_i$ .

$$n_v = \sum_i \sum_l \#\{v_l \in V_i\} \quad (6.8)$$

- The expired volume ( $\text{expVol}_{ml/kg}$ ) is the median expired volume from all ventilation events, divided by the newborns birth weight (BW) in kg.

$$\text{expVol}_{ml/kg} = \text{median} \left( \frac{\text{expVol}}{\text{BW}/1000} \right) \quad (6.9)$$

- The average ventilation rate ( $V_R$ ) is  $n_V$  over the total duration of all ventilation sequences in one episode.

$$V_R = \frac{n_V}{\sum_i V_i/60} \quad [\text{inflations}/\text{min}] \quad (6.10)$$

The features  $ST_{PRC}$ ,  $VT_{PRC}$ , and  $HOT_{PRC}$ , reflect therapeutic strategies, while the remaining treatment parameters are associated to quality of ventilation.

One can argue that the time from birth to start of bag-mask ventilation,  $t_{BMW}$ , can be seen as both a treatment parameter and an initial condition

for the bag-mask ventilation parameter. However, we have defined treatment parameters as parameters possible to extract from the measured sensor signals. Median expired volume can be both a treatment and response parameter. We have chosen to label this feature as a treatment parameter, as it is an efficient way to identify the ventilation quality. A poor ventilation with severe mask leakage will result in a high measured inflated volume, but a low measured expired volume. A good ventilation will result in a sufficiently high expired volume.

### Early response parameters

The following parameters are extracted from the sensor data, and can be regarded as response parameters:

- Heart rate when first ventilation sequence ends  $hr_{V_1,end}$

$$hr_{V_1,end} = HR(V_{1,end}) \quad (6.11)$$

- Heart rate 30 seconds after first ventilation sequence  $hr_{V_1,30s}$

$$hr_{V_1,30s} = HR(V_{1,end} + 30) \quad (6.12)$$

- Heart rate 60 seconds after first ventilation sequence  $hr_{V_1,60s}$

$$hr_{V_1,60s} = HR(V_{1,end} + 60) \quad (6.13)$$

### Classification

Due to the low number of episodes ending in death, there is a large class imbalance when trying to distinguish episodes in the normal group from the group ending in death. To alleviate this imbalance, the RUSboost classifier was chosen. Details of the classifier can be found in section 3.4.1.

A nested-CV scheme, presented in section 3.4.4, is used to estimate the performance. In the nested CV, the internal loop is used for feature selection and reduction of dimensionality of the feature set, using the external loop for validation. Feature selection is done using a wrapper method [90] with a forward selection approach, presented in section 3.4.4, maximizing the true (positive+negative) rate. The data set,  $R_{D1}$ , is further divided into three non-overlapping sections in both the internal and external loop. System performance is computed using the summed confusion matrices from all

external folds. The same type of classifier is used in both the internal and external loops. Feature normalization is applied in each CV fold, where the mean,  $\mu$ , and standard deviation,  $\sigma$ , is found for the current training set and then applied to the normalization of the current test set.

The same feature selection method is applied to the entire data set,  $R_D1$ , to identify the most crucial features corresponding to the performance found in the nested CV.

### 6.3.2 Result

Three experiments were conducted and validated:

- (i) Identification of neonatal outcome 24 hours after birth using all features available
- (ii) Identification of neonatal outcome using all features for newborns initially in a poor condition
- (iii) Identification of neonatal outcome using only initial- and treatment parameters for newborns initially in a poor condition.

Newborns with an initial heart rate below 120 beats per minute was defined to be in a poor condition.

Each experiment was run for all four class combinations: Normal - Death, Normal - NCU, NCU - Death, and the three-class problem Normal - NCU - Death. The performance of each outcome combination and experiments are shown in table 6.1. The label 1 is used for class normal, the label 2 for class NCU, and the label 3 for class death.

Using all 14 features, heart rate after the first ventilation, number of ventilations, total resuscitation time, ventilation time, and hands-off time percentage are found to be good identifiers of resuscitation outcome. For identification of neonatal outcome of episodes with an initial heart rate below 120, heart rate, ventilation time, stimulation time, and hands-off time percentage are found to be good identifiers. Total resuscitation duration is, however, only found as a good feature for one class combination. For newborns with an initial heart rate below 120, *identification of all class combinations includes at least two out of three parameters characterizing the therapeutic strategies*. Indicating that the therapeutic strategies are crucial for the newborn survival. A full overview of the found features in each experiment can be found in Paper 4, reprinted on page 167.

**Table 6.1:** Performance using feature selection in nested CV. P = precision, R = recall. 1 for class Normal, 2 for class NCU, and 3 for class Death

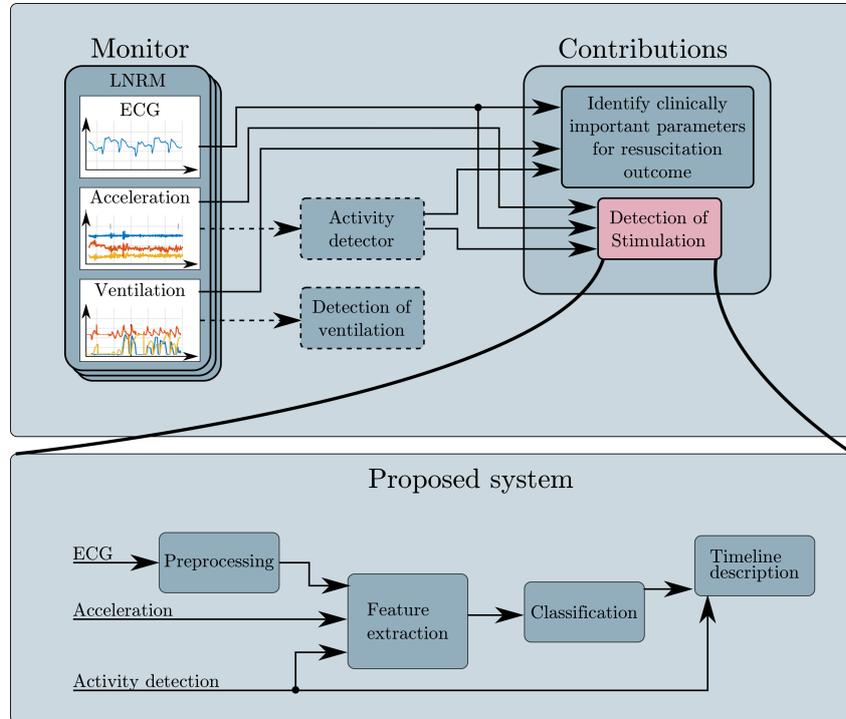
	Exp 1			Exp 2			Exp 3		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
	R1	R2	R3	R1	R2	R3	R1	R2	R3
Normal/Death	0.96		0.47	0.88		0.55	0.91		0.54
	0.89		0.74	0.87		0.57	0.83		0.71
Normal/NCU	0.84	0.42		0.74	0.46		0.72	0.45	
	0.71	0.60		0.71	0.50		0.74	0.43	
NCU/Death		0.85	0.60		0.78	0.69		0.71	0.51
		0.85	0.59		0.86	0.57		0.78	0.43
Normal/NCU/ Death	0.77	0.26	0.31	0.73	0.39	0.41	0.70	0.43	0.35
	0.64	0.28	0.65	0.74	0.24	0.69	0.75	0.16	0.67

## 6.4 Automatic annotation of Stimulation (Paper 5)

In this section, the contribution in the pink *Detection of Stimulation* block in Figure 6.3 is presented. Full details of this system can be found in paper 5.

During a resuscitation, the newborn will be moved, covered, and uncovered etc. Such activities will be visible on the accelerometer signals from the attached NeoBeat prototype but are not considered therapeutic activities. In VuClassifier, presented in section 6.1.2, our research group proposed a first attempt of classification of the detected activities based on ECG and accelerometer signals. The VuClassifier was, however, based on signal features extracted from detected activity events of variable duration. It needed statistics from the entire resuscitation episode, and as such, only suitable for retrospective analyses.

In this work we validate VuClassifier, and propose a new causal system, NBstim, for detecting time periods of stimulation activities based on the signals recorded by the NeoBeat prototype. In combination with VuVentilation [84], useful timelines can be created to illustrate the amount, duration and order of ventilations and stimulations performed in real world newborn resuscitation episodes. The generated timelines would be highly beneficial to evaluate which resuscitation activities are improving the state of the newborn. The timelines may be used to determine if current guidelines are adequate, and if they are followed.



**Figure 6.3:** Overview of the contributions in newborn resuscitation. Using raw inputs directly from the Laerdal NeoBeat prototype, a timeline describing the resuscitation event can be computed. The dotted line and boxes illustrate methods previously proposed by our research group.

### 6.4.1 Method

Using a larger data set, validation of the VuClassifier is first performed. To allow for either real-time analysis during resuscitation, or on request to obtain more details of a given resuscitation episode at a later time, a causal feature extraction is desired. A resolution of 10Hz in the classification is chosen. The features at index  $i$  are causally extracted from index  $i - k + 1$  to  $i$ , where  $k$  is the window size. The feature set are based on the 46 features proposed by Vu et al. in the VuClassifier [85]. As two of the features required information from the entire episode, they were excluded. The two features were the maximum value in the acceleration energy, and the maximum value in the ECG signal.

Stimulation activities often consist of repetitive movements, such as rubbing the back of the newborn. To represent these repetitive movements in

the analysis, three new features are defined for each axis in the accelerometer signal. Resulting in a total of 9 new features. Let  $P_{Acc,d}^i(f)$  denote the Short Time Fourier Transform (STFT) of window  $i$  in the accelerometer signal in axis  $d \in \{x, y, z\}$  as a function of the frequency,  $f$ . Using a window of size 1 second, equal to size of  $L = 100$  samples, the mean is first removed. The STFT is then computed using a  $N$ -point DFT, where  $N = 3 \cdot L$ . The first feature,  $A_{Pmax,d}$ , describes the maximum amplitude in the frequency domain,

$$A_{Pmax,d}(i) = \max\left(P_{Acc,d}^i(f)\right) \quad (6.14)$$

The second feature,  $A_{f,d}(i)$ , describes the frequency this maximum occurs at, according to

$$A_{f,d}(i) = \operatorname{argmax}\left(P_{Acc,d}^i(f)\right) \quad (6.15)$$

The third feature describes the highest frequency with an amplitude above a set threshold, given by

$$A_{fT,d}(i) = \max\{f : P_{Acc,d}^i(f) > T\} \quad (6.16)$$

The threshold  $T$  was set to 20, based on empirical observations to capture the highest frequency peak in the transformed signal.

As the ECG sensor in the LNRM is susceptible to noise from the power grid, a 50Hz notch filter is applied to the signal prior to extracting any of the features from the ECG. An overview of all features used can be found in Urdal et al. [91].

### 6.4.2 Result

Validation of the VuClassifier was done using the original feature set,  $fSet1$ . The classifier was trained on the original data described in Vu et al. [85], and then used to classify the subset  $R_D2_{blk}$ . The performance can be seen in table 6.2.

**Table 6.2:** Performance using the features and classifier proposed by Vu et al. [85] on the data set  $R_D2_{blk}$ .

Method	#features	Sensitivity	Specificity	Accuracy
VuClassifier [85]	46	88.4%	1.7%	50.7%

The performance using the 44 causal features,  $fSet2$ , on the data set  $R_D2_{win}$ , can be seen in table 6.3.

**Table 6.3:** Performance of various feature sets computed using sliding window on time points where an activity is found by VuDetector and either stimulation or no activity is manually annotated,  $R_{D2_{win}}$ . The performance is computed using a nested cross-validation. 3 folds were used in both the inner and the outer loop.

Feature set	#features	Sensitivity	Specificity	Accuracy
$fSet2$	44	61.7%	65.5%	61.7%
$fSet3$	24	63.3%	65.5%	63.7%
$fSet4$	9	75.6%	43.7%	69.7%
$fSet3 \cup fSet4$	33	57.4%	69.8%	59.7%
$fSet5$	23	67.3%	62.1%	66.4%

**Table 6.4:** Performance of NBstim when distinguishing between stimulation and non-stimulation on full resuscitation episodes,  $R_{D2}$ , with a maximum length of 7 minutes,

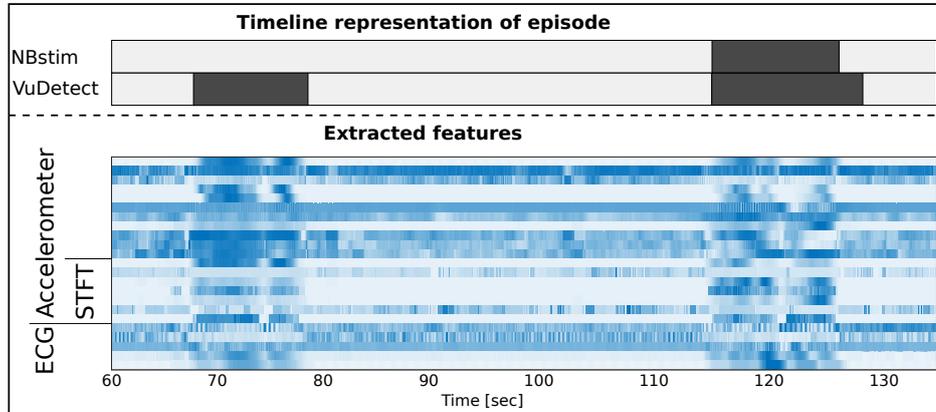
Method	Sensitivity	Specificity	Accuracy
No postprocessing	68.3%	93.1%	88.6%
Median filtering	69.2%	94.8%	90.3%

Using a greedy forward feature selection, the number of features were reduced down to 24,  $fSet3$ . A final feature selection was performed on the subset consisting of  $fSet3$  and the newly proposed STFT features,  $fSet4$ . The final feature set is denoted  $fSet5$ . The performance for each step is shown in table 6.3.

A visualization of the computed values for the final feature set,  $fSet5$ , is shown in Figure 6.4. Each row corresponds to a given feature, and a darker colour indicate a higher value in the computed feature value. For visualization purposes, all features in Figure 6.4 are normalized to  $[0, 1]$ .

To estimate the performance of NBstim, with the feature set  $fSet5$ , on complete episodes, a leave-one-out cross-validation is used. As we expect the therapeutic activities to last longer than the 0.1 second sample period, a median filtering is applied as a post processing scheme. A length of 11 samples is used for the median filter. The performance with- and without the post processing scheme is shown in table 6.4.

The proposed NBstim classifier can be combined with VuVentilation to create a complete system for detecting both stimulation and ventilation activities during newborn resuscitation. An example of how the graphical

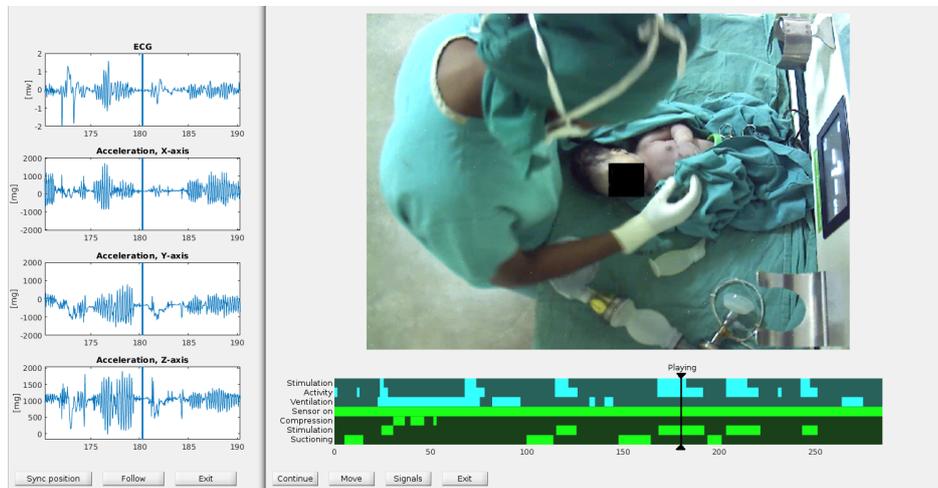


**Figure 6.4:** Overview of the 23 final features,  $fSet5$  computed for an example sequence where VuDetect identify two activities. Each row in the heat map illustrate the values of a given feature over time. A darker colour indicates a higher value. All features are normalized in the region  $[0, 1]$  for visualization purposes. NBstim evaluates the regions found by VuDetect and classifies them, one as stimulation and one as a non-stimulation activity.

user interface for this system could look like, is shown in Figure 6.5. The green timelines indicate manually annotated data, while the cyan timelines indicate automatically classified timelines. When used on new data, only the cyan timelines would be available to the user.

## 6. NEWBORN RESUSCITATION

---



**Figure 6.5:** Example graphical user interface for NBstim combined with VuVentilation. In this example, the timeline includes manually annotated data, shown in green, and automatically classified data, shown in cyan. ECG and acceleration signals are included to visualize the current measurements to the user.

## 6. NEWBORN RESUSCITATION

---

## Chapter 7

# Discussion and conclusion

In the following chapter, we first discuss our contributions and the challenges related to work presented on FHR and labour analysis and a conclusion and future work on the topic. We then discuss our contribution and the challenges related to newborn resuscitation. Finally, we conclude and discuss our plan for continuing this work.

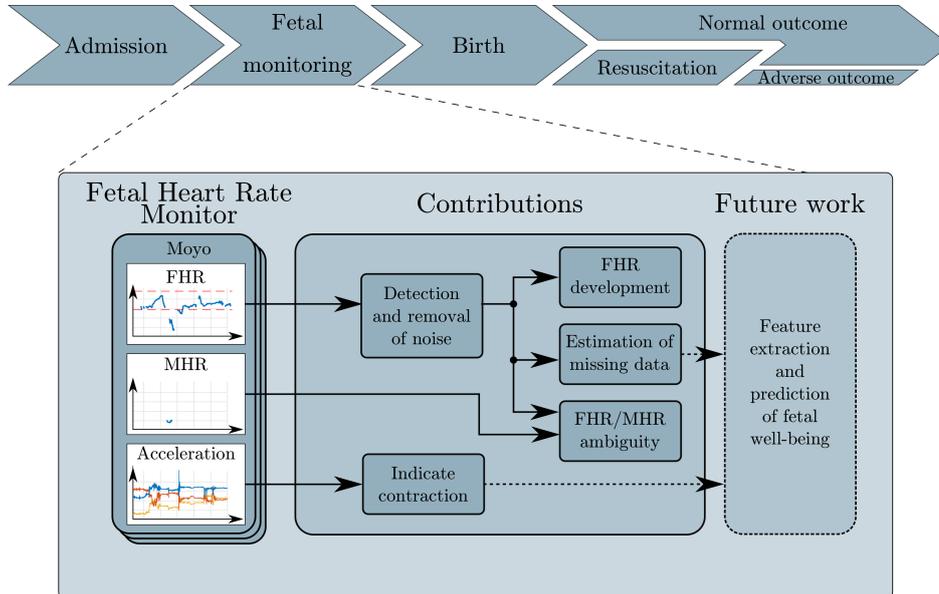
### 7.1 Fetal heart rate and labour analysis

The contributions to FHR and labour analyses, originally presented in Figure 5.1, are reprinted in Figure 7.1 to remind the reader of the relationship between the contributions. The main achievements can be summarized as:

- A1: Removal of less trustworthy segments in the measured FHR signal.
- A2: Estimation of missing data in FHR signals using shift-invariant dictionary learning.
- A3: Indication of when uterine contractions occur using an accelerometer mounted in proximity of the Doppler ultrasound sensor.
- A4: Explored FHR patterns during labour, and how these patterns differs based on the perinatal outcome of the newborn.

#### 7.1.1 Results and challenges

In the following subsections, each of the four achievements will be discussed.



**Figure 7.1:** Overview and relationship of the contributions related to fetal and heart rate and labour analysis presented in this work. This is a reprint of Figure 5.1.

### Noise detection

The noise-detection algorithm identifies many small sections of the FHR signal as noise. By removing these, a cleaner version of the FHR signal, and thereby a trend can be obtained. This opens for automated signal analysis and feature extraction in the future. A challenge in detecting time periods with noise in the data acquired using the Moyo, is that there is no truth data available. As it is difficult to determine with certainty which part of the measured FHR signal is noise, only time periods where the signal is very unlikely to contain information of the fetal status is removed. While some time period containing noise may be kept using this conservative approach, we consider it a better option over removing potential vital information.

### Uterine contractions

Information of when uterine contractions occur, can sometimes be found by looking for decelerations in the measured FHR signal. A challenge in this approach, is that an increased uterine activity may result in an increased movement of the mother, and thus the sensor. A result of this increased

movement, is more missing data. In our experiment with low amount of energy in the acceleration signal, the proposed method accurately identifies contractions corresponding to all large decelerations. Experienced midwives confirmed that the results resemble typical examples of decelerations caused by uterine contractions. However, when the amount of acceleration energy increases, it is challenging to determine if the detected contractions are correct. While the highest peaks are categorized as movement and therefore excluded, it is challenging to categorize remaining peaks as contractions and not smaller artefacts due to the movement. In cases where the FHR signal contains a large amount of missing data, the corresponding acceleration signal often contains more maternal movement. Potentially resulting in a lower identification of uterine contractions. The two thresholds used in the algorithm is defined conservatively based on empirical observation, and the chosen values might be suboptimal. A search for the optimal parameters is therefore required before the method is implemented in other systems.

While our findings of uterine contractions have been discussed with trained healthcare personnel during the study, the main limitation is the lack of truth data. A new data set should ideally be collected containing both Moyo signals as well as information of the uterine activity for validation of the proposed method. Uterine contractions may also in some cases occur at a higher rate than 5 per 10-minute window, known as tachysystole. As a threshold of minimum 2 minutes between the onset of two consecutive contractions is used in this method, detection of tachysystole may be limited.

### **Estimation of missing data**

Time periods with missing data in the FHR signal pose a challenge when extracting features in a continuous analysis. Simple methods, such as linear and spline interpolation are often used, but these affect computation of FHR variability. When the missing data percentage is low, both MOD and K-SVD achieve high performance, while SI-FSDL outperforms the other methods when the missing percentage is increased. A high SNR is also observed for linear interpolation. Reconstruction using dictionary learning is, however, shown to be closer to the true signal in terms of the spectral content of the signal. In order to have reliable information, less artefacts are crucial when performing further analysis on the data.

It is worthwhile to note that while estimations using dictionary learning can be utilized to reconstruct the gaps, the methods might miss some

details if the gaps become too large. In addition, if any of the missing time periods would have included a deceleration or acceleration in the FHR signal, we would need to know the duration of these patterns in order to be able to reconstruct the true signal.

### **Fetal heart rate development**

The heart rate of newborns have previously been reported to increase shortly after birth [22], and a corresponding drop in the FHR is observed close to the time of birth in this work. The behaviour is similar for all outcome groups, with a drop of the measured FHR shortly before birth. For labours in the perinatal mortality group, the drop in the measured FHR is larger, and it occur earlier relative to the time of birth. The variance in the estimated probability density function (pdf) has a larger increase for the perinatal mortality group, and the pdf peak has a larger shift. This difference may indicate that these newborns struggle to endure the physical strain of the labour, and that an earlier intervention could potentially save lives.

A limitation in this study is the low number of newborns in perinatal mortality group. In addition, not all recordings include data up until the time of birth, further reducing the number of episodes included in the study. To increase the amount of recordings, a longer data collection period would be desired. Ideally, data from additional hospitals would be used.

A second limitation of this work is that the internal clock in the first generation Moyo monitor was shown to be drifting. If not calibrated often enough this can result in inaccuracies of up to 30 minutes in the logged time stamp. A result of this, is that the FHR zero to one minute before birth, may in worst-case be recorded 30 minutes before birth for some episodes. In cases with drift in the opposite direction, an FHR may be present in the signal after the defined time of birth. These episodes are corrected by adjusting the time of birth to the time of the last recorded FHR.

### **7.1.2 Conclusion**

The contributions to FHR and labour analyses can be distinguished into two main categories: 1) A basis to allow for continuous analysis of the fetal well-being during labour, and 2) Exploration of how the FHR develops during labour based on the 24-hour outcome.

With the use of our proposed methods, less trustworthy time periods in the measured FHR signal can be removed. Missing data, or time periods where we have removed less trustworthy measurements, can be estimated using a shift-invariant dictionary. The dictionary achieves a high SNR and retains a lot of the spectral content in the estimations. With the use of an accelerometer mounted in proximity to the Doppler ultrasound sensor, we indicate when uterine contractions occur.

By observing the FHR, the last 150 minutes before birth, we found that the FHR drops for all labours close to birth. The drop was, however, larger and it occurred a longer time before birth for labours with an adverse outcome. This earlier drop in FHR, could indicate that these lives may have been saved with earlier interventions.

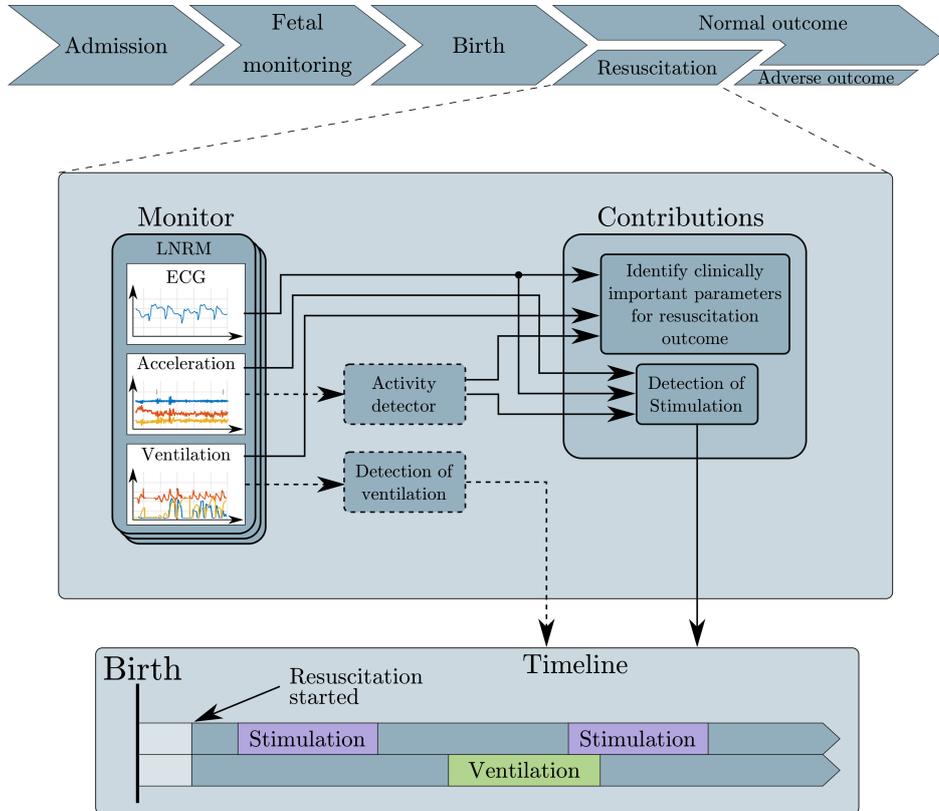
### 7.1.3 Future work

We plan to expand the work on FHR and labour analysis and develop a system to indicate the fetal well-being. Such a system can potentially alert the healthcare personnel when a risk is detected. The healthcare personnel can then perform a full assessment to determine if any interventions are required.

Before we start development on a system to indicate the fetal well-being, we plan to validate the results and methods presented in this work to ensure we base such a system on validated features and methods. While data collection in the Safer Birth project has reached its end, data collection at HLH continues with the Safer Haydom project. In addition, the scaling up birth bundle through quality improvement in Nepal (SUSTAIN) project has been started. This project aim to improve the intrapartum care using a set of quality improvement interventions and will collect data using both Moyo and NeoBeat. With the increased amount of data from these new projects, we plan to further explore and validate the differences between FHR development of labours with a normal and with an adverse outcome.

## 7.2 Newborn resuscitation

Our contributions on newborn resuscitation, originally presented in Figure 6.1, is reprinted in Figure 7.2 to remind the reader of the relationship between the two contributions. The main achievements can be summarized by:



**Figure 7.2:** Overview of the contributions in newborn resuscitation. The dotted lines indicate modules previously proposed by our research group. With the use of a system to detect stimulation and ventilation during resuscitation, timelines can be created to describe the resuscitation event. This is a reprint of Figure 6.1.

- A5: Identification of important parameters to describe the resuscitation outcome.
- A6: Proposed an improved system, NBstim, for automatic annotation of stimulation during newborn resuscitation using acceleration and ECG signals.

### 7.2.1 Result and challenges

In the following subsections, each of the two achievements will be discussed.

### **Identification of critically important parameters**

The results presented in this work suggest that parameters describing the initial status, therapeutic strategies, quality of ventilation, and early response parameters are crucial factors for distinguishing the 24-hour outcome for newborns identified as normal or dead at this point in time. These parameters do, however, struggle to identify newborns still in the neonatal care unit 24 hours after birth, indicating that additional features are required. To provide information back to the healthcare workers, the work must be extended to find features for identifying the newborns who are alive, but still in a critical condition 24 hours after birth.

Identification of features which can be utilized on all resuscitations is highly desirable. However, we acknowledge that newborns have a wide range of initial conditions, and that newborns in a poor initial condition may require a different therapy than newborns in a better condition.

