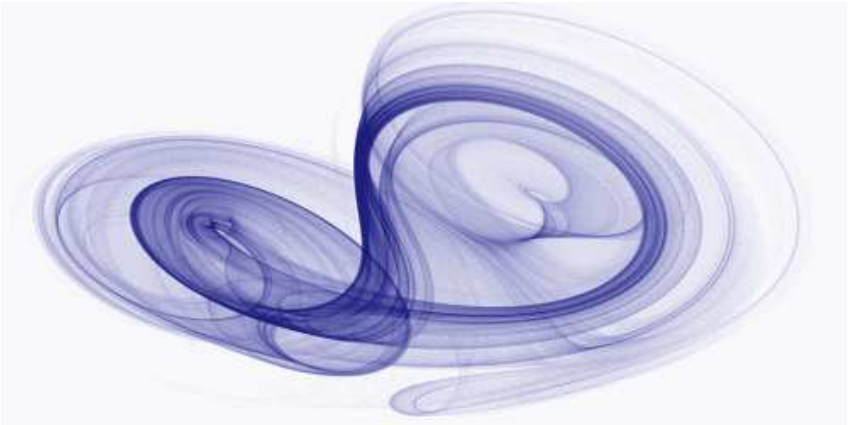
CZECH TECHNICAL UNIVERSITY IN PRAGUE
FACULTY OF ELECTRICAL ENGINEERING
DEPARTMENT OF CYBERNETICS

Fetal Electrocardiogram Analysis

Diploma Thesis



Supervisor: Ing. Václav Chudáček

Prague, 25th May 2009 Bc. Jiří Spilka

Czech Technical University in Prague
Faculty of Electrical Engineering

Department of Cybernetics

DIPLOMA THESIS ASSIGNMENT

Student: Bc. Jiří S p i l k a

Study programme: Electrical Engineering and Information Technology

Specialisation: Biomedical Engineering

Title of Diploma Thesis: Fetal ECG Analysis

Guidelines:

1. Familiarize yourself with fetal ECG problematics.
2. Investigate possibility to use non-linear features for classification of heart rate variability (HRV).
3. Suggest several possible features that could be useful for description of fetal ECG.
4. Classify fetal ECG records from database of the Clinic of obstetrics and gynecology, 1 st Medical Faculty Charles University using your suggested features.

Bibliography/Sources: Will be provided by the supervisor.

Diploma Thesis Supervisor: Ing. Václav Chudáček

Valid until: the end of the winter semester of academic year 2009/2010

L.S.

prof. Ing. Vladimír Mařík, DrSc.
Head of Department

doc. Ing. Boris Šimák, CSc.
Head

Prague, September 3, 2008

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra kybernetiky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Jiří S p i l k a

Studijní program: Elektrotechnika a informatika (magisterský), strukturovaný

Obor: Biomedicínské inženýrství

Název tématu: Analýza fetálního EKG

Pokyny pro vypracování:

1. Seznamte se s problematikou analýzy fetálního EKG.
2. Seznamte se s využitím nelineárních přízanků pro klasifikaci variability srdečního rytmu (HRV).
3. Navrhněte několik možných příznaků využitelných pro popis fetálního EKG.
4. S pomocí příznaků klasifikujte záznamy fetálního EKG z databáze Gynekologicko - porodnické kliniky 1.LF UK.

Seznam odborné literatury: Dodá vedoucí práce.

Vedoucí diplomové práce: Ing. Václav Chudáček

Platnost zadání: do konce zimního semestru 2009/2010

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE Fakulta elektrotechnická L.S.

prof. Ing. Vladimír Mařík, DrSc.
vedoucí katedry

doc. Ing. Boris Šimák, CSc.
děkan

V Praze dne 3. 9. 2008

Abstract

Cardiotocography (CTG) consists of fetal heart rate (FHR) and uterine pressure measurements. CTG has become a standard in obstetrics and is widely used in clinical settings nowadays. Fetal distress during labour causes firstly a loss of FHR variations, and secondly an occurrence of long lasting decelerations after uterine contractions. Automatic assessment of these changes is, however, not fully reliable. This thesis focused on assessment of fetal behaviour using nonlinear methods, such as fractal dimension, entropy and complexity measures. Performance of these methods was examined on short segments of fetal heart rate and evaluated in terms of ability to distinguish between normal and pathological fetuses. Using Mann-Whitney U test, it was found that all employed methods revealed a difference in feature's median values of significance $p < 0.01$. Performed feature selection rated following methods as the most suitable for further classification task: Lempel Ziv complexity, waveform fractal dimension, and sample entropy. The classification results proved that nonlinear methods provide useful information of fetus status even for fetal heart rate segment of short duration. The achieved classification performance, by support vector machine, was in terms of sensitivity and specificity 78 % and 70 %, respectively.

Keywords Cardiotocography, Fetal heart rate, Time series analysis, Chaos, Nonlinear methods, Feature selection, Classification

Abstrakt

Kardiotokografie se stala nedílnou součástí při monitorování plodu v průběhu porodu. Jedná se o záznam srdeční variability plodu a děložních kontrakcí. V průběhu porodu se nedostatečné zasobování plodu kyslíkem projevuje změnami v srdečním rytmu, a to jednak ztrátou variability či dlouho trvajícími deceleracemi po stahu dělohy. Vyhodnocování srdeční variability je subjektivní a nespolehlivé. Proto se tato práce zabývá automatickým hodnocením srdečního rytmu plodu. Pro analýzu srdečního rytmu byly použity nelineární metody fraktální dimenze, entropie a komplexity. Výsledky těchto metod byly porovnány z hlediska schopnosti rozlišení normálních a patologických plodů. Pomocí Mann-Whitney U statistického testu byla na hladině významnosti $p < 0.01$ zamítnuta nulová hypotéza, že mediány stanovených příznaků normálních a patologických plodů jsou stejné. Dále byla vykonána selekce příznaků, která ukázala, že nejdůležitější příznaky pro následnou klasifikaci srdečního rytmu plodu jsou: Lempel Ziv komplexita, fraktální dimenze signálu a sample entropie. Výsledky klasifikace dokazují, že nelineární příznaky jsou vhodné i pro analýzu krátkého úseku srdečního rytmu plodu. Klasifikací pomocí support vector machine bylo dosaženo sensitivity 78% a specificity 70%.

Klíčová slova Kardiotokografie, Srdeční rytmus plodu, Analýza časových řad, Chaos, Nelineární metody, Selekce příznaků, Klasifikace

Acknowledgements

Let me thank here all the people who helped me with this work. My thanks are especially to my supervisor, Václav Chudáček, for his friendship, patience, suggestions, and neverending optimism.
I am also grateful to Göran Salerud, my supervisor during my stay in Sweden, who let and help me to pursue my own ideas and directions. I would like to thank to Christer Ahlström whose work on phonocardiographic signals introduced me to the field of nonlinear time series analysis.
I am grateful to Jana, my sister, for the proofreading and valuable comments on the manuscript prior to publication.
At last but not at least I would like to express my gratitude to my parents Jirka and Vera who have supported me during my studies. Without them this work would not have been possible, literally.

Declaration

With my signature I certify that this thesis is my own work. All external materials, i.e. anything not being my own reasoning and written not by myself, is distinctly quoted or paraphrased and provided with a clear reference to the source. This applies to literature and software as well as other people's ideas and results.

Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil pouze podklady (literaturu, projekty, software atd.) uvedené v přiloženém seznamu.

Prague, May 20, 2009

Jiří Spilka

Contents

List of Abbreviations and Symbols

Abbreviations

AMI	Auto Mutual Information
ApEn	Approximate Entropy
$\mathbf{BD_{ecf}}$	Base Deficit
BPM	Beats Per Minute
CAFE	Computer Aided Fetal Evaluation
CS	Cesarean Section
CTG	Cardiotocogram
DFA	Detrend Flucations Analysis
ECG	Electrocardiogram
EFM	Electronic Fetal Monitoring
fBm	fractional Brownian motion
FECG	Fetal Electrocardiogram
fGn	fractional Gaussian noise
FHR	Fetal Heart Rate
FHRV	Fetal Heart Rate Variability
FSB	Fetal Blood Sampling
IUP	Intrauterine Pressure
KS	Kolmogorov-Sinai Entropy
LZC	Lempel Ziv Complexity
NICU	Neonatal Intensive Care Unit
NST	Non-Stress Test

PSD	Power Spectral Density
SampEn	Sample Entropy
SMOTE	Synthetic Minority Over-sampling Technique
STAN	ST ANalysis

Symbols

α	parameter of detrend fluctuation analysis
β	spectral index
$C(r)$	correlation sum
D_0	box counting dimension
D_1	information dimension
D_2	correlation dimension
D_σ	variance fractal dimension
Θ	Heaviside function
H	Hurst exponent
ϵ	length of box
m	embedding dimension
r	radius for correlation sum

List of Figures

List of Tables

Chapter 1

Introduction

"I am not discouraged because every wrong attempt discarded is another step forward."
Thomas Alva Edison

Being born is one of the most crucial events in our life. After intrauterine growth and development a baby is going to establish itself as an independent individual. To handle the labour stress a fetus has to be equipped with a defence mechanism. Good understanding of how an individual fetus reacts to the labour stress helps us to intervene in appropriate manner when the fetal defense has been activated but before risk of long-term consequences increases. To be more specific, during labour, fetus can repetitively suffer from oxygen insufficiency and, as a consequence, a metabolic acidosis could be developed. Severe hypoxic injuries can lead to neuro-development disability, cerebral palsy, neonatal encephalopathy or even death. We are trying to prevent these adverse outcomes using methods of electronic fetal monitoring (EFM).

The predecessor of EFM was auscultation where fetal heart rate (FHR) was sensed by fetal stethoscope. However, the stethoscope could not detect subtle changes in FHR and continuous monitoring was also impracticable. Introduction of EFM overcomes these disadvantages and offers continuous fetal surveillance. Nowadays, EFM has become a generally accepted method used during pregnancy and labour and gives important information about fetal behaviour. EFM involves recording of the cardiotocogram (CTG) and, recently introduced, ST analysis of the fetal electrocardiogram (FECG). The CTG consists of two signals the instantaneous activity of the fetal heart and uterine pressure.

The introduction of CTG monitoring in clinical settings significantly reduced the incidence of birth asphyxia, however, it has also contributed to the rise in cesarean sections and, as a consequence, potential maternal morbidity (Steer, 2008). The main reason of the failure to meet expectations was poor interpretation of CTG; therefore, guidelines for CTG were introduced in order to improve interpretation and thus to lower the number of asphyxiated neonates and also to decrease incidence of cesarean sections (FIGO, 1986; NIH, 1997). Although the guidelines are available a poor interpretation of CTG still persists (Doria et al., 2007; Westerhuis et al., 2007a) and variations in assessment are not only inter-observer but also intra-observer (Bernardes et al., 1997; Palomaki et al., 2006). Moreover, it has been reported that guidelines are sometimes violated which results in inconsistent assessment and possibly incorrect interventions. The assessment of CTG still remains subjective and difficult to reproduce and standardize. The introduction of ST analysis of FECG, provided by STAN (Neoventa Medical, Moelndal, Sweden), in clinical practice improved the labour outcomes (Amer-Wahlin et al., 2001; Noren et al., 2003). This method involves evaluation of morphological changes in fetal ECG. More specifi-

cally, it has been observed that fetal oxygen deficiency is pronounced in increased T/QRS[1] ratio and changes in ST segment morphology (Amer-Wahlin, 2003). Despite the proven benefit of this approach, it has also been reported that ST analysis not always decreases the number of neonates with developed acidemia (Ojala et al., 2006). Nonetheless, only this study disproved the contribution of ST analysis.

A recently published paper (Steer, 2008) concluded that weakness of EFM still lies in generally poor standard of CTG interpretation and the contribution of human factor as well. Therefore, a more education and intense training on CTG interpretation should be performed (Doria et al., 2007; Westerhuis et al., 2007a) or one should use more cost-effective solution by developing an expert system serving as source of additional information helping to validate or invalidate a hypothesis about fetus status (Steer, 2008).

1.1 The goals of thesis

As being explained above, an automatic assessment and interpretation of labour progress is of major importance. Therefore, this thesis focuses on the automatic analysis of signals recorded during labour. The thesis purpose is to analyse fetal heart rate by nonlinear methods and then select the most suitable methods for fetal heart rate. In addition, the results of analysis are classified and decision whether is fetal behaviour normal or abnormal is performed. This information could be in future provided to obstetricians to help them with labour progress. According to the CTG assessment and other aspects such as fetal blood samples (FBS), ST segment analysis, and mother's behaviour, obstetricians decide if the labour progress continues in order or if an intervention is needed. The aims of this work and steps to successful implementation are as follows:

- familiarization with fetal physiology and behaviour during labour (chapter 2)
- review of current methodologies for computerized CTG assessment (chapter 3)
- data processing and analysis
  - pre-processing – noise reduction, resampling, and interpolation (chapter 4)
  - nonlinear methods – choice of suitable and descriptive nonlinear methods, application to fetal heart rate (chapter 4 and 5)
  - data analysis, feature selection, and classification – analysis of fetal heart rate, selection of the best performing methods, classification (chapter 6)

The classification was performed on database of CTG recordings acquired at Charles University hospital in Prague during June 15th 2006 - January 12th 2009. This database contains pregnancies with different risk of possible labour complications. The broad spectrum of labour outcomes was necessary to be able to prove statistical significance of our solution.

[1] Waves in ECG signal are marked in alphabetical order (P,Q,R,S,T). QRS complex consists of Q,R, and S wave and represents depolarization of the fetal heart while repolarization is described by T wave

Chapter 2

General background

Labour is a very stressful period for fetus as well as for mother. Fetus is affected by mother's behaviour and condition. The way fetus reacts to its changing environment gives an important information about its status. For instance, a change in fetal heart rate can be caused by nervous system that is activated by receptors reacting to the change of internal environment.

One of the major fetus's tasks is to handle reoccurring hypoxic events that could lead to severe consequences for further child development. Fetus has its own physiological protective mechanism able to sustain repetitive hypoxic episodes. However, if the fetus is not able to adequately response or to recover from hypoxic stage, the hypoxia could be developed into the next stage of oxygen deficiency called asphyxia that could lead to cerebral palsy, neonatal encephalopathy or even to death. (Pierrat et al., 2005) reported prevalence of birth asphyxia to be 0.86 per 1000 term live births. Moreover, they examined ninety neonates with moderate or severe newborn encephalopathy with prevalence 1.64 per 1000 term live birth. The main cause of newborn encephalopathy was birth asphyxia, diagnosed in 52% cases. From these cases, asphyxia was caused intrapartum in 56% of cases, antepartum in 13%, ante-intrapartum in 10%, and postpartum in 2%. In 19% of cases, no underlying cause was identified during the neonatal course. The following risk factors, occurring antepartum or intrapartum, contribute to high prevalence of neonatal encephalopathy and to perinatal mortality as well. Antepartum risk factors are: preeclampsia, fetal hypertrophy, prematurity[1], multiple pregnancies, and diabetes mellitus. More important factors are those occurred intrapartum because 56% of cases with born asphyxia were caused intrapartum. These factors involve: bleeding during labour, epidural analgesia, intrauterine infection, meconium-stained liquor, post-term delivery, induced labour, and cesarean section (NIHCE, 2007). In presence of some risk factors, electronic fetal monitoring is necessary for fetal surveillance. On the other hand, for low risk pregnancies the use of electronic fetal monitoring does not offer significant contribution to fetal outcomes (Cunningham, 2005).

This chapter is organized as follows: first, we outline the basics of fetal physiology and fetus response to different stages of oxygen deficiency - hypoxanemia, hypoxia, and asphyxia. Next, we describe an interaction between mother and fetus during gestation with emphasis on the antepartum and intrapartum period. Finally, we introduce methods for the fetal hypoxia diagnostics with focus on electronic fetal monitoring that involves observation of CTG or FECG changes. We stress the significance of signal interpretation and describe advantages and disadvantages of respective methods.