### **Automatic annotation of stimulation**

A system for automatic identification of stimulation during newborn resuscitation is proposed. The system consists of an activity detector, and the proposed NBstim classifier with 23 features. Of these, 18 features are defined using the 100Hz accelerometer signals in X, Y, and Z-directions and 5 features are from the 500Hz dry-electrode ECG signal. The features are computed using a rectangular sliding window of 1 second with 900ms overlap. NBstim achieves high performance, with an accuracy of 90.3% in identifying stimulation, and could therefore be used as a replacement of time-consuming manual annotation, or as an initial step in an interactive tool. It can also be used with the recently released Laerdal NeoBeat newborn heart rate meter. But a validation using a larger data set is required before implementing the method in clinical practice.

Due to the small data set of only 74 resuscitations episodes, and a total of 21830 seconds, the proposed system may be seen as a feasibility study. When identifying the feature set, a smaller subset of only 1961 seconds was utilized. The reduction was performed by only including time periods where some movement occur, and only stimulation or no therapeutic activity is performed. The advantage of using the smaller subset, is that the method will identify features which are crucial in distinguishing stimulation and non-stimulation activities instead of focusing on patterns from other activities. Due to the limited data set with ground truth, further validation is required before applying this method in clinical practice.

### 7.2.2 Conclusion

The ultimate objective is to save lives at birth, and more specifically by studying what activities are performed by healthcare providers during resuscitation of asphyxiated newborns. Are the guidelines followed, and are the current guidelines effective in saving lives?

By identifying critically important parameters during newborn resuscitation, we found that therapeutic strategies such as stimulation, ventilation and hold-off were found to be important parameters in most of the classifications. While guidelines exist for how ventilation should be performed, guidelines on how stimulation should be applied are lacking. To aid in this work, we have proposed a system for automatically annotating stimulation using ECG and accelerometer signals of the newborn. With information on how stimulation is being applied during newborn resuscitation, we can potentially identify which stimulation strategy should be recommended.

### 7.2.3 Future work

Two main directions are planned for expansion of this work. We first plan to expand the work of finding critical features to also identify newborns who are alive, but still in a critical condition 24 hours after birth. We then plan to validate the NBstim classifier using a larger data set. While data collection in the Safer Births has reached its end, data collection of newborn resuscitation will continue at the Safer Haydom and SUSTAIN projects.

The next step for our research group is to develop a user-friendly standalone tool that can be utilized by other researchers to automatically annotate newborn resuscitation episodes. This tool can be used to create timelines for thousands of newborn resuscitation episodes. In combination with the immediate and 24-hour outcome, available in the Safer Births project, vital statistics can be extracted to potentially get a greater understanding of how stimulation activities affect resuscitation procedures and newborn outcomes.

**Paper 1:  
Noise and contraction  
detection using fetal heart  
rate and accelerometer  
signals during labour**



---

## Noise and contraction detection using fetal heart rate and accelerometer signals during labour

J. Urdal<sup>1</sup>, K. Engan<sup>1</sup>, Trygve Eftestøl<sup>1</sup>, Ladislaus Blacy Yarrot<sup>2</sup>, Kidanto Hussein<sup>3</sup>, Hege Ersdal<sup>4,5</sup>

<sup>1</sup> Department of Electrical Engineering and Computer Science, University of Stavanger, Norway

<sup>2</sup> Research Institute, Haydom Lutheran Hospital, Haydom, Manyara, Tanzania

<sup>3</sup> School of Medicine, Aga Khan University, Dar es Salaam, Tanzania

<sup>4</sup> Department of Anesthesiology and Intensive Care, Stavanger University Hospital, Norway

<sup>5</sup> Faculty of Health Sciences, University of Stavanger, Norway

**Published in the Scandinavian Health Informatics Conference 2019, SHI 2019.**

**Abstract:**

Fresh stillbirths and early neonatal deaths due to birth asphyxia are global challenges with an estimated 1.3 and 1.0 million deaths respectively every year. Adequate fetal monitoring during labour to prevent these deaths, is challenging, and regular assessment of fetal heart rate (FHR) in relation to uterine contractions is a key factor. A multi-crystal strap-on low-cost Doppler device, including an accelerometer, is recently developed to improve FHR monitoring in lower resource settings. In this work, we propose a method to increase interpretability of FHR Doppler signals by reducing noise, and a method to utilize accelerometer signals to estimate uterine contractions.

## 8.1 Introduction

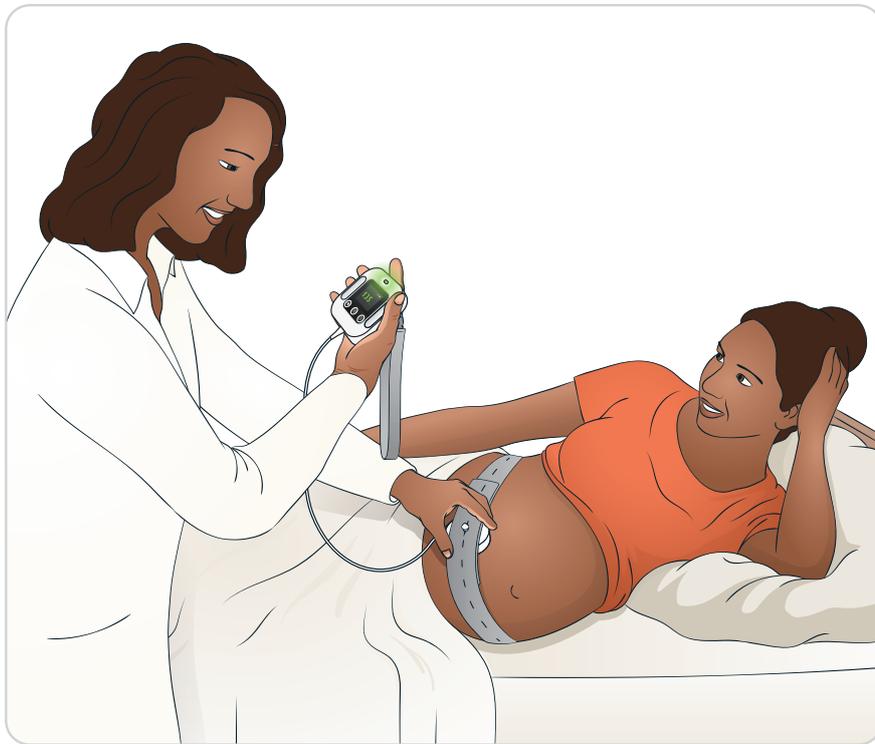
Fetal heart rate (FHR) monitoring is a widely used method to assess the status of a fetus during pregnancy and labour. In high resource countries, cardiotocography (CTG) is normally used for all labours assessed as high risk. This measuring technique normally includes an external Doppler based FHR sensor and a tocometer to measure uterine contractions. In cases where the Doppler based sensor is insufficient in obtaining a good quality measurement, an alternative FHR sensor can be attached directly to the scalp of the fetus. In low resource settings, however, assessment of the FHR is often conducted manually using either a fetoscope or intermittent Doppler. As these techniques does not include information of the uterine contractions, the FHR is often not assessed in relation to the contractions.

Fresh stillbirths and asphyxia-related newborn deaths, meaning the fetus dies during labour or soon after birth, are global challenges with an estimated 1.3 and 1.0 million deaths respectively every year [1]. The vast majority of these, 98%, occurs in low resource settings [1], and the primary cause of these deaths is interruption of placental blood flow with ensuing changes in FHR patterns [1, 2, 92]. Optimal FHR monitoring should detect such changes at an early stage to facilitate adequate obstetric interventions.

The introduction of a portable, low-cost, multi-crystal Doppler continuous FHR monitoring device (Moyo, Laerdal Global Health, Stavanger, Norway) at several sites in Tanzania, provides the opportunity to study the FHR changes and patterns without relying on human interventions to conduct periodic measurements. Well-known problems with such continuous Doppler devices are both noise and missing signal data. This can be caused by sensor movement, suboptimal placement of the sensor, maternal heart rate, doubling and halving of the FHR signal caused by the Doppler principle. Missing data can be estimated to resemble the measured data using dictionary learning [73, 93]. Artefacts due to noise may affect the interpretability and should be removed for both visual interpretation and further digital analysis. Methods for classification and suppression of this noise [71] and removal of the maternal heart rate [72] have previously been used on electrocardiography (ECG) signals from CTG. A system utilizing the sampled heart rate is, however, desired for low-cost continuous FHR monitoring devices.

Interpretation of the FHR signal during labour is normally conducted in relation to the corresponding uterine contraction, if this measurement is available. Accelerometers have previously been used to monitor muscle

contractions [77], and muscular fatigue [78]. Signals from an accelerometer attached to the abdomen during labour has been shown to correlate to uterine contractions [79]. By utilizing an accelerometer mounted in close proximity of the Doppler sensor, indications of when contractions occur can potentially be extracted. In this work, we have studied Doppler and accelerometer signals from Moyo and identified time periods in the measured FHR where the signal is likely to be noise. Using the three-axes accelerometer, we indicate the position where uterine contractions occur.



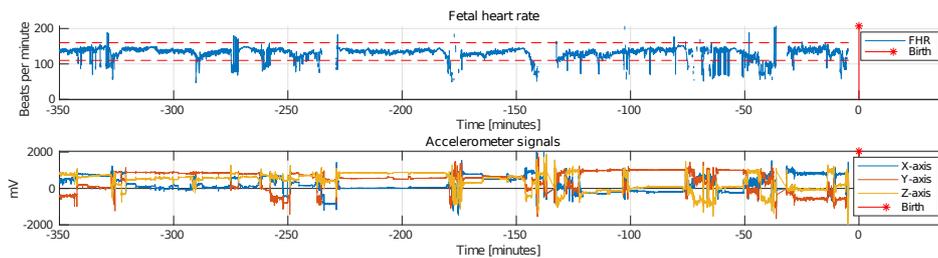
**Figure 8.1:** The Laerdal Moyo fetal heart rate monitor. Reprinted with permission [49].

## 8.2 Data material

The data is collected as part of the Safer Births research project, which is a research collaboration between multiple international research institutions, and hospitals in Tanzania. Data is collected at two urban and one rural

hospital in Tanzanian between October 2015 and June 2018. In total, 3807 labours were recorded. Of these, 3593 were classified as normal 24 hours after birth, 184 were still admitted to a neonatal care unit, 18 died during the first 24 hours, and 12 died during labour. Only labours which were assessed as normal on admission to the hospital were included in the study.

Data collection was done using the Laerdal Moyo fetal heart rate monitor [94], illustrated in Figure 8.1. The device consists of a main unit with a display presenting the measured heart rate to the health care personnel, and a sensor unit with a Doppler ultrasound sensor and an accelerometer. The sensor unit is attached to the mother using an elastic strap. If the detected FHR stays outside the 110-160 range for 10 minutes, or outside the 100-180 range for 3 minutes, an alarm will sound to alert the health care personnel. The FHR is measured using a 9-crystal pulsed wave Doppler ultrasound sensor operating at a frequency of 1MHz and an intensity of less than  $5\text{mW}/\text{cm}^2$ . The computed FHR is sampled at 2Hz. Movement of the sensor unit is measured using a three-axes accelerometer, sampled at 50Hz.



**Figure 8.2:** Signal example recorded using the Moyo fetal heart rate monitor. The red dashed lines indicate the normal region of the fetal heart rate. In the bottom plot, the three acceleration axes can be observed.

The project was ethically approved prior to implementation by the National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434) and the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest).

### 8.3 Method

This section first introduces a method to identify time regions in the FHR measurement where the heart rate is less trustworthy, and thus should be removed. A proposed method of estimating the point in time when

contractions occur based on acceleration signal follows. An example of the recorded signals is shown in Figure 8.2. The upper plot shows the FHR signal, and the lower plot shows the corresponding accelerometer. In the following we will use the notation  $fhr(n)$  to refer to a discrete FHR signal, and  $Acc_x(n)$ ,  $Acc_y(n)$  and  $Acc_z(n)$  to refer to the measured acceleration signals in the  $x$ ,  $y$ , and  $z$  directions respectively.  $\dot{x}(n)$  denotes the discrete derivative of the signal  $x(n)$ .

### 8.3.1 Noise detection

Noise introduced in the measured FHR,  $fhr(n)$ , can affect the visual interpretation conducted by medical personnel as well as introduce undesired artefacts in a continuous digital analysis. To identify time periods, hereafter called segments, where variations in the FHR cannot be explained from a physiological perspective, we first fill missing data in the FHR using forward replication, given by

$$fhr_{rep}(n+1) = fhr(n) : fhr(n+1) = 0 \quad \forall n \quad (8.1)$$

Let  $s$  be a pair of time indexes  $(t_s, k)$  representing the start time and length of a segment. Let  $A$  be a set of  $s$ ,

$$A = \{s : |\dot{fhr}_{rep}(t_s)| > c \cap |\dot{fhr}_{rep}(t_s + k)| > c \cap k < T_k\} \quad (8.2)$$

As the measured FHR is a result of a biological process, physiological limitations exist for how fast the heart rate can change, the threshold  $c$  is set to 30 beats per minute. The segments are thereafter checked in order from the shortest to the largest, to see if the large signal variation is a doubling or halving caused by a Doppler shift error. Let  $fhr_d(n)$  and  $fhr_h(n)$  denote the intersample variation, and be defined by:

$$fhr_d(n) = |2 \cdot fhr(n) - fhr(n-1)| \quad (8.3)$$

$$fhr_h(n) = |0.5 \cdot fhr(n) - fhr(n-1)| \quad (8.4)$$

The shift errors are identified by comparing the intersample variation to a threshold  $T_D$ , allowing for some intersample variability. The shift errors are corrected using:

$$fhr_c(n) = 2 \cdot fhr(n) : fhr_h(n) < T_D \quad (8.5)$$

$$fhr_c(n) = 0.5 \cdot fhr(n) : fhr_d(n) < T_D \quad (8.6)$$

$T_D$  is set to 5 based on empirical observation. If the sharp variations do not correspond to doubling or halving, the segment is considered as noise. When all segments of length  $< T_K$  are checked, the process is repeated using backward replication as some segments may be  $> T_K$  due to replication of missing data in the end of the segment. Based on findings from our previous work [5], the threshold  $T_K$  is set to 25 seconds. A cleaned FHR signal is returned. An overview of the method is shown in Algorithm 1.

---

**Algorithm 3:** noisedetection
 

---

**Input:** fetal heart rate,  $fhr$   
 Variation threshold,  $c$   
 Maximum length of segment,  $T_K$   
 Doubling/halving variation threshold,  $T_D$

**Output:** cleaned fetal heart rate,  $fhr_c$   
 $fhr_c = fhr$

**for**  $direction \in \{forward, backward\}$  **do**  
    $fhr_{rep}(n) = fillGaps(fhr_c(n), direction)$   
    $A = \{s : |fhr_{rep}(t_s)| > c \cap |fhr_{rep}(t_s + k)| > c \cap k < T_k\}$   
   **for** all  $s \in A$  sorted from smallest  $k$  **do**  
     **for** all  $i \in \{t_s, t_s + k\}$  **do**  
        $fhr_d(n) = |2 \cdot fhr(n) - fhr(n - 1)|$   
        $fhr_d(n) = |0.5 \cdot fhr(n) - fhr(n - 1)|$   
        $fhr_c(n) = 2 \cdot fhr(n) : fhr_h(n) < T_D$   
        $fhr_c(n) = 0.5 \cdot fhr(n) : fhr_d(n) < T_D$   
        $fhr_c(i) = 0 : |fhr(i)| > T_D$   
        $fhr_{rep}(n) = fillGaps(fhr_c(n), direction)$

**def**  $fillGaps(fhr_{rep}, direction)$ :  
   **if**  $direction = forward$  **then**  
      $fhr_{rep}(n + 1) = fhr(n) : fhr(n + 1) = 0 \quad \forall n$   
   **else**  
      $fhr_{rep}(n - 1) = fhr(n) : fhr(n - 1) = 0 \quad \forall n$   
   return  $fhr_{gaps}$

---

### 8.3.2 Estimation of contractions

An advantage of indicating the positions of the uterine contractions based only on the acceleration signal, allows the algorithm to run on recordings independent of missing FHR. The accelerometer captures small movements in the abdomen muscle as well as larger movements due to the mother changing positions. The acceleration signal amplitude of these movements is, however, typically vastly different. As the sensor location and orientation may be different between each labour, a trend describing the movement is computed using the acceleration energy,  $Acc_E(n)$ , given by:

$$Acc_E(n) = \sqrt{Acc_x^2(n) + Acc_y^2(n) + Acc_z^2(n)} \quad (8.7)$$

As the acceleration energy signal contains high frequency components, an upper envelope is computed to obtain the movement trend. The envelope of the acceleration energy,  $Acc_{env}(n)$ , is computed using a 20 second window. A set of positions,  $C$ , indicating contractions at time points,  $t_c$ , are found as local peaks of the envelope, given by

$$C = \{t_c : \dot{Acc}_{env}(t_c) = 0 \cap T_1 < Acc_{env}(t_c) < T_2\} \quad (8.8)$$

Where the thresholds  $T_1$  and  $T_2$  are set to  $10^{-2}$  and  $10^{-1}$  standard gravity,  $g_0$ , correspondingly, to avoid detecting small movements, and movements due to the mother changing position as contractions. As the intrapartum fetal monitoring guidelines from the International Federation of Gynecology and Obstetrics (FIGO) [3] states that  $< 5$  per 10-minute window averaged over 30 minutes is considered normal, the onset of two consecutive indicated contractions must occur at least 2 minutes from each other. The indicated contractions are hereafter called detected contractions. A pseudocode of the proposed contraction detection is depicted in Algorithm 2.

---

**Algorithm 4:** contractions

---

**Input:** Acceleration signals,  $Acc_x, Acc_y, Acc_z$

**Output:** Set positions for detected contractions,  $C$

$$Acc_E(n) = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2}$$

$$Acc_{env}(n) = envelope(Acc_E(n))$$

$$C = \{t_c : \dot{Acc}_{env}(t_c) = 0 \cap T_1 < Acc_{env}(t_c) < T_2\}$$


---

## 8.4 Experiments and results

As the dataset does not include measurements or registrations describing when uterine contractions or noise on the FHR signal occurs, experiments were devised to utilize both visual interpretation and statistics from the complete dataset to assess if the results from the proposed algorithms are reasonable. Experiments with visual interpretation of detected contractions on signals with low, medium, and high amounts of energy in the acceleration signal were chosen. The visual interpretation is based on discussions with trained midwives and the FIGO guidelines [3].

### 8.4.1 Noise removal

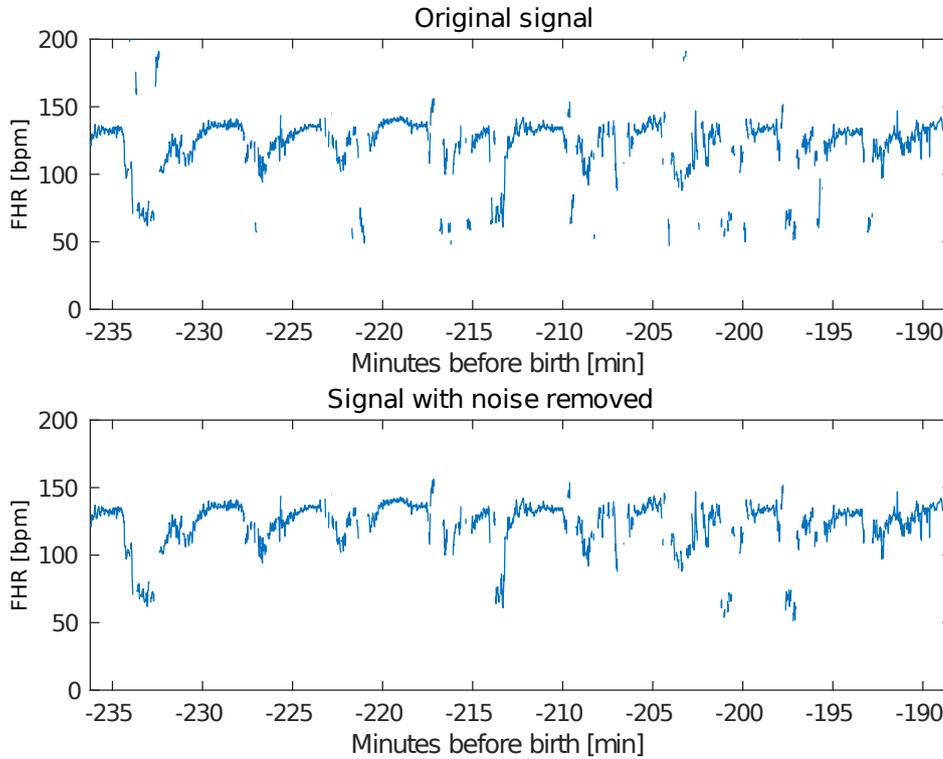
An example illustrating an example FHR signal, and the corresponding signal after noise removal is removed is shown in Figure 8.3. The method successfully identifies many of the outliers as noise, while some segments in the 75bpm region is kept. At the first stage of the data collection, the first generation Moyo was used. At a later stage, a second generation Moyo was used, and the percentage of missing data as well as noise was decreased. The algorithm was run on the complete dataset. An overview of the amount of detected noise is shown in table 1.

**Table 8.1:** Overview of the detected noise in the complete dataset.

Number of episodes	3807
Total duration of all episodes	14201 hours
Percentage of all samples with detected, and corrected, Doppler shift error	0.22
Percentage of all samples removed	2.73

### 8.4.2 Contractions on signals with low energy in the acceleration signal

Detection of contractions were conducted on a recording with low amount of energy in the acceleration signal extracted from the dataset, Figure 8.4. The FHR signal shows decelerations, which typically occur as a fetal response to a contraction. In the Figure we show the time points of detected contractions

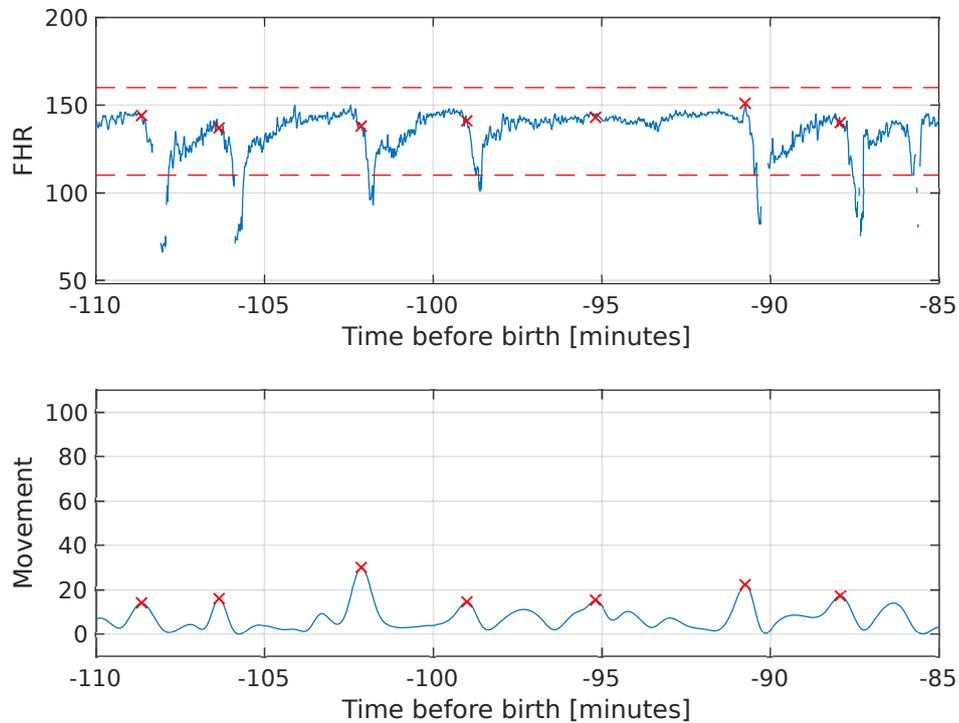


**Figure 8.3:** Example of noise detection and removal. Original signal on top, with some artefacts. Filtered signal on the bottom.

using red markers. It is easily seen that contractions corresponding to the 6 largest decelerations are detected. The contraction associated to the deceleration with a smaller drop in heart rate, at approximately 86 minutes before birth, is not considered to be caused by a contraction as it is too close to the previous detected contraction. An additional uterine contraction is detected at approximately 95 minutes before birth, without a corresponding deceleration in the FHR.

#### 8.4.3 Contractions on signals with medium energy in the acceleration signal

Detection of contractions were conducted on a recording with medium amount of energy in the acceleration signal from the dataset, Figure 8.5. Contractions are detected periodically in the first half of the signal, while

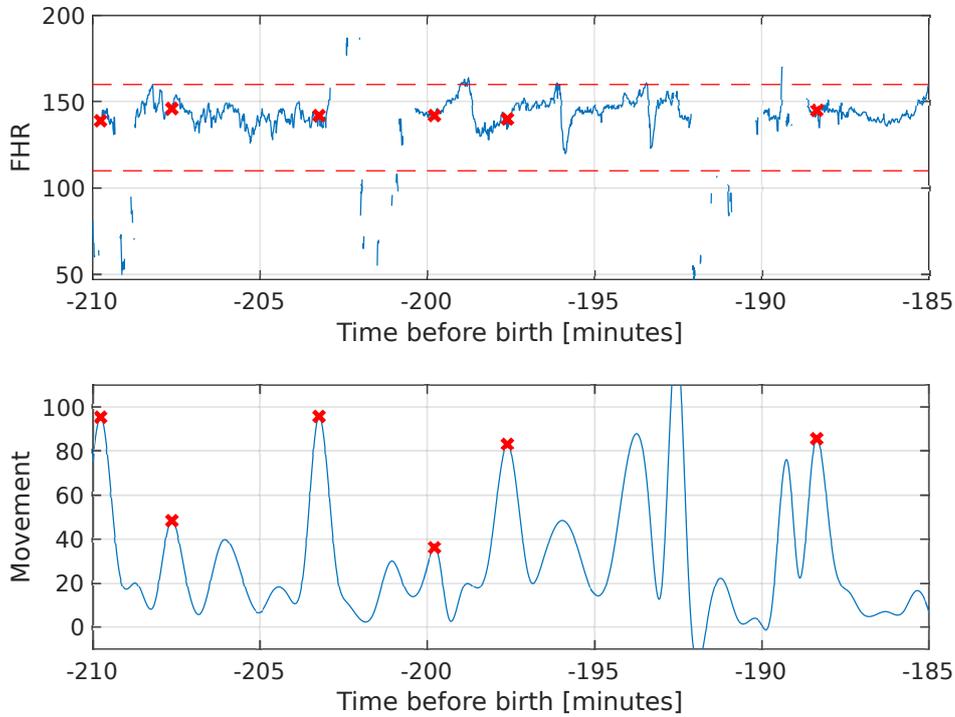


**Figure 8.4:** Detected contractions on a signal with low energy in the acceleration signal. The red dashed lines indicate the normal range of the FHR. The red crosses indicate the detected contractions.

only one contraction are detected in the second half. Due to the quality of the FHR signal, it is challenging to assess if these are actual uterine contractions.

#### 8.4.4 Contractions on signals with high energy in the acceleration signal

Detection of contractions were conducted on a recording with high amount of energy in the acceleration signal from the dataset, Figure 8.6. Four uterine contractions are detected in the 25-minute window, but it is challenging to assess if these are actual contractions due to the FHR signal quality.



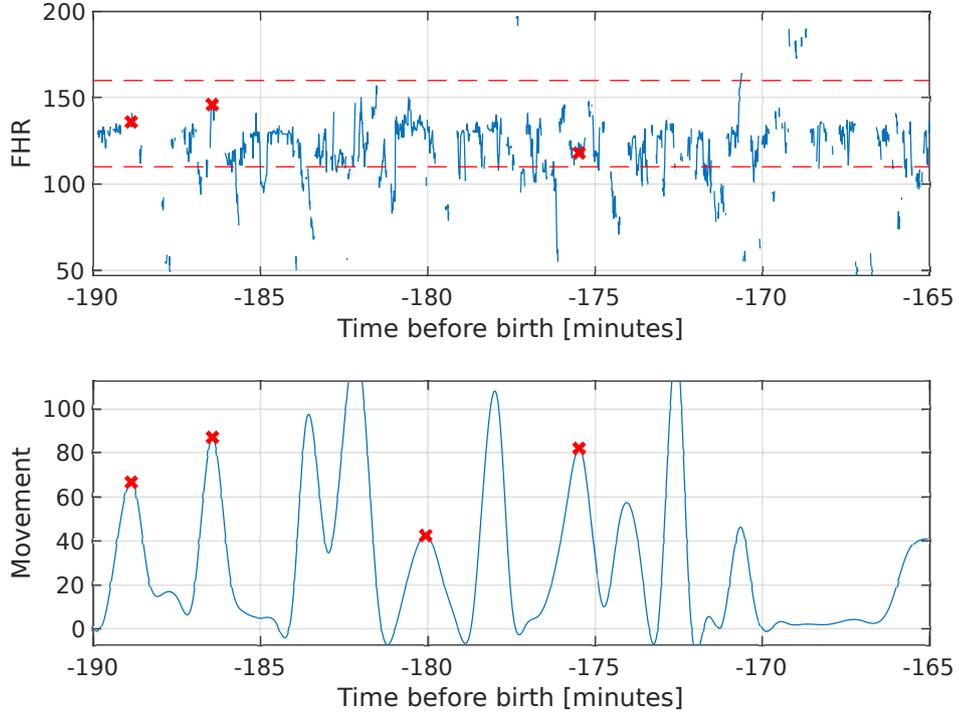
**Figure 8.5:** Detected contractions on a signal with a medium energy in the acceleration signal. The red dashed lines indicate the normal range of the FHR. The red crosses indicate the detected contractions.