[1] the birth of a premature infant

2.1 Fetal physiology

Fetal development lasts about 40 weeks. Complex systems, such as circulatory, respiratory, nervous, gastrointestinal, etc. are being developed during that time. In this work we discuss in detail only the circulatory system; the rest are mentioned only to give overall insight into fetal behaviour.

Fetal heart begins beating approximately at 4^{th} week of pregnancy with frequency about 65 beats per minute (BPM). This frequency increases during a gestation up to 140 bpm before delivery. The main function of fetal heart is to pump oxygenated blood from placenta to the organs and, in turn, to carry carbon dioxide back to placenta, where an exchange between mother and fetus is maintained. The exchange is not limited to gases only, but is performed for all substances such as nutrition and fetus' waste products.

Fetal circulation The oxygenated blood from mother's aorta is distributed to the uterine arteries and further to the spiral arteries that deliver blood to placenta. Here, in the thin capillaries membranes, the exchange of gases and substrates is performed. The fetus respiration system is non-functional and placenta works as the fetal lungs. Therefore, blood flows bypass lungs by ductus arteriosus. The same situation applies for liver, only with the difference, that liver are partially functioning and blood is not completely bypassed by ductus venous. The whole organization of the fetal circulation is illustrated in figure 2.1. The oxygenated blood from placenta enters the right atrium and continues directly to the left atrium throughout foramen ovale. From there it is pumped into aorta and then back to placenta via umbilical arteries. The deoxygenated blood returning from the upper part of the body enters the right atrium and is pumped into the right ventricle. Then, after ventricle contraction, blood is pumped through ductus arterious into the descending aorta.

2.1.1 Energy metabolism

Placenta maintains an exchange of oxygen and carbon dioxide between mother and fetus. This exchange can only be performed due to different partial pressures of gases. In placenta, oxygen is bound to haemoglobin and released in the capillaries in the fetal circulatory system. There the carbon dioxide replaces the oxygen and is carried back to placenta. Depending on oxygen availability we distinguish aerobic and anaerobic metabolism. These are illustrated in figure 2.2. The aerobic metabolism utilises glycogen (or fatty acids), oxygen, adenosine diphosphate (ADP), and phosphate (P) in order to create adenosine triphosphate (ATP) which serves as energy source. The waste product are carbon dioxide and water.

$$glycogen\ or\ fatty\ acids + P + ADP + O_2 \longrightarrow CO_2 + H_2O + ATP$$

The anaerobic metabolism is also used for glycolysis but with the difference that the oxygen is not available and cannot be used. The reaction for creating adenosine triphosphate is as follows:

$$glycogen\ or\ glucose + P + ADP \longrightarrow lactic\ acid + ATP$$

The waste product of anaerobic metabolism is lactic acid. The anaerobic metabolism only provides energy for basal (vital) activity and, as a consequence, fetus growth is restricted.

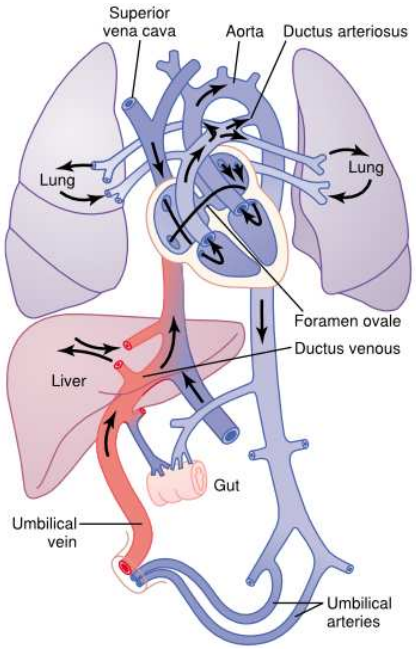
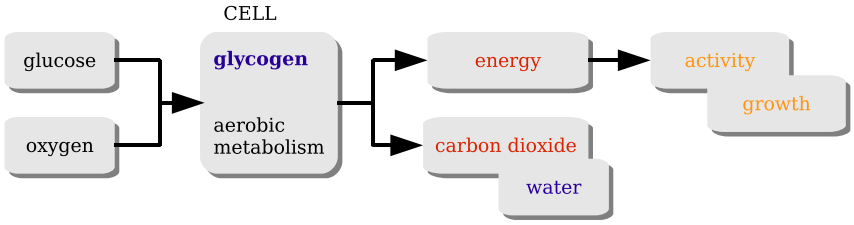
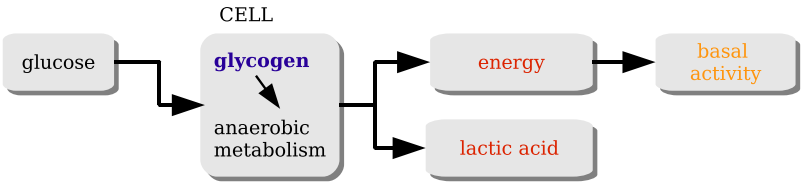


Figure 2.1: Organization of the fetal circulation. The difference between fetal and neonatal circulation lies in so called "blood shortcuts". It involves ductus venous, ductus arterious and foramen ovale. If these are not closed at the first breath, there is a serious risk for new born child and his/hers further development (Guyton 2005).



(a) aerobic metabolism



(b) anaerobic metabolism

Figure 2.2: Energy metabolism. The aerobic metabolism is oxygen dependent. In cases of oxygen insufficiency, the so called anaerobic metabolism produces enough energy to cover basal activity (modified from (Sundstrom et al., 2000)).

Therefore, the anaerobic metabolism should not last for hours. If the supply of oxygen is not re-established, hypoxanemia, hypoxia, and sequentially asphyxia are developed. These terms express different stages of decreased oxygen saturation of the fetal artery blood. Asphyxia is the last and worst stage that might occur. Before describing the individual stages, it is neces-

sary to explain autonomic nervous system and its reaction to oxygen deficiency. This system adapts fetal heart rate to changing environment and regulates blood distribution. It consists of humoral and neural (parasympathetic and sympathetic) systems that function antagonistically. Parasympathetic system reacts rapidly on abrupt changes, whereas the sympathoadrenal system works at more fundamental level prevailing during stage of fetal hypoxia (Amer-Wahlin, 2003). Parasympathetic activation causes reduction in fetal heart rate called bradycardia, while sympathetic activation leads to surge of stress hormones from the adrenals and FHR may increase up to tachycardia. It is worth to mention that transition between sympathetic and parasympathetic system is not linear, i.e. changing constantly in time, but rather shows non-linear behaviour (Goldberger et al., 2002). In figure 2.3 is illustrated how the nervous systems reflect a change in blood gases.

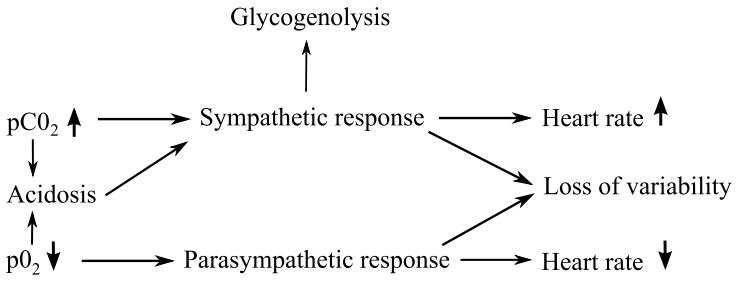


Figure 2.3: Relationship between blood gases and heart function (Amer-Wahlin, 2003).

2.1.2 Hypoxanemia

Hypoxanemia is an initial stage of oxygen deficiency. The oxygen is depleted in the arterial blood at the periphery. Central organs and peripheral tissues are intact and enough oxygen is provided to maintain aerobic metabolism. The fetal response is activated by chemoreceptors located in major vessels. It involves several safety precautions. First, the more efficient uptake of oxygen is performed by increased blood flow or increased number of erythrocytes. Second, the fetal movements are reduced and also growth is restricted in order to save the oxygen. The fetus can sustain hypoxanemia for days and weeks. However, in presence of fetal hypoxanemia before labour, fetus has less ability to handle labour stress because of restriction of energetic reserves.

2.1.3 Hypoxia

Hypoxia represents second stage of oxygen deficiency when the peripheral tissues are affected. Blood flow is redistributed in favour of central organs guaranteeing aerobic metabolism. On the contrary, anaerobic metabolism is utilised at peripheral tissues. The prime reaction to hypoxia is surge of stress hormones (adrenalin, noradrenalin) and sympathetic activation. Without any damage to fetus, hypoxia can last several hours.

2.1.4 Asphyxia

Asphyxia is the most critical stage. The oxygen is depleted and high priority organs utilise anaerobic metabolism. The energy is created from glucose stored in liver and myocardium. Brain has very low glucose level, therefore glucose is supplied by liver. The fetal response to asphyxia involves release of stress hormones and activation of sympathetic nervous system. The fetus attempts to maintain function of central organs as long as possible. The final stage of asphyxia is the collapse of system with brain and heart failure. Asphyxia that lasts only few minutes might cause irrecoverable damage.

2.2 Diagnosis of hypoxia

In the previous section we described the fetal behaviour. The reliable assessment and diagnosis of changes in fetus condition is of major importance. The fetus hypoxia activate defense mechanism and anaerobic metabolism is utilised at the peripheral tissues. Using diagnostic tools we detect and evaluate these changes. The diagnosis can be roughly split into two groups: fetal blood measurement (fetal blood sampling, pulse oximetry) and electronic fetal monitoring (cardiotocogram, fetal electrocardiogram).

Let us now focus on the invasive (fetal blood sampling) and non-invasive (pulse oximetry) measurement of oxygen content in fetal blood. The EFM is described in detail later in this section.

Fetal blood sample FBS is used in conjunction with EFM and serves as an accurate tool for measurement of metabolic acidosis. In case of non-reassuring patterns on CTG or FECG the FBS might be performed in order to acquire precise value of pH. The small sample is obtained from fetal scalp capillary. The amnioscope is inserted into the vagina and fetal scalp is cleaned. Then a small stab incision is made and fetal blood sample obtained. Eventually, the automatic blood-gas analysis is performed and value of pH is obtained. If abnormalities of EFM persist, the whole procedure is repeated. The whole process of fetal blood sampling is presented in figure 2.4.

Figure 2.4: The technique of fetal blood sampling using amnioscope (Cunningham, 2005).

The normal pH values of blood samples from fetal scalp capillary ranges from 7.25 to 7.35. The borderline, i.e. low normal pH is between 7.20 – 7.25. When the pH value is below 7.20, the intervention is needed and fetus should be immediately delivered. The fetal blood sampling

requires expertise and is time-consuming. It may also cause complications (Cunningham, 2005) but it is generally considered to be safe (Ojala et al., 2006).

Pulse oximetry This method uses light reflection from blood where light is differently reflected or inhibited depending on the oxygen saturation ($\mathrm{FSp0_2}$) in fetal blood. The electrode emitting and receiving light is placed against fetal scalp and continuous $\mathrm{FSp0_2}$ is acquired. However, as it was noted by (Cunningham, 2005; Steer, 2008), low oxygen saturation has poor specificity for acidosis. Therefore, application of pulse oximetry made no significant contribution to any measures of fetal outcome and is not commonly used for fetal surveillance.

2.2.1 Cardiotocogram

As we already pointed out, the fetal heart rate reflects changes in fetal behaviour and condition. The electronic fetal monitoring was introduced in 1960s and is a successor of auscultation method where the FHR was monitored periodically by stethoscope. Cardiotocogram (CTG) involves monitoring of fetal heart rate and uterine pressure. It offers valuable insight into fetal condition and serves intrapartum as well as peripartum (the admission CTG) when it might diagnose potential fetal compromise.

Cardiotocogram recording

We distinguish two types of CTG monitoring based on different stages of labour. Before rupture of membranes the external ultrasound probe and transducer are used to acquire FHR and uterine pressures (called TOCO), respectively. After the rupture of membranes the electrod is attached at fetus scalp and FHR is computed directly from ECG's R-R intervals. The uterine pressures are obtained using internal electrode placed in vagina. The record is called intrauterine pressure (IUP). The external and internal monitoring is shown in figure 2.5.

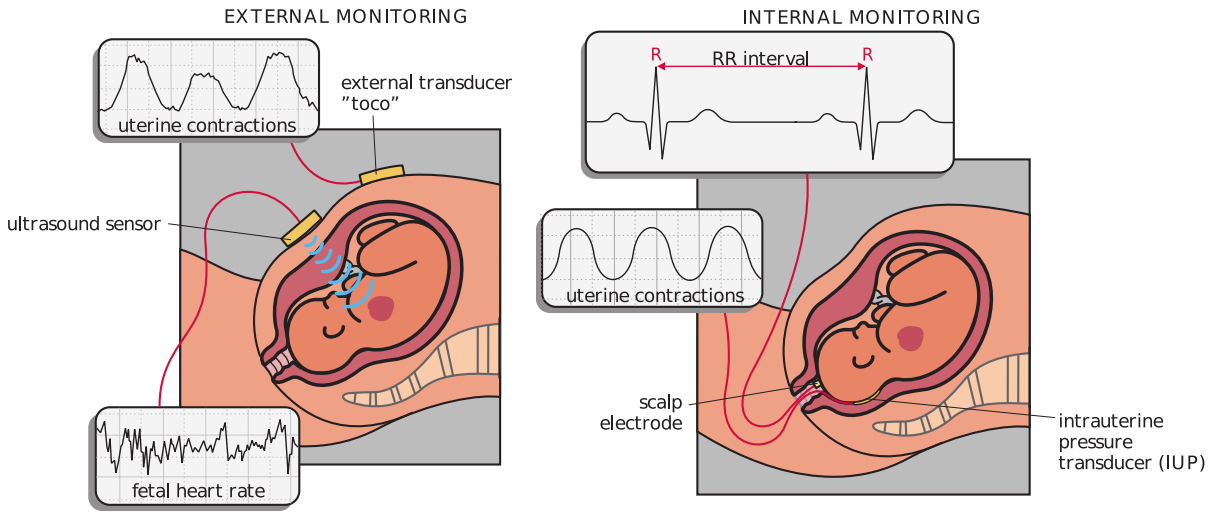


Figure 2.5: Recording of fetal heart rate and uterine activity (Sundstrom et al., 2000).

External monitoring has certain limitations in comparison to internal. In external monitoring the ultrasound Doppler principle is utilised to detect fetal heart pulsations. Therefore the

ultrasound probe must be located precisely at the position of fetal heart and any movement either mother's or fetus' may cause distortions. The probe is usually held in the position by a belt. Further problem is the separation of fetal and maternal heart. The fetal heart may be confused by maternal aortic pulsation resulting in misleading monitoring. The great advantage of external monitoring lies in easy application and non-invasibility. Internal monitoring is invasive and can be used only when the fetal position is normal, i.e. head first presentation, and after fetal membranes' rupture. The electrode is screwed to fetal scalp without any damage to fetus and complete electrocardiogram is acquired. Then the fetal heart rate is computed as difference of successive beats. The pressure transducer is placed in vagina and intrauterine pressure (IUP) is recorded. The internal monitoring has a higher signal to noise ratio than the external one and, in addition, FECG and its morphological changes can be examined (Sundstrom et al., 2000).