#### 8.4.5 Overview of contractions on complete dataset

The algorithm was run on all 3807 recordings in the dataset to indicate how many contractions were found, the mean time between contractions and other performance metrics. The results are shown in table 2.

### 8.5 Discussion

The noise-detection algorithm identifies many small sections of the FHR signal as noise. By removing these, a cleaner version of the FHR signal, and thereby the trend can be obtained. This may result in improved visual interpretation as well as it opens for automated signal analysis and feature extraction for future work. As it is difficult to determine with certainty which part of the measured FHR signal that is noise, only time periods



**Figure 8.6:** Detected contractions on a signal with a high energy in the acceleration signal. The red dashed lines indicate the normal range of the FHR. The red crosses indicate the detected contractions.

**Table 8.2:** Overview of the detected uterine contractions in the complete dataset.

Episodes with detected contractions	3753
Episodes without detected contractions	54
Median number of detected contractions per episode	29 [14, 51]
Median length of episode	171 [90, 304]
Mean time between contractions	6.27 minutes

where the signal is very unlikely to contain information of the fetal status is removed. This conservative approach results in that some time periods containing noise may be kept.

Information of when uterine contractions occur can sometimes be found by

studying the FHR signal itself, as the fetus might respond to a contraction by a deceleration. A challenge in this approach is that uterine contractions may cause increased movement of the mother and sensor, thus increasing the amount of missing data in the FHR.

The proposed method correctly identifies contraction waveforms corresponding to all six large decelerations in the example with low amount of movement, seen in Figure 8.4. These decelerations are confirmed by experienced midwives to resemble typical examples of decelerations caused by uterine contractions. The detected contraction at 95 minutes before birth may still be an actual uterine contraction, even if it does not have a deceleration in the measured FHR. The time periods in between the detected contractions resembles typical labour, and it would be less typical if there was not detected a contraction at the 95-minute point. When the energy in the acceleration signal increase, as seen in Figure 8.5, less contractions are detected. As the number of contractions during a 10-minute window varies from labour to labour, it is difficult to do a direct comparison between recordings. In cases with a high energy in the acceleration signal, Figure 8.6, the movement create peaks with a higher amplitude than contractions. While the highest peaks, categorized as movement and therefore excluded, is not detected as contractions it is challenging to categorize remaining peaks as contractions and not artefacts due to the movement. In cases where the FHR signal contains a large amount of missing data, the corresponding acceleration signal often contains more maternal movement. That is resulting in a lower identification of uterine contractions. In addition, real contractions may in some cases occur at a higher rate than 5 per 10-minute windows, known as tachysystole. In the proposed algorithm, a threshold of minimum 2 minutes between the onset of two concurring uterine contractions is used, and this may be a limiting factor to detect tachysystole.

### 8.5.1 Limitations

A limitation of this work is the lack of tocometer measurements and manual annotations of the positions where uterine contractions occur in the dataset. To overcome this challenge, discussions regarding noise removal and indication of likely uterine contractions has been conducted during the study with trained health care personnel.

## 8.6 Conclusion

The work presented indicates that a large portion of the noise present in a FHR signal from Moyo can be removed utilizing only the sampled heart rate. It also indicates that a three-axes accelerometer mounted in proximity of the Doppler sensor, i.e. Moyo Fetal Heart Rate Monitor, can be used to estimate the point in time where contractions occur when the maternal movement is low. Further work validating indication positions of contractions with the use of a tocometer or manually annotated data must be conducted to determine the real performance.

## 8.7 Acknowledgement

This work is part of the Safer Births project which has received funding from Laerdal Foundation, Laerdal Global Health, Skattefunn, Norwegian Ministry of Education and USAID. The work was partly supported by the Research Council of Norway through the Global Health and Vaccination Programme (GLOBVAC) project number 228203. Validation of the dataset has been conducted by Sara Brunner and Solveig Haukås Haaland at Laerdal Medical AS.



**Paper 2:**  
**Estimation of Missing data  
in fetal heart rate signals  
using shift-invariant  
dictionary**



---

## Estimation of Missing data in fetal heart rate signals using shift-invariant dictionary

**F. Barzideh<sup>1</sup>, J. Urdal<sup>1</sup>, K. Engan<sup>1</sup>, Karl Skretting<sup>1</sup>, Paschal Mdoe<sup>2</sup>, Benjamin Kamala<sup>3,4</sup>, Sara Brunner<sup>5</sup>, Kidanto Hussein<sup>6</sup>**

<sup>1</sup> Department of Electrical Engineering and Computer Science, University of Stavanger, Norway

<sup>2</sup> Research Institute, Haydom Lutheran Hospital, Haydom, Manyara, Tanzania

<sup>3</sup> Faculty of Health Sciences, University of Stavanger, Norway

<sup>4</sup> Department of Obstetrics and Gynecology, Muhimbili National Hospital, Tanzania

<sup>5</sup> Strategic Research, Laerdal Medical AS, Stavanger, Norway

<sup>6</sup> Muhimbili University of Health and Allied Sciences, Tanzania

**Published in the 26th edition of the European Signal Processing Conference, EUSIPCO 2018.**

<https://doi.org/10.23919/EUSIPCO.2018.8553110>

**Abstract:**

In 2015, an estimated 1.3 million intrapartum stillbirths occurred, meaning that the fetus died during labour. The majority of these stillbirths occurred in low and middle income countries. With the introduction of affordable continuous fetal heart rate (FHR) monitors for use in these settings, the fetal well-being can be better monitored and health care personnel can potentially intervene at an earlier time if abnormalities in the FHR signal are detected. Additional information about the fetal health can be extracted from the fetal heart rate signals through signal processing and analysis. A challenge is, however, the large number of missing samples in the recorded FHR as fetal and maternal movement in addition to sensor displacement can cause data dropouts. Previously proposed methods perform well on estimation of short dropouts, but struggle with data from wearable devices with longer dropouts. Sparse representation and dictionary learning have been shown to be useful in the related problem of image inpainting. The recently proposed dictionary learning algorithm, SI-FDSL, learns shift-invariant dictionaries with long atoms, which could be beneficial for such time series signals with large dropout gaps. In this paper it is shown that using sparse representation with dictionaries learned by SI-FDSL on the FHR signals with missing samples, provides a reconstruction with improved properties compared to previously used techniques.

## 9.1 Introduction

Fetal heart rate (FHR) monitoring is a widely used method to assess the fetal well-being during labour. FHR monitoring, used by trained health care professionals, allows for early detection of a fetus at risk and consequently appropriate and timely action to prevent further harm to fetus and mother.

In high income countries, FHR monitoring in labours with high risk are usually measured using continuous Doppler ultrasound in cardiotocography (CTG), while low risk labours are monitored intermittent with handheld Doppler devices. In low resource settings, FHR is often accessed intermittent with a fetal stethoscope or handheld Doppler. In 2015, there were an estimated of 2.6 [uncertainty range 2.4-3.0] million stillbirths [1, 2], with 1.3 [uncertainty range 1.2-1.6] million deaths occurring during labour [1]. The vast majority (98%) of these occur in low and middle income countries [1]. Improved care at birth, including continuous FHR monitoring, is the key to reduce the number of stillbirths. Abnormalities in the FHR signal can be detected earlier with the use of continuous monitoring. If any abnormalities are detected, an alarm can be used to alert qualified health care personnel to assess the situation. Additional information of the fetal well-being can be extracted from the FHR signals through signal processing and analysis. With the introduction of affordable devices for continuous FHR monitoring, such as Moyo Fetal Heart Rate Monitor, used in this study, new opportunities arises as these devices are also obtainable in low resource settings.

A well-known problem measuring FHR using Doppler ultrasound are signal dropouts due to both fetal and maternal movement in addition to sensor displacement. With the introduction of wearable devices for continuous FHR monitoring, allowing the mother to move freely while the device is attached, an increase in both the number and length of signal dropouts are expected. These missing samples are a challenge when determining traditional features used to assess the fetal well-being, such as the short and long time variability of the FHR, as well as when doing time-frequency analysis on the heart rate signal.

Simple methods such as linear interpolation[61] and cubic Hermite spline interpolation[62] and more complex methods such as Gaussian processes [63] and K-SVD [29] have previously been used to estimate the missing samples on FHR recorded by CTG. However, depending on the length of the gaps, these methods affect computation of the traditional heart rate *features*.

As more and longer dropouts are expected when using wearable monitors, better estimations are desired.

Dictionary learning and sparse approximation have been shown to produce state of the art results in estimation of missing data [27, 28, 29]. An important advantage of using dictionary learning over methods such as linear or spline interpolation is that through learning from the signal class, a learned dictionary introduces less artefacts during processing, feature extraction, and time-frequency analysis.

For an inpainting problem with large gaps, unstructured dictionaries, produced by general dictionary learning methods such as MOD [30] or K-SVD [31] require large atom lengths which means increase in number of free variables. This leads to slow training and usage as well as the possibility of overfitting.

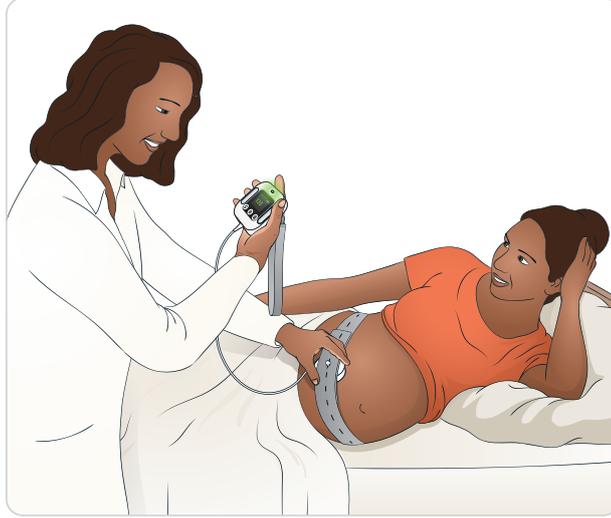
In this work we propose to use shift-invariant dictionary by utilizing SI-FSDL, a dictionary learning method for shift invariant dictionaries recently proposed by our group [80]. SI-FSDL is capable of handling variable shifts and length of the atoms.

## 9.2 Data material

The data used in this study are collected by the Safer Births Research Project, which is a research collaboration with partners including, but not limited to, University of Stavanger, Laerdal Global Health, and partner hospitals in Tanzania. The data were collected at Haydom Lutheran Hospital (HLH), which is a rural hospital, and Muhimbili National Hospital (MNH), and Temeke Regional Referral Hospital (TRRH) which are both urban hospitals. 85 labours were monitored and recorded at HLH between February 1st. and March 18th. 2016, 227 labours were recorded at MNH between March 15th. and July 13th. 2016, and 1087 births were recorded at TRRH between June 4th. and October 1st. 2016. All data were anonymized prior to transfer to researchers.

The project was ethically approved prior to implementation by the National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434) and the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest) before the start of the study.

Moyo Fetal Heart Rate Monitor, Fig. 9.1, is used to record the fetal heart rate and is developed by Laerdal Global Health [94] as an affordable



**Figure 9.1:** Moyo Fetal Heart Rate Monitor, Laerdal Global Health AS, Norway. Illustration reproduced with permission [49]

FHR monitoring device for both intermittent and continuous monitoring for use in low resource settings. The device consists of a small handheld unit with display and a sensor unit. For continuous monitoring, the sensor can be applied to the maternal abdomen using an elastic band, shown in Fig. 9.1. The sensor unit includes a 9-crystal pulsed wave Doppler ultrasound sensor operating at a frequency of 1 MHz and an intensity of less than:  $5mW/cm^2$ . The detected fetal heart rate is logged at 2Hz. The sensor unit also includes a 3-axes accelerometer sampled at 50Hz used to describe maternal movement, and a temperature sensor, sampled at 2Hz. FHR and maternal HR as well as accelerometer and temperature values are stored to files and can be accessed via USB connection.

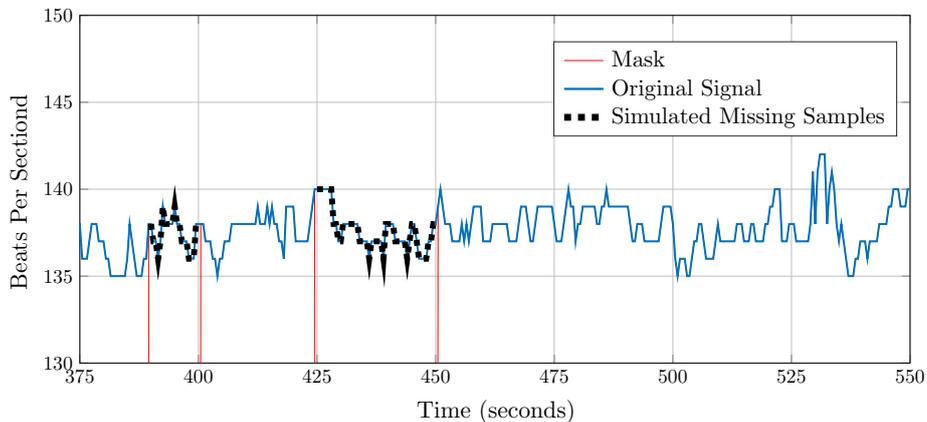
### 9.3 The proposed method

Sparse representation and dictionary learning is based on the idea that it is possible to represent a signal class sparsely in some domain, and that a learned dictionary can represent this domain. Let an  $N \times 1$  signal vector be denoted by  $\mathbf{x}$ , and its approximation as  $\hat{\mathbf{x}} = \mathbf{D}\mathbf{w}$ , where  $\mathbf{D}$  is the dictionary matrix of size  $N \times K$ , with the columns ( $\mathbf{d}_i$  s) forming dictionary *atoms*,

and  $\mathbf{w}$ ,  $K \times 1$ , is the vector of *sparse* coefficients. The dictionary learning problem is formulated as follows:

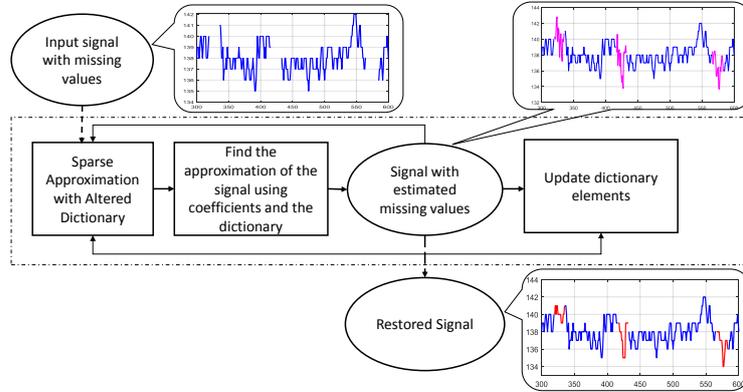
$$\mathbf{W}, \mathbf{D} = \underset{\mathbf{W}, \mathbf{D}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 \quad s.t \quad \begin{cases} \mathbf{w}_i \text{ is sparse} \\ \|\mathbf{d}_i\|_2 = 1. \end{cases} \quad (9.1)$$

where  $\mathbf{W}$  and  $\mathbf{X}$  are formed from concatenation of coefficient  $\mathbf{w}_i$  and signal vectors  $\mathbf{x}_i$  respectively. Since equation 9.1 is not tractable, it is usually broken into two steps: in the first step, *sparse coding*, one would find  $\mathbf{W}$  while fixing  $\mathbf{D}$ . In the second step, *dictionary update*,  $\mathbf{D}$  is found while keeping  $\mathbf{W}$  constant. MOD [30] and K-SVD [31] are examples of dictionary learning methods using these steps.



**Figure 9.2:** A sample FHR signal, its masked version and simulated missing samples. The masked signal has zeros as value where there are missing samples and is equal to the original signal elsewhere.

In this paper we are dealing with recovering of missing data or inpainting where the location of missing data is known beforehand. Since signals and the location of missing samples differ from patient to patient, and these analyses are done in retrospect, we wish to learn a dictionary on the signals with missing data to tailor the dictionary to the person, before performing the reconstruction. For this reason, the information about the missing samples is built into a mask matrix. One possible mask matrix is an identity matrix with its rows removed on locations where the data is missing. This means that applying it to a vector or a matrix, removes the rows corresponding to missing samples.



**Figure 9.3:** Different steps of inpainting. The input contains missing samples. Each iteration of this method replaced the missing values with their approximated version. This is done until some criteria is satisfied.

Another type of mask, identity matrix with zeros on the locations of missing samples, only zeros out the missing samples. In this paper, the later type of mask is used to help visualize the gaps while the first type is used for sparse approximation. When having multiple vectors such as in  $\mathbf{X}$ , there is a mask matrix for every vector in  $\mathbf{X}$ . Fig. 9.2 shows a sample FHR signal along with the masked signal and the missing samples.

The literature describes two ways to incorporate the mask information into the dictionary learning steps. One method was used in [28, 95, 96] which alters both dictionary learning steps. An alternative approach was briefly discussed in [29] which requires only changes to the sparse coding step. In this paper, we apply the second method for inpainting using shift-invariant dictionary.

The altered steps required for the second method are as follows:

- (i) While the dictionary entries are fixed, remove the rows of the dictionary and the signal vectors corresponding to the missing samples for each signal vectors.
- (ii) Normalize the dictionary columns to 1.
- (iii) Find the coefficients using their own respective dictionaries.
- (iv) Find the approximation of the signal vectors by multiplying the coefficients with the full dictionary.

- (v) Reconstruct the signal by replacing the missing samples with their approximated resulted from above.
- (vi) Update the dictionary elements and normalize the columns to 1.

The first three steps describe the sparse coding stage for inpainting. Fig. 9.3 illustrates the different steps of this solution.

For imposing the shift-invariant structure onto the dictionary we utilized our previously proposed method, SI-FSDL[80]. One of the benefits of shift-invariant dictionaries is that we can address larger gaps (by using larger shift-invariant atoms) than a general dictionary while keeping the number of free variables fixed.

$$\begin{bmatrix} \star_0 & 0 & \star_0 & 0 & 0 & \alpha_0 & 0 \\ \star_1 & \star_0 & \star_1 & \star_0 & 0 & \alpha_1 & 0 \\ \star_2 & \star_1 & 0 & \star_1 & \star_0 & 0 & \alpha_0 \\ 0 & \star_2 & 0 & 0 & \star_1 & 0 & \alpha_1 \end{bmatrix}$$

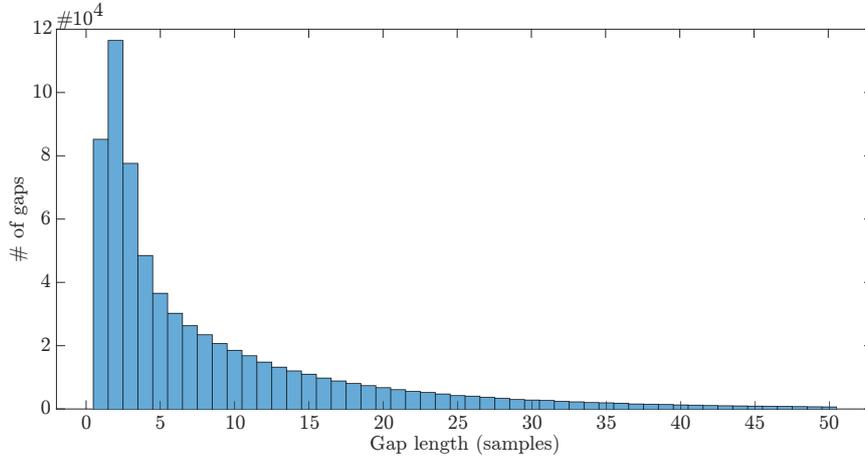
**Figure 9.4:** A simple shift-invariant dictionary with 3 shift-invariant atoms(SIAs). The first two SIAs have 1 shift while the last one has 2 shifts.

An example of a small shift-invariant dictionary with three shift-invariant atoms or SIAs, is depicted in Fig. 9.4. This example does not have circular shifts, i.e. the the shifting ends as the last non-zero element of a SIA reaches the bottom row of the dictionary matrix. SI-FSDL handles variable length and variable shift atoms as Fig. 9.4 illustrates.

## 9.4 Experiments

A total of 691400 segments of missing samples are found in the 1399 recordings in the dataset, with an average of 494 missing segments in each recording. In total, the missing percentage of data is 36.4%. However, 96.9% of the missing data gaps are less than 50 samples in length. The distribution of the length in these gaps from 1 to 50 samples is shown in Fig. 9.5.

We have chosen a signal *without* missing samples from our database and randomly removed parts of it so that the true signal is available to evaluate the recovery results. The used recovery methods are SI-FSDL, MOD, K-SVD, spline and linear interpolation. During the experiments the number



**Figure 9.5:** Distribution of gap lengths

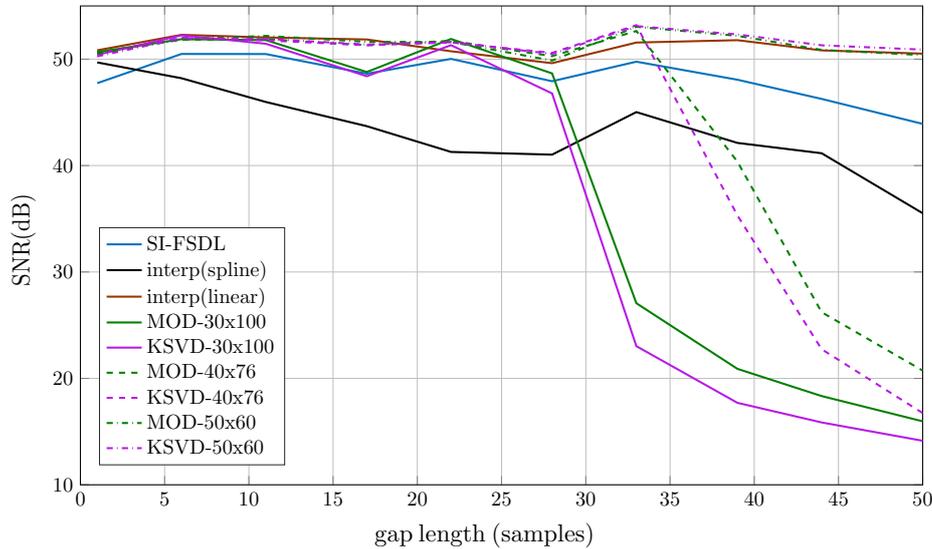
of free variables in the dictionaries are kept constant at approximately 3000. The ratio of non-zeros coefficients to the number of elements in the signal is 0.1. Signal blocks are chosen in an overlapping fashion for the dictionary learning methods. Block lengths,  $N$ , of 30, 40 and 50 are used for MOD and K-SVD, and 500 for SI-FSDL.

#### 9.4.1 Experiment 1

The first experiment is designed to evaluate the average performance of each method when the missing percentage is fixed, but the gap lengths changes. In order to have a realistic scenario, the fixed missing percentages are set to 10 and 30. The length of the gaps ranges from 1 sample to 50 samples. To find the average performance for each gap length, different randomly created masks are used.

Performance of the tested methods for 10% missing data are shown in Fig. 9.6. All methods achieves similar performance for short interval lengths, with the exception of spline interpolation. As the gap lengths increases, the performance of MOD and K-SVD decreases. The exception to this is when the segment length is 50. Their performance is at least as good as linear interpolation and always better than SI-FSDL.

Performance of the tested methods for 30% missing data are shown in Fig. 9.7. All methods achieves similar performance for short interval lengths. With higher ratio of missing data, the performance for MOD and



**Figure 9.6:** Recovery performance for different methods when the missing interval changes but the missing percentage stays the same (10%)

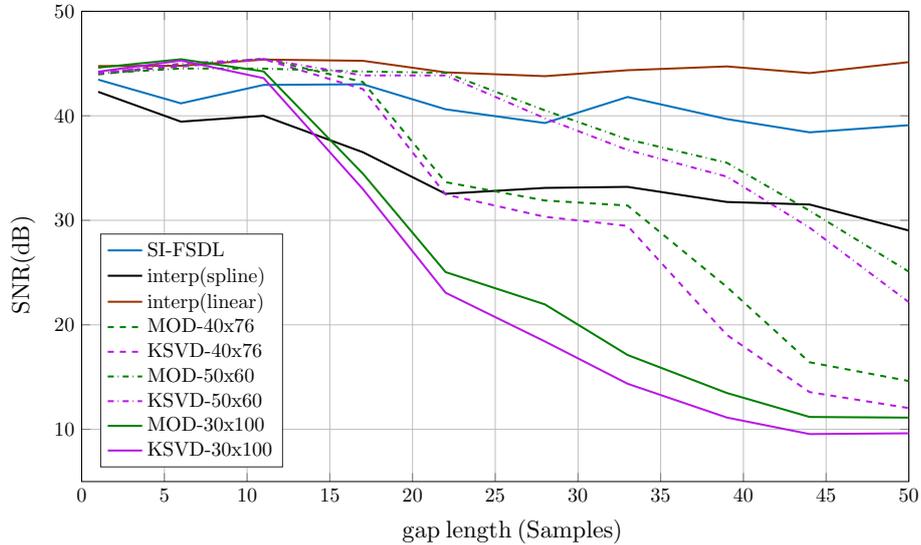
K-SVD for all segment lengths decrease faster than the case of 10%. The performance of linear interpolation and SI-FSDL, remain almost the same regardless of length of missing sample interval.

### 9.4.2 Experiment 2

The experiment is devised to have a closer look at the best performing methods of last experiment when the missing percentage is 30%, which is close to the percentage for our dataset. These methods are linear interpolation and SI-FSDL. The intent is to look at the continuous wavelet transform of their reconstructed signals and see how similar the time-frequency distribution of the reconstructed signal is to the original signal.

The time-frequency response for a short section of FHR with 3 missing gaps are shown in Fig. 9.8. In close-ups of the signal around each mask, the original signal is shown in blue, estimations using linear interpolation in dotted red, and SI-FSDL in dashed black.

It can be seen by visual inspection that SI-FSDL restores the time-frequency properties better than linear interpolation even though the later has higher SNR.

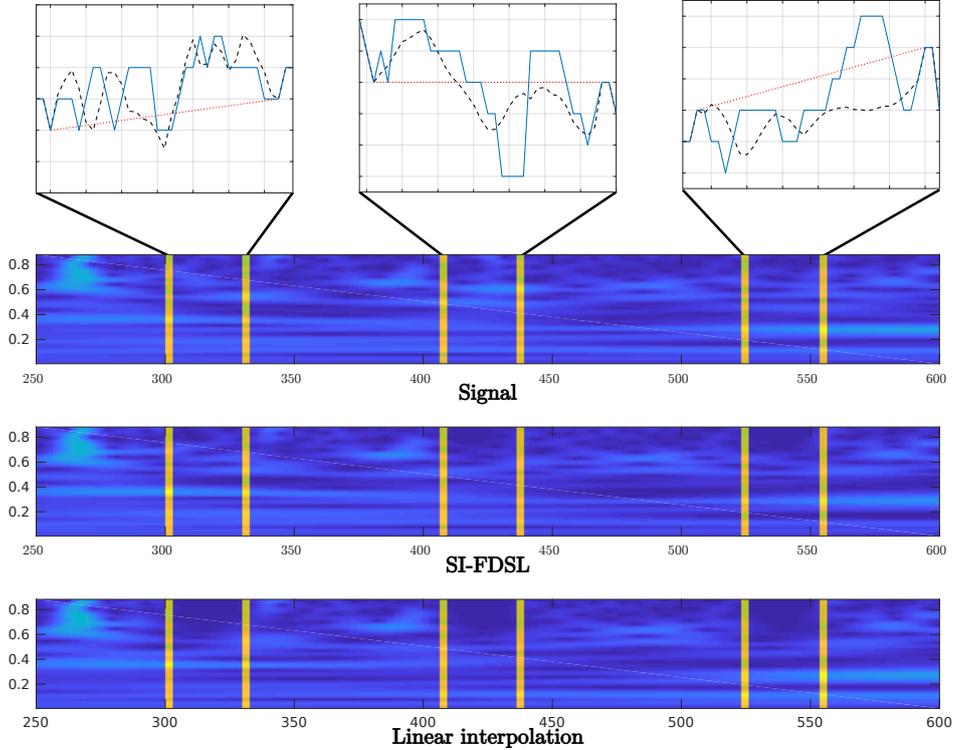


**Figure 9.7:** Recovery performance for different methods when the missing interval changes but the missing percentage stays the same (30%)

## 9.5 Discussion

It is worthwhile to note that while inpainting methods can be utilized to reconstruct the gaps, they might miss some details if the gaps become too large. In the case of FHR signals temporary increases or decreases, known as accelerations and decelerations in the heart rate, are important details when determining the fetal well-being. In order to recover such information, we need to know the duration of these patterns. In an abrupt accelerations and decelerations the FHR has a change of 15 beats per minute with a time from onset to extremum of  $\leq 30$  seconds and total duration of less than 2 minutes. Based on this, it is safe to reconstruct segments with a maximum length of 25 seconds, corresponding to 50 samples. Fig. 9.5, shows that most of the gaps are short in length, with 96.9% of the gaps below our upper limit.

On data where 10% of the samples are missing, Fig. 9.6, linear interpolation and all dictionary learning techniques achieve similar SNR for gaps  $\leq 28$  samples in length. MOD and K-SVD with block length 50 show the best performance for all the missing gaps. Increasing the missing percentage to 30%, Fig. 9.7, a large drop in SNR is seen in both MOD and



**Figure 9.8:** Continuous wavelet transform over a short signal with 3 gaps missing samples. Original signal shown in blue, linear interpolation shown as dotted red line, SI-FSDL shown as dashed black line.

K-SVD. Depending on the gap size, this dropout occurs for gaps larger than 10, 17 and 23 samples. Since MOD and K-SVD have block lengths of 30 to 50, they cannot restore gaps close or larger than their size. A possible solution to this is to increase the block length in MOD and K-SVD. However, this usually means increasing the overall number of free variables as well and learning a larger dictionary which requires more data and processing time. Due to its structure, however, SI-FSD can reconstruct larger gaps by adjusting the length and number of SIAs.

A high SNR is seen for linear interpolation and SI-FSDL for all gaps. The challenge of using linear interpolation, however, is that it introduces artefacts, as seen in Fig 9.8. In the first and third gaps, samples 300-330 and 525-555, linear interpolation introduces artefacts by removing high frequencies. In the same gaps, SI-FSDL shows more fidelity to the original

signal. Both methods perform similarly in the second gap, samples 405-435, and introduce artefacts.

## 9.6 Conclusion

The results presented in this work indicate that for dictionary learning based methods, gap interval and missing percentage are important parameters when attempting to recover missing data in the signal. When the missing percentage is low, MOD and K-SVD achieve the highest performance, while SI-FSDL outperforms the other methods when the missing percentage is increased. A high SNR is also observed for linear interpolation. Reconstruction based on dictionary learning methods, however, are shown to be closer to the true signal in terms of the spectral content of the signal. In order to have reliable information, having less artefacts is crucial when performing further analysis on the data.

## 9.7 Conflict of interest

The study was supported by the Laerdal Foundation and the Research Council of Norway through the Global Health and Vaccination Program (GLOBVAC), project no. 228203

Sara Brunner is an employee of Laerdal Medical AS.



**Paper 3:**  
**Fetal heart rate**  
**development during labour**



---

## Fetal heart rate development during labour

**J. Urdal<sup>1</sup>, K. Engan<sup>1</sup>, Trygve Eftestøl<sup>1</sup>, Solveig H. Haaland<sup>5</sup>, Benjamin Kamala<sup>3,4</sup>, Paschal Mdoe<sup>2</sup>, Kidanto Hussein<sup>3</sup>, Hege Ersdal<sup>4,5</sup>**

<sup>1</sup> Department of Electrical Engineering and Computer Science, University of Stavanger, Norway

<sup>2</sup> Research Institute, Haydom Lutheran Hospital, Haydom, Manyara, Tanzania

<sup>3</sup> School of Medicine, Aga Khan University, Dar es Salaam, Tanzania

<sup>4</sup> Department of Anesthesiology and Intensive Care, Stavanger University Hospital, Norway

<sup>5</sup> Faculty of Health Sciences, University of Stavanger, Norway

**The paper is currently under review.**

**Abstract:**

**Background** Fresh stillbirths (FSB) and very early neonatal deaths (VEND) are important global challenges with 2.6 million deaths annually. The vast majority of these deaths occur in low and low-middle income countries. Assessment of the fetal well-being during pregnancy, labour, and birth is normally conducted by monitoring the fetal heart rate (FHR). The heart rate of newborns is reported to increase shortly after birth, but a corresponding trend in how FHR changes just before birth for normal and adverse outcomes has not been studied. In high resource settings, monitoring of the FHR is done using cardiotocography for labours assessed as high risk, but the number of labours with an adverse outcome are low. In this work we utilize FHR measurements collected from 3711 labours from a low and low-middle income country to study if there are trends and patterns in the FHR development close to the time of birth related to the neonatal well-being 24-hours after birth. A signal pre-processing method was applied to identify and remove time periods in the FHR signal where the signal is less trustworthy. We suggest an analysis framework to study the FHR development. The FHR trend is found for labours with a normal outcome, neonates still admitted for observation, FSB and VEND. Finally, we study how the spread of the FHR changes over time during labour.

**Results** A drop in median FHR as well as an increased spread in FHR is observed for all defined outcome groups near the time of birth. For labours ending with FSB or VEND, the drop in median FHR is larger, it occurs longer time before birth, and the spread is higher compared to labours with a normal outcome or for neonates admitted to neonatal care unit.

**Conclusion** The observed difference in the drop of median FHR between the outcome groups indicate that neonates in the VEND/FSB group struggle to endure the physical strain of the labour, and that an earlier intervention could potentially save lives.

## 10.1 Background

Fetal heart rate (FHR) monitoring is a widely used method to assess the status of the fetus during pregnancy, labour and birth. In high resource countries, continuous monitoring of the FHR is done using cardiotocography (CTG) for labours categorized as high risk. In low income and low-middle income countries (LMIC), an intermittent measurement is the norm for all labours. The intermittent measurement is normally conducted using a Pinard stethoscope or hand-held Doppler. Guidelines state that auscultation of FHR should be conducted every 15-30 minutes during the first stage of labour, and every 5-15 minutes during the second stage of labour. Each auscultation should also last for at least one minute [5]. The intervals defined by the guidelines is not possible without a nurse:patient ratio of 1:1 [6] and will be a challenge in LMIC where the ratio of health care workers to the number of labours is much lower. A limitation of intermittent auscultation, independent of the device used, is that the status of the fetus is only checked during a specific point in time. When the time between each auscultation increase, the possibility of detecting an abnormal FHR may be reduced.

Stillbirths is a worldwide challenge, with an estimated 2.6 million [uncertainty range 2.4-3.0] stillbirths in 2015 [1], of these 1.3 million is estimated to have died during labour and birth, i.e. fresh stillbirth (FSB). In addition, one million newborns die within their first and only day of life [1, 2]. Asphyxia and prematurity complications during labour are the primary causes of death. The vast majority, 98%, of stillbirth and early neonatal death are found in LMIC settings [1]. Current guidelines states that a FHR in the range 110-160 beats per minute (bpm) during labour is considered normal [3, 4]. The use of continuous FHR monitoring devices in LMIC settings, may help the health care workers detect abnormalities in FHR at an earlier stage, allowing the health care personnel time to intervene before it is so too late.

In high income settings, continuous FHR monitoring is primarily used for high risk labours. This study, however, include labours assessed as low-risk with normal FHR on admission. This inclusion criteria gives us the possibility to study how FHR development for labours with a normal outcome differ from labours with an adverse outcome. The heart rate of newborns is earlier reported to increase shortly after birth [22]. But to the authors knowledge, a corresponding development of the FHR just before birth has not been studied previously. The use of an embedded system

analyzing the FHR development during labour, can potentially reduce the number of labours with a severe outcome by alerting the health care personnel if the FHR deviates from the expected trend.

This work is a part of the larger Safer Birth project <sup>1</sup>, a collaboration between multiple Norwegian and international research institutions as well as hospitals in Tanzania. The aim of Safer Births is to increase newborn survival by gaining new knowledge and developing innovative products to aid the health care workers. Among the data that has been collected through the Safer Births project, FHR from 3711 labours in Tanzania was collected using the Moyo fetal heart rate monitor <sup>2</sup>, a doppler based ultrasound sensor developed for continuous or intermittent FHR monitoring, useful also in LMIC settings.

The objective of this paper is to present a framework that can be used to analyze fetal heart rate collected continuously using Moyo (or potentially other FHR devices) and to use that on this unique data material to show how the FHR develops during labour for neonates with normal outcome, compared to neonates with abnormal or severe outcomes. Both the development as a function of time of the median heart rate and the spread is reported. Another objective is to study the maternal-fetal heart rate ambiguity for the Moyo device, and compare that to reported ambiguities using a traditional CTG.

## 10.2 Results

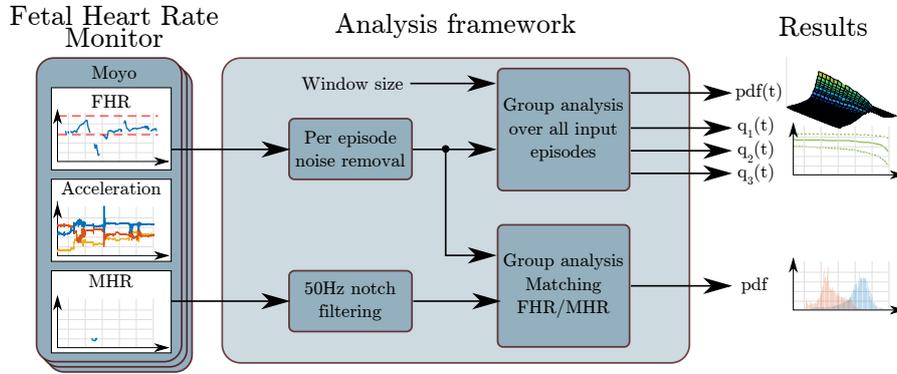
The proposed FHR analysis framework is shown in Figure 10.1. Each episode is first processed to remove noise, before group analyses over all inputs are conducted. Both MHR/FHR ambiguity and FHR development for different neonatal outcome groups are analysed. Further details of these steps can be found in the Methods section.

Noisy signal samples in the Moyo FHR was identified and removed by FhrClean [97]. For the following experiments, the data set were divided into subsets based on the newborn outcome 24 hours after birth. These outcomes were *Normal*, *still admitted to neonatal care unit (NCU)*, *very early neonatal death (VEND)* and *FSB*. An overview of the subsets are shown in table 10.1. Due to the low number of episodes in  $s_4$  and  $s_5$ ,

---

<sup>1</sup><http://www.saferbirths.com>

<sup>2</sup><https://laerdalglobalhealth.com/products/moyo-fetal-heart-rate-monitor/>



**Figure 10.1:** Signals acquired from the Moyo Fetal Heart Rate Monitor is used as the input for our analysis framework in this work. Noise is first removed on the FHR signals, before group analysis is conducted. The acceleration signal present in the Moyo has previously been shown to identify contractions [97], but it is not used in the work presented in this paper.

these subsets were combine to describe all episodes ending with death,  $s_3$ . The subsequent experiments on development of FHR trend were conducted using the cleaned version of the FHR signals. In experiment 3-5, the defined starting point,  $t_0$ , is set to -9000 seconds, equivalent of 150 minutes before birth.

### 10.2.1 Noise removal

The noise removal algorithm, FhrClean [97], was run on the entire data set. An example of the original- and cleaned FHR signal is shown in Figure 10.2.

The mean and standard deviation of the % of sample points that is removed as artefacts for each of the four outcome groups are shown in table 10.1.

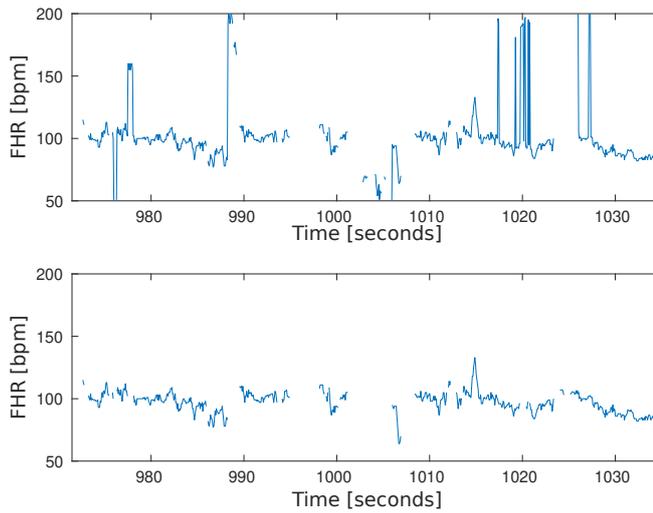
### 10.2.2 Experiment 1: Maternal heart rate

The MHR and FHR measurements are extracted from all episodes in the data set. The dry-electrode ECG sensor for MHR is used in 30.54% of the episodes in the data set. In these episodes, the MHR is measured in  $0.412 \pm 0.542$  percent of the episode duration.

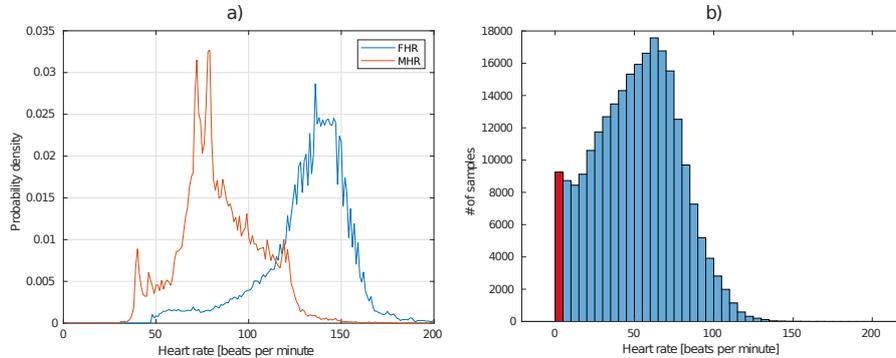
The distribution of the measured FHR and MHR in all time points where both values are measured are shown in Figure 10.3a. The red indicates

**Table 10.1:** Overview of data set and the removed noise from FHR signal in each labour outcome group.

Outcome	#episodes	Subset	Missing data [%]	Removed data [%]
All	3705	$s_0$		
Normal	3490	$s_1$	$27.83 \pm 19.87$	$1.79 \pm 1.35$
NCU	185	$s_2$	$31.28 \pm 20.38$	$1.82 \pm 1.31$
VEND	18	$s_4$	$29.22 \pm 24.34$	$1.31 \pm 0.83$
FSB	12	$s_5$	$40.50 \pm 28.60$	$1.92 \pm 1.61$
VEND/FSB	30	$s_3$		

**Figure 10.2:** Example of the result using FhrClean. The original FHR signal with noise on top, and the corresponding signal with noise removed, bottom.

measured MHR, and blue indicates the measured FHR. The absolute difference between the measured FHR and FHR at each time point is shown in Figure 10.3b. The red bar indicates the amount of MHR/FHR ambiguity. The MHR/FHR ambiguity in time points where both heart rates are measured is 4.53 percent. The similarity threshold,  $T_{mhr}$ , is set to 5 when computing the ambiguity, according to the study of Reinhard et al. [76].



**Figure 10.3:** a) Distribution of the MHR and FHR from all time points in the data set where both values are measured. b) The absolute difference between the corresponding pairs of MHR and FHR. The found MHR/FHR ambiguity of 4.53 percent is illustrated with the red bar.

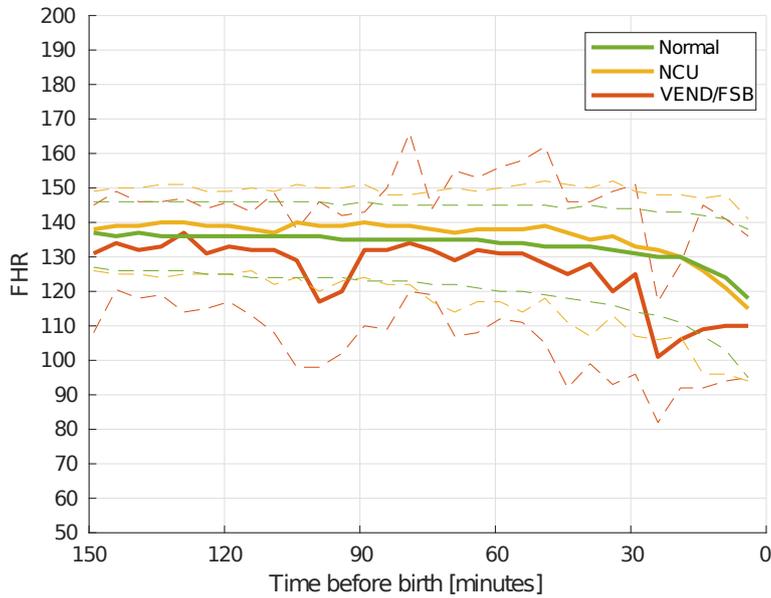
### 10.2.3 Experiment 2: Fetal Heart Rate Development

The FHR development is found as the median FHR,  $mFHR(t)$ , using non-overlapping intervals of fixed size. Intervals of 5 and 10 minute duration were used to obtain multiple resolutions of the heart rate trend in the period from 150 minutes before birth until the time of birth. An overview of  $mFHR(t)$  using 5- and 10-minute intervals are shown in Figures 10.4 and 10.5 respectively. The solid lines indicate  $mFHR(t)$ , and dashed lines the  $HR_{q_1}(t)$  and  $HR_{q_3}(t)$ . The green lines shows the trend for neonates with a normal outcome,  $s_1$ , yellow neonates still admitted to NCU,  $s_2$ , and red newborns either defined as fresh stillborn or died during the first 24-hours,  $s_3$ .

The found trend indicates a reduction in  $mFHR$  the last 30 minutes before birth, with the reduction for the VEND/FSB,  $s_3$ , group occurring longer time before before birth than for  $s_1$  and  $s_2$ .

### 10.2.4 Experiment 3: Fetal Heart Rate Distribution

An estimate of the probability density function (pdf) for all  $fhr(n)$  over all episodes in each sub group was found for the last 30-minutes before birth, and the two preceding 30-minute intervals, were computed to study how the estimated pdf of FHR changes before and after the drop in the  $mFHR$  observed in experiment 3. The estimated pdf for the last 90 to 60-minutes before birth are shown in Figure 10.6a, the last 60 to 30-minutes before



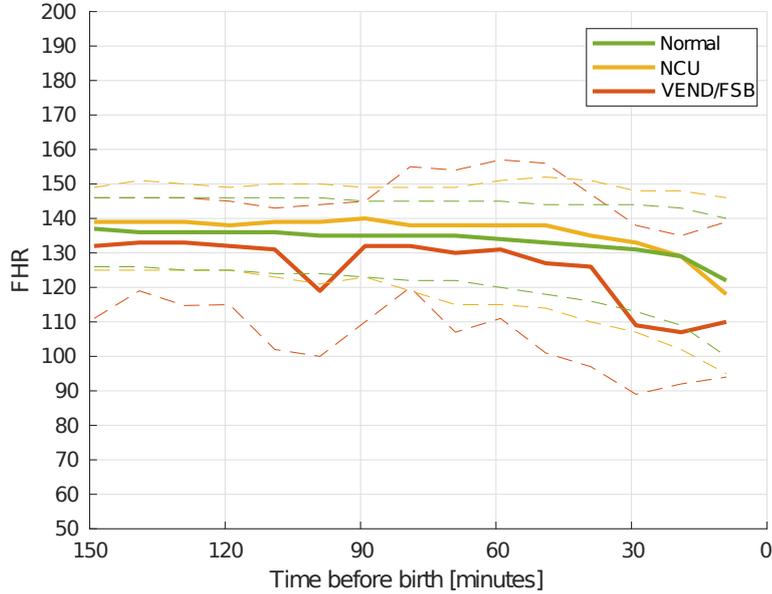
**Figure 10.4:** Trend of the FHR the last 150 minutes before birth. Solid lines indicate the median heart rate, and the matching dashed lines the 25 and 75 percentiles. Computed using non-overlapping 5-minute intervals.

birth in Figure 10.6b, and the final 30-minutes before birth in Figure 10.6c. The blue line indicates the estimated pdf for normal outcome, red line is the the NCU group and black for the VEND/FSB group.

90 to 60-minutes before birth, Figure 10.6a, all outcome groups have a peak in the 135-145 bpm region. 60 to 30-minutes before birth, a similar peak is found for the normal and NCU groups. The VEND/FSB group still has its peak at 135 bpm, although not as distinct, and the variance of the pdf is increased. In the last 30 minutes before birth, the pdf variance increases for both the normal group and the NCU group, but the peak stays within the same 135-145 bpm region. For the VEND/FSB group, the variance is high, and the peak has now shifted down to 110 bpm.

### 10.2.5 Experiment 4: Fetal Heart Rate Distribution Over Time

To increase the visual interpretability of how the trend and spread changes over time, an estimated pdf of the heart rate was computed using 10-minutes



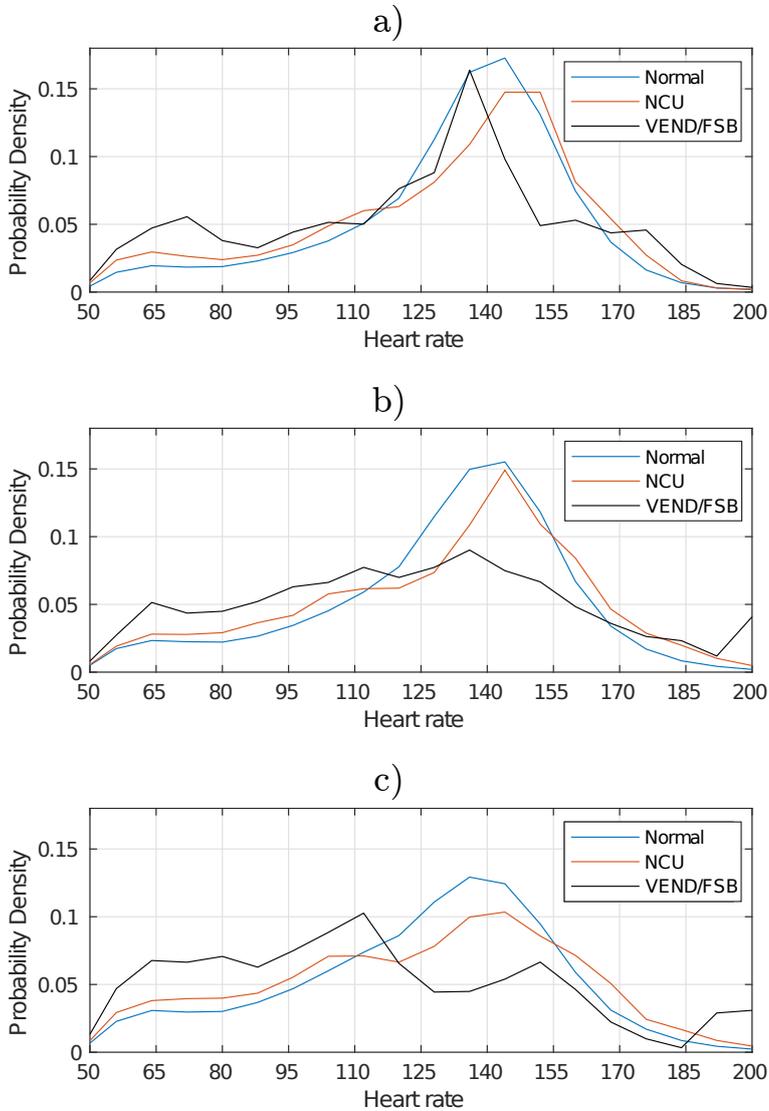
**Figure 10.5:** Trend of the FHR the last 150 minutes before birth. Solid lines indicate the median heart rate, and the matching dashed lines the 25 and 75 percentiles. Computed using non-overlapping 10-minute intervals.

non-overlapping intervals for the last 150 minutes before birth, plotted together in a 3D surface plot.

The estimated pdf of the heart rate over time for neonates defined as normal is shown in Figure 10.7, and for neonates still admitted to NCU after 24 hours in Figure 10.8. The red line indicate the number of episodes containing any measured FHR signal in the corresponding time interval. For neonates identified as normal, Figure 10.7, and still admitted to NCU, Figure 10.8, the variance increase closer to birth. The number of episodes contributing the analysis are, however, lower for the NCU group than the normal group. For the VEND/FSB group it is even smaller, and we have not included the pdf.

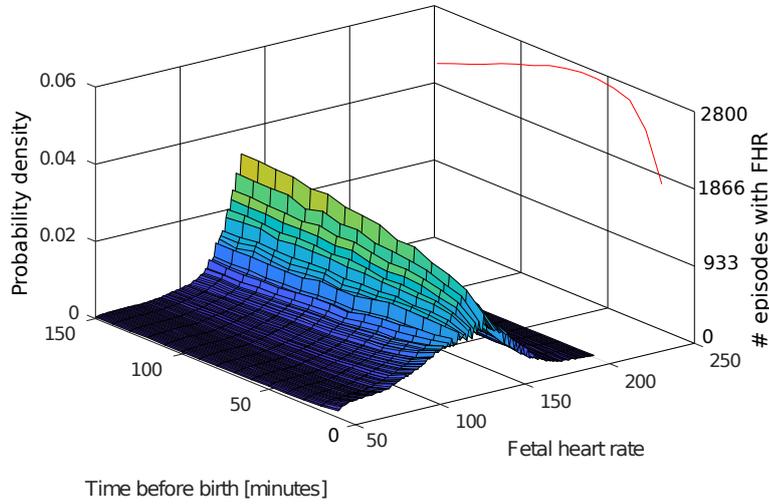
### 10.3 Discussion

All subgroups in the data set contain a relatively high percentage of missing data. A mean of approximately 30% missing data is seen in the normal, NCU and VEND groups, while episodes in the FSB group has a mean of

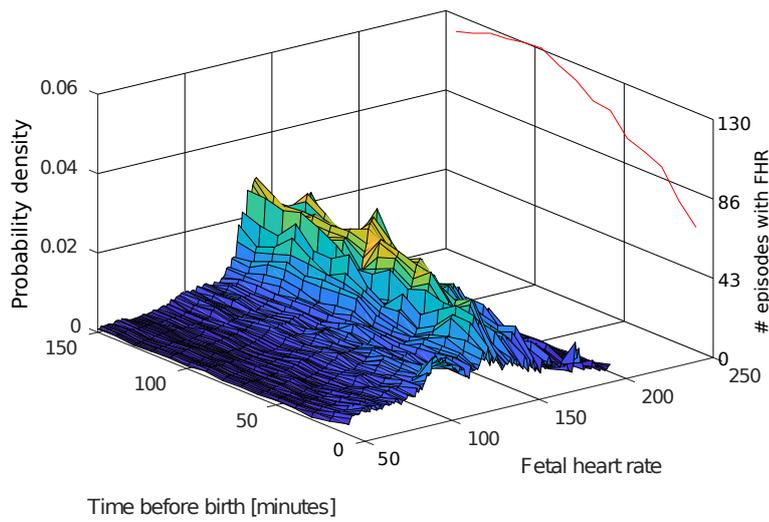


**Figure 10.6:** Estimated pdf of the measured heart rate the last 90 minutes before birth, divided in 30-minute windows. **a)** Estimated pdf for the last 90 to 60 minutes before birth. **b)** Estimated pdf for the last 60 to 30 minutes before birth. **c)** Estimated pdf for the last 30 minutes before birth.

40% missing data points. The spread is however large, with a standard deviation of approximately 20% for all groups. FhrClean [97] identifies and removes 1-2% of the data, further increasing the amount of missing data.



**Figure 10.7:** Estimated pdf changes over time, neonates defined as normal 24 hours after birth



**Figure 10.8:** Estimated pdf changes over time, neonates admitted to NCU 24 hours after birth

This removal is, however, desired as these data points are considered as noise or MHR. The missing data points may introduce an uncertainty in the analyses, but the impact is expected to be relatively small due to the large number of labours included in the study.

The major findings in this work can be summarized into five points:

- 1: MHR/FHR ambiguity stays within the same region observed on conventional CTG.
- 2a: A reduction in mFHR is observed for all subgroups close to birth.
- 2b: The reduction in mFHR for VEND/FSB is larger, and occurs longer time before birth.
- 3a: The variance of the estimated pdf increases for all subgroups close to birth.
- 3b: An larger increase in the variance, as well as shift in the peak is observed for VEND/FSB.

The MHR/FHR ambiguity found at time points where both measurements are available stays within the same region as observed by Reinhard et. al. [76]. Indicating that the Moyo Fetal Heart Rate monitor does not pick more of the MHR than conventional CTG.

The FHR development the last 150 minutes before birth, shown in Figure 10.5, shows a decrease in the measured heart rate for all defined subset in the study, finding 2a. Labours in the normal and the NCU subsets follow the same trend, where a small decrease can be seen from 40 minutes before birth and at the onset of a larger drop at 10-20 minutes before birth. For labours in the VEND/FSB subset, the onset of a larger decrease occurs already at 40 minutes before birth. The observed drop in the normal and NCU group is likely to be caused by an increased frequency and intensity of the uterine contractions as the labour progresses. The drop is larger and occur longer time before birth in the VEND/FSB group, finding 2b, may also be caused by the increased frequency and intensity of the uterine contractions, and may indicate that the fetus is unable to cope with this increased intensity.

By studying the difference in estimated pdf in 30-minute intervals during the last 90-minutes before birth, we observe that the variance increase for all subgroups as the labour progresses towards birth, finding 3a. The VEND/FSB has also a larger increase in the variance than the other two groups, and the shift in peak down to 110bpm in the last 30-minutes is also larger than for the other subgroups, finding 3b. As the estimated pdf for VEND/FSB in Figure 10.6a resembles the normal and NCU in Figure 10.6c,

lives could potentially been saved if interventions had been conducted at an earlier point in time.

By introducing the same 10-minute intervals from the FHR development experiment to the computation of the estimated pdf of the heart rate, we can see that the variance gradually increases, Figure 10.7. The probability for heart rates below 100 bpm also increases during the last 10-minute interval before birth. A similar trend is seen for labours where the neonate is still admitted to neonatal care unit 24-hours after birth, Figure 10.8. The jagged shape of the pdf in Figure 10.8 may be a result of less data available. A similar illustration for VEND/FSB is challenging due to the low amount of episodes with this outcome.