Changes in fetal heart rate Let us focus on changes that affect fetal heart rate. These changes either may occur during oxygen insufficiency or could be caused by aspects, such as mother behaviour or external influences. The FHR changes and its causes are as follows:

- Normal changes – the FHR is different during quiet and active sleep (REM[2]). There are rapid shifts in autonomic nervous system resulting in accelerations and increased heart variability during active sleep.
- Changes in placental blood flow – mainly due to cord compression. When the cord is compressed, the blood is pushed into fetus. The heart must pump more blood and the heart rate increases. The increase in blood volume results to increase in blood pressure. Hence, sensitive baro-receptors are activated and cause decrease in fetal heart rate. When the compressed cord is released, the FHR returns to normal.
- Adaptation to oxygen insufficiency - when oxygen content decrease, chemo-receptors are activated and stimulate sympathetic and parasympathetic nervous system. The changes in fetal heart rate depends on the stages of hypoxia. In case of acute hypoxanemia, immediate fall in FHR occurs while gradually developing hypoxia causes increase in FHR.
- External stimuli – due to the contraction there is an increase of head pressure that may cause deceleration. Also pressure on eye bulb might induce bradycardia.
- Increase in mother's temperature – in case of mother fever, the fetal metabolism increase which leads to higher oxygen consumption and may result in fetal tachycardia.
- The effect of drugs – the fetus could be affected by various drugs and the ability to handle labour stress may decrease, e.g. mother over-stimulation with oxycitocin results in increased uterine activity and fetus is affected by more intensive contraction.

Assessment of fetal heart rate changes

As mentioned above, CTG records have certain patterns and features that are important for fetus assessment and labour evaluation. Observation of these patterns and their correct interpretation is of major task in CTG monitoring. The following patterns are usually assessed in CTG records: baseline rate, variability, acceleration, and deceleration. These patterns and their

[2]rapid eyes movement

properties are strictly defined in guidelines for fetal monitoring (FIGO, 1986; NIH, 1997) and according to their occurrence the appropriate reaction is suggested. The normal CTG record is presented in figure 2.6. It shows accelerations and normal heart variability that are markers of fetal well-being.

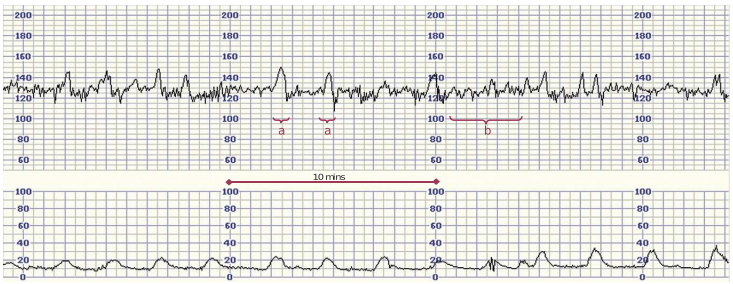


Figure 2.6: Normal reactive trace. (a) Accelerations; (b) normal variability (Hinshaw 2005).

Baseline heart rate Baseline fetal heart rate is determined over time period of 5 or 10 minutes when acceleration and deceleration are absent. Normal baseline rate is in range of 110 – 160 bpm. The decrease of heart rate below 110 bpm is called bradycardia and the increase of heart rate up to 150 bpm is called tachycardia, see figure 2.7.

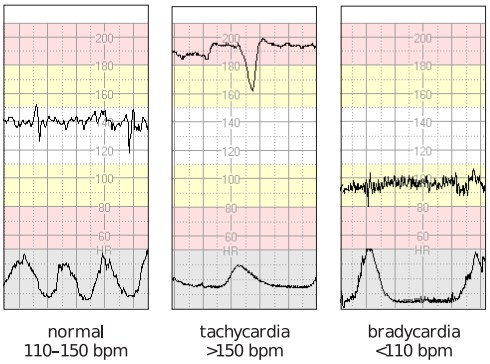


Figure 2.7: Normal baseline rate, bradycardia, and tachycardia (Sundstrom et al., 2000).

Variability FHR variability is defined as amplitude oscillations around baseline heart rate. Normal values are between 5–25 bpm. Example of normal variability is shown in figure 2.6. The so called saltatory pattern is an increase in variability of more than 25 bpm. Complete loss of variability for more than 40 minutes is the most abnormal sign and fetus may no longer finetune its circulation. The FHR could also have sinusoidal pattern with smooth, undulating sinewave. In case of sinusoidal pattern, immediate intervention is required.

Accelerations Acceleration is a transient increase in the heart rate of more than 15 bpm lasting 15 seconds or more. This is associated with fetal movements or stimulation, and indicates fetal well-being, see figure 2.6.

Decelerations Deceleration is characterized as a transient decrease of FHR below the baseline level of more than 15 bpm lasting at least 10 seconds. The decelerations are linked to uterine activity and distinguished as uniform or variable. Uniform deceleration has the same pattern and shape from one deceleration to another, whereas the variable decelerations might vary from one contraction to another; for illustration see figure 2.8.

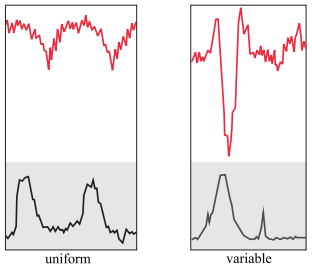


Figure 2.8: Uniform (rounded pattern, shape is similar) and variable (rapid loss of beats, pattern may vary) decelerations (Sundstrom et al., 2000).

Uniform decelerations can be further divided into early and late depending on time of occurrence. Early deceleration represents transient decrease in FHR when the drop in FHR matches the onset of contraction. On the contrary, late decelerations are characterized as those with different onset of the contraction and deceleration, see figure 2.9. Note that only late decelerations are connected with hypoxia.

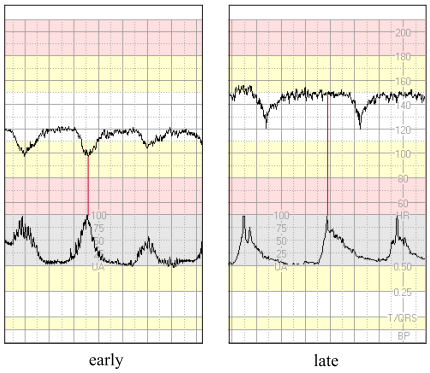


Figure 2.9: Early and late decelerations. For late deceleration the onset of contraction and drop in FHR differs (Sundstrom et al., 2000).

The variable decelerations have different shape from one deceleration to another. As for uniform deceleration, the variable can be also split into two groups: uncomplicated and complicated. Uncomplicated deceleration are defined as deceleration lasting less than 60 seconds; below this time fetus is able to sustain it. When the deceleration is complicated, i.e. the duration exceeds 60 seconds, there is an increased risk of fetal hypoxia. Uncomplicated and complicated deceleration are shown in figure 2.10.

Guidelines for CTG categorization The CTG interpretation and classification is very subjective. To be able to characterize important patterns and features, the doctors have to be well

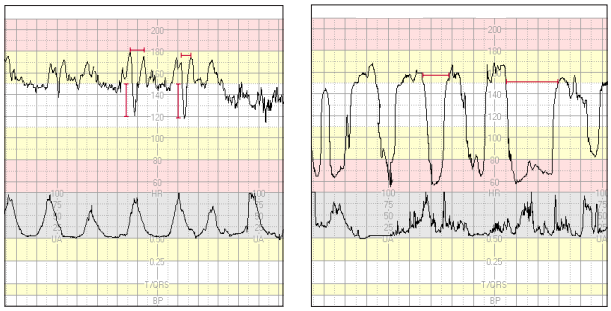


Figure 2.10: Uncomplicated and complicated variable deceleration. A deceleration of duration longer than 60 seconds is considered as complicated (Sundstrom et al., 2000).

educated and experienced. In order to standardize CTG interpretation, the status of fetus condition was divided into four categories: normal, intermediary, abnormal, and preterminal (FIGO, 1986; NIH, 1997; Sundstrom et al., 2000). These categories were developed during years and are commonly applied. The classification of CTG is based on patterns and features explained above. The occurrence of particular features or their combinations indicate either fetal well-being or its adaptation to changing environment. The complete description of fetus status is presented in table 2.1. When the CTG is intermediary, the fetus is suspected to compromise. If the fetus' condition changes and CTG features are evaluated as abnormal or preterminal, an immediate intervention is needed.

CTG classification	Baseline heart frequency	Variability Reactivity	Decelerations
Normal CTG	• 110–150 bpm	• 5–25 bpm • Accelerations	• Early decelerations • Uncomplicated variable decel. with a duration of <60 sec and loss of <60 beats
Intermediary CTG	• 100–110 bpm • 150–170 bpm • Short bradycardia episode	• >25 bpm without accelerations • <5 bpm for >40 min	• Uncomplicated variable decel. with a duration <60 sec and loss of >60 beats
	• A combination of several intermediary observations will result in an abnormal CTG		
Abnormal CTG	• 150–170 bpm and reduced variability • >170 bpm • Persistent bradycardia	• <5 bpm for >60 min • Sinusoidal pattern	• Complicated variable decel. with a duration of >60 sec • Repeated late decelerations
Preterminal CTG	• Total lack of variability and reactivity with or without decelerations or bradycardia		

Table 2.1: Categorization of CTG trace features into normal, intermediary, abnormal, and preterminal (Sundstrom et al., 2000).

2.2.2 Fetal electrocardiogram analysis

ST analysis of fetal electrocardiogram was successfully introduced into clinical practise by Neoventa Medical, Moelndal, Sweden. This technique is commonly referred to as STAN®(ST ANalysis). Contrary to CTG, the complete ECG curve is used to examine and evaluate morphological changes. The ST analysis is not intended to be used autonomously but only as addition to standard CTG. It serves as source of additional information validating or invalidating hypothesis of fetal condition and behaviour observed on CTG. The analysis of ST segment is well established in detecting and monitoring of myocardial insufficiency in adults cardiology and the development of ST analysis of fetal ECG has been based on this experience and knowledge. The fetal brain and heart are equally sensitive to changes in oxygen content; therefore, myocardial function serves as indirect measurement of brain condition.

The ECG signal is acquired by internal electrodes screwed into the fetal scalp without any damage to fetus. The continuous ECG is displayed and important markers of ECG are automatically computed. These markers involve changes in T wave amplitude and ST segment. For illustration of important ECG waves and intervals see figure 2.11.

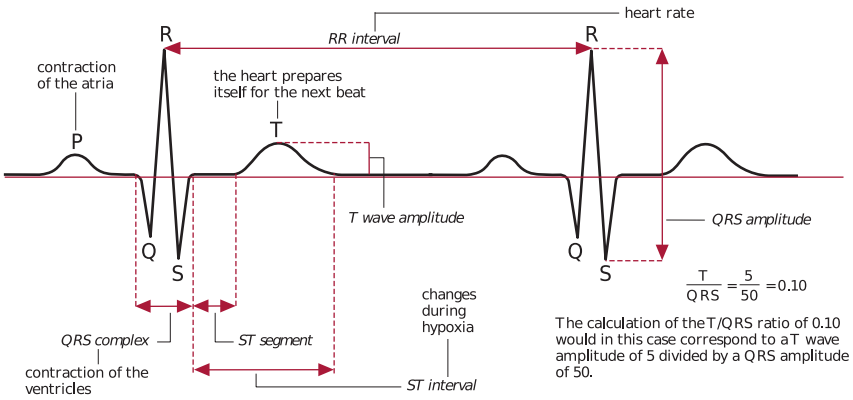


Figure 2.11: The representation of ECG curve and its important features (Sundstrom et al., 2000).

The T wave amplitude is used for computation of T/QRS ratio. This is performed periodically on ensamble average of several consequent beats. An increase in T wave reflects to fetus hypoxia and the degree of rise corresponds to degree of hypoxia. The second important feature of ECG is ST segment where its changes are examined. The biphasic ST is defined as a downward-leaning ST segment. We distinguish different degrees of biphasic ST segment starting at Grade 1 and continuing to Grade 2 and 3. With progression of disturbance in myocardial function, there is a shift in degree from Grade 1 to Grade 2 or even to the worst Grade 3. The morphologies of particular biphasic degrees are shown in figure 2.12.

As mentioned above, the ST analysis should be performed after occurrence of suspected patterns on CTG. The sole assessment of ST segment could lead to misleading results and rise in the labour intervention (Sundstrom et al., 2000) and potential adverse outcomes for fetus as well as for mother. As for the CTG, the interpretation of ST segment was standardized and guidelines were created in order to avoid subjective assessment of ST changes. In this guidelines, see table 2.2, we distinguish three types of events: episodic T/QRS rise, baseline T/QRS rise, and biphasic ST. The T/QRS rise is considered as episodic when the T/QRS rises and returns to the baseline in time period no longer than 10 minutes. The degree of change indicates the fetal stress

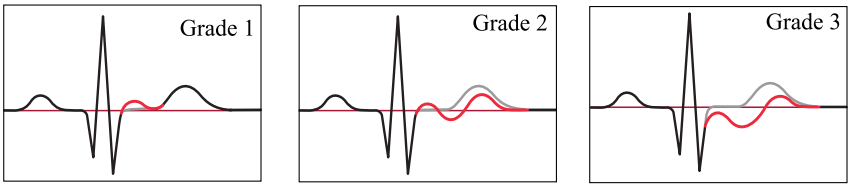


Figure 2.12: The morphology of biphasic ST. In Grade 1 the ST segment is above isoelectric line whereas in Grade 3 is completely below (Sundstrom et al., 2000).

and corresponds to short lasting hypoxia. The T/QRS increased of more than 0.10, in connection with abnormal CTG, is considered as significant and registered as an ST event. Baseline T/QRS rise is similar to episodic rise with exception that increase of T/QRS has duration longer than 10 minutes. The baseline T/QRS rise of 0.05 with CTG classified as abnormal, is consider as significant and indicates persistent stress and zero opportunity to recover. The last event assessed is the biphasic ST with different degrees where the degree corresponds to the degree of abnormality. The grade 2 and 3 are generally considered as abnormal. The above mentioned events are connected with CTG interpretation in guidelines that are called STAN simplified guidelines; see table 2.2.

ST Events	Normal CTG	Intermediary CTG	Abnormal CTG	Preterminal CTG
Episodic T/QRS rise	• Expectant management • Continued observation	• >0.15	• >0.10	• Immediate delivery
Baseline T/QRS rise		• >0.10	• >0.05	
Biphasic ST		• 3 Biphasic log messages	• 2 Biphasic log messages	

Table 2.2: ST analysis guidelines in association with CTG classification (Sundstrom et al., 2000).

2.3 Assessment of labour and neonate outcome

In the previous section we introduced electronic fetal monitoring as the methodology to identify fetal distress and oxygen insufficiency. When child is born, we need to assess its status in order to acquire additional information wheather to what extent baby suffered. The commonly used methods for assessment are Apgar score, cord acid-base analysis, and the occurrence of neonatal complications.

2.3.1 Apgar score

This method was devised by Virginia Apgar in 1953. It was not initially intended to asses neonates that suffered from asphyxia. However, this methodology was established in clinical settings and is widely used. The Apgar score includes five parametres that are examined at the neonate's age of 1, 5, and 10 minutes. These parametres are heart rate, breathing, skin color, muscular tone, and excitability. Each parameter is given score in range of 0 - 2 points and then

all parametres are summed up giving the score at particular child's age. The maximum score that could be achieved is 10 points. Note that the assessment of child is subjective.

There is a high correlation between low Apgar score and neonates that suffered from asphyxia during labour. However, there are also many reasons for low Apgar score that are not related to asphyxia, such as immaturity, labour trauma, drugs, infection, the activation of reflexes through manipulation of the upper airways, meconium aspiration or carbon dioxide narcosis (Sundstrom et al., 2000).

The Apgar score below or equal to 7 at 5 minutes is generally considered as an indicator of metabolic acidosis (Doria et al., 2007; Westerhuis et al., 2007b).

2.3.2 Acid-base analysis

The acid-base analysis is used to directly estimate the degree of metabolic acidosis. When child is born, the cord is immediately clamped and samples are taken from artery and vein. From these samples the value of artery and vein pH is calculated. The value of pH is evaluated solely or could be also used with value of carbon dioxide content for further analysis of base deficit ($\mathrm{BD_{ecf}}$). This marker is suggested in addition to pH and expresses the base deficit in the extracellular fluid. $\mathrm{BD_{ecf}}$ indicates the amount of buffers[3] in both blood and tissues which have been utilised in order to buffer hydrogen ions that had caused the drop in pH.

The normal artery pH ranges from 7.05 up to 7.38 and value of $\mathrm{BD_{ecf}}$ from -2.5 to 12.0 mmol/l. The neonatal acidemia is defined as pH below 7.05 and $\mathrm{BD_{ecf}}$ exceeding 12 mmol/l. These values were suggested in the following works (Sundstrom et al., 2000; Westerhuis et al., 2007b). Nevertheless, long term observation is highly desireable to more precisely connect values of pH and $\mathrm{BD_{ecf}}$ with severe consequences for further child development. (Pierrat et al., 2005) examined neonates with newborn encephalopathy and assessed their outcomes at 2 years of age, though, they did not included pH values into the study. Based on the computerized data analysis, (Salamalekis et al., 2002) suggest the reasonable pH borderline to be at 7.15. This value was also used previously in the work of (Chung et al., 1995). Considering these facts and on recommendation by experienced obstetricians at the 1st Faculty of Medicine, Charles University, Prague we used pH borderline of 7.15.

[3] haemoglobin, protein, and bicarbonate

Chapter 3

Automatic assessment of CTG – a review

A lot of attempts have been made to tackle the unresolved problem of reliable automatic analysis of CTG signal but, unfortunately, none of them were successful enough to be able to meet demands and expectation of clinicans. The automatic classification of fetus behaviour and condition is still challenge for many researches. In this chapter we briefly introduce solutions that were developed and used for automatic assessment of CTG records. The description of particular methods is concise and should serve rather as an introduction and overview than an exhaustive description. It is important to mention that none of the complete systems we are going to describe is widely applied in clinical settings. Each system is used merely used in the place or in the country where it was developed.

A first attempt for automatic CTG analysis was to follow the clinical guidelines used for CTG assessment (FIGO, 1986). These guidelines involved morphological changes in CTG and (Mantel et al., 1990a,b) developed an iterating procedure for their extraction. Note that these morphological features became fundamental for almost all methods that try to classify fetal status. The extraction of morphological features were improved by (Bernardes et al., 1991) and resulted in development of automatic system, SisPorto, for CTG analysis. This system is briefly described below. Linear and nonlinear methods used for FHR analysis were mostly derived from adults HRV research. This field was thoroughly investigated and a general agreement on HRV analysis exists (Task-Force, 1996). The statistical description of CTG tracings was employed in work of (Magenes et al., 2000) and then in following study of (Goncalves et al., 2006). Different approach of FHR analysis is to examine frequency content by spectral analysis. This analysis was performed by many research groups. The recent paper (Laar et al., 2008) gives a short overview of papers which analyzed spectrum to FHR either antepartum or intrapartum. Here, let us mention only the first attempt of (Cerutti et al., 1989). The wavelet transform can be considered as an extension of spectral analysis. The FHR is analyzed by different wavelets with different properties. This approach was utilised by (Salamalekis et al., 2002) and (Georgoulas et al., 2005).

Use of nonlinear methods for FHR analysis has also its roots in adults HRV research where these methods has proven their usefulness. The measure of fractal dimension of reconstructed attractor was performed by (Chaffin et al., 1991) and (Felgueiras et al., 1998). The later paper also examined waveform fractal dimension. A slightly different approach was applied by (Gough, 1993) who measured the length of FHR at different scales and thus estimated fractal dimension. Another attempt to measure length of FHR curve but with Higuchi method was made by (Kikuchi et al., 2005). However, they estimated fractal dimension during pregnancy

and not during labour. The different estimation of fractal dimension were reviewed by (Hopkins et al., 2006).

Probably the most successful nonlinear methods for FHR analysis are approximate entropy (ApEn) and sample entropy (SampEn). They are widely used for examination of nonlinear systems and also proved their applicability in FHR analysis. Let us mention only few studies that employed ApEn or SampEn (Pincus and Viscarello, 1992; Lake et al., 2002; Goncalves et al., 2006; Georgoulas et al., 2006). Another methods for nonlinear analysis are detrend fluctuation analysis applied by (Echeverria et al., 2004) and Lempel Ziv complexity used by (Ferrario et al., 2005).

So far we briefly introduced linear and nonlinear methods for FHR analysis. These methods are used either solely or in the combination. It is convenient to combine some methods together and then use a classifier to discriminate individual instances. Note that the combination of features forms a feature space that could have very high dimension. This problem is referred to as a curse of dimensionality. In order to reduce the dimension the feature selection or extraction are usually performed prior to the classification. (Georgoulas et al., 2006) employed principal component analysis (PCA) for dimensionality reduction and then used support vector machine for classification. (Magenes et al., 2000) employed artificial neural networks for the classification of linear features. (Chung et al., 1995) devised an algorithm for analysis and prediction of fetal acidosis. The exhaustive work of CTG analysis were performed by Georgoulas et al. For CTG classification they used Markov Models (Georgoulas et al., 2004), support vector machines (Georgoulas et al., 2005), and recently a hybrid approach utilising grammatical evolution (Georgoulas et al., 2007). They compared the classification performance of respective methods to conventional methods, such as k-nn (k-nearest neighbors), qdc (quadratic discriminant classifier), and ldc (linear discriminant classifier). Moreover, in the work of (Georgoulas et al., 2007) a synthetic minority oversampling technique (SMOTE) was used in order to balance number of normal and abnormal fetus' outcomes.

The complete systems used for fetal assessment mostly employ an expert system. The brief description of each system follows. *NST-Expert* (Non-Stress Test) (Alonso-Betanzos et al., 1995) is a noninvasive method used for fetal assessment by analysing fetal heart rate and uterine pressure. The main part of this software is an expert system that is capable of proposing a diagnose and treatment. Moreover, it might also estimate the potential problems of neonate. *CAFE* (Computer Aided Fetal Evaluation) (Guijarro-Berdinas et al., 2002) is successor of NST-Expert. It is an intelligent and hybrid system developed for CTG analysis in La Coruna, Spain. A neural net (NN) and rule based system are used for records classification. Moreover, they are also utilised for artefacts elimination and recognition of decelerations. *SisPorto* system has being developed by prof. Bernandes at al. at University of Porto, Portugal, since 1990. It consists of expert system which evaluates individual features described according to guidelines for CTG assessment. The description of system could be found in (de Campos et al., 2000). Nowadays, the SisPorto is in its 3$^{\text{rd}}$ version. *K2 Medical Systems* (Greene and Keith, 2002) has been developed at Plymouth University in Great Britain. It is distributed system consisting of central PC and local units that are situated at the patient's bed and gathering information, such as CTG and results of blood sample analysis. The alarm is evoked in case of abnormalities. The advantage of K2MS is its distributive character when mother can comfortably lie in the bed and still be under control.

Chapter 4

Signal processing and analysis

One of the most important aspects of signal processing is the quality of input data collection. In order to develop system for reliable classification of fetus status, the experimental data has to contain wide spectrum of fetal outcomes ranging from normal, healthy fetuses, to abnormal with serious metabolic acidosis. The further processing stages strongly depend on available data set. The data, we had available, includes variety of fetal outcomes and was collected during June 15th 2006 – January 12th 2009 at Charles University in Prague. The clinical description, fetus assessment and labour evaluation are available. CTG data usually contains artefacts caused by mother or fetal movements and displacements of the ultrasound transducer.

The traditional approach to CTG analysis is to study morphological changes of signal, i.e. baseline, variability, accelerations, and decelerations. The more advanced methods use either statistical theory of linear process or some transformation to have better insight into the data structure. Statistical analysis, used in time domain, examine changes in RR intervals and evaluate them either as a function of time or as a scalar value. The very useful insight into data, specifically to its frequency content, offer signal decomposition by Fourier transform. Frequencies are extracted from signal and can be individually analysed. Usually the frequency spectrum is divided into the frequency bands where each band corresponds to a particular physiological fetus behaviour.

Next approach, that has been recently introduced, is to consider FHR as a fractional Brownian motion (fBm). We would like to remark here that almost all methods used for FHR analysis have their roots in adults HRV research. (Goldberg et al., 1985) observed that human heart beat is self-similar, i.e. a signal observed at different scales has the same properties irrespective on the scale. Therefore, tools for examination of chaos and its dynamics can be used for measuring properties of FHR. Note that the time series can be analysed as monofractal or multifractal. The monofractal has same fractal properties irrespective on the time, whereas for multifractal properties are varying over time. Despite the fact that multifractal approach could be useful, in our work we considered FHR as monofractal since we are mainly focused on minutes preceding the delivery.

The above mentioned linear methods have proven their usefullness either in analysis of FHR or HRV. Nevertheless, in our work we especially focus on nonlinear methods that could reveal important clinical information hidden in time and frequency domain.

Estimated values of linear and nonlinear methods are, in machine learning field, called features. Individual methods are called attributes. The combination of features creates something what is called feature space with dimension equal to the number of features. A classifier oper-

ates in this space and try to discriminate two different classes, normal and pathological fetuses. The whole scheme of signal processing is shown in figure 4.1.

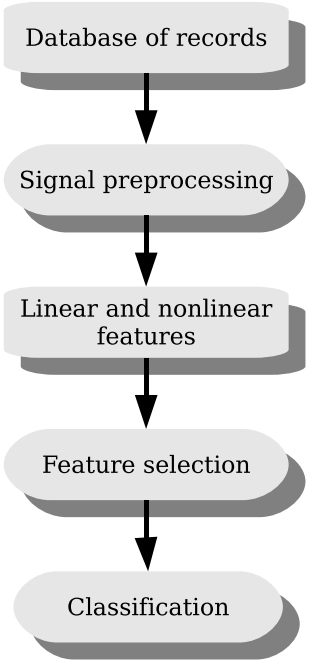


Figure 4.1: The overall scheme of signal processing.

Chapter at a glance. First, we present the available data set from medical and technical point of view. Second, we outline the preprocessing stage and then describe linear and nonliner methods for fetal heart rate analysis. The nonlinear methods are explained in more detail since are fundamental in this work. Next, we describe the concept of state space reconstruction and methods for dimension analysis in this space. We also introduce methods that estimate fractal dimension directly in the time domain. Other methods, useful for fetal heart rate analysis, are described as well. At the end of this chapter we introduce methods for feature selection and classification.

4.1 Data collection

For this work we used data set collected during June 15th 2006 – January 12th 2009 at the 1st Faculty of Medicine, Charles University, Prague. These records were acquired by device STAN S21 (Neoventa Medical, Moelndal, Sweden) that is capable to acquire either external or internal records. In this section we describe the used data from medical and technical point of view. In figure 4.2 we present sample record of fetal heart rate and uterine pressure. The record is shown in the raw form without any preprocessing.

Medical point of view

The used data set contains 189 records and clinical description is available for each record describing the labour progress and newborn with the following parametres: maternal age, gestational age, gestational pathology, weight, sex, Apgard score, artery pH, and vein pH. These

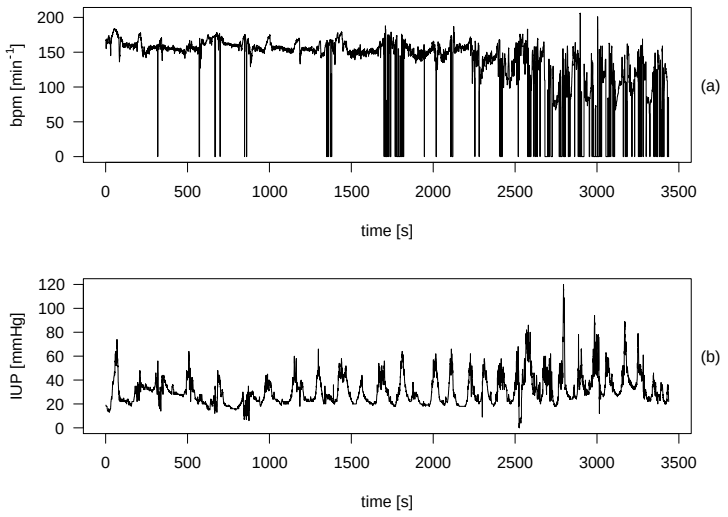


Figure 4.2: Raw record of CTG. (a) fetal heart rate, (b) intrauterine pressure.

clinical parameters and their association to adverse fetal outcomes were thoroughly described in chapter 2. Note that the clinical description is not complete for all records, though, our main interest is centered on the artery pH value which is presented in all clinical records. Hence this inconsistency of clinical description can be abandoned as insignificant.

In the clinical description the artery pH and Apgar score at 5 minutes are of major importance because it divides the data set into two groups: normal and abnormal. Fetuses are classified as normal if having pH above or equal to 7.15 or Apgar score at 5 min. higher than 7 and abnormal if having pH lower than 7.15 or Apgar score below or equal to 7. For more information about the pH borderline see section 2.3.2 where we referenced several studies that examined pH of newborns.

In table 4.1 we present summary of outcomes measures of all records (n = 189). From these records 94 are considered as normal and 95 as abnormal. Values of gestational age, Apgar score and pH are expressed as mean $\pm$ standard deviation. Statistical analysis showed that Apgar score and pH are significant on $p < 0.01$ confidence level.

	normal (n = 94)	abnormal (n = 95)
Gestational age (weeks)	39.2 ± 2	39.8 ± 1.3
Apgar at 1 min.^{a}	8.53 ± 1.25	5.62 ± 2.82
Apgar at 5 min.^{a}	9.66 ± 0.56	7.93 ± 2.33
Apgar at 10 min.^{a}	9.88 ± 0.33	8.91 ± 1.77
pH - artery^a	7.25 ± 0.06	7.04 ± 0.09
pH - vein^a	7.32 ± 0.06	7.14 ± 0.1

Table 4.1: Summary of outcome measures of neonates. Data are presented as mean $\pm$ standard deviation. ^{a}values are significant on $p < 0.01$ confidence level.

Technical point of view

STAN S21 automatically records all signals during a labour. If the labour progress is in order, FHR and uterine pressures are acquired using Doppler ultrasound. If suspicious, non-reassuring patterns occurs a scalp electrode is screwed into the fetal scalp and fetal electrocardiogram (FECG) is recorded. The uterine pressure is acquired using pressure sensors placed directly in vagina. Note that the scalp electrode can be attached only after fetal membrane rupture. In some records FECG has short duration or was not acquired, hence, we used ultrasound signals to properly classify these records. Note that STAN device is also equipped with second ultrasound probe for twins monitoring.

The acquired fetal hear rate by Doppler ultrasound is non-uniformly sampled. STAN S21 re-sample FHR with the sampling frequency $f_s = 2$ Hz and uniformly sampled FHR is provided. On the other hand, the computed RR intervals from FECG are provided with non-uniformly spaced time instances.

4.2 Data preprocessing

Preprocessing is the main part in every signal processing task and is always the first step to be made. Values of extracted features and further classification are highly dependent on the preprocessing quality. For instance preprocessing steps could distorts the deterministic nature of the data and add some stochastic components. The ideal signals for analysis would be those measured directly in the heart. This is, however, not possible and signals are measured either externally using Doppler ultrasound or internally by scalp electrode. As mentioned above, signals recorded externally have lower signal to noise ratio than those recorded internally but even internal records are contaminant with noise and artefacts.

In our case, the preprocessing is divided into four steps: artefacts removal, interpolation, choice of appropriate segment, and detrend. The segment should be chosen as close as possible to delivery because during last minutes a major changes in fetal condition occurs. However, as it is shown in figure 4.3, FHR directly preceding the delivery is largely contaminant with artefacts and noise. Therefore, we evaluate signal in terms of quality and choose the segment with sufficient quality that is closest to delivery. The segments are 20 minutes long which means 4800 samples for 4 Hz sampled signal.