A limitation in this study is the low number of neonates in the VEND/FSB subset. In addition, not all recordings include data up until the time of birth, further reducing the amount of episodes included in the study. The drop in number of episodes with data close to birth can be seen on the red line in Figure 10.7 and 10.8. To increase the amount of data, a longer data collection period would be desired. Ideally, data collection from multiple hospitals would be used.

A second limitation of this work is that the internal clock in the Moyo monitor has shown to be drifting. If not calibrated often enough this can result in inaccuracies of up to 30 minutes in the logged time stamp. The result of this is that the heart rate presented zero to one minute before birth may in worst-case be recorded 30 minutes before birth for some episodes. In cases with drift in the opposite direction, a FHR may be present in the signal after the defined time of birth. Episodes with a measured FHR after the time of birth are corrected by adjusting the time of birth to the time of the last found FHR.

## 10.4 Conclusion

In this work, we first remove time periods with less trust-worthy signal from the measured FHR signal prior to further analysis. The dry-electrode ECG sensors on Moyo is used to intermittently measure the MHR in 30.5% of the episodes, and the observed MHR/FHR ambiguity of 4.53% is within the same area as previously reported on on CTG [76].

The heart rate of newborns have previously been reported to increase shortly after birth [22], and a corresponding drop in the measured FHR close to the time of birth is for the first time observed and reported in this

work. A drop in the FHR shortly before birth is observed for all outcome groups. However, for labours ending in very early neonatal death or fresh stillbirth, this drop in the measured FHR is larger, and it occurs longer time before the time of birth. The variance in the estimated FHR pdf increases for all subgroups as the labour get closer to the time of birth. Specifically, in the FSB/VEND group there is a higher increase, and the corresponding pdf peak also has a larger shift. This may indicate that these neonates struggle to endure the physical strain of the labour, and that an earlier intervention could potentially save lives.

Further investigations with more cases of VEND and FSB are required to validate this difference in the FHR development.

## 10.5 Methods

### 10.5.1 Data material

The data was collected at three hospitals in Tanzania between October 2015 and June 2018. Haydom Lutheran Hospital is located in a rural part of the country, while the Muhimbili National Hospital and Temeke Referral Hospital are located in Dar es Salaam. During the study period, data from 3711 labours were collected. At 24 hours after birth, 3490 neonates were defined as normal, 185 were still admitted to neonatal care unit, 18 died within the first 24 hours, and 12 were classified as fresh stillborn. 6 labours were not associated with any of the four outcomes above, and were therefore excluded from this work.

The data was collected using the Moyo Fetal Heart Rate Monitor, a small handheld device, illustrated in Figure 10.9. The monitor is developed by Laerdal Global Health <sup>3</sup> for FHR monitoring in LMIC settings.

#### Moyo Fetal Heart Rate Monitor

The Moyo Fetal Heart Rate Monitor measures FHR using a 9-crystal pulsed wave Doppler ultrasound sensor operating at a frequency of 1MHz and an intensity of less than 5mW/cm<sup>2</sup>. The heart rate is computed and logged twice per second, equivalent of a sampling rate of 2Hz. In addition, the device is equipped with an accelerometer, sampled at 50Hz, a temperature sensor and an dry-electrode ECG sensor for measurement of the maternal

<sup>3</sup><https://laerdalglobalhealth.com>

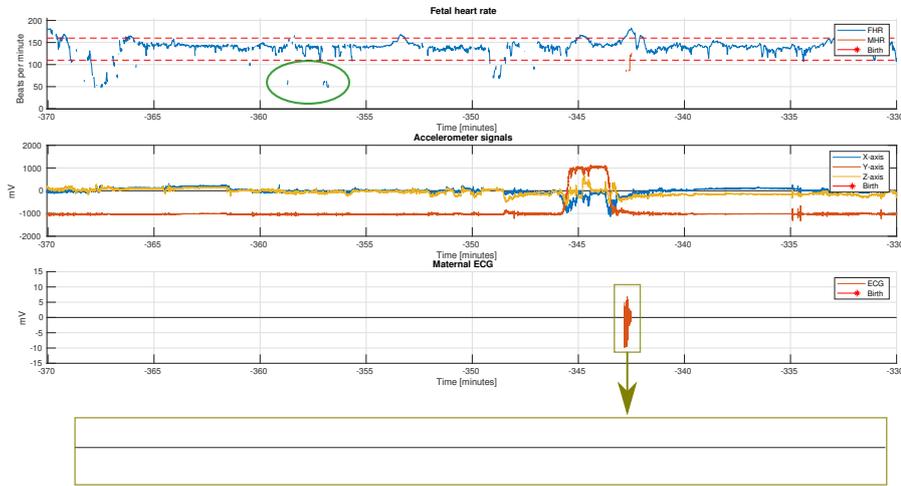


**Figure 10.9:** Moyo Fetal Heart Rate Monitor, Laerdal Global Health AS, Norway. Illustration reproduced with permission [49]

heart rate (MHR). The ECG sensor, used to measure the MHR requires the mother to keep one finger from each hand on the monitor. It is therefore suitable to intermittently assess the MHR, or to determine if the Doppler measurement captures the true FHR or if it falsely detects the MHR.

The Moyo FHR monitor is similar to conventional CTG with external Doppler for measurement of the FHR, but it lacks a sensor to detect uterine activity. To overcome this, an approach of using the accelerometer measurements from Moyo to estimate the uterine contractions has been proposed by our research group [97]. While measurements of the MHR is typically done using a separate device in high-resource settings, the inclusion of the ECG sensor is an advantage in LMIC settings as the availability of other devices may be limited. The small size of the Moyo also allows the mother to move more freely around while the device is attached. An overview of the differences and similarities between Moyo and conventional CTG can be seen in Table 10.2.

A segment from the signals collected using Moyo during a labour example episode is shown in Figure 10.10. The top plot shows FHR and MHR, in blue and red, in relation to the time of birth. The normal region, for FHR, of 110 – 160 bpm is indicated by red dashed lines. The second subplot shows movement of the sensor measured by the three axes accelerometer. The MHR is computed using the ECG signal shown in the third subplot.



**Figure 10.10:** Example signals from the Moyo fetal heart rate monitor. The top plot includes the FHR, shown in blue, and MHR, shown in red. The red dashed lines indicate the normal FHR region during labour. An example of the noise removed by FhrClean is illustrated in the green circle. The second plot shows the three axes measured from the accelerometer. The MHR is intermittently measured at -343 minutes. A zoomed version of the maternal ECG, with visible R-waves is shown on the bottom.

### 10.5.2 Pre-processing of data

Doppler measurements of the FHR is known to be noisy [98]. To allow for a computerized analysis of the FHR, the acquired measurements should first be preprocessed to remove unwanted artefacts, such as for example interference from the MHR [98, 99], misinterpreted harmonics [100]. The use of a portable device for FHR monitoring allows for more movement, thus potentially increasing the amount of noise. In conventional CTG, an internal transducer can be attached directly to the fetal scalp in case of poor signal quality from the Doppler sensor on the abdomen. However, this is not possible using the Moyo FHR monitor alone.

There are three distinct patterns of noise, i) short spikes with both higher and lower values than the baseline FHR, and ii) longer time periods deviating from the baseline FHR in a non-physical (impossible) way, iii) missing data points. Short spikes are relatively easy to detect, and examples can be seen the green circles in Figure 10.10 as short signal segments outside of the normal region of 110-160bpm, illustrated using the dashed red lines. Time periods with such noise in the measured FHR signal are identified using our recently proposed method for noise identification in

**Table 10.2:** Comparison between Moyo Fetal Heart Rate Monitor and conventional CTG

	Moyo	CTG	Comments
FHR	Doppler	Doppler	
Contractions	No*	Yes	*Contractions can be identified
Acceleration	Yes*	No	using accelerometer [97]
Moveable	Yes	No	
Maternal HR	Yes	No	

FHR signals [97]. The noise removal method, denoted FhrClean [97], is run on the complete data set prior to further analysis, as illustrated by the box in Figure 10.1. FhrClean first utilize forward- and backward replication to fill any missing data in the measured signal. Less trustworthy time periods, meaning periods where the measurement is not likely to consist of the true FHR, are found by identifying temporary drops or peaks in the signal. A more in-depth view of the used method can be found in Urdal et. al. [97].

### 10.5.3 Maternal heart rate

Doppler based FHR measurements are susceptible of incorrectly picking up the MHR due to sub-optimal sensor placement [74]. If the FHR is within 5 bpm of the MHR, it can be classified as an MHR/FHR ambiguity. Since the amount of MHR/FHR ambiguity in Doppler CTG is found to be  $1.22 \pm 1.9$  percent during the first stage of labour and  $6.2 \pm 9.0$  percent during the second stage or labour [76], it may cause unwanted artefacts in digital analysis of the FHR. The option of measuring MHR using Moyo is intended to be intermittent and not continuous, the amount of measured MHR varies from labour to labour. Thus the possibility of verifying whether the measured HR from the ultrasound Doppler is maternal or fetal is therefore limited. The MHR can also mimic an expected FHR, making it challenging to distinguish true MHR from true FHR signals [75].

To study the MHR/FHR ambiguity on data acquired using the Moyo Fetal Heart Rate Monitor, all time points where both signals exist are studied, indicated as the *Group analysis matching FHR/MHR* box in Figure 10.1. Let  $h_t$  be a vector of a FHR  $fhr(n_t)$  sample and a MHR sample  $mhr(m)$

$$h_t = [fhr(n_t), mhr(m_t)] \quad (10.1)$$

sampled with different sampling rate, here represented with a sample at the corresponding time point,  $t$ . Let  $H$  be the set of all such matching heart rate pairs,  $h_t$

$$H = \{h_t : fhr(n_t) > 0 \cap mhr(m_t) > 0\} \quad (10.2)$$

The MHR/FHR ambiguity,  $mhr_{amb} \in \{0, 1\}$ , in an episode is calculated as a fraction of the time where both signals are present, defined as

$$mhr_{amb} = \frac{1}{N_H} \sum_{h_t \in H} \mathbf{I}(h_t) \quad (10.3)$$

where  $N_H$  is the number of vectors,  $h_t$ , i.e. the number of matching time points, in  $H$ , and  $\mathbf{I}(h_t)$  is an indication function given by

$$\mathbf{I}(h_t) = \begin{cases} 1; & \text{if } |h_t(1) - h_t(2)| \leq T_{mhr} \\ 0; & \text{if } |h_t(1) - h_t(2)| > T_{mhr} \end{cases} \quad (10.4)$$

and  $T_{mhr}$  is a threshold to allow some inequalities due to the different measurement techniques.

#### 10.5.4 Fetal heart rate

Labour is normally a physical strain on both the mother and the fetus. As the labour progresses, this strain may affect the physical condition of the fetus, and can potentially be observed on the measured FHR. Analysing continuous FHR measurements from a large number of labours assessed as low-risk on admission, can potentially be used to determine if differences exist in the heart rate development between neonates with normal or adverse outcomes. As the time period where the FHR is measured vary from labour to labour, we define the sample index,  $n$ , in the measured FHR signal based on the measured FHR sample rate, 2Hz, the elapsed time,  $t$ , and a defined start point before birth,  $t_0$ , such as

$$n = 2(t + t_0) \quad (10.5)$$

In the following sections, we describe the *group analysis over all episodes* box in Figure 10.1, utilizing subsets of the data set based on neonatal status 24-hours after birth. The subset  $s_1$  includes all labours where the neonate was assessed as normal, the subset  $s_2$  include all labours where the neonate was still admitted to neonatal care unit (NCU) at 24 hours, and the subset  $s_3$  all episodes where the neonate in the VEND and FSB outcome groups. These two outcomes are grouped together in subset  $s_3$  due to the low number of episodes within each outcome.

### Fetal Heart Rate Development

With the use of continuous FHR monitoring in a large number of labours, it is possible to determine how the heart rate develops during labour and develop trend models. This can in turn be useful to determine how new labours progress compared to the known trend models.

The measured FHR within a defined interval,  $\Delta$ , in an episode,  $i$ , from the subset,  $s_k$ , is extracted from the start time,  $t$ , and throughout the duration,  $t + \Delta$ . Let the trend,  $mFHR_s(p)$ , be defined as the median of all measured heart rates in the interval, of all episodes in the subset,  $s$

$$mFHR_s(p) = median \left( \left( fhr_{1,s}^t(n), \dots, fhr_{L_s,s}^t(n) \right) \right) \quad n \in \{t, t + \Delta\} \quad (10.6)$$

Where  $L_s$  is the number of episodes in the subset  $s$ , and the sampling index  $p$  is given by

$$p = \frac{1}{\Delta}(t + t_0) \quad (10.7)$$

To describe the spread at each interval, the 1st and 3rd quartiles,  $q_1, q_3$ , called  $HR_{q_1}(t)$  and  $HR_{q_3}(t)$ , are computed using the concatenated vector of all FHR in the interval,  $\left( fhr_{1,s}^t(n), \dots, fhr_{L_k,s}^t(n) \right) \forall n \in \{t, t + \Delta\}$ .

### Fetal Heart Rate Distribution

To illustrate changes in the estimated probability density function (pdf) over time, we utilize a normalized histogram to estimate the pdf in an interval defined by the start point,  $t$  and end point  $t + \Delta$  for all episodes in a subset. When computing two or more distributions, these can be used to identify how the distribution changes over time. Let  $h_i^t(l)$  be the histogram of the measured FHR in episode  $i$ , in the interval with start point  $t$  and end point  $t + \Delta$ .

$$h_s^t(l) = \sum_{i \in s} h_i^t(l) \quad \forall \quad l = \{50, 51, \dots, 200\} \quad (10.8)$$

Where  $l$  indicates the histogram variable, heart rate from 50 to 200. The normalized histogram,  $\bar{h}_s(l)$ , is then given by

$$\bar{h}_s^t(l) = \frac{1}{N} h_s^t(l) \quad (10.9)$$

Where  $N$  is the total count in  $h_s^t(l)$

By combining multiple normalized histograms using continuous non-overlapping intervals, both the change in trend and spread of the FHR can be visualized simultaneously in a 3D surface plot. A peak in the computed histograms will result in a visible ridge in the 3D-visualization.

## Abbreviations

FHR: fetal heart rate; CTG: cardiotocography; mFHR: median Fetal Heart Rate; MHR: maternal heart rate; VEND: very early neonatal death; FSB: fresh stillbirth; NCU: neonatal care unit

## Acknowledgements

The authors would like to acknowledge the health care personnel Haydom Lutheran Hospital, Temeke Referral Hospital and Muhimbili National Hospital for all the work done at the labour wards.

## Authors' contributions

JU, KE, TE, HE designed the study; BK, PM, HK and SHH performed data collection and quality assurance. JU performed the analysis. JU drafted the manuscript, KE, TE, SHH, BK, PM, HK and HE revised the manuscript. All authors read and approved the final manuscript.

## Funding

This work is part of the Safer Births project which has received funding from Laerdal Foundation, Laerdal Global Health, Skattefunn, Norwegian Ministry of Education and USAID. The work was partly supported by the Research Council of Norway through the Global Health and Vaccination Programme (GLOBVAC) project number 228203.

## Availability of data and materials

The datasets analysed during the current study are not publicly available due National Tanzanian regulations, but may be available from the corresponding author on reasonable request.

## **Ethics approval and consent to participate**

The project was ethically approved prior to implementation by the National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434) and the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest).

## **Consent for publication**

The participants acknowledged their consent to publish the acquired data.

## **Competing interests**

Solveig Haukås Haaland is an employee of Laerdal Medical AS.



**Paper 4:  
Signal processing and  
classification for  
identification of clinically  
important parameters  
during neonatal  
resuscitation**



---

## Signal processing and classification for identification of clinically important parameters during neonatal resuscitation

J. Urdal<sup>1</sup>, K. Engan<sup>1</sup>, Trygve Eftestøl<sup>1</sup>, Hussein Kidanto<sup>2</sup>, Ladislaus Blacy Yarrot<sup>3</sup>, Joar Eilevstjønn<sup>4</sup>, Hege Ersdal<sup>5,6</sup>

<sup>1</sup> Department of Electrical Engineering and Computer Science, University of Stavanger, Norway

<sup>2</sup> Muhimbili University of Health and Allied Sciences, Tanzania

<sup>3</sup> Research Institute, Haydom Lutheran Hospital, Haydom, Manyara, Tanzania

<sup>4</sup> Strategic Research, Laerdal Medical AS, Stavanger, Norway

<sup>5</sup> Department of Anesthesiology and Intensive Care, Stavanger University Hospital, Norway

<sup>6</sup> Faculty of Health Sciences, University of Stavanger, Norway

**Published by the IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017.**

<https://doi.org/10.1109/ICSIPA.2017.8120672>

**Abstract:**

Neonatal mortality is a global challenge. One million newborns die each year within their first 24 hours as a result of complications during labour and birth asphyxia. Most of these deaths happen in low resource settings. However, basic resuscitation at birth can increase newborn survival. Identification of initial factors and simple therapeutic strategies determinant for neonatal outcome can aid health care workers provide the best follow-up during resuscitation. In this work, the initial condition of the newborn, the treatment given, and early heart rate response from manual bag mask ventilation are parameterized. The features are investigated in a machine learning framework to identify which features are determinant for the different outcomes. Using a selection of the defined features, an identification rate of 89% for newborns in the normal group, and an identification rate of 74% for episodes ending in death was found. This points to the direction of identifying the important factors of newborn survival.

## 11.1 Introduction

In 2012, 2.9 million newborns died worldwide within their first 28 days of life, counting for two out of every five deaths for children under the age of five. Of these, one million died within their first 24 hours of life [101]. The primary causes for these deaths are complications during labour, and birth asphyxia [101, 102]. An increased challenge is found in low resource settings, as the number of newborn deaths is four times as high in Africa as it is in Europe [101]. To reduce the newborn mortality, it is crucial to ensure that the optimal treatment is available and is provided during labour, delivery, and immediately after birth when the mortality risk is highest.

By identifying treatment factors determinant for the 24-hour neonatal outcome, better treatment and feedback solutions to guide the best therapy, can be sought. In the case where newborns are unable to start breathing by themselves, guidelines published by the World Health Organization and others [7, 103] recommend neonatal resuscitation should start within the first minute after birth, known as the golden minute [9]. While factors for prediction of an increased need for neonatal resuscitation [88], and the relationship between ventilation performance and response of newborns in term of apgar score [89] have been explored. There is still an uncertainty of which therapeutic strategies are determinant for outcome in neonatal resuscitation - what characterize a good resuscitation?

The Safer Births Project<sup>4</sup> is a research and development collaboration between Norwegian, Tanzanian, American and Irish research institutions collecting and analyzing data of neonatal resuscitations to find new information with the ultimate goal of improving the survival rate of newborns. In collaboration with the Safer Births project, multiple signals are measured during resuscitation of newborns at Haydom Lutheran Hospital in Tanzania including; ECG, acceleration (reflecting movement) of the newborn, and flow, pressure, and expired CO<sub>2</sub> during ventilation. In addition, initial conditions are manually observed and logged by research assistants continuously present in the labour ward.

Detection of physiological events associated with manual bag-mask ventilation during resuscitation have previously been proposed by our research group [84], utilizing signal processing on the physiological signals. Identification of therapeutic strategies is done using the ECG signal in combination

---

<sup>4</sup><http://www.saferbirths.com>

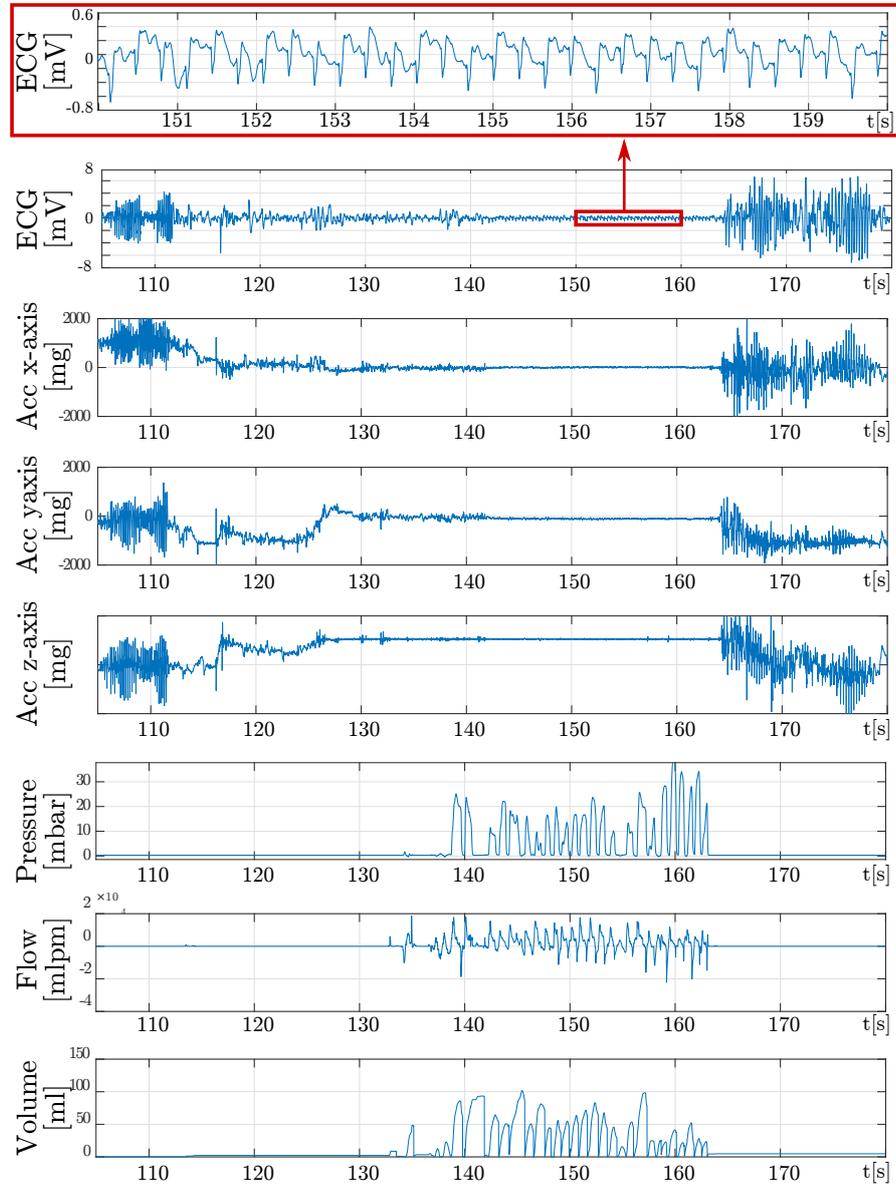
with the short-time energy of the acceleration signals of an accelerometer mounted together with the ECG sensor, proposed by our research group in [52, 85].

In this work, we attempt to identify vital and predictive factors for determining outcome during resuscitation of newborns from features derived from the given therapy and response using a machine learning framework. We define 14 features to characterize the initial condition of the newborn, the treatment given, and early heart rate response from manual bag mask ventilation. Evaluation of the effectiveness of the treatment is based on the neonatal status 24 hours after birth. The possible outcomes are defined as: normal, still admitted in neonatal care unit (NCU), and death. A wrapper based nested cross-validation classification scheme is used to identify the most important features for achieving high classification performance which can indicate that these features are important for the chance of survival. However, as the number of newborns surviving are expected to greatly outnumber newborns dying, classification can be challenging as traditional methods tend to create models featuring the largest class [43]. Multiple algorithms have previously been proposed to alleviate this challenge, including SMOTEBoost [45] and RUSBoost [42].

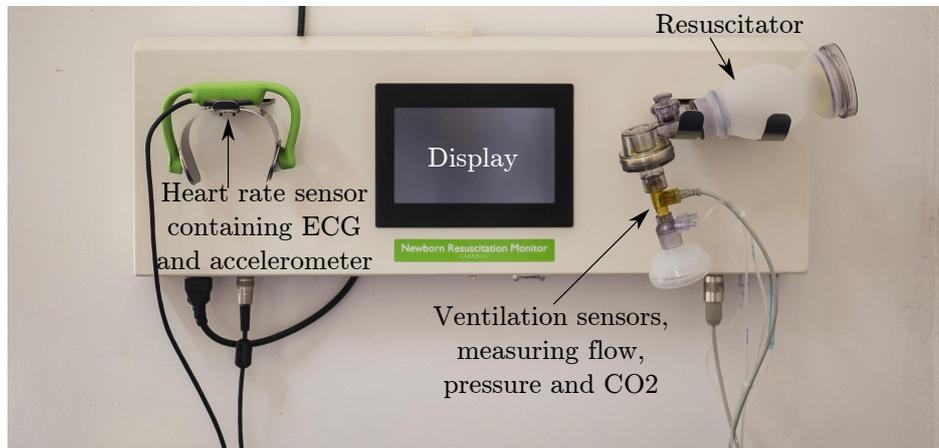
The study was supported by the Laerdal Foundation and the Research Council of Norway through the Global Health and Vaccination Program (GLOBVAC), project no. 228203.

## 11.2 Data material

The data set presented in this work is collected at Haydom Lutheran Hospital in collaboration with the Safer Births project. Haydom Lutheran Hospital is a resource limited hospital in rural Tanzania with a shortage of health care staff. Assistance during labour and treatment, including necessary resuscitations, of the newborns are primarily conducted by the midwives. All events at the labour ward are observed by trained research staff, logging predefined data on forms. Implementation of the research project was approved by the National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434) and the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest) before the start of the study.



**Figure 11.1:** Example signals of ECG (50 Hz filtered), acceleration in three axis, ventilation, pressure, flow, and volume (integrated from flow). A magnified section of the ECG is included to illustrate the dynamic range of measurements with little noise. An intervention from the health care workers are seen during the first 15 seconds, followed by a movement of the newborn. A ventilation sequence is seen from 135 to 165 seconds, followed immediately by a new intervention from the health care workers.



**Figure 11.2:** The Laerdal Newborn Resuscitation Monitor

The Laerdal Newborn Resuscitation Monitor (LNRM) is developed for research in a low resource setting where newborn resuscitation is usually performed by a single health care provider. Example signals measured by the LNRM is presented in Figure 11.1, and a photo of the LNRM can be seen in Figure 11.2. The monitor provides heart rate to the health care provider during a resuscitation as well as collecting data for research purposes. The heart rate sensor, green sensor shown in Figure 11.2, is placed across the abdomen or thorax of the newborn. The heart rate sensor contains dry-electrode ECG, sampled at 500 Hz, and a 3-axis accelerometer to monitor movement of the newborn during the resuscitation event, sampled at 100 Hz. An extra preprocessing step, using a 50 Hz notch filter, is used on the ECG signal. Additional sensors for monitoring pressure, flow, and expired CO<sub>2</sub> are mounted between the resuscitator bag and mask for research purposes. Pressure and flow are both sampled at 100 Hz, while CO<sub>2</sub> is sampled at 20 Hz. The measured signals and the corresponding time of each measurement is stored on the LNRM, ensuring synchronization between the various signals. The stored data is then transferred to a computer or external hard drive using the embedded USB interface after the resuscitation event is finished. The example signals in Figure 11.1 shows the ECG, acceleration, pressure, flow and volume (integrated from volume) measured using the LNRM. A magnified section of the ECG signal is included in Figure 11.1 to illustrate the dynamic range of measurements with little noise. The resuscitation monitors were installed in all labour

**Table 11.1:** Distribution of 24-hour outcome for the 530 episodes.

Outcome label	Normal	NCU	Death
# of patients	361	123	46

rooms as well as the operating theatre at Haydom Lutheran Hospital to measure data for research purposes during all resuscitation events.

A total of 916 resuscitation episodes were recorded between October 2013 and August 2016. Outcome at 24 hours included, 617 labelled as normal, 194 still admitted to NCU, 48 deaths, and 57 episodes classified as stillborn. Heart rate is however observed on 27 episodes identified as stillborn, these episodes are therefore reclassified and included in the study, thus the group with heart rate ending in death includes  $(48 + 27 =)75$  episodes. The data set was further reduced by only using episodes where all features used throughout this paper could be derived. Using this criteria, the data set was reduced to 530 episodes, where the distribution is shown in Table 11.1.