4.2.1 Artefacts removal

The FHR signal contains a lot of artefacts caused by mother and fetal movements or displacements of the transducer. In general, the amount of data being removed as artefacts are in the range between 20% – 40% of all data. We employed the artefacts removal developed by (Bernardes et al., 1991). The algorithm for artefacts replacement is as follows: first, the successive five beats with difference lower than 10 bpm among them are considered as a stable segment. Then, whenever the difference between adjacent beats is higher than 25 bpm, the sample is substituted by linear interpolation between previous beat and new stable segment. Thus, all abrupt changes in FHR are removed and replaced by line. The result of artefacts removal is present in figure 4.3. Notice that the artefacts occur mostly at the end of labour.

We shall emphasize here the importance of missing data replacement. The results of analysis are affected by the way we treat gaps. We used Hermit interpolation of missing data which is

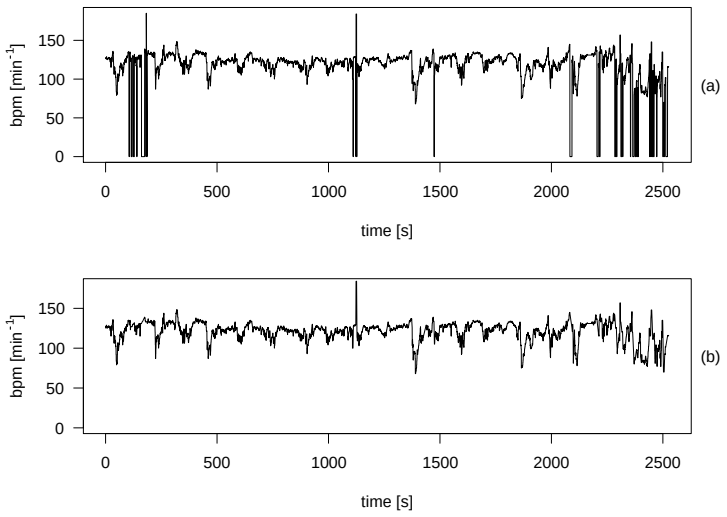


Figure 4.3: Removing of artefacts. (a) the raw signal with artefacts, (b) signal after artefacts removal.

possibly correct for fetal heart rate but holds only for gaps of short duration $t_{gap} \leq 20$ s. Instead of Hermit interpolation we could employ nonlinear prediction which is the best way for data replacement (Sprott, 2003). If the gap is of long duration, $t_{gap} > 20$ s, none of these methods work, and we have to skip the data and do not compute across the gap (Sprott, 2003).

4.2.2 Resampling the fetal heart rate

As mentioned above, the fetal heart rate is non-uniformly sampled. This might affect the results of nonlinear methods, such as fractal dimension and entropy, though, since sampling is deterministically non-uniform, everything is possibly correct (Sprott, 2003). Note that fractal dimension may increase because of additional equations that comes into play. (Eke et al., 2002) also stated that nonlinear methods work on non-uniformly sampled time series too but great caution is required. In order to examine effect of non-uniform sampling, we analysed results for both equidistant and non-equidistant data sets.

In order to resample data we applied Cubic Hermite Spline Interpolation (Store and Bulirsch, 1993). This method allows to find piecewise approximation of signal $x_1, x_2, \ldots, x_n$ by polynomial $H(x)$. First we find $H(x)$ such that

$$\begin{aligned} H_i(x_1) &= f(x_1) \\ H_i(x_2) &= f(x_2) \\ H_i'(x_1) &= f'(x_1) \\ H_i'(x_2) &= f'(x_2) \end{aligned} \tag{4.1}$$

Then we need a cubic polynomial to fit the four degreeds of freedom:

$$\begin{aligned} H(x) &= a + b(x - x_1) + c(x - x_1)^2 + d(x - x_1)^2(x - x_2) \\ H'(x) &= b + 2c(x - x_1) + d[2(x - x_1)(x - x_2) + (x - x_1)^2] \end{aligned} \tag{4.2}$$

now consider $x = x_1$ and $x = x_2$, i.e.:

$$\begin{aligned} y_1 &= H(x_1) = a \\ y_1' &= H'(x_1) = b \\ y_2 &= H(x_2) = a + b(x_2 - x_1) + c(x_2 - x - 1)^2 \\ y_2' &= H'(x_2) = b + 2c(x_2 - x_1) + d(x_2 - x_1)^2 \end{aligned} \tag{4.3}$$

a, b, c, d could be solved. The approximation is then sampled with equidistantly time space instances.

4.2.3 Detrend

Physiological time series are generally considered as nonstationary, i.e. statistical properties of physiological signal (mean, variance, and correlation structure) vary during time. We describe stationarity and nonstationarity in section 4.4.1. Here let us mention that we work with segments of short duration. Therefore, we can carefully detrend signal using third order polynomial and consider it as stationary. This polynomial is estimated, such that uninteresting trend is removed but interesting dynamics preserved.

4.3 Linear time series analysis

We examine oscillation in intervals between consecutive beats and also variations in difference of adjacent beats. For data analysis we use statistical methods in the time domain, such as first and second order statistics (Task-Force, 1996). Another approach is to examine frequency spectrum by Fourier transform. A signal is decomposed to its single frequencies where each frequency is represented either by amplitude or power.

4.3.1 Time domain

There exists a large amount of statistical methods used in heart rate variability analysis. In our work we mainly adopt those methods that have proven their usefulness in fetal heart rate analysis (Magenes et al., 2000). Let $x(i)$ be defined as the FHR signal for $n = 1, 2, \ldots, N$ where N is a length of FHR. The time domain features representing the variation between consecutive R-R intervals are as follows:

- The mean heart rate:

$$\mathrm{mFHR} = \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x(i) \tag{4.4}$$

- Standard deviation of the FHR:

$$\mathrm{SD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x(i) - \bar{x})^2} \tag{4.5}$$

- The Delta value:

$$\Delta = \frac{1}{M}\sum_{i=1}^{M}\left[\max_{i\in M}(x(i)) - \min_{i\in M}(x(i))\right], \tag{4.6}$$

  where M is the number of minutes of a signal.

- Short term variability:

$$\frac{1}{24}\sum_{i=1}^{24}|(sm(i+1) - sm(i))|, \tag{4.7}$$

  where $sm(i)$ is a value of $x(i)$ taken every 2.5 s.

- Long term irregularity:

$$\mathrm{LTI} = \mathrm{IQR}\sqrt{x^2(i) - x^2(i+1)}, \tag{4.8}$$

  where IQR is inter-quartile range with $i = 1, \ldots, N-1$.

- Interval index:

$$\mathrm{II} = \frac{\mathrm{STV}}{\mathrm{STD[sm(i)]}}, \tag{4.9}$$

  where STD expresses standard deviation and $sm(i)$ is again a value of $x(i)$ taken every 2.5 s.

- Total value of the Delta:

$$\Delta_{total} = \max_{i\in[1,N]}(x(i)) - \min_{i\in[1,N]}(x(i)), \tag{4.10}$$

4.3.2 Frequency domain

Signal decomposition into frequency components is very useful in signal processing field. With this approach we lose the notion of time and only frequency components of signal are provided. The power as a function of frequency constitutes to what is known as power spectral density (PSD). The PSD could be estimated by various methods. One of them is Fourier transformation which considers signal as a composition of cosine waves with different amplitudes, phases, and frequencies.

The PSD is usually divided into non-overlapping energy bands. These bands represent underlying physiological activity of either mother or fetus. The division of power spectrum into individual bands is not such straightforward as for adult heart rate variability and exact bands for fetal monitoring still remain unknown (Laar et al., 2008). Slightly different spectral bands were examined and described by (Sibony et al., 1994) and (Signorini et al., 2003). The former approach divides spectra into four bands: very low frequency VLF : 0 – 0.03 Hz, low frequency LF: 0.03 – 0.15 Hz that reflects sympathetic activity, mild frequency MF: 0.15 – 0.5 Hz which is associated with fetal movement and maternal breathing, high frequency HF: 0.5 – 1 Hz that represents fetal breathing[1], and LF/(MF + HF) ratio that corresponds with balance of two autonomous systems. Other frequency bands were proposed by Sibony et.al. They partitioned spectra similarly as Signorini et.al., with the modification that number of bands was

[1]Note that fetal lungs are non-functional and only movements are performed

reduced into three and boundaries of bands changed : very low frequency VLF : 0 – 0.05 Hz, low frequency LF: 0.05 – 0.15 Hz, high frequency HF: 0.15 – 0.5 Hz, and LF/HF ratio.

We shall note here that power spectral density of fetal heart rate has power law scaling relationship. The energy as a function of frequency decreases in power low fashion $1/f^{\beta}$. The spectral index β is estimated as a slope of line fitted to the spectrum estimate. The β equals 0 for white noise, 1 for pink noise, and 2 for fBm (Eke et al., 2002). Spectral analysis performed on the whole record obscures detailed information about autonomic modulation of RR intervals (Furlan et al., 1990). Nevertheless, since we focus on the FHR segment directly before delivery, the spectral analysis may reveal usefull clinical information about the fetus condition.

4.3.3 Morphological features

The folowing group of descriptive features is based on guidelines for CTG evaluation (FIGO, 1986; NIH, 1997). These features and patterns are used by clinicans for CTG assessment and were previously described in section 2.2.1. The set of features is defined as follows:

- baseline – the mean level of fetal heart rate where acceleration and deceleration are absent
- number of accelerations
- number of decelerations
- percentage of time occupied by decelerations
- number of uncomplicated decelerations
- number of complicated decelerations
- number of early decelerations
- number of late decelerations

Baseline is the most fundamental morphological feature. The improper baseline estimation destroys subsequent analysis of accelerations and decelerations. (Taylor et al., 2000) developed an iterative procedure for determination of baseline and extraction of acceleration and decelerations.

4.4 Nonlinear time series analysis

Fetal heart is controlled by both linear and nonlinear mechanisms and as a consequence the FHR signal contains both linear and nonlinear components. Therefore, it is eligible and natural to use nonlinear methods for dynamics estimation of the FHR. The nonlinear approach may also reveal relevant clinical information of FHR, not apparent in the time and frequency domain (Van Leeuwen and Bettermann, 2000).

Since heart beat fluctuates on different time scales and is self-similar, fractal dimension is useful estimator of FHR dynamics. Methods for estimation of fractal dimension are described in this section and brief introduction to fractal theory is given as well. Furthermore, we concentrate

on other nonlinear methods, such as entropy and complexity measures. It is worth of mentioning that application of nonlinear methods is not such straightforward as linear one. Differences between nonlinear methods are sometimes subtle but could turn out to be crucial.

4.4.1 Chaos dynamics and fractal properties

Fractals geometry is closely related to chaos dynamics where chaos always results in the formation of fractal but not all fractals are associated with chaos. In physiological time series, in our case FHR, fluctuations exhibit a long-range correlation that extends across many time scales and underlying dynamics is nonlinear "driven by" deterministic chaos (Van Leeuwen and Bettermann, 2000).

Fractals are complex geometrical objects which can be iteratively generated from simple structures, such as a line, a triangle, a square, or a cube resulting to the ideal, mathematical fractals. Simple forms are called initiators and rules transforming them are called generators. Examples of such structures are well known Koch curve, Sierpinski gasket, and Menger sponge presented in figure 4.4.

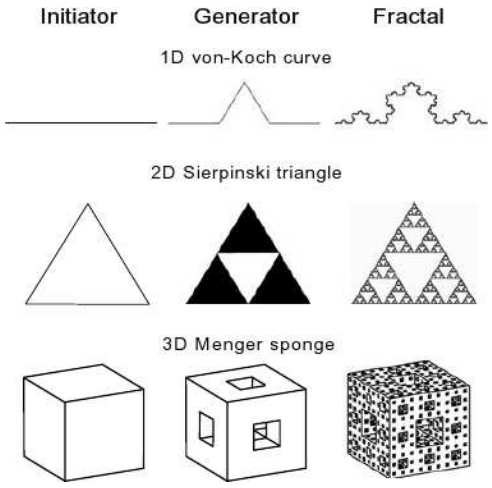


Figure 4.4: Ideal, mathematical fractals. These structures are generated by iterative application of generation rule to the initiator object (Eke et al., 2002).

These objects can be described, in traditional way, by Euclidian geometry but this description is cumbersome since the complex structure has to be broken down into large number of Euclidian objects assembled according to an equally large set of corresponding spatial coordinates. Thus, use of Euclidean geometry does not grab the essence of object design neither characterize its complexity.

Unlike Euclidean geometry, the complex structures can be characterized using dimension of non-integer range. Mandelbrot named these structures fractals and defined them not by axioms, but with set of properties instead, whose presence indicates that a structure is fractal.

Self-similarity

Self-similarity is essential property of fractals. Basically, it means that same features of complex object are observed independently on scale. If observed features are exactly the same a fractal

is ideal, exact. When statistical properties at smaller scale are geometrically similar to those at a bigger scale, a fractal is statistical.

Note that physiological time series are usually self-affine, instead of self-similar. When a structure is self-affine the scaling is anisotropic. In one direction the proportion of enlarged pieces does not coincide from those in the other. This distinction is, however, often smeared and term self-similarity is used when self-affinity is meant (Eke et al., 2002). Figure 4.5 shows the example of self-affinity in fetal heart rate.

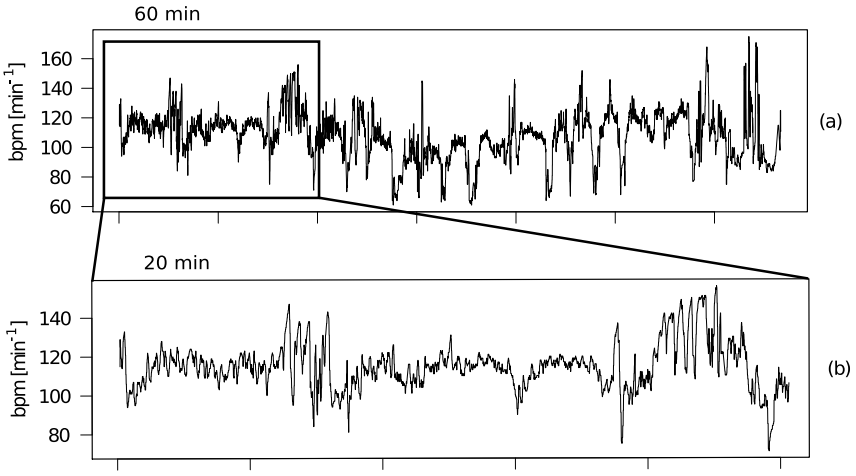


Figure 4.5: Self-affinity of fetal heart rate. Fluctuations of FHR at different time scales that are statistically self-similar (self-affine).

Scaling

Self-similarity of fractals is inherent and is behind the scaling relationship. In principle, a measured quantitative property q is dependent on the scale s according to following scaling relationship:

$$q = f(s) \tag{4.11}$$

In case of nonfractal object, estimation of q at finer scale s converge to a single value. However, for fractals q exhibits a power law scaling relationship with s. For finner scale it increases without any limit (Eke et al., 2002)

$$q = ps^{\epsilon} \tag{4.12}$$

where p is a proportionality factor and ϵ is the scaling exponent. The value of ϵ can be estimated as the slope of linear regression line fit to a data points on the plot of $\log p$ versus $\log s$

$$\log q = \log p + \epsilon \log s \tag{4.13}$$

Another important term is scale-invariance. It is addition to the power scale relationship and basically says that two estimates q_1 and q_2 at the two different scales s_1 and s_2 depends merely on the scales ratio and not on the absolute scale

$$q_2/q_1 = ps_2^{\epsilon}/ps_1^{\epsilon} = (s_2/s_1)^{\epsilon} \tag{4.14}$$

This equation is commonly referred as the scale-invariant property of fractals.

Stationarity and nonstationarity - a dichotomous model

The properties of fractals described above are fundamental for their characterisation. Let us consider stationary and nonstanionary time series. The discrimination to stationary and nonstationary is useful not only in fractal analysis but is essential in every signal processing. Each signal, either stationary or nonstationary, requires different method for the analysis. For stationary time series the statistical measures, i.e. mean, variance, and correlation structure, are the same irrespective of time. On the other hand, the nonstationary time series do not possess this property and statistical measures are fluctuates over time.