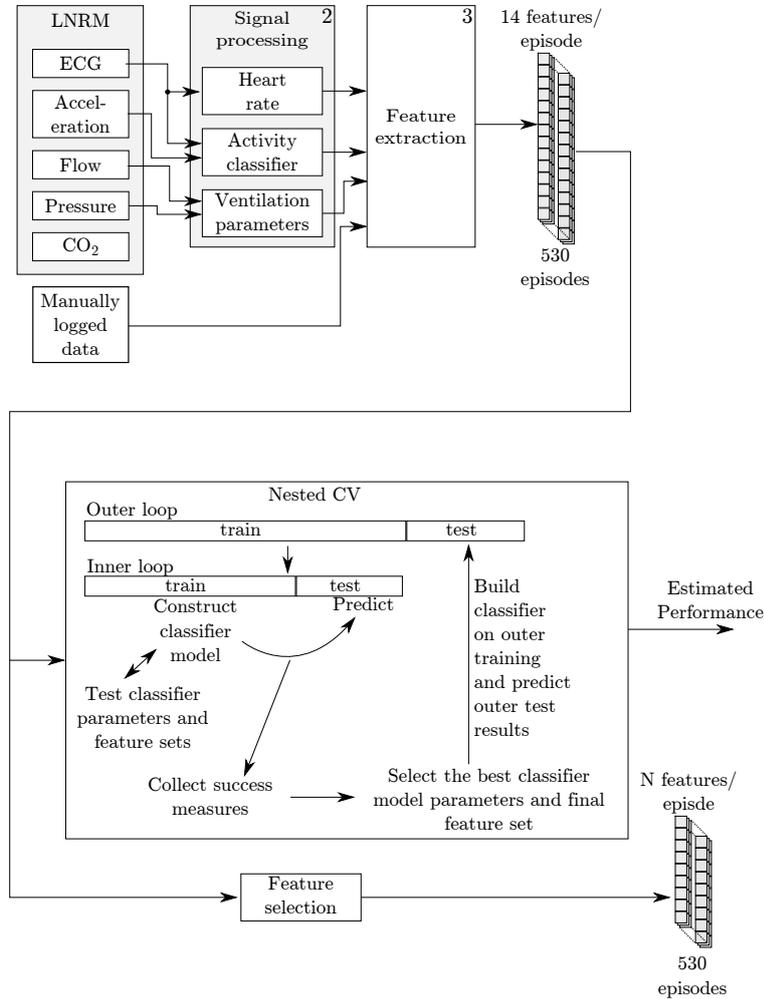
### 11.3 Proposed method

A block diagram of the proposed system is shown in Figure 11.3. Features are extracted from LNRM data as well as manually logged values during labour. The features are used in a nested cross-validation to estimate performance of a reduced feature set, and in a feature selection to identify the most vital features for determining neonatal outcome.

#### 11.3.1 Signal processing and feature extraction

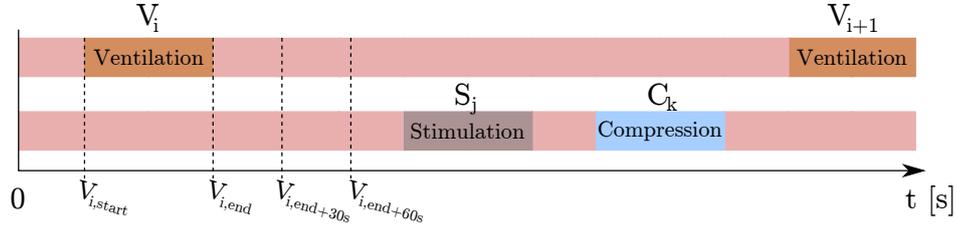
For parameterization of the resuscitation episode we distinguish between three feature categories: initial conditions, treatment, and early response parameters. Treatment parameters are further divided into two subgroups, those reflecting therapeutic strategies, like stimulation vs ventilation, and those reflecting the quality of the ventilations.

During the therapeutic strategy of ventilation, the actual quality of ventilations performed can vary. The quality of ventilations might be of crucial importance for neonatal asphyxia and thus have to be assessed. The flow and pressure signals measured by the LNRM are used to extract features defined by the work of this group in Vu et al. [84] where the details



**Figure 11.3:** Overview of the proposed system. Data from the Laerdal Newborn Resuscitation Monitor is used to compute heart rate, ventilation parameters and therapeutic strategies applied during the resuscitation episode. The computed parameters are used in combination with manually logged data as features in the nested cross-validation.

can be found. In short, ventilation events ( $v_l$ ) are defined by thresholding the pressure signals after baseline wander removal, expired volume is found by extracting start of inspiration and expiration periods from the flow signal, and integrating the flow signal (as seen in the volume signal of Figure 11.1) over the relevant period. An example of pressure, flow, and volume (integrated from flow) of a ventilation sequence is shown in the



**Figure 11.4:** Timeline representation of therapeutic strategies during resuscitation. When no activity is registered in either timeline, the activity is defined as hands off.

period from 135 to 165 seconds in Figure 11.1.

Identification of therapeutic strategies during the resuscitation episode is implemented (box 2 in Figure 11.3) by extracting features from the energy of the different axes in the accelerometer signals, shown in Figure 11.1, as well as energy of different subbands after a discrete wavelet decomposition of the the ECG signals, shown in Figure 11.1, from the LNRM. This identification is proposed and described by this research group in Vu et al. [52, 85] where more details can be found. The features are fed to a decision tree classifier using the activity classifier proposed by Vu et al. [85]. The activity classifier identifies therapeutic strategies in the following categories: ventilation, stimulation, suction, chest compressions, drying, tactile stimulation, moving heart rate sensor, hands off, and other, illustrated in the timeline of Figure 11.4. In this work we group compression, stimulation, tactile stimulation, and suction together as stimulation, and drying, moving sensor, other, and hands off together as hands off. In Figure 11.4,  $V_i$ ,  $S_j$  and  $C_k$  illustrate sequence  $i/j/k$  of the corresponding activity carried out by the health care provider.  $V_{i,start}$  defines the time when ventilation sequence  $i$  starts, and  $V_{i,end}$  the time when ventilation sequence  $i$  ends.  $V_{i,end+30s}$  and  $V_{i,end+60s}$  corresponds to the time 30 and 60 seconds after the ventilation sequence  $i$  ends. These time points are chosen to measure the heart rate response a fixed delay after the ventilation sequence.

Heart rate (HR) (box 2 in Figure 11.3) is computed using a median filter of instantaneous heart rate found from the ECG, shown in Figure 11.1. As intervention from health care workers occur during resuscitation, computation of the instantaneous heart rate is challenging due to the noise. An example of such noise can be observed in the period 100 to 115 seconds and 165 to 180 seconds in the ECG signal, shown in Figure 11.1. By observing the energy in the acceleration signals in the same time periods, it is reason-

able to assume that this noise is added as a result of intervention from the health care workers. A linear interpolation is therefore used to estimate heart rate between each computed heart rate point (with a resolution of 0.1 seconds). If no instantaneous heart rate is found in a 10 seconds period, the heart rate is considered difficult to estimate and is therefore set as undefined.

### Initial condition parameters

The following parameters were used to describe the initial condition of the newborn prior to bag-mask ventilation:

- Birth weight (BW) in grams.
- Gestational Age (GA) in weeks, estimated at birth.
- Time from birth to start of bag-mask ventilation ( $t_{BMW}$ ). This time is logged by research assistants continuously present during labour.
- Initial heart rate ( $hr_{V_i}$ ) is the heart rate when the first ventilation sequence starts ( $V_{1,start}$ ).

$$hr_{V_1} = HR(V_{1,start}) \quad (11.1)$$

### Treatment parameters

The following parameters are extracted from the sensor data:

- Total duration of the resuscitation episode ( $T$ ). Defined as the period from the start, until the end of the final stimulation or ventilation sequence.
- Stimulation time in percent ( $ST_{PRC}$ ) is the ratio of the time of all stimulation sequences ( $S_j$ ) over  $T$ .

$$ST_{PRC} = \frac{1}{T} \sum_i S_j \cdot 100\% \quad (11.2)$$

- Ventilation time in percent ( $VT_{PRC}$ ) is the ratio of the time of all ventilation sequences ( $V_i$ ) over  $T$ .

$$VT_{PRC} = \frac{1}{T} \sum_i V_i \cdot 100\% \quad (11.3)$$

- Hands off time in percent ( $\text{HOT}_{PRC}$ ) is the ratio of the time of all hands off sequences  $\text{HO}_k$  over  $T$ .

$$\text{HOT}_{PRC} = \frac{1}{T} \sum_i \text{HO}_k \cdot 100\% \quad (11.4)$$

- The total number of ventilations ( $n_V$ ) is the sum of all ventilation events ( $v_l$ ) in all  $V_i$ .

$$n_v = \sum_i \sum_l \#\{v_l \in V_i\} \quad (11.5)$$

- The expired volume ( $\text{expVol}_{ml/kg}$ ) is the median expired volume from all ventilation events ( $v_l$ ), divided by the newborns birth weight (BW) in kg.

$$\text{expVol}_{ml/kg} = \text{median} \left( \frac{\text{expVol}}{\text{BW}/1000} \right) \quad (11.6)$$

- The average ventilation rate ( $V_R$ ) is  $n_V$  over the total duration of all ventilation sequences in one episode.

$$V_R = \frac{n_V}{\sum_i V_i/60} \quad [\text{inflations}/\text{min}] \quad (11.7)$$

$\text{ST}_{PRC}$ ,  $\text{VT}_{PRC}$ , and  $\text{HOT}_{PRC}$ , reflect therapeutic strategies. Remaining treatment parameters are associated to quality of ventilations.

One can argue that the time from birth to start of bag-mask ventilation ( $t_{BMW}$ ) can be seen as both a treatment parameter and an initial condition for the bag-mask ventilation parameter. However, we have defined the treatment parameters as parameters possible to extract from the measured sensor signals during the bag-mask ventilation episode. Median expired volume can be seen as both a treatment and response parameter. In this work, we have chosen to label this feature as a treatment parameter, as it is an efficient way to identify the ventilation quality. A poor ventilation with severe mask leakage will result in a high measured inflated volume, and a low measured expired volume. A good ventilation will result in a sufficiently high expired volume.

### Early response parameters

The following parameters are extracted from the sensor data, and can be regarded as response parameters:

- Heart rate when first ventilation sequence ends  $hr_{V_{1,end}}$

$$hr_{V_{1,end}} = HR(V_{1,end}) \quad (11.8)$$

- Heart rate 30 seconds after first ventilation sequence  $hr_{V_{1,30s}}$

$$hr_{V_{1,30s}} = HR(V_{1,end} + 30) \quad (11.9)$$

- Heart rate 60 seconds after first ventilation sequence  $hr_{V_{1,60s}}$

$$hr_{V_{1,60s}} = HR(V_{1,end} + 60) \quad (11.10)$$

### 11.3.2 Classification

Due to the low number of episodes ending in death, there is a large class imbalance when trying to distinguish episodes in the normal from the group ending in death. To alleviate this imbalance, we have chosen to use the RUSBoost classifier. While this is a simpler and faster technique, it performs comparably to SMOTEBoost [42]. A nested cross-validation (CV) [104, 105] scheme is chosen instead of using a dedicated training and validation set. In the nested CV, the internal loop is used for feature selection and reduction of dimensionality of the feature set, using the external loop for validation. Feature selection is done using a wrapper method [90] with a forward selection approach, maximizing the true (positive+negative) rate. In the external loop, the data set is divided into three non-overlapping sections (folds) where 66% is used for training and the remaining 33% is used for validation. For each fold in the external loop, the internal loop is run on the training set with the same 66%/33% divide between training and test (3 folds). System performance is computed using the summed confusion matrices from all external folds. The same type of classifier is used in both the internal and external loops. Feature normalization is applied in each CV fold, where  $\mu$  and  $\sigma$  is found for the current training set and then applied to the normalization of the current test set. A result of using independent feature selection in each fold of the external loop, is that the used features might vary between the folds. Nested CV does, however,

give a measure of how well the model performs on the data set, where the model includes feature selection and the used classifier. The same feature selection is applied to the entire data set to identify the most crucial features corresponding to the performance found in the nested CV.

## 11.4 Experiments and results

**Table 11.2:** Performance using feature selection in nested CV. P = precision, R = recall. 1 for class Normal, 2 for class NCU, and 3 for class Death

	Exp 1			Exp 2			Exp 3		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
	R1	R2	R3	R1	R2	R3	R1	R2	R3
Normal/Death	0.96		0.47	0.88		0.55	0.91		0.54
	0.89		0.74	0.87		0.57	0.83		0.71
Normal/NCU	0.84	0.42		0.74	0.46		0.72	0.45	
	0.71	0.60		0.71	0.50		0.74	0.43	
NCU/Death		0.85	0.60		0.78	0.69		0.71	0.51
		0.85	0.59		0.86	0.57		0.78	0.43
Normal/NCU/ Death	0.77	0.26	0.31	0.73	0.39	0.41	0.70	0.43	0.35
	0.64	0.28	0.65	0.74	0.24	0.69	0.75	0.16	0.67

Identification of factors which can be utilized on all resuscitations is highly desirable. However, we acknowledge that newborns have a wide range of initial conditions, and that newborns in a poor initial condition may require a different therapy than newborns in a better condition. It was also considered interesting to study if the outcome could be described using only the initial status and the treatment given, as this could potentially improve identification of the best treatment factors. Three experiments were conducted and validated:

**Exp 1:** Identification of neonatal outcome 24 hours after birth using all features available

**Exp 2:** identification of neonatal outcome using all features for newborns initially in a poor condition

**Exp 3:** identification of neonatal outcome using only initial- and treatment parameters for newborns initially in a poor condition.

Each experiment was run for all four class combinations: Normal - Death, Normal - NCU, NCU - Death, and the three-class problem Normal - NCU - Death. To identify newborns in a poor condition, initial heart rate was chosen as an identifier. A threshold was then used to only study resuscitation episodes with an initial heart rate below 120.

**Table 11.3:** Mean and standard deviation of each class for all features

	Normal	NCU	Death
BW	3229.9±494.5	2905.7±675.8	2777.1±643.0
GA	38.2±1.7	37.0±2.8	36.9±2.9
$t_{BMW}$	125.1±67.8	136.8±82.9	142.8±106.5
$hr_{V_i}$	125.8±44.2	107.5±40.0	74.9±34.4
T	365.7±348.9	500.2±424.9	913.7±780.3
$ST_{PRC}$	24.7±14.4	19.2±12.7	12.5±10.4
$VT_{PRC}$	26.0±18.1	35.9±22.4	50.0±22.2
$HOT_{PRC}$	49.2±18.3	44.9±19.9	37.5±17.5
$n_V$	76.2±103.7	146.2±168.1	438.5±546.6
$expVol_{ml/kg}$	7.0±5.3	9.0±7.4	9.3±5.4
$V_R$	55.5±21.6	54.1±22.5	54.1±18.0
$hr_{V_{1,end}}$	130.8±42.6	113.3±40.4	78.9±39.4
$hr_{V_{1,30s}}$	142.8±37.0	131.1±34.2	87.8±44.9
$hr_{V_{1,60s}}$	148.9±34.0	138.7±35.2	91.5±50.0

Performance was evaluated using precision and recall, these performance metrics were chosen to keep the same evaluation parameters for both the two- and three-class problems. Precision is defined as the fraction of correctly classified examples over all examples classified as this class, and recall is defined as the fraction of correctly classified examples over all examples really belonging to this class [106]. The performance metrics were computed from the summed confusion matrices from all folds in the external loop in the nested CV. The use of only forward feature selection was chosen due to the computational requirement of nested CV. The approach does however, have its limitations, with the possibility of only finding a local maximum. Results for identifying the various outcomes in all three experiments are shown in Table 11.2. P denotes precision and R denotes recall. 1 is used for class normal, 2 for class NCU, and 3 for class death. An overview of mean and standard deviation of the features are given in Table 11.3.

**Table 11.4:** The selected features using all available data and all features (Exp 1).

Normal/Death	Normal/NCU	NCU/Death	Normal/NCU/Death
$hr_{V_{1,60s}}$	$n_V$	$hr_{V_{1,60s}}$	$n_V$
T	GA	BW	$hr_{V_{1,30s}}$
$hr_{V_{1,end}}$	T	T	$V_R$
$expVol_{ml/kg}$	$hr_{V_{1,60s}}$	$hr_{V_{1,30s}}$	T
$n_V$	$ST_{PRC}$	$hr_{V_i}$	$HOT_{PRC}$
BW	$hr_{V_{1,30s}}$	$expVol_{ml/kg}$	$VT_{PRC}$
$hr_{V_i}$	$t_{BMW}$	$HOT_{PRC}$	
$t_{BMW}$	BW	$t_{BMW}$	
$VT_{PRC}$	$hr_{V_{1,end}}$	GA	
$hr_{V_{1,30s}}$	$VT_{PRC}$	$V_R$	
	$expVol_{ml/kg}$	$hr_{V_{1,end}}$	
		$n_V$	

A corresponding feature selection was done on each experiment using the entire data set in combination with the same forward feature selection scheme as used in the previous experiments. The selected feature order is shown in Table 11.4 using all features, in Table 11.5 using all features for newborns initially in a poor condition, and in Table 11.6 using only initial- and treatment parameters for newborns initially in a poor condition.

## 11.5 Discussion

In Table 11.3, a mean expired volume of 9 ml/kg can be seen for both NCU and death, higher than episodes with a normal outcome. In combination with the increased number of ventilations given for NCU and Death, we assume that the health care workers perform a high number of ventilations of sufficient quality to the newborns in need. The resuscitation duration shows an expected increase from the normal class to both NCU and Death. The therapeutic strategies show a shift in focus for the health care providers for newborns in a critical condition. While ventilations are given 26% of

**Table 11.5:** The selected features using all defined features for newborns with an initial heart rate below 120 (Exp 2).

Normal/Death	Normal/NCU	NCU/Death	Normal/NCU/Death
$hr_{V_{1,60s}}$	$VT_{PRC}$	$hr_{V_{1,60s}}$	$hr_{V_{1,60s}}$
T	GA	BW	$VT_{PRC}$
BW	$HOT_{PRC}$	$hr_{V_{1,30s}}$	$hr_{V_i}$
$expVol_{ml/kg}$	$n_V$	$n_V$	$hr_{V_{1,end}}$
$V_R$	$hr_{V_{1,30s}}$		$hr_{V_{1,30s}}$
$hr_{V_i}$	$ST_{PRC}$		$HOT_{PRC}$
$t_{BMW}$	$t_{BMW}$		GA
GA			BW
$VT_{PRC}$			$t_{BMW}$
$HOT_{PRC}$			
$ST_{PRC}$			
$hr_{V_{1,30s}}$			

the resuscitation period for newborns in the normal class, an increased focus on ventilations is observed in the NCU and Death classes, where 50% of the treatment duration is used for ventilations in the latter. A reduction in stimulation is also observed from normal to NCU and death. This reduction is not necessarily a critical factor, as stimulation alone often will be insufficient treatment for newborns in a poor initial condition.

For the two-class problems, Table 11.2, a precision of 96% and a recall of 89% for normal, and a precision of 47% and a recall of 74% for episodes ending in death were achieved (Exp 1), indicating that the features can be used to describe the 24-hour outcome. Exp 3, Table 11.2, achieves a similar results, while Exp 2, Table 11.2, achieves a lower performance. This could be a local maximum due to the forward feature selection, as Exp 2 contains all features used in Exp 3 and should therefore be able to achieve similar performance. The high precision in one class and low precision in the other class, can be explained due to the large class imbalance. In the case of Exp 1 for normal vs death, the 11% misclassification of the normal class (38 episodes), is almost the same number as the entire death

**Table 11.6:** The selected features using initial conditions and treatment parameters for newborns with an initial heart rate below 120 (Exp 3).

Normal/Death	Normal/NCU	NCU/Death	Normal/NCU/Death
$n_V$	$VT_{PRC}$	$ST_{PRC}$	$n_V$
$VT_{PRC}$	BW	$hr_{V_i}$	$t_{BMW}$
$HOT_{PRC}$	GA	T	T
$V_R$	$n_V$	BW	$ST_{PRC}$
GA	T	$\text{expVol}_{\text{ml/kg}}$	$VT_{PRC}$
$ST_{PRC}$	$ST_{PRC}$	$HOT_{PRC}$	GA
	$hr_{V_i}$	$V_R$	
	$V_R$	$t_{BMW}$	
	$t_{BMW}$	$VT_{PRC}$	

class (46 episodes), thus a low precision for the smaller class is obtained. Distinguishing between episodes in normal and NCU achieves a recall above 70% for the normal class in all three experiments, the NCU class is however close to random guessing. Classifying NCU vs death achieve similar performance both in Exp 1 and Exp 2. Studying the feature selection, two of the early response features are chosen in Exp 1 and Exp 2, indicating that *early response parameters are crucial for distinguishing these two classes*. In the three-class problem, identification of NCU is low in all three experiments. This is considered a consequence of the class temporary state, as most episodes in this class will eventually change into normal or death. A second challenge is the large standard deviation of all features while the feature mean difference is fairly low between the three outcomes.

Using all 14 features, heart rate after the first ventilation, number of ventilations, total resuscitation time, ventilation time, and hands off time percentage are found to be good identifiers of resuscitation outcome. For identification of neonatal outcome of episodes with an initial heart rate below 120, heart rate, ventilation time, and hands off time percentage are found to be good identifiers. Total resuscitation duration is, however, only found as a good feature for one class combination. When studying newborns with an initial heart rate below 120 using only initial condition and treatment parameters, *identification of all class combinations includes*

*at least two out of three parameters characterizing the therapeutic strategies.* This tells us that the therapeutic strategies are crucial for the newborn survival.

## **11.6 Conclusion and future work**

The results presented in this paper suggest that parameters describing the initial status, therapeutic strategies, quality of ventilations, and early response parameters are crucial factors for distinguishing the 24-hour outcome for newborns identified as normal or dead at this point in time. With the inclusion of newborns still in neonatal care unit 24 hours after birth, the identification rate of all outcomes are reduced, indicating additional features are required.

To provide additional information to the health care workers, the work has to be extended to find critical features for identifying the newborns who are alive, but still in a critical condition 24 hours after birth.

**Paper 5:**  
**Automatic identification of  
stimulation activities during  
newborn resuscitation using  
ECG and accelerometer  
signals**



---

## **Automatic identification of stimulation activities during newborn resuscitation using ECG and accelerometer signals**

**J. Urdal<sup>1</sup>, K. Engan<sup>1</sup>, Trygve Eftestøl<sup>1</sup>, Valery Naranjo<sup>2</sup>, Ingunn Anda Haug<sup>3</sup>, Anita Yeconia<sup>4</sup>, Kidanto Hussein<sup>5</sup>, Hege Ersdal<sup>6,7</sup>**

<sup>1</sup> Department of Electrical Engineering and Computer Science, University of Stavanger, Norway

<sup>2</sup> Instituto de Investigación e Innovación en Bioingeniería (I3B), Universitat Politècnica de València, Valencia, Spain

<sup>3</sup> Research Institute, Haydom Lutheran Hospital, Haydom, Manyara, Tanzania

<sup>4</sup> School of Medicine, Aga Khan University, Dar es Salaam, Tanzania

<sup>5</sup> Department of Anesthesiology and Intensive Care, Stavanger University Hospital, Norway

<sup>6</sup> Faculty of Health Sciences, University of Stavanger, Norway

**Published in Computer Methods and Programs in Biomedicine, Elsevier.**

<https://doi.org/10.1016/j.cmpb.2020.105445>

**Abstract:**

*Background and Objective:* Early neonatal death is a worldwide challenge with 1 million newborn deaths every year. The primary cause of these deaths are complications during labour and birth asphyxia. The majority of these newborns could have been saved with adequate resuscitation at birth. Newborn resuscitation guidelines recommend immediate drying, stimulation, suctioning if indicated, and ventilation of non-breathing newborns. A system that will automatically detect and extract time periods where different resuscitation activities are performed, would be highly beneficial to evaluate what resuscitation activities that are improving the state of the newborn, and if current guidelines are good and if they are followed. The potential effects of especially stimulation are not very well documented as it has been difficult to investigate through observations. In this paper the main objective is to identify stimulation activities, regardless if the state of the newborn is changed or not, and produce timelines of the resuscitation episode with the identified stimulations. *Methods:* Data is collected by utilizing a new heart rate device, NeoBeat, with dry-electrode ECG and accelerometer sensors placed on the abdomen of the newborn. We propose a method, NBstim, based on time domain and frequency domain features from the accelerometer signals and ECG signals from NeoBeat, to detect time periods of stimulation. NBstim use causal features from a gliding window of the signals, thus it can potentially be used in future realtime systems. A high performing feature subset is found using feature selection. System performance is computed using a leave-one-out cross-validation and compared with manual annotations. *Results:* The system achieves an overall accuracy of 90.3% when identifying regions with stimulation activities. *Conclusion:* The performance indicates that the proposed NBstim, used with signals from the NeoBeat can be used to determine when stimulation is performed. The provided activity timelines, in combination with the status of the newborn, for example the heart rate, at different time points, can be studied further to investigate both the time spent and the effect of different newborn resuscitation parameters.

## 12.1 Introduction

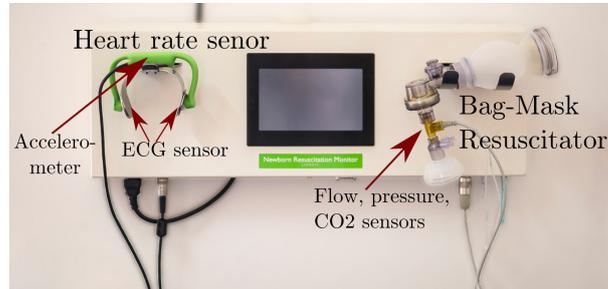
Early neonatal death is a worldwide challenge with 1 million newborn deaths every year, and the vast majority of these are found in low and low-middle income countries [101]. The primary cause of these deaths are complications during labour and birth asphyxia [101, 102]. Guidelines on newborn resuscitation are published by both the World Health Organization and others [7, 8]. The general guideline is to start resuscitation within the first minute after birth if the newborn is unable to start breathing [9]. A gap between the medical guidelines and what is actually performed in practice has also been observed [10]. While resuscitation immediately after birth is a crucial part of saving these lives, the full understanding of how to best apply therapeutic activities is not well documented. Therapeutic activities includes stimulation of the newborn, like firmly rubbing the back and drying, removal of mucus and obstructions in the airways by suction, and bag mask ventilation. The amount of activities performed during resuscitation, such as tactile stimulation and bag-mask ventilation, has been shown to be correlated with the 24-hour outcome of the newborn [107]. Further analyses should be conducted to study the importance of factors like duration and order of these therapeutic activities.

Safer Births<sup>5</sup> is a large and collaborative research project with the goal of establishing new knowledge and develop new innovative products to save lives at birth. One of the goals of this collaborative project is to construct a system that can automatically detect time periods where different resuscitation activities are performed. Such a system can be used as part of a debriefing system, and will make it possible to evaluate a large number of episodes to find out which resuscitation activities that are improving the state of the newborn, if current guidelines are good and if they are followed. The state of the newborn can effectively be evaluated by assessing the heart rate [8], and a change in the observed heart rate may be the result of prolonged resuscitation activities. There might as well be potential for real time decision support during resuscitation. A number of sensor data have been collected during newborn resuscitation at partner hospitals in Tanzania during the research project; pressure and flow from the bag-mask resuscitator (BMR), dry-electrode electrocardiogram (ECG) signals and signals from an accelerometer using a prototype of the NeoBeat<sup>6</sup>, attached over the abdomen of the newborn.

---

<sup>5</sup><http://www.saferbirths.com/>

<sup>6</sup><https://laerdalglobalhealth.com/products/neobeat-newborn-heart-rate-meter/>

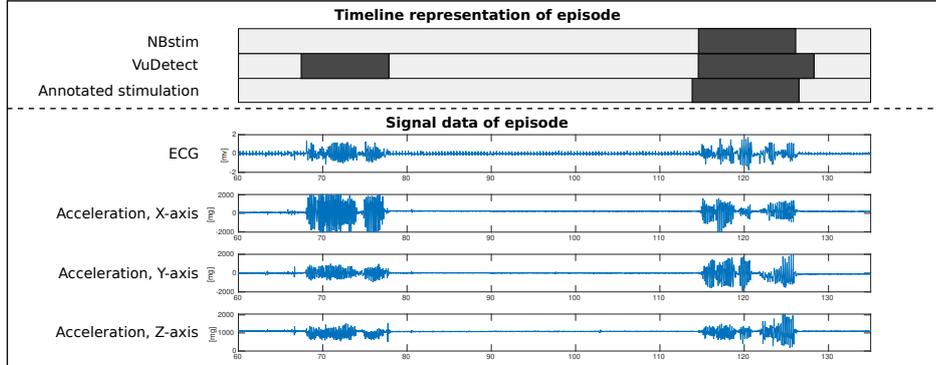


**Figure 12.1:** Laerdal Newborn Resuscitation Monitor with the various sensors indicated. The measured heart rate is shown on the LCD to give feedback to the health care personnel. The green buckle with accelerometer and dry-electrode ECG is a prototype of the NeoBeat,

The detection of bag-mask ventilation can relatively easily be performed using the flow and pressure signals from sensors mounted in the BMR [84], and detection and recognition of treatment activities during newborn resuscitation using deep neural networks on videos of resuscitation [86] has been described in earlier work from this research group. Videos of the resuscitation is often not available, or the view can be blocked by some of the activities. Thus it would be beneficial to be able to detect *stimulation* based on the NeoBeat signals. As tactile stimulation during newborn resuscitation involves some kind of repetitive movement, we hypothesize that these activities can be picked up using an accelerometer attached to the newborn through the use of NeoBeat.

Detection and recognition of activities using data from an accelerometer have previously been explored on healthy adults: static and dynamic activities such as sitting positions, walking versus running were found using an accelerometer mounted on the subjects back [81] and the subjects waist [82]. With the rise of wearable technology in every day life such as sport watches and cell phones, accelerometers are now available for activity recognition using commercially available devices [83]. To the authors knowledge, there are no reported correlation between the ECG morphology [108] and external stimulation, and no reported works utilizing accelerometer and ECG signals of newborns to automatically classify therapeutic activities, except from this research group.