According to the dichotomous model (Eke et al., 2002), signals are seen as realization of one of two temporal processes: fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). The fBm signal is nonstationary with stationary increments. Physiological signals are generally consider as fBm, e.g. see figure 4.5, where statistical properties of FHR varies over time. The fGn is considered as stationary. The transition between these two models was proposed by (Eke et al., 2000) where summation of stationary fGn signal leads to nonstationary fBm and in turn, the increments $y_i = x_i - x_{i-1}$ of nonstationary fBm result in stationary fGn. In figure 4.6 the stationary and nonstationary time series are illustrated.

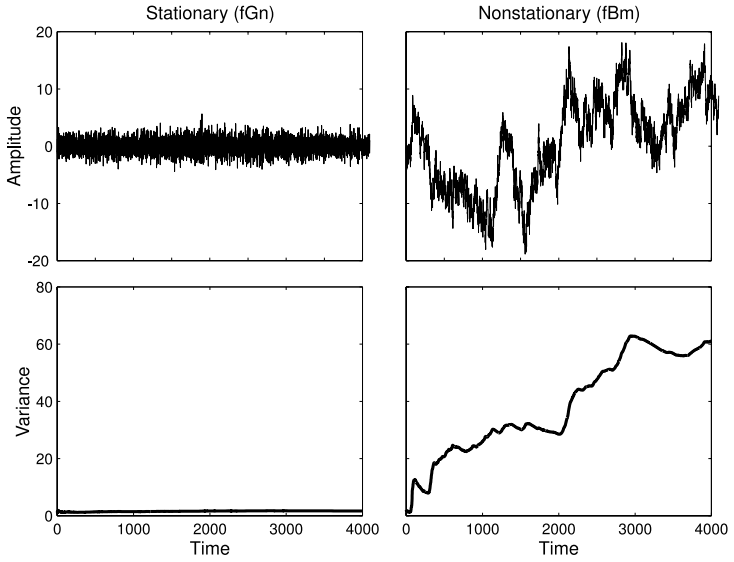


Figure 4.6: The stationary and nonstationary time series. Presented signals are pure-monofractal. These two signals differ in variance shown in lower panels. Stationary fGn has constant variance, unlike fBm which variance increases with time (Eke et al., 2002).

Since FHR is generally accepted as to be fBm methods able to overcome long-term statistical fluctuation should be applied. Nevertheless, we focus on signal segments directly preceding the delivery and due to this truncation we consider FHR as fGn.

4.4.2 State space reconstruction

Before we get to particular methods used for fractal dimension analysis, we have to define in which space is fractal dimension estimated. There are two principal approaches to estimate dimension of time series either by direct measurement of waveform or by operating in reconstructed state space. The former approach considers signal in $\mathbb{R}^2$ as a geometric object and directly uses it without any further transformation. Thus estimated dimension is always in range between 1 to 2 because the geometrical representation of signal is more complicated than line but never covers the whole 2D space. In this work, we used box counting, Higuchi, and variance method for estimating waveform fractal dimension.

The estimation of fractal dimension in reconstructed state space is similar to waveform, however, the state space reconstruction is not so straightforward; therefore, we properly define state space and its reconstruction in order to underline its importance for fractal dimension analysis. Note that waveform fractal dimension and dimension of reconstructed state space are usually different and should not be reversed.

The state space is sometimes called phase space. It is an abstract mathematical space in which coordinates represent the variables needed to specify the state of dynamical system. As time evolves, a system moves from one state to another creating a trajectory which provides a geometrical interpretation of system dynamics. The trajectories that never intersect and touch each other are called strange attractors and are typical for chaotic systems. In figure we illustrate the state space for the famous Lorenz system 4.7.

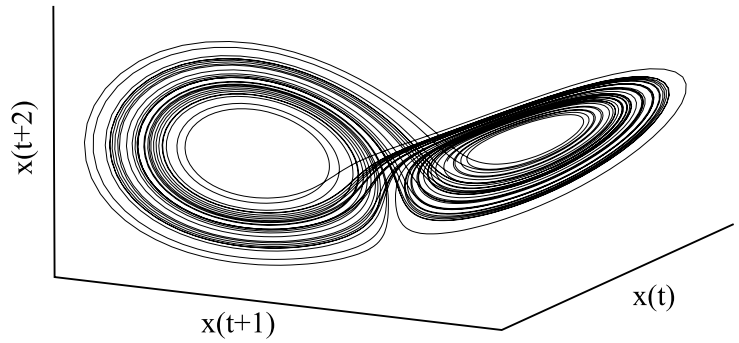


Figure 4.7: The state space reconstruction of Lorenz system. The attractor of chaotic systems are called strange.

The structure of dynamical system that is not seen in plot of time series comes out usually in the state space plot. In order to study dynamics of a system, we first need to reconstruct the state space. As there is no mathematical definition of the underlying dynamics of the heart, the state space has to be reconstructed from the time series. (Packard et al., 1980) showed that it is possible to reconstruct state space from scalar time series and this reconstructed space is diffeomorphically[2] equivalent to the original state space. The reconstruction of the state space is defined by embedding theorems (Takens, 1981). Taken's embedding theorem says that it is possible to reconstruct state space from time series $x(t)$ delayed by time τ as long as the embedding dimension m is larger than $2d + 1$, where d is a box counting dimension of strange attractor. In other words, the time delay embedding provides one-to-one image of the original

[2] A diffeomorphism is a map between manifolds which is differentiable and has a differentiable inverse

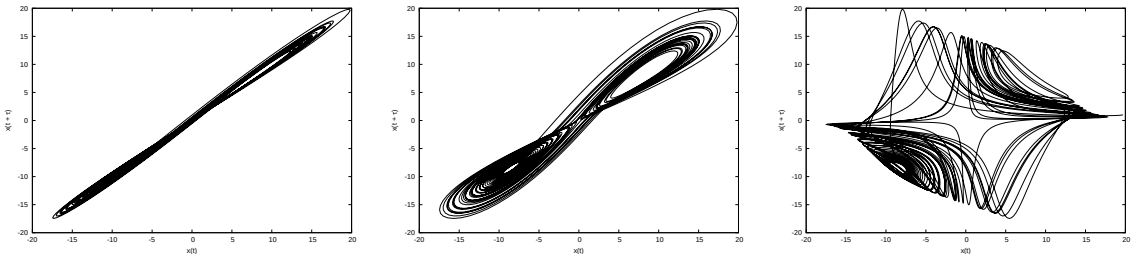
time series, provided m is large enough. The time delayed sequence $y(t)$ is computed according to following equation:

$$x(t) \rightarrow y(t) = [x(t), x(t+\tau), \ldots, x(t+(m+1)\cdot\tau)] \tag{4.15}$$

Different choice of τ and m leads to different reconstruction. Optimal embedding parametres cannot be established in general but are connected to specific application. This means that no single method is superior to others. However, regarding our application, a suitable search is performed and sufficient parameters are found. We adopted mutual information approach (Fraser and Swinney, 1986) to search the time delay and Cao's method (Cao, 1997) for examination of the embedding dimension. For other methods see any literature about nonlinear time series analysis, e.g. (Kantz and Schreiber, 2004).

The time delay

The terms redundancy and irrelevance are used to characterize the problem of choosing time delay. If τ is too short, vector elements are strongly correlated and reconstruction is concentrated around the diagonal of the embedding space (redundancy). On the other hand, long delays yield to (almost) uncorrelated vector elements and attractor is excessively overfolded (irrelevance). In figure 4.8 we show reconstruction of Lorenz attractor with different time delays.



(a) (b) (c)

Figure 4.8: Dependence of time-delay reconstruction of the Lorenz system on the choice of the time delay. (a) Small τ leads to redundancy, (b) proper choice of τ, the attractor is successfully reconstructed. However, this is not an embedding because embedding dimension is small $m = 2$. The embedding dimension of $m \geq 3$ is required. (c) very large τ results in irrelevance.

An auto mutual information (AMI), used for finding optimal time delay, was proposed by (Fraser and Swinney, 1986) and is similar to autocorrelation function with difference that account also nonlinear correlation. The AMI expresses the average information gained when the time series is delayed by time τ. The formula for AMI computation is following:

$$S = \sum_t p_{t\tau}(x_t, x_{t+\tau}) \log \frac{p_{t\tau}(x_t, x_{t+\tau})}{p_t(x_t) p_\tau(x_{t+\tau})}, \tag{4.16}$$

where p_t is marginal probability of time series being in t-th interval and $p_{t\tau}$ is the joint probability that an observation falls into the t-th interval and the observation time τ later falls into τ-th interval. The optimal time delay is suggested as a first marked minimum of the auto mutual information when the $x(t+\tau)$ adds a maximal information to the knowledge based solely on $x(t)$.

The embedding dimension

Recall that to ensure one-to-one reconstruction the embedding dimension m should be $m \geq 2d + 1$, where d is a box counting dimension. Unfortunately, we have no idea of value of d and we have to find an optimal m.

The Cao's method searches for minimal sufficient embedding. It is based on nearest neighbor approach (Kennel et al., 1992) where false nearest neighbors are defined as a set of points that are close to each other in dimension m but are apart in dimension $m + 1$. The false neighbors are close to each other due to projection rather than actual closeness. To define the distance between points we used Euclidean distance measure $\| \cdot \|$. The distance between points in m is defined as $\|\mathbf{y}^m(t) - \mathbf{y}^m_{NN}(t)\|$ and in $m + 1$ as $\|\mathbf{y}^{m+1}(t) - \mathbf{y}^{m+1}_{NN}(t)\|$. Then the ratio $R(t, m)$ between the two different m is computed and optimal embedding is determined as a value of $R(t, m)$ smaller than an empirical threshold.

$$R(t,m) = \frac{\|\mathbf{y}^{m+1}(t) - \mathbf{y}^{m+1}_{NN}(t)\|}{\|\mathbf{y}^{m}(t) - \mathbf{y}^{m}_{NN}(t)\|} \tag{4.17}$$

Setting of threshold value is sometimes cumbersome and subjective; therefore, Cao proposed an extension when the mean of $R(t, m)$, eq (4.18), is computed over all t. He introduced a new variable $E(m)$ and stated that ratio of values $E(m)$ and $E(m + 1)$ stops changing when the number of nearest neighbors is constant.

$$E(m) = \frac{1}{N - d\tau} \sum_{t=1}^{N-d\tau} R(t,m) \tag{4.18}$$

$$E_1(m) = \frac{E(m+1)}{E(m)} \tag{4.19}$$

Note that embedding dimension could be estimated as saturation in correlation dimension for increasing m. For details on correlation dimension see section 4.4.3 and figure 4.13.

4.4.3 Fractal dimension

The term fractal dimension was briefly introduced above but without proper description. Fractal dimension gives objective measure of structure complexity and is useful for fractals comparison. For instance, fractal dimension could be estimated for various natural fractals: coastline, clouds, trees, and physiological systems, such as networks of neurons, lungs and for temporal signals reflecting physiological processes.

There are slightly different meanings of dimension usage (Kantz and Schreiber, 2004). The dimensions measurements ordinary used are Euclidian and topological. For measurement of fractals a non-integer "exponent" dimension is used and relates to a number of increments. This exponent relationship is apparent below in the self-similarity dimension, equation (4.20). A lot of methods exist for fractal dimension estimation. Some of these methods are introduced in this section but before their presentation let us describe Euclidian and topological dimension. Euclidian dimension is used in every day live and indicates that object is on a line ($D_E = 1$), on a plane ($D_E = 2$), or in three dimensional space ($D_E = 3$). Euclidian dimension is basically a space occupied by an object and says how many coordinates are needed to properly determine position of a point of object. Topological dimension (D_T) indicates numbers needed

to determine the position of a point on the actual geometrical structure. The difference between D_T and D_E is apparent from figure 4.9.

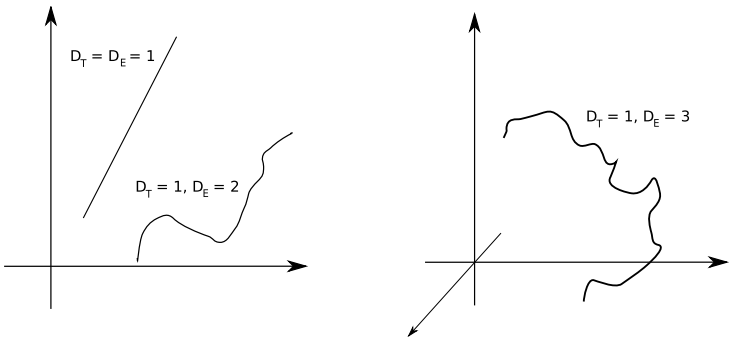


Figure 4.9: The difference between topological, D_T, and Euclidian, D_E, dimension.

As we have said, fractal dimension D is non-integer number. According to Mandelbrot, a structure is fractal when it satisfies condition that $D > D_T$.

Basic fractal dimension is self-similarity dimension D_{ss} which involves principles that are essential for all types of exponent dimensions. It holds only for exact fractals and tells us how many structural units N of the observed object, are seen at given resolution, $r = 1/s$

$$N = r^{D_{ss}} \quad \text{this leads to} \quad D_{ss} = \frac{\log N}{\log r} \tag{4.20}$$

The ideal mathematical fractals were introduced above and showed in figure 4.4. For inexact fractals, not exactly geometrically self-similar, Hausdorff introduced dimension measure which has become one of the most important in classifying fractals. This dimension is rather theoretical and is practically impossible to calculate it for real data. We therefore elucidate Hausdorff dimension in order to give insight into dimension measure and later on we rather concentrate on other fractals dimensions. Closely related to Hausdorff dimension is capacitance dimension estimated using box counting method. Note that it is better to begin description with capacitance dimension and then continue with Hausdorff dimension and others.

Box counting dimension

The capacitance dimension is based on evaluation of signal's capacity by covering it by N boxes of the side length ϵ. A minimal number of boxes needed to cover whole signal is counted and then the side length of boxes is increased. Thus, repetitively increasing size of boxes and counting them we are able to estimate box counting dimension D_0 as a slope of a linear regression fit to pairs on a log-log plot of N as a function of ϵ

$$\begin{aligned} N &= \epsilon^{D_B} \\ D_B &= \lim_{\epsilon \to 0} \frac{\log N}{\log(1/\epsilon)}. \end{aligned} \tag{4.21}$$

Example of covering signal by boxes and linear regression fit line are shown in figure 4.10 and 4.11, respectively.

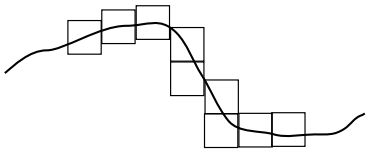


Figure 4.10: Box counting method. Covering the signal by N boxes of side length ϵ.

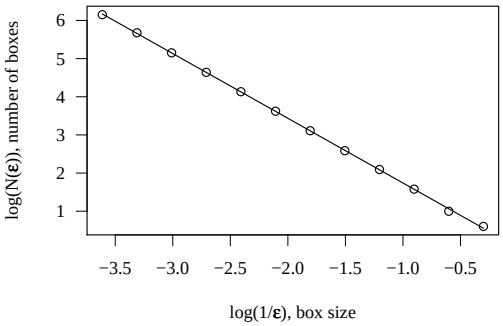


Figure 4.11: Box counting method. Linear regression fit to data points.

In real world time series, it is impossible to decrease length of box's side towards zero $\epsilon \approx 0$, hence the box counting method provides only estimation.

Computation of box counting dimension is slow and ineffective but if we limit ourselves for functions only, which is exactly the waveform of FHR, the computational demands could be significantly lowered. We describe the simplifications yielding to effective estimate of box counting dimension in the next chapter 5.

Hausdorff dimension

This type of dimension is usually referred to when mathematicians talk about fractal dimension. Hausdorff dimension is similar to capacitance dimension but instead of repetitively covering object by boxes with increasing size, an object is covered with sets, e.g. balls with diameter r. Then we compute sum of function applied to the diameter r of each covering set A_i. This function raise the diameter of each set to the power s:

$$H(s,r) = \inf \sum_i (\text{diameter} A_i)^s. \tag{4.22}$$

Let us examine behaviour of this sum when $r \to 0$. The sum will grow very large if s is smaller than certain number D_H, and on the other hand, the sum approaches 0 if s is greater than D_H; see expressions (4.23) and figure 4.12. The number D_H which separates these two states is called Hausdorff dimension.

$$\begin{aligned} &\lim_{r\to 0} H(s,r) = \infty \quad \text{for all } s < D_H \\ &\lim_{r\to 0} H(s,r) = 0 \quad \text{for all } s > D_H. \end{aligned} \tag{4.23}$$

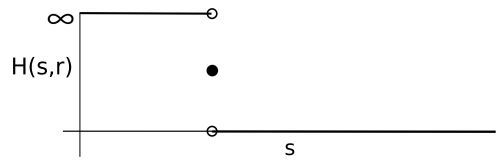


Figure 4.12: The Hausdorff dimension, D_H, is the critical value of s where the dimension jumps from ∞ to 0.

In practice, it is impossible to probe real world signal with non-integer objects and also estimate the boundary when r approaches 0. Therefore, Hausdorff dimension is useful for theoretical, mathematical context but not in practice.

Information dimension

The information dimension is basically extension of box counting method. Instead of simply computing the number of points that fall in particular hypercube of side length ϵ, the approach of information dimension weights number of points within hypercube by total number of points that construct an attractor. The information dimension is formulated using Shannon entropy:

$$I(\epsilon) = -\sum_i p_i(\epsilon) \log p_i(\epsilon), \tag{4.24}$$

where p_i describes the probability of some point being in the i-th hypercube of side ϵ. The information dimension is expressed as:

$$D_1 = \lim_{\epsilon \to 0} \frac{I(\epsilon)}{\log(1/\epsilon)} \tag{4.25}$$

Correlation dimension

The correlation dimension D_2 is by far the most important dimension for numerical applications. It is based on estimation of correlation sum $C(r)$, equation (4.26), which gives the probability that two randomly chosen points are close to each other with distance smaller than r.

$$C(m, r) = \frac{1}{N(N-1)} \sum_i^N \sum_{i>j}^N \Theta(r - \|\mathbf{y_i} - \mathbf{y_j}\|), \tag{4.26}$$

where $\mathbf{y}$ are m-dimensional delay vectors, N the number of points, and Θ is Heaviside function

$$\Theta(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1, & \text{for } x \geq 0 \end{cases}, \tag{4.27}$$

Note that in correlation sum, by inequality $j > i$, we omitted the "self-distance" between a point and the same point itself. If we consider the limit when r approaches zero (theoretically) the correlation dimension is estimated as:

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r} \tag{4.28}$$

This equation, of course, can not be applied for finite data set. The correlation dimension is usually estimated as a slope on log-log plot where $\log C(r)$ is plotted as a function of $\log r$. The correlation dimension can not be estimated for all radii r but proper region of r have to be chosen. The local slope approach offers a good insight into proper scaling region. The equation 4.29 is plotted as a function of $\log r$ and D_2 is estimated as a marked plateau.

$$D_2 = \frac{d \log C(r)}{d \log r} \tag{4.29}$$

In addition, the value of embedding dimension m can be altered and thus obtained several curves for different embedding, see figure 4.13. Three distinct intervals are generally distinguished on this plot depending on the scale of r. At very large r, the limited size of reconstructed attractor causes the slope approach zero. At small r the measurement noise smears correct scaling. Due to infinite dimension of noise the embedding space is spanned by noise in all directions tending to correlation dimension of m. Even without noise, quantization errors and finite sampling introduce errors for small r. In order to avoid incorrect dimension estimation, the proper scaling region has to be chosen. This regions are determined by r_l and r_u which stand for lower and upper radius r, respectively.

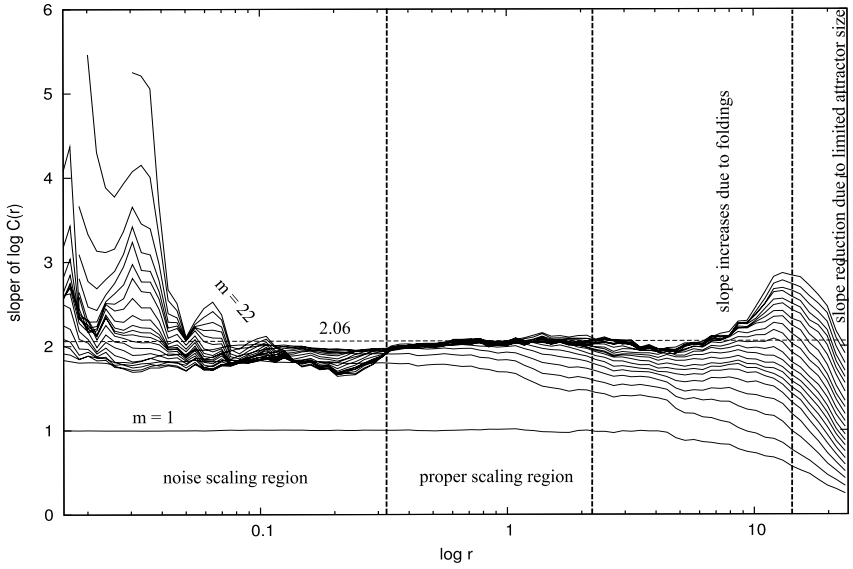


Figure 4.13: Correlation dimension for z- coordinate of Lorenz attractor with additional noise calculated for $m = 1, 2, \ldots, 22$ embedding dimensions. The proper scaling region is plateau of the slope of correlation sum $C(r)$.

Although the local-slope approach is useful and proper scaling region is easily estimated. In real world application the derivative in equation 4.29 have to be numerically approximated. In order to avoid heavy computation demands radius is sampled at few values. As a results, the approximation of derivatives suffers from large fluctuations. To overcome this problem, the correlation dimension can be estimated using $\log C(r)$ plotted as function of $\log r$ or by the Takens estimator using maximum likelihood approach (Takens, 1984). Later, Takens estimator was modified by (Theiler, 1988) and (Ellner, 1988) and resulted to the following formula:

$$D_2 = \frac{C(r_u) - C(r_l)}{\int_{r_l}^{r_u} \frac{C(r)}{r} dr} \tag{4.30}$$

There are some aspects that have to be considered when using correlation dimension. The correlation sum should cover the random sample of points drawn independently but successive samples of signal are not independent. To overcome this problem it has been proposed (Theiler, 1990) that only pairs of points should be account whose indices is more than w, where w is known as Theiler window and should be chosen generously. (Kantz and Schreiber, 2004) suggested space-time separation plot that helps to determine a sufficient value of w.

Higuchi's dimension

The Higuchi method (Higuchi, 1988) calculates fractal dimension from estimated length of curve. As a curve we consider temporal signal, fetal hear rate. Let us define time series $x(1), x(2), \ldots, x(N)$ with length N. Then we construct a new time series, X_k^m such that:

$$X_k^m; X(m), X(m+k), X(m+2k), \ldots, X(m + \left[\frac{N-m}{k}\right]) \quad (m = 1, 2, \ldots, k),$$

where [] denotes the Gauss' notation, m defines the initial time, and k the time interval. The k represents time displacement and number of new created subsets is equal to k. For example, for $k = 3$ and $N = 100$ we create following sequences:

$$\begin{aligned} X_3^1; X(1), X(4), X(7), \ldots, X(97), X(100), \\ X_3^2; X(2), X(5), X(8), \ldots, X(98), \\ X_3^3; X(3), X(6), X(9), \ldots, X(99). \end{aligned}$$

The length of curve, X_k^m, is defined as follows:

$$L_m(k) = \left\{ \left(\sum_{i=1}^{[\frac{N-m}{k}]} |X(m+ik) - X(m+(i-1)\cdot k)| \right) \frac{N-1}{[\frac{N-m}{k} \cdot k]} \right\} \Big/ k, \tag{4.31}$$

where $N - 1/[(N-m)/k] \cdot k$ represent the normalization factor for the curve length of subset time series. Then the length of curve for time interval k, $\langle L(k) \rangle$, is defined as average value over k set of $L_m(k)$:

$$\langle L(k) \rangle = \frac{\sum_{m=1}^{k} L_m(k)}{k} \tag{4.32}$$

The computed curve length $\langle L(k) \rangle$ for different k is related to the fractal dimension D by exponential formula:

$$\langle L(k) \rangle \propto k^{-D} \tag{4.33}$$

The fractal dimension is estimated as a slope of fitted regression to log-log plot of $\langle L(k) \rangle$ versus k.

Dimension of Variance

The variance technique of fractal dimension calculation is based on properties of fractional Brownian motion (fBm). It is a very useful approach because is robust to noise. Let the signal $X(t)$ be continuous in the time t and Δt is the time increment. The variance σ^2 is then related to the Δt according to the power law (Kinsner, 1994).

$$Var\{\Delta X(t_n, \Delta t)\} = \langle X^2(t_n, \Delta t)\rangle \propto |\Delta t|^{2H} \tag{4.34}$$

where $\Delta X(t_n, \Delta t) = X(t_n + \Delta t) - X(t_n)$ and H is the Hurst exponent computed from a log-log plot using

$$H = \lim_{\Delta t \to 0} \frac{\log Var\{\Delta X(t_n, \Delta t)\}}{\log(\Delta t)} \tag{4.35}$$

Finally, the variance dimension is defined as

$$D_\sigma = E + 1 - H \tag{4.36}$$

where E is the Euclidian dimension which equals to one for time series. The variance dimension is robust to noise and suitable for our application.

4.4.4 Detrend Fluctuations Analysis

The detrend fluctuation analysis (DFA) was proposed by (Peng et al., 1995) and probes the signal at different time scales. The result of the DFA is the fractal scaling exponent α. The whole process of estimating α is as follows. First, the time series $\{X_i\} = \{X_1, X_2, \ldots, X_N\}$ is summed giving

$$Y(k) = \sum_{i=1}^{k} \left[X(i) - \bar{X}\right], \tag{4.37}$$

where $Y(k)$ is the k-th of resulting series ($k = 1, 2, \ldots, N$), and $\bar{X}$ is the averaged value of the entire signal. Then $Y(k)$ is divided into boxes, or windows, of equal size n. For each box a least square line, $Y(n)$, representing the trend in the box is then calculated. This line is subtracted from summed $Y(k)$ in order to reduce possible nonstationarity. The formula for computation of fluctuations $F(n)$ in given box is following:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \left[Y(k) - Y_n(k)\right]^2} \tag{4.38}$$

This procedure is repeated for all time scale (different sizes of box n). Then the $F(n)$ is plotted on log-log graph against all size of box n. Typically, the relationship between $F(n)$ and n is exponential $F(n) \sim n^\alpha$. This indicates the presence of self-similarity, i.e. for small boxes size n the fluctuations are similar to those for large n.

The resulting scaling exponent α gives us information about origin of time series. For instance, $\alpha = 0$ indicates random process (white noise), $1/f$ pink noise has $\alpha = 1$, and $\alpha = 1.5$ indicates Brownian noise. Note the relation between α and spectral index $\beta = 2\alpha - 1$. Also note the relationship to the Hurst exponent $H = \alpha - 1$ (Eke et al., 2002).

4.4.5 Entropy

Entropy describes behaviour of system in terms of randomness, and quantifies information about underlying dynamics. Entropy is simply a fancy word for the "disorder". A stochastic, irregular, and less predictable signal has higher entropy than a completely deterministic. In other words, entropy is amount of a energy in system that is unable to do work (Eckmann and Ruelle, 1985).

There are number of definitions available to estimate entropy. The very basic one is Shannon information entropy $H(X)$ calculated by the equation (Shannon, 1948)

$$H(X) = -\sum_{x_i \in \Theta} p(x_i) \log p(x_i), \tag{4.39}$$

where X is a single discreate random variable with a set of values Θ and probability mass function $p(x_i)$. It is obvious from the equation (4.39) that Shannon entropy depends on probabilities of information contained in the signal.

Now let us consider time series which represent output of stochastic process. It is same as for single variable but with difference that joint entropy for each representative is computed. The time series consits of sequence of n random variables, $\{X_i\} = \{X_1, X_2, \ldots, X_n\}$, with set of values $\Theta_1, \ldots, \Theta_n$, respectively, the joint entropy is calculated as

$$\begin{aligned} H_n &= -H(X_1, X_2, \ldots, X_n) \\ &= -\sum_{x_1 \in \Theta_1} \cdots \sum_{x_n \in \Theta_n} p(x_1, x_2, \ldots, x_n) \log p(x_1, x_2, \ldots, x_n), \end{aligned} \tag{4.40}$$

where $p(x_1, x_2, \ldots, x_n)$ is the joint probability for the n variables $\{X_1, \ldots, X_n\}$. Furthermore, by application of the chain rule to eq. (4.40), the joint entropy can be rewritten as summation of conditional entropy

$$H_n = -\sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1). \tag{4.41}$$

It is apparent that instance X_n is partially deduced from its history, $X_1, X_2, \ldots, X_{n-1}$. Thus, each state has particular information of the past variables and carries a certain amount of new information as well.

The Kolmogorov–Sinai (KS) entropy is used to characterize system dynamics and it is an extension of information entropy. KS expresses the mean rate of creation of information. Let us divide state space into $\mathcal{D}$-dimensional hypercubes of content $\epsilon^{\mathcal{D}}$. The term $p(k_1 \ldots, k_n)$ represents joint probability that system is at hypercube k_1 at time δ, in the cube k_2 at $t = 2\delta$ and etc. The ϵ is a content in each hypercube. Then, KS entropy is calculated according to following equation (Eckmann and Ruelle, 1985)

$$\begin{aligned} H_{KS} &= \lim_{\delta \to \infty} \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n\delta} \sum_{k_1, \ldots, k_n} p(k_1 \ldots, k_n) \log p(k_1, \ldots, k_n) \\ &= \lim_{\delta \to \infty} \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n\delta} H_n \\ &= \lim_{\delta \to \infty} \lim_{\epsilon \to 0} \lim_{n \to \infty} (H_{n+1} - H_n). \end{aligned} \tag{4.42}$$

The last equality in the previous equations (4.42) is derived for stationary process and after application of the chain rule. From equation 4.42 we can conclude that KS entropy can be estimated only for n going to infinity and ϵ approaching zero. As a result KS entropy is limited of use for finite time series because it cannot be estimated precisely. For short noisy time series, (Pincus, 1995) devised a theory and a method for entropy estimation. It belongs to a family of statistics named approximate entropy (ApEn) with roots in work of (Grassberger and Procaccia, 1983) and (Eckmann and Ruelle, 1985).

Before explaining the approximate entropy, let us introduce several formulas which were proposed by above mentioned authors. Grassberger and Procaccia suggested estimation of lower boundary of KS entropy. They named the estimation K_2 and showed its applicability for characterization of chaos of dynamic system.