Automatic detection of some sort of activity performed by the health care personnel, VuDetector, has previously been proposed by members of our research group [52]. VuDetector achieved a sensitivity of 90% and



**Figure 12.2:** Example segment of the measured ECG and accelerometer signals with the corresponding timeline representation. The signal with a lower amplitude in the beginning of the ECG signal corresponds to the expected QRS-complex. Two distinct regions with movement can be seen in the signals, and VuDetect identifies both activities. NBstim, proposed in this work, evaluates the regions found by VuDetect and classify them as stimulation or non-stimulation activities.

a specificity of 80%. During a resuscitation, the newborn will be moved, covered and uncovered etc. and such activities will also be visible on the accelerometer signals from the attached NeoBeat, but are not considered therapeutic activities. In VuClassifier [85], we also proposed a first attempt of classification of the detected activities based on ECG and accelerometer signals, reported with an accuracy of 79.8 % when distinguishing stimulation and chest compression from other activities. The VuClassifier was, however, based on signal features extracted from detected activity events of variable duration, it needed statistics from the entire resuscitation episode, and as such, only suitable for retrospective analyses, and it was trained and tested on relatively few episodes, and needed further verification.

In this work the main objective is to propose a system, NBstim, for detecting time periods of *stimulation activities* based on the signals recorded by the NeoBeat placed on the abdomen of the newborn, using causal signal features, and as such, suitable for real time analysis. In combination with a method of detected bag-mask ventilation sequences [84], this can be used to create useful timelines illustrating the amount, duration and order of ventilation and stimulation performed in real world newborn resuscitation episodes, see figure 12.2. In the rest of the paper we start by explaining the data material and the manual annotations in section 12.2. In Section 12.3 the proposed NBstim is explain in context with the larger system, thereafter

the signal features are defined. In the experimental section, VuClassifier and NBstim is tested using a larger dataset and compared with manually annotated activities.

## 12.2 Data material

The data material used in this work was collected at Haydom Lutheran Hospital (HLH), a rural hospital in Tanzania, between October 2013 and September 2016 by the Safer Births project. The research project was approved by the Regional Committee for Medical and Health Research Ethics (REK) in Norway (2013/110/REK vest) and National Institute for Medical Research (NIMR) in Tanzania (NIMR/HQ/R.8a/Vol. IX/1434). Parenteral verbal consent was obtained for all resuscitated newborns. Within this research project, subprojects have been subject to randomized trials. For this particular work, the data collection has been part of an observational study, not an intervention study.

The data were collected using the Laerdal NeoBeat prototype, Figure 12.1, which is part of the research device Laerdal Newborn Resuscitation Monitor [52, 84, 85, 89]. The NeoBeat prototype measures the heart rate using two dry-electrode ECG sensors attached to a buckle, which is placed over the abdomen of the newborn. This design allows the health care personnel to quickly attach the ECG sensor to the newborn and monitor the heart rate, and can therefore focus on giving the best treatment possible without struggling with gel and placement of the ECG sensors. An example of ECG and accelerometer signals measured using the NeoBeat prototype is shown in Figure 12.2. Due to the combination of dry-electrode ECG sensors and an environment with a lot of movement, the measured signal contains more noise than what is seen when using traditional ECG in settings with less movement. In HLH, a resuscitation monitor is installed in each of the labour rooms and the midwives are primarily responsible for the health care both during labours and potential resuscitation immediately after birth. The health care workers involved in the data collection were trained to follow the existing Helping Babies Breathe (HBB) guidelines<sup>7</sup> for newborn resuscitation. These guidelines state what should be checked, and what action to perform if the newborn is asphyxiated and need help to start breathing. The guidelines were posted on the wall above each

<sup>7</sup><https://shop.aap.org/helping-babies-breathe-2nd-ed-action-plan-wall-poster-paperback/>

resuscitation bed to remind the health care personnel to follow them. A limitation of the guidelines is, however, that they only defines what activity to perform, and not the amount, length, or how often the activity should be performed. Additional clinical data related to the labour and resuscitation was logged by designated research assistants present at the labour ward for the research project.

The resuscitation monitor consists of a main processing unit with a display to show the measured heart rate, as well as the heart rate sensor and a bag-mask resuscitator (BMR). The NeoBeat prototype seen in green in Figure 12.1, contains dry-electrode ECG, sampled at 500Hz, and a three-axes accelerometer to monitor movement of the newborn, sampled at 100Hz. The ventilation bag includes pressure and flow sensors, sampled at 100Hz, as well as a  $CO_2$  sensor sampled at 20Hz.

A total of 916 resuscitation episodes were recorded during the data collection period. A set of 76 randomly selected videos were annotated to obtain a timeline description of the resuscitation for further evaluation.

### 12.2.1 Annotations

Videos of the resuscitation were annotated by two independent reviewers; one neonatologist and one human factors engineer. In cases with agreement score  $< 80\%$ , the two reviewers sat together and obtained consensus. The following categories were annotated: 1) *stimulation*, 2) *suction*, 3) *uncovered*, 4) *other*, 5) *obscured view*, and 6) *start/stop of resuscitation*. If the resuscitation lasted longer than seven minutes, only the first seven minutes were annotated. Stimulation and suction are considered two of the three primary treatment events performed during resuscitation in addition to ventilation of the newborn. *Uncovered* describes how much of the newborn covered by a blanket, this is considered an important information, as covering more of the newborn will result in a lower heat loss. The fourth category is all other activities that are considered as relevant for the treatment. This can for example be clamping of umbilical cord and injections.

The heart rate sensor is sometimes detached and later reattached during a resuscitation episode. As this will contribute to artefact's and missing data in the dataset, the author has manually annotated attachment of the heart rate sensor. Only time regions where the heart rate sensor is fully attached to the newborn will be used in the analysis.

**Table 12.1:** Overview of the data subsets used in experiments. All subsets is based on 74 episodes of newborn resuscitation.

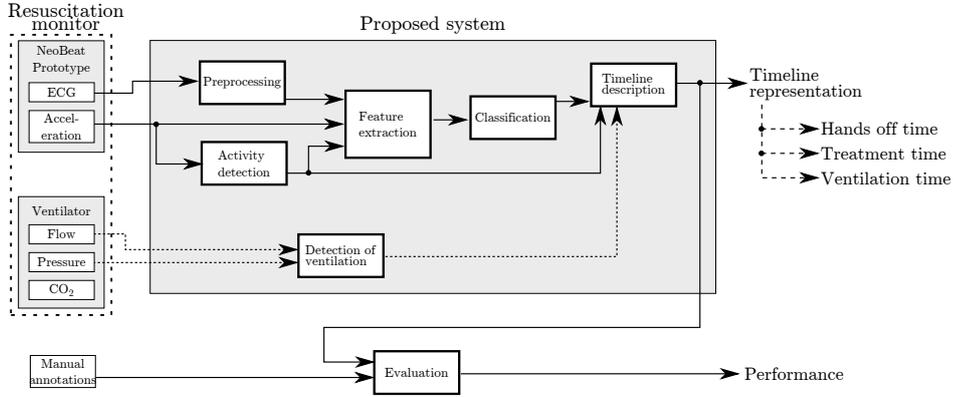
Data set	Method	Inclusion criteria	Total duration of data set
<i>D1</i>	Sliding window	Full episodes	21830 seconds
<i>D2</i>	Sliding window	VuDetect and manually annotated stimulation or hold off	1961 seconds
<i>D3</i>	Blocks of variable size	VuDetect and manually annotated stimulation or hold off	1961 seconds

### 12.2.2 Dataset

Two episodes were excluded due to corrupted data. The dataset used in this work therefore consists of 74 episodes of newborn resuscitation, called *D1* in this paper.

The example signals, seen in Figure 12.2, include two distinct areas where some activity clearly is happening. VuDetector identifies both areas in the timeline representation, shown as *detected activity*. The manual annotations, does however only recognize one of the regions as a stimulation activity.

Newborn resuscitation often involves multiple health care providers, resulting in multiple activities being performed at the same point in time. A data subset is created to study regions where only one activity is being performed on the newborn. This data subset, *D2*, consists of regions in the resuscitation where VuDetector has identified an activity, and where either only stimulation or no therapeutic activity are manually annotated. As the detected activity regions will not overlap perfectly with the annotated data, non-overlapping regions are removed. The subset, *D2*, consists of 15958 time points of stimulation and 3653 time points of non-therapeutic activities. Grouping these based on the manual annotations, we obtain *D3* with 464 regions with stimulation, and 357 regions with non-therapeutic activities. An overview of the data subsets are shown in table 12.1.



**Figure 12.3:** Block diagram of the proposed systems. Using raw inputs directly from the Laerdal NeoBeat prototype, a timeline describing the resuscitation are computed and presented to the user for further analysis of the event. The dotted line to and from the block *detection of ventilation* illustrates an expansion of the system using a previously proposed method from our research group.

### 12.3 Activity recognition system

A block diagram of the proposed system for detecting and recognizing regions with activities during newborn resuscitation is shown in Figure 12.3. An example graphical user interface (GUI) of the system is shown in Figure 12.5. The system takes raw input from the NeoBeat prototype, do necessary processing and classifications on the signals, and present a timeline to the user describing what events occurs at various times during the resuscitation. The analysis is designed to be able to run in real time during resuscitation, or on request to obtain more details of a given resuscitation episode at a later time. To obtain a high resolution in the classified activity and allow for real time operation, stimulation is classified as time series signal of 10Hz. Where the features at index  $i$  are causally extracted from index  $i - k + 1$  to  $i$ , where  $k$  is the window size. As ventilation and stimulation activities can occur at the same time, the two classes are differentiated when presented to the user. The GUI, Figure 12.5, shows the manually annotated data in green and the classified annotation in cyan. When run on new unannotated data, only the cyan timelines will be visible to the user.

The following subsections will describe in more detail the various parts of the system, shown in Figure 12.3.

### 12.3.1 Activity detection

Detection of time periods during resuscitation where activities are likely to be performed on the newborn has previously been proposed by our research group, VuDetector [52]. The method detects time regions based on the short time energy (STE) of the acceleration energy signal. The acceleration energy,  $Acc(n)$  is found by

$$Acc(n) = \sqrt{Acc_x^2(n) + Acc_y^2(n) + Acc_z^2(n)} \quad (12.1)$$

Where  $Acc_d$  is a low pass filtered version of the measured acceleration in axis  $d \in \{x, y, z\}$ , and  $n$  is the index in the acceleration signal. The STE,  $E(i)$ , is then found by

$$E_{Acc}(i) = \sum_{n=i-N+1}^i (Acc(n) \cdot w(i-n))^2 \quad (12.2)$$

Where the STE at index  $i$  is computed using samples from the window of length  $N$ . The STE is thresholded to determine if an activity occurs at the current window. The method achieves a sensitivity of 90% and a specificity of 80%, both with a standard deviation of 6%. More details of VuDetector can be found in [52].

### 12.3.2 Detection of ventilation

A method for detecting ventilation during newborn resuscitation based on the measured pressure signal in the BMR has previously been proposed by our research group with an accuracy of 95%, VuVentilation [84]. As ventilation and stimulation events can occur at the same time, ventilation is not taken into account when trying to recognize stimulation.

### 12.3.3 Preprocessing of data

The recorded ECG signal is susceptible to noise from the power grid, and is therefore first filtered using a 50Hz notch filter prior to any analysis. In addition, the QRS wave amplitude is affected both by the condition of the heart as well as the sensor placement. Variations in the amplitude due to inferior sensor placement are not desired, and the ECG signal is therefore normalized based on the median R-height of the signal. R-waves in the

ECG signal are found using a discrete wavelet transform with the *sym4* wavelet. The signal is then normalized using the formula:

$$ECG_{norm}(m) = \frac{ECG(m)}{\text{median}(R_h)} \quad (12.3)$$

Where  $R_h$  is a vector containing the height of the detected R-waves, and  $m$  the index in the ECG signal.

#### 12.3.4 Feature extraction

The proposed system utilize a subset of the ECG and acceleration features used by VuClassifier [85]. An overview of the features used in VuClassifier can be found in column *fSet1* in table 12.2. Features in the time domain, such as energy, RMS, entropy, and in the frequency domain, wavelet, were defined for both the acceleration and ECG signals by Vu [85]. The wavelet features were extracted using a 6-level decomposition using the Daubechies mother wavelet. These features are denoted  $Ea$  for the energy corresponding to the approximation and  $D1 - D6$  for the energy corresponding to the detail at each level. More details of these features can be found in [85].

VuClassifier [85] considers the entire region detected by VuDetector as a single class. This approach will introduce unwanted misclassifications in all cases where the detected activity region does not perfectly match the true stimulation activity in both time and length. As two or more events performed by the health care workers are unlikely to have the same duration, this approach also extracts features from windows at various lengths throughout each episode which is also not desirable. To handle these challenges, a sliding window of fixed size is introduced. Initial experiments were conducted using various length of the sliding window, with only minor differences between window sizes. A sliding window of 1 second, with a 900ms overlap were chosen to achieve a high resolution in the classification, so that all feature values are recalculated every 0.1 second, and as such can be seen as a function of a time index,  $i$ , at a sample rate of 10 Hz.

Based on visual observation of the resuscitation activities applied by the health care personnel, it is clear that stimulation often contains some repetitive movements, i.e. rubbing the back of the newborn. We therefore consider the accelerometer signals to be the most important signals for describing this repetitive movement. To represent these movements in the analysis, three new features are defined for each axes in the accelerometer

signal. Resulting in a total of 9 new features. Let  $P_{Acc,d}^i(f)$  denote the Short Time Fourier Transform (STFT) of window  $i$  in the accelerometer signal in axis  $d \in \{x, y, z\}$  as a function of the frequency,  $f$ . The first feature,  $A_{Pmax,d}$ , describes the maximum amplitude in the frequency domain,

$$A_{Pmax,d}(i) = \max \left( P_{Acc,d}^i(f) \right) \quad (12.4)$$

The second feature,  $A_{f,d}(i)$ , describes the frequency this maximum occurs at, according to

$$A_{f,d}(i) = \operatorname{argmax} \left( P_{Acc,d}^i(f) \right) \quad (12.5)$$

The third feature describes the highest frequency with an amplitude above a set threshold, given by

$$A_{fT,d}(i) = \max \{ f : P_{Acc,d}^i(f) > T \} \quad (12.6)$$

An overview of the STFT features are seen in column *fSet4* in table 12.2.

### 12.3.5 Classification

Initial tests were conducted using classifiers such as Naive Bayes, SVM [109], and RUSBoost [42]. Due to only minor differences in performance, the Naive Bayes classifier is used throughout the experiments due to its low computational complexity. The classifier is designed to distinguish between stimulation and non-stimulation activities. Discrimination between these two classes are conducted on all regions identified by VuDetector.

## 12.4 Experiments

Three experiments were designed. The first experiment was conducted to validate the previously published activity classifier. In the second experiment, all features were computed using a sliding window of fixed size. New features are added, and a nested cross-validation with feature selection are conducted to illustrate the performance which is possible. A reduced feature set is then found using feature selection. The final experiment studies performance of NBstim on full episodes, and how post processing can be utilized to increase the performance. The first experiment utilize the dataset *D3*, the second experiment uses the corresponding points extracted using a sliding window defined in *D2*. The third experiment utilize *D1* of 74 resuscitation episodes of up to 7 minutes each.

**Table 12.2:** Overview of all features. The dashed line separate features from Vu [85] and new features proposed in this work for the accelerometer signals.

Feature number	Feature name	Description	fSet 1, VuClassifier [85]	fSet 2, Vu window	fSet 3, Vu reduced	fSet 4, STFT features	fSet 5, Final set
1	$A_{te}(\hat{t})$	Total energy	1	1	1	-	1
2	$A_{globalMax}(\hat{t})$	Max acceleration value in episode	1	-	-	-	-
3	$A_{Ea}(\hat{t})$	Ea, energy approximation	1	1	1	-	1
4	$A_{Ed1}(\hat{t})$	Ed1, energy detail, level 1	1	1	1	-	1
5	$A_{Ed2}(\hat{t})$	Ed2, energy detail, level 2	1	1	-	-	-
6	$A_{Ed3}(\hat{t})$	Ed3, energy detail, level 3	1	1	-	-	-
7	$A_{Ed4}(\hat{t})$	Ed4, energy detail, level 4	1	1	-	-	-
8	$A_{Ed5}(\hat{t})$	Ed5, energy detail, level 5	1	1	-	-	-
9	$A_{Ed6}(\hat{t})$	Ed6, energy detail, level 6	1	1	-	-	-
10	$A_{autocorr}(\hat{t})$	Energy of the auto-correlation signal	1	1	1	-	1
11	$A_{winMax}(\hat{t})$	Max acceleration value in window	1	1	1	-	1
22	$A_{valley}(\hat{t})$	Valley energy	1	1	1	-	1
13	$\bar{A}(\hat{t})$	Mean of acceleration energy	1	1	-	-	-
14	$\bar{A}_x(\hat{t})$	Mean of acceleration <sub>x</sub>	1	1	-	-	-
15	$\bar{A}_y(\hat{t})$	Mean of acceleration <sub>y</sub>	1	1	1	-	1
16	$\bar{A}_z(\hat{t})$	Mean of acceleration <sub>z</sub>	1	1	-	-	-
17	$A_{\sigma}(\hat{t})$	Standard deviation of acc energy	1	1	1	-	1
18	$A_{\sigma,x}(\hat{t})$	Standard deviation of acc <sub>x</sub>	1	1	1	-	1
19	$A_{\sigma,y}(\hat{t})$	Standard deviation of acc <sub>y</sub>	1	1	1	-	1
20	$A_{\sigma,z}(\hat{t})$	Standard deviation of acc <sub>z</sub>	1	1	-	-	-
21	$A_e(\hat{t})$	Entropy of acc energy	1	1	-	-	-
22	$A_{H,x}(\hat{t})$	Entropy of acc <sub>x</sub>	1	1	1	-	1
23	$A_{H,y}(\hat{t})$	Entropy of acc <sub>y</sub>	1	1	-	-	-
24	$A_{H,z}(\hat{t})$	Entropy of acc <sub>z</sub>	1	1	1	-	1
25	$ARM_{S,e}(\hat{t})$	RMS of acc energy	1	1	-	-	-
26	$ARM_{S,x}(\hat{t})$	RMS of acc <sub>x</sub>	1	1	-	-	-
27	$ARM_{S,y}(\hat{t})$	RMS of acc <sub>y</sub>	1	1	1	-	1
28	$ARM_{S,z}(\hat{t})$	RMS of acc <sub>z</sub>	1	1	-	-	-
29	$A_{corr,xy}(\hat{t})$	Correlation, acc <sub>x</sub> , acc <sub>y</sub>	1	1	1	-	1
30	$A_{corr,xz}(\hat{t})$	Correlation, acc <sub>x</sub> , acc <sub>z</sub>	1	1	1	-	1
31	$A_{corr,yz}(\hat{t})$	Correlation, acc <sub>y</sub> , acc <sub>z</sub>	1	1	1	-	1

Table 12.3: Continuation of Table 12.2.

Feature number	Feature name	Description	fSet 1, ViClassifier [85]	fSet 2, Vu window	fSet 3, Vu reduced	fSet 4, STFT features	fSet 5, Final set
32	$A_{Fmax,x}(i)$	Highest amplitude in STFT x-axis	-	-	-	1	1
33	$A_{Fmax,y}(i)$	Highest amplitude in STFT y-axis	-	-	-	1	1
34	$A_{Fmax,z}(i)$	Highest amplitude in STFT z-axis	-	-	-	1	1
35	$A_{F,x}(i)$	Frequency of peak amplitude, x-axis	-	-	-	1	1
36	$A_{F,y}(i)$	Frequency of peak amplitude, y-axis	-	-	-	1	-
37	$A_{F,z}(i)$	Frequency of peak amplitude, z-axis	-	-	-	1	1
38	$A_{Fthresh,x}(i)$	Highest frequency over a threshold, x-axis	-	-	-	1	-
39	$A_{Fthresh,y}(i)$	Highest frequency over a threshold, y-axis	-	-	-	1	1
40	$A_{Fthresh,z}(i)$	Highest frequency over a threshold, z-axis	-	-	-	1	1
41	$ECG_{te}(i)$	Total energy	1	1	-	-	-
42	$ECG_{globalMax}(i)$	Max eeg value in episode	1	-	-	-	-
43	$ECG_{Ea}(i)$	Ea, energy approximation	1	1	1	-	1
44	$ECG_{Ed1}(i)$	Ed1, detail, level 1	1	1	1	-	-
45	$ECG_{Ed2}(i)$	Ed2, energy detail, level 2	1	1	1	-	-
46	$ECG_{Ed3}(i)$	Ed3, energy detail, level 3	1	1	1	-	-
47	$ECG_{Ed4}(i)$	Ed4, energy detail, level 4	1	1	1	-	1
48	$ECG_{Ed5}(i)$	Ed5, energy detail, level 5	1	1	-	-	-
49	$ECG_{Ed6}(i)$	Ed6, energy detail, level 6	1	1	-	-	-
50	$ECG_{corr}(i)$	Energy of the auto-correlation signal	1	1	-	-	-
51	$ECG_{\mu}(i)$	Mean value of ECG	1	1	1	-	1
52	$ECG_{\sigma}(i)$	Standard deviation of eeg	1	1	-	-	-
53	$ECG_{E}(i)$	Entropy of the ECG	1	1	-	-	-
54	$ECG_{RMS}(i)$	RMS of the ECG	1	1	1	-	1
55	$ECG_{winMax}(i)$	Max ECG value in window	1	1	1	-	1

### 12.4.1 Experiment 1: Validation of previous work

In VuClassifier [85] the activities were divided into three classes, 1) *chest compression*, 2) *stimulation*, and 3) *other*. The first two classes were proposed to be combined to obtain a classification of treatment versus non-treatment. In the present work, we focused on distinguishing stimulation from non-stimulation activities, thus the two first classes were combined, and considered stimulation. The third class was interpreted as all non-stimulation activities. Validation of VuClassifier was conducted using the dataset *D3*.

### 12.4.2 Experiment 2: Improvement of activity classifier

To improve the usability of an improved version of the classifier, and facilitate for real-time classification, only causal features were implemented. The feature set *fSet1* include the two features,  $ECG_{globalMax}$  and  $A_{globalMax}$ , which are computed using entire episodes, and were therefore omitted.

Selecting a smaller feature subset, with a high performance, from a larger set can be achieved using multiple approaches. Exhaustive search is rarely used in datasets with many features due to the heavy computational cost related to do validation of every possible feature combination in a dataset. A common approach is to use a greedy method, such as a forward selection or backward elimination, as they are fast and robust against overfitting [110]. A wrapper based nested cross-validation [90] with a modified feed forward approach, where the 5 best features in an iteration was used in this work to determine the best feature combination in the next iteration.

The nested cross-validation with feature extraction scheme is used to determine the performance which can be obtained using a feature subset. A new feature selection is then conducted to identify the optimal subset, *fSet3*, from *fSet2*. The new feature set, *fSet4*, are then be included. A second round of nested cross-validation and feature selection will be performed to identify the potential performance and the feature set, *fSet5*, from  $fSet3 \cup fSet4$ .

### 12.4.3 Experiment 3: Full episodes

As the proposed system is designed to annotate full episodes, it is important to present the performance which the end user will see. As a result of

this, the performance on full episodes are computed using a leave-one-out validation on full episodes using the feature set  $fSet5$ .

The complete system classifies stimulation with a resolution of  $10Hz$ . It is however a reasonable assumption that activities performed by health care workers do not change at such a speed, nor do they last as short time. By taking these two factors into consideration, a post processing scheme were introduced, with the potential of eliminating short segments where the activity was misclassified.

One of the most basic post processing schemes consists of doing a majority voting within a detected activity region, this approach has the same challenges as VuClassifier [85], and will therefore not be considered for further analysis. As one of the challenges of this approach is misclassifications at the borders, a second post processing scheme consists of classifying the edge regions alone, while leaving the bigger middle section to be classified as a single activity. This scheme can solve the edge problem, but determining the ideal size of these edges can pose a challenge. An alternative processing scheme is based on the idea that the detected activity region could include one or more areas with actual stimulation. The change between stimulation and non-stimulation activities should, however, not be able to change as fast as original classification. This post processing scheme can easily be implemented using a median filter on the classified timeline.

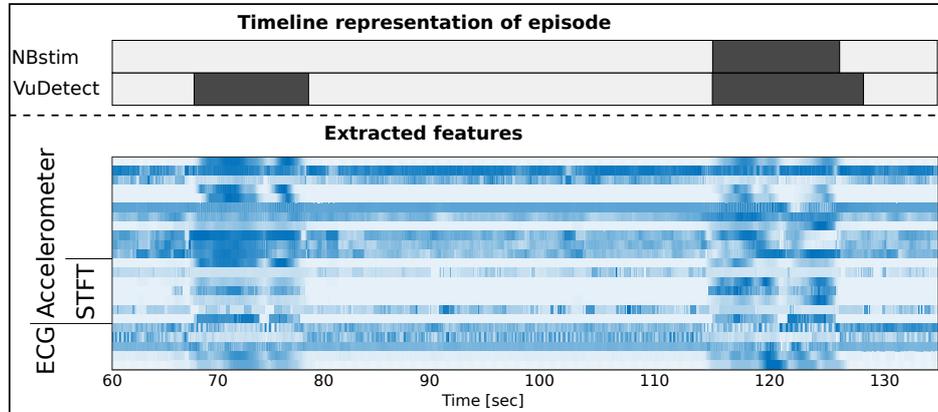
## 12.5 Results

### 12.5.1 Experiment 1

The VuClassifier was trained on the original dataset described in [85], and then used to classify the subset  $D3$ . The performance can be seen in table 12.4, and the used features can be seen in column  $fSet1$  in table 12.2.

**Table 12.4:** Performance using the features and classifier proposed by Vu et al. on blocks with variable length found as time segments where an activity is found by VuDetector and either stimulation or no activity is manually annotated,  $D3$

Method	#features	Sensitivity	Specificity	Accuracy
VuClassifier [85]	46	88.4%	1.7%	50.7%



**Figure 12.4:** Overview of the computed feature values for the same time period illustrated in Figure 12.2 using the 23 final features,  $fSet5$ . Each row in the heat map illustrate the values of a given feature over time. A darker color indicates a higher value. The feature order is the same as shown in table 12.5. All features are normalized in the region  $[0, 1]$  for visualization purposes. NBstim evaluates the regions found by VuDetect, and classify them as stimulation or non-stimulation activities.

### 12.5.2 Experiment 2

The performance using all 44 causal window based features in  $fSet2$  on the dataset  $D2$ , can be seen in table 12.5. The 44 features are then reduced to 24 features using feature selection. The chosen features can be seen in column  $fSet3$  in table 12.2, and the performance of the reduced feature set in table 12.5. A feature extraction is conducted the feature subset consisting of  $fSet3 \cup fSet4$  resulting in  $fSet5$  as seen in table 12.2. Performance for each subset is computed using a nested cross-validation with feature extraction. The performance of these three feature sets are seen in table 12.5.

A visualization of the computed values for the final feature set,  $fSet5$ , is shown in Figure 12.4. Each row corresponds to a given feature, and a darker color indicate a higher value in the computed feature value. For visualization purposes, all features are normalized to  $[0, 1]$ .

### 12.5.3 Experiment 3

The performance when distinguishing between stimulation and non-stimulation activities in entire resuscitation episodes,  $D1$ , with and without a post processing scheme are shown in table 12.6. A leave-one-out

**Table 12.5:** Performance of various feature sets computed using sliding window on time points where an activity is found by VuDetector and either stimulation of no activity is manually annotated, *D2*. The performance is computed using a 3-fold nested cross-validation

Feature set	#features	Sensitivity	Specificity	Accuracy
<i>fSet2</i>	44	61.7%	65.5%	61.7%
<i>fSet3</i>	24	63.3%	65.5%	63.7%
<i>fSet4</i>	9	75.6%	43.7%	69.7%
<i>fSet3</i> $\cup$ <i>fSet4</i>	33	57.4%	69.8%	59.7%
<i>fSet5</i>	23	67.3%	62.1%	66.4%

**Table 12.6:** Performance of NBstim when distinguishing between stimulation and non-stimulation in full resuscitation episodes with a maximum length of 7 minutes, *D1*

Method	Sensitivity	Specificity	Accuracy
No postprocessing	68.3%	93.1%	88.6%
Median filtering	69.2%	94.8%	90.3%

cross-validation is used, and the classified timeline is compared to when stimulation or no-stimulation is manually annotated.

## 12.6 Discussion

Validation of VuClassifier achieves a high sensitivity and a low specificity in distinguishing between stimulation and non-stimulation events on the data set annotated by a neonatologist, more details in section 12.2.1. This performance does not correspond to the accuracy of 79.8%, sensitivity of 84% and specificity of 72.6% reported in Vu et al. [85]. The degradation in performance could be a result of how the new dataset is defined. The increased size should not affect performance, but how manual annotations are found may result in a change. In the original publication, Vu et al [85], the data was annotated by the author using several additional categories, and may therefore differ from how a trained clinician would annotate it.

By combining features in *fSet1* with a sliding window, *fSet2*, and a simple classifier, an accuracy of 61.7% is achieved, with a sensitivity and a specificity both above 60%. By adding more features to a system, the

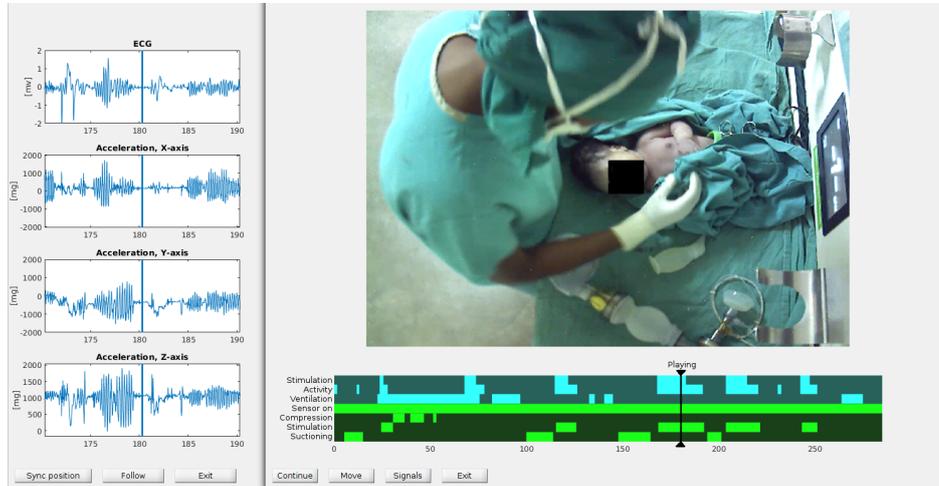
performance does not necessarily increase. The final feature set,  $fSet5$ , found using a feature selection approach, outperforms the original feature set,  $fSet2$ .

The performance when using the STFT features,  $fSet4$ , to distinguish between stimulation and non-stimulation activities achieve a high sensitivity and accuracy. The specificity is, however, reduced compared to the reduced feature set from Vu et al [85],  $fSet3$ . The final subset,  $fSet5$ , consisting of 23 features, achieves a sensitivity and accuracy which is lower, but the specificity has a large increase compared to using only the new features. While correct identification of the stimulation events are important, we want to keep the false positive rate as low as possible to better indicate how much stimulation is actually given during the resuscitation.

Using the obtained feature subset,  $fSet5$ , on complete episodes,  $D1$ , a large increase in performance is seen with the accuracy increasing from 66.4% in table 12.5 to 88.6% in table 12.6. This increase is a result of complete resuscitation episodes often include large time periods where no stimulation occurs. As many of these periods are not identified as interesting by VuDetector, the specificity will increase. By applying a median filter post processing scheme, the accuracy is further increased to 90.3%.

While the addition of new STFT features may be considered a small expansion of previous work, we consider the proposed causal system with a reduced feature set to be an important step towards utilizing this work for automatic annotation of newborn stimulation. The new features were proposed to describe repetitive movement observed when health care personnel performed stimulation, i.e. rubbing the back of the newborn. In the visualization of the final feature set,  $fSet5$ , we can see some differences between the feature values in the stimulation and non-stimulation regions. It is, however, challenging to determine how each feature is physiological linked to the performed activity.

The signal composition, consisting of accelerometer and dry electrode ECG, as available from the NeoBeat has never been available before. A method utilizing these signals to automatically annotate when stimulation is being performed during resuscitation can greatly impact future studies of how stimulation activities affect the resuscitation process and newborn outcome. With such a system, like we propose with NBstim, studies exploring how stimulation activities affects the resuscitation process will no longer be limited to using a low number of manually annotated data, but



**Figure 12.5:** Example graphical user interface for the proposed system. In this example, the timeline includes manually annotated data, shown in green, and automatically classified data, shown in cyan. ECG and acceleration signals are included to visualize the current measurements to the user.

information from larger data sets can be extracted. Statistics on how guidelines are followed can be extracted. It will also allow to extract information of the duration of stimulation activities during a resuscitation. Was it a continuous stimulation or multiple? When the newborn was not ventilated, was that due to hands-off time or stimulation? Such information may be vital when exploring how resuscitation outcomes are correlated with the stimulation activities during resuscitation. In a future scenario we might have NeoBeat available in some hospitals, video signals in other hospitals, ventilation data from some hospitals, or several of these components. We are planning to fuse the output of the NBstim algorithm with the output of the automated video analysis [86], and potentially output from ventilator signals [84] to produce more reliable timelines of activities, also including ventilation and suction when possible.

For further quality assurance and truth marking for validation, an interface similar to what is shown in Figure 12.5 can be used as an interactive annotation tool with the automatic detections as a first step, and with the option of refining these manually if needed.

### 12.6.1 Limitations

Due to the small data set of only 74 resuscitations episodes, a total of 21830 seconds, the proposed system may be seen as a feasibility study of the possibility of annotating stimulation based on the measured accelerometer and ECG signals. When identifying the feature set, a smaller subset of only 1961 seconds is utilized. This reduction is performed by only including time periods where some movement occur, and VuDetector identifies the movement as an activity, and either only stimulation or no therapeutic activity is performed. The advantage of using this smaller subset for the feature selection is that the method will identify features which are crucial in distinguishing stimulation and non-stimulation activities instead of focusing on patterns from other activities. Because of this limited data set with ground truth, further validation is required before applying the method in clinical practice.

## 12.7 Conclusion

In this work, we present a complete system for automatic identification of stimulation during newborn resuscitation. The system consists of an activity detector, and the proposed NBstim classifier with 23 features, 18 from the 100Hz accelerometer signals in X,Y, and Z-directions and 5 from the 500Hz dry-electrode ECG signal. Features are computed using a sliding window of 1 second with 900ms overlap. NBstim achieves a high performance, with an accuracy of 90.3% in identifying stimulation, and could therefore be used as a replacement of time consuming manual annotation, or as an initial step in an interactive tool. The ultimate objective is to save lives at birth, and more specifically by studying what activities are performed by health care providers during resuscitation of asphyxiated newborns, if guidelines are followed, and if current guidelines are effective in saving lives.

The system can be used with the newly released Laerdal NeoBeat Newborn Heart Rate Meter, but a validation using a larger data set is required before implementing the method in clinical practice. In the Safer Births project, we are currently working on expand our data collection of newborn resuscitation, and we want to increase the number of manually annotated data.

In future work, we want to utilize NBstim for creating timelines for thousands of newborn resuscitation episodes. In combination with the

immediate and 24-hour outcome, available in the Safer Births project, we can extract vital statistics and potentially get a greater understanding of how stimulation activities affect resuscitation procedures and newborn outcomes.

## **12.8 Conflict of Interest**

This work is part of the Safer Births project which has received funding from Laerdal Foundation, Laerdal Global Health, Skattefunn, Norwegian Ministry of Education and USAID. The work was partly supported by the Research Council of Norway through the Global Health and Vaccination Programme (GLOBVAC) project number 228203.

Ingunn Anda Haug is an employee of Laerdal Medical AS.

# Bibliography

- [1] Joy E Lawn, Hannah Blencowe, Peter Waiswa, Agbessi Amouzou, Colin Mathers, Dan Hogan, Vicki Flenady, J Frederik Frøen, Zeshan U Qureshi, Claire Calderwood, et al. Stillbirths: rates, risk factors, and acceleration towards 2030. *The Lancet*, 387(10018):587–603, 2016.
- [2] Hannah Blencowe, Simon Cousens, Fiorella Bianchi Jassir, Lale Say, Doris Chou, Colin Mathers, Dan Hogan, Suhail Shiekh, Zeshan U Qureshi, Danzhen You, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *The Lancet Global Health*, 4(2):e98–e108, 2016.
- [3] Diogo Ayres-de Campos, Catherine Y Spong, Edwin Chandrachan, and FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. Figo consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *International Journal of Gynecology & Obstetrics*, 131(1):13–24, 2015.
- [4] World Health Organization et al. WHO recommendations: intrapartum care for a positive childbirth experience. 2018.
- [5] Debrah Lewis, Soo Downe, and FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. Figo consensus guidelines on intrapartum fetal monitoring: Intermittent auscultation. *International Journal of Gynecology & Obstetrics*, 131(1):9–12, 2015.
- [6] John C Morrison, Bonnie F Chez, Ivory D Davis, Rick W Martin, William E Roberts, James N Martin Jr, and Randall C Floyd. Intrapartum fetal heart rate assessment: monitoring by auscultation or electronic means. *American journal of obstetrics and gynecology*, 168(1):63–66, 1993.
- [7] WHO. Guidelines on basic newborn resuscitation, 2012.
- [8] Jeffrey M Perlman et al. Part 7: neonatal resuscitation: 2015 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. *Circulation*, 132(16\_suppl\_1):S204–S241, 2015.
- [9] Hege Langli Erdsdal, Estomih Mduma, Erling Svensen, and Jeffrey M. Perlman. Early initiation of basic resuscitation interventions including face mask ventilation may reduce birth asphyxia related mortality in low-income countries: A prospective descriptive observational study. *Resuscitation*, 83(7):869 – 873, 2012.

## BIBLIOGRAPHY

---

- [10] M Iriundo, M Thió, E Burón, E Salguero, J Aguayo, M Vento, and Neonatal Resuscitation Group (NRG) of the Spanish Neonatal Society (SEN). A survey of neonatal resuscitation in Spain: gaps between guidelines and practice. *Acta paediatrica*, 98(5):786–791, 2009.
- [11] Acog practice bulletin no. 106: Intrapartum fetal heart rate monitoring: Nomenclature, interpretation, and general management principles, 2009.
- [12] NICE. Intrapartum care for healthy women and babies: Clinical guideline, 2017.
- [13] WHO recommendation on intermittent fetal heart rate auscultation during labour. page 13, 2018.
- [14] Vicky Chapman and Cathy Charles. *The midwife’s labour and birth handbook*. John Wiley & Sons, 2018.
- [15] Tammy Y Euliano, Minh Tam Nguyen, Shalom Darmanjian, Susan P McGorray, Neil Euliano, Allison Onkala, and Anthony R Gregg. Monitoring uterine activity during labor: a comparison of 3 methods. *American journal of obstetrics and gynecology*, 208(1):66–e1, 2013.
- [16] Gabi Haran, Michal Elbaz, Moshe D Fejgin, and Tal Biron-Shental. A comparison of surface acquired uterine electromyography and intrauterine pressure catheter to assess uterine activity. *American journal of obstetrics and gynecology*, 206(5):412–e1, 2012.
- [17] Tammy Y Euliano, Minh Tam Nguyen, Shalom Darmanjian, John D Bussowski, Neil Euliano, and Anthony R Gregg. Monitoring uterine activity during labor: clinician interpretation of electrohysterography versus intrauterine pressure catheter and tocodynamometry. *American journal of perinatology*, 33(09):831–838, 2016.
- [18] Jørgen Linde. *Normal heart rate transition after birth and heart rate responses and newborn outcome after resuscitation*. PhD thesis, University of Stavanger, November 2018.
- [19] C Omar F Kamlin, Jennifer A Dawson, Colm PF O’Donnell, Colin J Morley, Susan M Donath, Jasbir Sekhon, and Peter G Davis. Accuracy of pulse oximetry measurement of heart rate of newborn infants in the delivery room. *The Journal of pediatrics*, 152(6):756–760, 2008.
- [20] Myra H Wyckoff, Khalid Aziz, Marilyn B Escobedo, Vishal S Kapadia, John Kattwinkel, Jeffrey M Perlman, Wendy M Simon, Gary M Weiner, and Jeanette G Zaichkin. Part 13: neonatal resuscitation: 2015 American heart association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation*, 132(18\_suppl\_2):S543–S560, 2015.
- [21] Anup Katheria, Wade Rich, and Neil Finer. Electrocardiogram provides a continuous heart rate faster than oximetry during neonatal resuscitation. *Pediatrics*, 130(5):e1177–e1181, 2012.

- 
- [22] Jørgen Erland Linde, Jörn Schulz, Jeffrey M Perlman, Knut Øymar, Fortunata Francis, Joar Eilevstjønn, and Hege Langli Ersdal. Normal newborn heart rate in the first five minutes of life assessed by dry-electrode electrocardiography. *Neonatology*, 110(3):231–237, 2016.
- [23] Janusz Jezewski, Tomasz Kupka, and Krzysztof Horoba. Extraction of fetal heart-rate signal as the time event series from evenly sampled data acquired using doppler ultrasound technique. *IEEE Transactions on Biomedical Engineering*, 55(2):805–810, 2008.
- [24] Saeed Abdulrahman Alnuaimi, Shihab Jimaa, and Ahsan H Khandoker. Fetal cardiac doppler signal processing techniques: challenges and future research directions. *Frontiers in bioengineering and biotechnology*, 5:82, 2017.
- [25] Alain Rakotomamonjy. Direct optimization of the dictionary learning problem. *IEEE Transactions on Signal Processing*, 61(22):5495–5506, 2013.
- [26] Chenglong Bao, Hui Ji, Yuhui Quan, and Zuowei Shen. Dictionary learning for sparse coding: Algorithms and convergence analysis. *IEEE transactions on pattern analysis and machine intelligence*, 38(7):1356–1369, 2015.
- [27] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, Jan 2008.
- [28] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley. Audio inpainting. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(3):922–932, March 2012.
- [29] V. P. Oikonomou, J. Spilka, C. Stylios, and L. Lhostka. An adaptive method for the recovery of missing samples from fhr time series. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 337–342, June 2013.
- [30] K. Engan, S.O. Aase, and J. Håkon Husøy. Method of optimal directions for frame design. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 5, pages 2443–2446 vol.5, 1999.
- [31] M. Aharon, M. Elad, and A. Bruckstein. K -SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, Nov 2006.
- [32] Sharat Chikkerur, Alexander N Cartwright, and Venu Govindaraju. Fingerprint enhancement using stft analysis. *Pattern recognition*, 40(1):198–211, 2007.
- [33] Jingshan Huang, Binqiang Chen, Bin Yao, and Wangpeng He. Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network. *IEEE Access*, 7:92871–92880, 2019.
- [34] Bryan E Usevitch. A tutorial on modern lossy wavelet image compression: foundations of jpeg 2000. *IEEE signal processing magazine*, 18(5):22–35, 2001.

## BIBLIOGRAPHY

---

- [35] Hasan Demirel and Gholamreza Anbarjafari. Discrete wavelet transform-based satellite image resolution enhancement. *IEEE transactions on geoscience and remote sensing*, 49(6):1997–2004, 2011.
- [36] Murugappan Murugappan, Nagarajan Ramachandran, Yaacob Sazali, et al. Classification of human emotion from eeg using discrete wavelet transform. *Journal of biomedical science and engineering*, 3(04):390, 2010.
- [37] Khalid Sayood. *Introduction to data compression*. Morgan Kaufmann, 2017.
- [38] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. Detection of eeg characteristic points using wavelet transforms. *IEEE Transactions on biomedical Engineering*, 42(1):21–28, 1995.
- [39] Kyriaki K Spyridou and Leontios J Hadjileontiadis. Analysis of fetal heart rate in healthy and pathological pregnancies using wavelet-based features. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1908–1911. IEEE, 2007.
- [40] Agnese Sbröllini, Annachiara Strazza, Manila Caragiuli, Claudia Mozzoni, Selene Tomassini, Angela Agostinelli, Micaela Morettini, Sandro Fioretti, Francesco Di Nardo, and Laura Burattini. Fetal phonocardiogram denoising by wavelet transformation: robustness to noise. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE, 2017.
- [41] Madhur Srivastava, C Lindsay Anderson, and Jack H Freed. A new wavelet denoising method for selecting decomposition levels and noise thresholds. *IEEE access*, 4:3862–3877, 2016.
- [42] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 40(1):185–197, Jan 2010.
- [43] G. M. Weiss. Mining with rarity: A unifying framework. *SIGKDD Explor. Newsl.*, 6(1):7–19, June 2004.
- [44] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995.
- [45] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer. Smoteboost: improving prediction of the minority class in boosting. In *In Proceedings of the Principles of Knowledge Discovery in Databases, PKDD-2003*, pages 107–119, 2003.
- [46] Sergios Theodoridis. *Machine learning: a Bayesian and optimization perspective*. Academic Press, 2015.
- [47] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7(1):91, 2006.
- [48] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.
- [49] Moyo fetal heart rate monitor, user guide. 20-08388/00026151 Rev E.

- 
- [50] Paschal Francis Mdoe. *Improving Fetal Heart Rate Monitoring in Low Resource Settings*. PhD thesis, University of Stavanger, April 2019.
- [51] Benjamin Anathory Kamala. *Assessment of foetal heart rate monitoring devices in referral hospitals in Tanzania*. PhD thesis, University of Stavanger, November 2019.
- [52] Huyen Vu, Trygve Eftestøl, Kjersti Engan, Joar Eilevstjønn, Ladislaus Blacy Yarrot, Jørgen E Linde, and Hege Ersdal. Detection of activities during newborn resuscitation based on short-time energy of acceleration signal. In *International Conference on Image and Signal Processing*, pages 262–270. Springer, 2016.
- [53] G Dawes, YJ Meir, and GP Mandruzzato. Computerized evaluation of fetal heart-rate patterns. *Journal of perinatal medicine*, 22(6):491–499, 1994.
- [54] Zarko Alfirevic, Gillian ML Gyte, Anna Cuthbert, and Declan Devane. Continuous cardiotocography (ctg) as a form of electronic fetal monitoring (efm) for fetal assessment during labour. *Cochrane database of systematic reviews*, (2), 2017.
- [55] Amir Sweha, Trevor W Hacker, and Jim Nuovo. Interpretation of the electronic fetal heart rate during labor. *American family physician*, 59(9):2487, 1999.
- [56] Ruth Martis, Ova Emilia, Detty S Nurdiati, and Julie Brown. Intermittent auscultation (ia) of fetal heart rate in labour for fetal well-being. *Cochrane Database of Systematic Reviews*, (2), 2017.
- [57] Antoniya Georgieva, Patrice Abry, Václav Chudáček, Petar M Djurić, Martin G Frasch, René Kok, Christopher A Lear, Sebastiaan N Lemmens, Inês Nunes, Aris T Papageorghiou, et al. Computer-based intrapartum fetal monitoring and beyond: A review of the 2nd workshop on signal processing and monitoring in labor (october 2017, oxford, uk). *Acta obstetricia et gynecologica Scandinavica*, 98(9):1207–1217, 2019.
- [58] Diogo Ayres-de Campos, Mariana Rei, Inês Nunes, Paulo Sousa, and Joao Bernardes. Sisporto 4.0—computer analysis following the 2015 figo guidelines for intrapartum fetal monitoring. *The Journal of Maternal-Fetal & Neonatal Medicine*, 30(1):62–67, 2017.
- [59] Antoniya Georgieva, Christopher WG Redman, and Aris T Papageorghiou. Computerized data-driven interpretation of the intrapartum cardiotocogram: a cohort study. *Acta obstetricia et gynecologica Scandinavica*, 96(7):883–891, 2017.
- [60] A Georgieva, AT Papageorghiou, SJ Payne, M Moulden, and CWG Redman. Phase-rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidaemia at birth. *BJOG: An International Journal of Obstetrics & Gynaecology*, 121(7):889–894, 2014.

## BIBLIOGRAPHY

---

- [61] Jirí Spilka, V Chudáček, Michal Koucký, Lenka Lhotská, Michal Huptych, P Janků, George Georgoulas, and Chrysostomos Stylios. Using nonlinear features for fetal heart rate classification. *Biomedical Signal Processing and Control*, 7(4):350–357, 2012.
- [62] Jirí Spilka, George K Georgoulas, Petros S Karvelis, Václav Chudáček, Chrysostomos D Stylios, and Lenka Lhotská. Discriminating normal from "abnormal" pregnancy cases using an automated fhr evaluation method. In *SETN*, pages 521–531. Springer, 2014.
- [63] Guanchao Feng, J Gerald Quirk, and Petar M Djurić. Recovery of missing samples in fetal heart rate recordings with gaussian processes. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, pages 261–265. IEEE, 2017.
- [64] Shunji Suzuki, Yuuri Sensu, and Mayumi Ueda. Maternal heart rate mistaken for the fetal heart rate during delivery. *Journal of Clinical Gynecology and Obstetrics*, 5(2):74–76, 2016.
- [65] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals. *Information Sciences*, 415:190–198, 2017.
- [66] U Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. *Information sciences*, 405:81–90, 2017.
- [67] Alessio Petrozziello, Ivan Jordanov, T Aris Papageorghiou, WG Christopher Redman, and Antoniya Georgieva. Deep learning for continuous electronic fetal monitoring in labor. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5866–5869. IEEE, 2018.
- [68] Antoniya Georgieva, Stephen J Payne, Mary Moulden, and Christopher WG Redman. Artificial neural networks applied to fetal monitoring in labour. *Neural Computing and Applications*, 22(1):85–93, 2013.
- [69] Steven L Clark, Emily F Hamilton, Thomas J Garite, Audra Timmins, Philip A Warrick, and Samuel Smith. The limits of electronic fetal heart rate monitoring in the prevention of neonatal metabolic acidemia. *American journal of obstetrics and gynecology*, 216(2):163–e1, 2017.
- [70] Antoniya Georgieva, Mary Moulden, and Christopher WG Redman. Umbilical cord gases in relation to the neonatal condition: the everest plot. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 168(2):155–160, 2013.
- [71] D Preethi and RS Valarmathi. Classification and suppression of noises in fetal heart rate monitoring: A survey. In *Microelectronics, Electromagnetics and Telecommunications*, pages 607–615. Springer, 2019.

- 
- [72] Rik Vullings, Chris Peters, Massimo Mischi, Guid Oei, and Jan Bergmans. Maternal ecg removal from non-invasive fetal ecg recordings. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1394–1397. IEEE, 2006.
- [73] Faraz Barzideh, Jarle Urdal, Kidanto Hussein, Kjersti Engan, Karl Skretting, Paschal Mdoe, Benjamin Kamala, and Sara Brunner. Estimation of missing data in fetal heart rate signals using shift-invariant dictionary. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 762–766. IEEE, 2018.
- [74] Chang Su Lee, Martin Masek, Chiou Peng Lam, and Keng T Tan. Towards higher accuracy and better noise-tolerance for fetal heart rate monitoring using doppler ultrasound. In *TENCON 2009-2009 IEEE Region 10 Conference*, pages 1–6. IEEE, 2009.
- [75] Michelle L Murray. Maternal or fetal heart rate? avoiding intrapartum misidentification. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 33(1):93–104, 2004.
- [76] Joscha Reinhard, Barrie Hayes-Gill, Sven Schiermeier, Hendrike Hatzmann, Tomas M Heinrich, and Frank Louwen. Intrapartum heart rate ambiguity: a comparison of cardiotocogram and abdominal fetal electrocardiogram with maternal electrocardiogram. *Gynecologic and obstetric investigation*, 75 2:101–8, 2013.
- [77] M Watakabe, K Mita, K Akataki, and K Ito. Reliability of the mechanomyogram detected with an accelerometer during voluntary contractions. *Medical and Biological Engineering and Computing*, 41(2):198–202, 2003.
- [78] M Tarata, A Spaepen, and R Puers. The accelerometer mmg measurement approach, in monitoring the muscular fatigue. *Measurement Science Review*, 1(1):47–50, 2001.
- [79] Hinke de Lau, Chiara Rabotti, Nicole Haazen, S Guid Oei, and Massimo Mischi. Towards improving uterine electrical activity modeling and electrohysterography: ultrasonic quantification of uterine movements during labor. *Acta obstetrica et gynecologica Scandinavica*, 92(11):1323–1326, 2013.
- [80] Faraz Barzideh, Karl Skretting, and Kjersti Engan. Imposing shift-invariance using flexible structure dictionary learning (fsdl). *Digital Signal Processing*, 69(Supplement C):162 – 173, 2017.
- [81] S. H. Lee et al. A study on the activity classification using a triaxial accelerometer. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, volume 3, pages 2941–2943 Vol.3, Sept 2003.
- [82] P. Gupta and T. Dallas. Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Transactions on Biomedical Engineering*, 61(6):1780–1786, June 2014.

- [83] F. Yang and L. Zhang. Real-time human activity classification by accelerometer embedded wearable devices. In *2017 4th International Conference on Systems and Informatics (ICSAI)*, pages 469–473, Nov 2017.
- [84] Huyen Vu, Trygve Eftestøl, Kjersti Engan, Joar Eilevstjønn, Ladislaus Yarrot, Jorgen Linde, and Hege Ersdal. Automatic detection and parameterization of manual bag-mask ventilation on newborns. *IEEE journal of biomedical and health informatics*, 2016.
- [85] Huyen Vu, Kjersti Engan, Trygve Eftestøl, Aggelos Katsaggelos, Samwel Jatosh, Simeon Kusulla, Estomih Mduma, Hussein Kidanto, and Hege Ersdal. Automatic classification of resuscitation activities on birth-asphyxiated newborns using acceleration and ECG signals. *Biomedical Signal Processing and Control*, 36:20 – 26, 2017.
- [86] Øyvind Meinich-Bache, Kjersti Engan, Ivar Austvoll, Trygve Eftestøl, Helge Myklebust, Ladislaus Yarrot, Hussein Kidanto, and Hege Ersdal. Object detection during newborn resuscitation activities. *IEEE journal of biomedical and health informatics*, 2019.
- [87] O Meinich-Bache, SL Austnes, K Engan, I Austvoll, T Eftestøl, H Myklebust, S Kusulla, H Kidanto, and H Ersdal. Activity recognition from newborn resuscitation videos. *IEEE journal of biomedical and health informatics*, 2020.
- [88] Khalid Aziz, Mairi Chadwick, Mary Baker, and Wayne Andrews. Ante- and intra-partum factors that predict increased need for neonatal resuscitation. *Resuscitation*, 79(3):444 – 452, 2008.
- [89] Huyen Vu, Trygve Eftestøl, Kjersti Engan, Joar Eilevstjønn, Ladislaus Blacy Yarrot, Jørgen E Linde, and Hege Ersdal. Exploring the relationship between characteristics of ventilation performance and response of newborns during resuscitation. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 275–290. Springer, 2015.
- [90] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2):273 – 324, 1997. Relevance.
- [91] Jarle Urdal, Kjersti Engan, Trygve Eftestøl, Valery Naranjo, Ingunn Anda Haug, Anita Yeconia, Hussein Kidanto, and Hege Ersdal. Automatic identification of stimulation activities during newborn resuscitation using ecg and accelerometer signals. *Computer Methods and Programs in Biomedicine*, page 105445, 2020.
- [92] Hege L Ersdal, Joar Eilevstjønn, Jørgen E Linde, Anita Yeconia, Estomih R Mduma, Hussein Kidanto, and Jeffrey Perlman. Fresh stillborn and severely asphyxiated neonates share a common hypoxic–ischemic pathway. *International Journal of Gynecology & Obstetrics*, 141(2):171–180, 2018.
- [93] Vangelis P Oikonomou, Jiri Spilka, Chrysostomos Stylios, and Lenka Lhostka. An adaptive method for the recovery of missing samples from fhr time series. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 337–342. IEEE, 2013.

- 
- [94] Laerdal Global Health. Moyo fetal heart rate monitor. Accessed 27 August 2019.
- [95] J. Sulam and M. Elad. Large inpainting of face images with trainlets. *IEEE Signal Processing Letters*, 23(12):1839–1843, Dec 2016.
- [96] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. springer, 2010.
- [97] Jarle Urdal, Kjersti Engan, Trygve Eftestøl, Ladislaus Blacy Yarrot, Hussein Kidanto, and Hege Ersdal. Noise and contraction detection using fetal heart rate and accelerometer signals during labour. In *Paper from the Scandinavian Conference on Health Informatics*,. Linköping University Electronic Press., 2018.
- [98] Philip A Warrick, Emily F Hamilton, Doina Precup, and Robert E Kearney. Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiocography. *IEEE Transactions on Biomedical Engineering*, 57(4):771–779, 2010.
- [99] Philip A Warrick, Emily F Hamilton, Doina Precup, and Robert E Kearney. Identification of the dynamic relationship between intrapartum uterine pressure and fetal heart rate for normal and hypoxic fetuses. *IEEE Transactions on Biomedical Engineering*, 56(6):1587–1597, 2009.
- [100] Janusz Jezewski, Janusz Wrobel, and Krzysztof Horoba. Comparison of doppler ultrasound and direct electrocardiography acquisition techniques for quantification of fetal heart rate variability. *IEEE Transactions on Biomedical Engineering*, 53(5):855–864, 2006.
- [101] S. Wright, K. Matheieson, L. Brearley, S. Jacobs, L Holly, and R. Wickremasinghe. Save the children: Ending newborn deaths: Ensuring every baby survives, 2014.
- [102] Haidong Wang, Chelsea A Liddell, Matthew M Coates, Meghan D Mooney, Carly E Levitz, Austin E Schumacher, Henry Apfel, Marissa Iannarone, Bryan Phillips, Katherine T Lofgren, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 384(9947):957–979, 2014.
- [103] Jeffrey M Perlman, Jonathan Wyllie, John Kattwinkel, Dianne L Atkins, Leon Chameides, Jay P Goldsmith, Ruth Guinsburg, Mary Fran Hazinski, Colin Morley, Sam Richmond, et al. Neonatal resuscitation: 2010 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. *Pediatrics*, 126(5):e1319–e1344, 2010.
- [104] Tobias Scheffer. *Error estimation and model selection*. PhD thesis, Technischen Universitat Berlin, School of Computer Science, 1999.
- [105] Sandrine Dudoit and Mark J. van der Laan. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, 2(2):131 – 154, 2005.

## BIBLIOGRAPHY

---

- [106] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, 4th edition, 2008.
- [107] J. Urdal et al. Signal processing and classification for identification of clinically important parameters during neonatal resuscitation. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 547–552, Sept 2017.
- [108] Antoni Bayés De Luna, Velislav N Batchvarov, and Marek Malik. The morphology of the electrocardiogram. *The ESC Textbook of Cardiovascular Medicine Blackwell Publishing*, pages 1–35, 2006.
- [109] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [110] Isabelle Guyon and André Elisseeff. An introduction to feature extraction. In *Feature extraction*, pages 1–25. Springer, 2006.