Let us consider a time series X_n of length N again. This series is divided into a set of m-length vectors $u_m(i)$. Then the number of vectors $u_m(i)$ and $u_m(j)$, which are close to each other in an Euclidean sense $d[u_m(i), u_m(j)] \leq r$, is expressed by the number $n_i^m(r)$. This number is used to calculate of the probability of vectors being close according to $C_i^m(r) = n_i^n/(N-m+1)$. Note that template length m is also referred as the embedding dimension. The final equation for estimation K_2 is defined as

$$K_2 = \lim_{N\to\infty} \lim_{m\to\infty} \lim_{r\to 0} -\ln[C^{m+1}(r) - C^m(r)], \tag{4.43}$$

where $C^m(r)$ represents the average probability that vectors $u_m(i)$ and $u_m(j)$ are within r in dimension m and $C^{m+1}(r)$ express also the average probability of close vectors but in dimension $m+1$. The $C^m(r)$ can be computed as

$$C^m(r) = \frac{1}{(N-m+1)} \sum_{i=1}^{N-m+1} C_i^m(r). \tag{4.44}$$

Eckmann and Ruelle showed that it is also possible to estimate KS entropy directly. Let us define the function $\Phi^m(r) = 1/(N-m+1)\sum_{i=1}^{N-m+1} \ln C_i^m(r)$. Consequently the expression $\Phi^{m+1}(r) - \Phi^m(r)$ is average of the natural logarithm that sequences of length m are close to each other even if new point is added. Eckmann and Ruelle introduced computation of KS entropy as

$$H_{ER} = \lim_{N\to\infty} \lim_{m\to\infty} \lim_{r\to 0} [\Phi^{m+1}(r) - \Phi^m(r)]. \tag{4.45}$$

Although the formula mentioned above has been useful in classifying dynamical chaotic system its limitation, for real-world data series, lies with necessity of a vast amount of input data. Moreover, these estimates badly compromised even with very small amounts of noise. Therefore, Pincus proposed way of calculation entropy (ApEn) for relatively short and noisy data sets.

Approximate entropy

In contrast to KS entropy, ApEn is not estimation of chaos dynamics but it is able to distinguish low-dimensional deterministic system, chaotic system, stochastic and mixed systems (Pincus, 1995). Considering ER entropy, let us now define ApEn entropy as family of measures

$$H_{AE}(m, r) = \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)]. \tag{4.46}$$

For a deterministic time series the is H_{AE} has lower entropy; whereas for stochastic time series H_{AE} is higher. The ApEn has its shortcomings in biased estimation of entropy and as a consequence give higher number of entropy. In addition, the ApEn is also strongly dependent on the series length.

Sample entropy

A slightly modified estimation of approximate entropy was proposed by (Richman and Moorman, 2000) and resulted in what is known as sample entropy (SampEn). This estimation overcame the shortcomings of the ApEn mainly because the self-matches are excluded. Secondly, conditional probabilities are not estimated by a template-wise approach. SampEn requires only that one template finds a match of length $m + 1$. The calculation of SampEn is as follows:

$$H_{SE}(m, r) = \lim_{N \to \infty} - \ln \frac{C^{m+1}(r)}{C^{m}(r)}. \tag{4.47}$$

In the following figure 4.14, we present simulated time series and procedure for calculating sample entropy. We define the template length $m = 2$ and r as a positive value (usually $r = (0.1 - 0.2) \cdot SD$, where SD stands for standard deviation). The samples similar to first sample $u[1]$ are marked by filled squares, to second sample $u[2]$ by filled square, and to third sample by filled triangle. Then we count occurrence of two-patterns and three-patterns. These are as follows: three two-patterns $(u[1], u[2]; u[9], u[10]; u[24], u[25])$ and two three-patterns $(u[1], u[2], u[3]; u[9], u[10], u[11])$. Since we do not count self-matches, they are reduced to two and one, respectively. This is repeated for all two-patterns and three-patterns in sequence and then computed using formula 4.47.

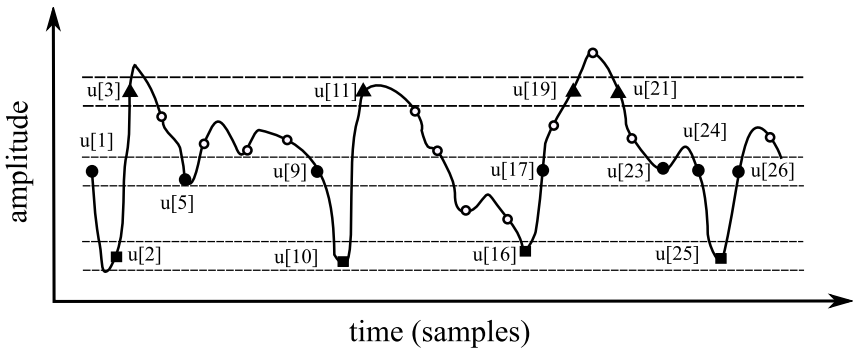


Figure 4.14: Simulated time series and its sample entropy estimation, for details see text. Modified from (Costa et al., 2005).

4.4.6 Lempel Ziv Complexity

The Lempel Ziv Complexity (LZC) (Lempel and Ziv, 1976) is widely used in data compression. It is based on information theory approach. The LZC estimates reoccurring patterns contained in the time series irrespective of time. A periodic signal has the same reoccurring patterns and low complexity while in random signal individual patterns are rarely repeated and signal complexity

is high. To be more precise, (Lempel and Ziv, 1976) defined complexity as "a measure on the extent to which the given sequence resembles a random one".

Let us consider time series $x(1), x(2), \ldots, x(n)$ and apply encoding procedure in order to form sequences S of strings. For the binary encoding this sequence contains only {0,1}. The increase in signal value, $x(i+1) > x(i)$, is encoded by 1 and decrease, $x(i+1) \leq x(i)$, by 0. To indicate that, substring of S starts at position i and ends at position j, we write $S(i,j)$. The vocabulary of the sequence $v(S)$ contains all substring of S, e.g. for $S = 101$ the $v(S) = \{1, 0, 10, 01, 101\}$. Let S and Q denotes two strings and SQ their concatenation. When the length of sequence is not specified a operator π is used to remove last string from concatenated SQ. The operator π comes as a postfix $SQ\pi$.

The whole procedure of computation complexity $c(n)$ is following: At the start the $c(n)$ is set to 1, $S = s_1$, $Q = s_2$, $SQ = s_1, s_2$, $SQ\pi = s_1$, and the vocabulary $v(SQ\pi)$ is empty. For generalization purpose, let us assume that we moved in sequence to sample r. The $v(SQ\pi)$ is not empty and strings S and Q are following $S = s_1, s_2, \ldots, s_r$, $Q = s_{r+1}$. If $Q \in SQ\pi$ then Q contains the substring of S and do not provide new information, therefore, the S remain unchanged and a new character s_{r+2} is add to Q. Again, we check if $Q \in SQ\pi$, if Q is not substring of $SQ\pi$ we increase $c(n)$ by one and concatenate S and Q, otherwise we continue in adding the new characters to Q until the end of the sequence is reached

At the end the number of different strings is equal to $c(n)$. By convention, when the sequence reaches its last element, the $c(n)$ is increased by 1. It is apparent that $c(n)$ is dependent on the length of original sequence n. We use the normalization form to avoid this dependence on the number of data points (Lempel and Ziv, 1976). The normalized $c(n)$ is defined as

$$C(n) = \frac{c(n)}{\frac{n}{\log 2n}} \tag{4.48}$$

Note that another coding scheme can be used in order to encode signal. The above described binary encoding can be extended to a ternary and even more quantizing encoding. However, as (Kaspar and Schuster, 1987) pointed out, the higher encoding should not be used in order to minimize the dependence of results on quantification criteria and normalization procedures.

4.5 Surrogate data test

So far we assumed that fetal heart rate is nonlinear driven by deterministic chaos. In order to verify this hypothesis we use surrogate data test. In this test we formulate a null hypothesis e.g. that data are generated by gaussian linear stochastic process. Then, if this hypothesis is rejected on some significance level, we can conclude that data do not origin from such process and nonlinear methods may reveal important information about underlying system dynamics. As the test statistics, in order to discriminate nonlinear time series from surrogate data, we use correlation dimension, see section 4.4.3.

There are many available null hypothesis against which we can test our time series. For example, null hypothesis could be that data are independent, identically distributed random variables of unspecified mean μ and variance σ^2. In our work, we employed general hypothesis that data are produced by gaussian linear stochastic process (AR(p) process). During data generation we required that surrogate and original data have the same power spectrum and probability density function. There exists broad area of methods one can use for data generation.

The main idea behind creation of surrogate data is following: first, apply Fourier transform to original time series. Second, replace the phases by random numbers ranging from $(-\pi, \pi)$. Finally, apply inverse Fourier transform to the Fourier coefficients. In our work, we used iteratively refined surrogates proposed by (Schreiber and Schmitz, 1996). For more information see referenced paper or book of (Kantz and Schreiber, 2004).

The level of significance is commonly set to be $p \leq 0.05$, therefore we need at least 19 or 39 surrogate data for one- and two-sided test, respectively. Lastly, we compute the test statistics. We use correlation dimension. If the results of correlation dimension are not different for nonlinear time series and surrogate data, the null hypothesis could not be rejected. The whole scheme of surrogate data test is presented in figure 4.15.

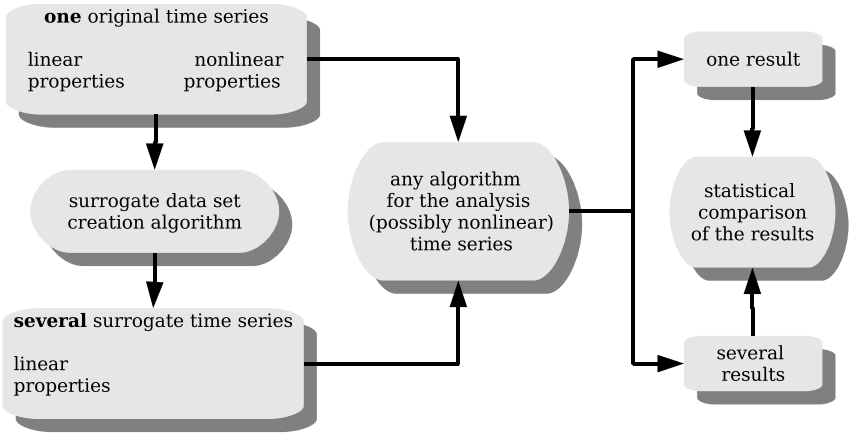


Figure 4.15: Scheme of surrogate date test, for the case of the null hypothesis of a linear process. Modified from (Galka, 2000).

4.6 Feature selection and classification

The estimated values of fractal dimension and entropy are in machine learning field called features. Combination of features create something what is called feature space with dimension equal to number of features. A classifier operates in this space and try to discriminate the two different classes normal and abnormal fetuses. We shall note here, that we employ nomenclature used commonly in practice and by normal FHR records we mean those records when fetuses were classified as healthy and with term at risk or abnormal FHR we consider records when fetuses have developed metabolic acidosis. For more details about classification of fetus status, see section 2.3.

The feature selection is considered as a reduction of dimension of feature space. By selecting appropriate features we first, trying to avoid the curse of dimensionality and second, choose most relevant features for further classification task.

4.6.1 Feature selection

First of all, it is necessary to mention that feature selection is often considered as an art than a science. There exist a lot of methods and algorithms one can use for feature selection; only

few of them were used in our work. We start with very simple filter method, correlation criteria, and continue with Correlation based feature selection, Information gain, and Wrapper of Naive Bayes implemented in Weka data mining software (Witten and Frank, 2005). According to (Guyon and Elisseeff, 2003) methods for feature selection can be divided into three groups: filters, wrappers, and embed methods. Filters use a criterion of selection that does not make use of the learning machine, whereas wrappers use learning machine for alternative subset evaluation. In contrast of filters and wrappers, embedded method perform feature selection in learning process. The output of this method are selected features and trained learning machine. In spite of their power we do not use embedded methods in this work. The cross-validation method was used for subset evaluation. We rank the features selected by different methods and then combine all results together and chose features with highest rank.

Correlation criteria

The linear relationship between two random variables could be expressed by correlation coefficient. There are a lots of ways of coefficient computation, each one suitable for different task. We choose Pearson correlation coefficient[3] for its general applicability. Let us consider a series of n measurements of variables X and Y written as x_i and y_i where $i = 1, 2, \ldots, n$. The Pearson correlation is defined as:

$$\mathcal{R} = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}} \tag{4.49}$$

where cov stands for the covariance and var for the variance. The estimate of R is given by

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}} \tag{4.50}$$

The coefficient ranges from -1 to 1. For coefficient values of $\langle -1, 0\rangle$ the two variables are uncorrelated and for values of $(0, 1\rangle$ are correlated. We divide the last interval into three equidistant subintervals and then describe correlation as weak $(0, 0.33\rangle$, moderate $(0.33, 0.66\rangle$, and strong $(0.66, 1\rangle$. The two variables are identical if $R = 1$. When the correlation coefficient is estimated we need a tool to determine if the correlation between variables is statistical significant. Therefore, we employ t-test for the following hypothesis:

$H_0 : R(i) = 0$ There is no correlation between the variables x and y.

$H_1 : R(i) \neq 0$ There is correlation between the variables x and y.

To perform t-test we compute t value for every coefficient:

$$t(i) = R(i)\sqrt{\frac{N-2}{1-R(i)^2}} \tag{4.51}$$

then using t we find p-value for $N - 2$ degrees of freedom. The p-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data. Note that $p = 0.05$ is generally considered as significant.

[3] often referred as sample correlation coefficient

Next we create a correlation matrix that describes relationship of all measured features by correlation coefficient. Moreover we can also include the resulting class (normal, abnormal FHR records) and easily observed which features are best correlated with class. Diagonal elements of correlation matrix consist of autocorrelated features which are always equal one. The off-diagonal elements are coefficients between different features.

Correlation based feature selection

Correlation based feature selection (CFS) uses a heuristic evaluation function to rank feature subsets. This algorithm chooses features that are in strong relationship with a class while having low intercorrelation (Hall, 1998). It is a similar method we already used but with difference that subsets of features are used.

Information gain

The information gain evaluates the attribute by measuring the amount of information gain with respect to class. The mutual information, termed $\mathrm{InfoGain}$, is computed using entropy H:

$$\mathrm{InfoGain}(Class, Attribute) = H(Class) - H(Class|Attribute) \tag{4.52}$$

The mutual information was already mentioned in section 4.4.2 where we used the auto mutual information for determining optimal time delay for attractor reconstruction.

Wrapper method

We used Naive Bayes as a learning machine for feature subset evaluation. Naive Bayes assumes that within each class the probability distributions of attributes are independent of each other. Note that its performance on domains with redundant features can be improved by removing such features (Hall, 1998).

4.6.2 Feature classification

Before we approach to classification, we have to emphasize that we work with balanced data set, i.e. the number of all cases in both classes is almost the same ($n_{normal} = 94$, $n_{pathological} = 95$). If the data set is unbalanced with favour to normal, healthy cases, there are two possible approaches to balance data: under sampling of the majority class or over sampling of the minority class. The method of over-sampling the minority class called Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) was utilised in work of (Georgoulas et al., 2007). Note that SMOTE operates in feature space and not in data space. The process of over-sampling is following: new synthetic instances are created along any/all lines connecting the particular instance from the minority class with its k nearest neighbors belonging also to the minority class. In other words, the new instaces represent interpolation of neighbors from minority class.

Another approach to classify unbalanced data set is to utilise a penalty function. For the pathological cases, the penalty of misclassification is higher than for normal case. Nevertheless, since we worked with balanced data set we have not used any of above mentioned method.

Below, we briefly introduce the three models that were used for classification of fetal heart rate records. Each method operates in the feature space. For more information about models, see e.g. (Duda et al., 2000).

Naive Bayes

Naive Bayes classifier is based on the Bayes theorem where posterior probability is computed as:

$$p(D|X_1,\ldots,X_n) = \frac{p(D)p(X_1,\ldots,X_n|D)}{p(X_1,\ldots,X_n)} \tag{4.53}$$

where D is the class variable and $X_1,\ldots,X_n$ are features. The naive Bayes uses this theorem but with strong (naive) assumption that features are conditionally independent given the class. Therefore, we can estimate posterior probability as:

$$p(D|X_1,\ldots,X_n) = \frac{p(D)}{p(X_1,\ldots,X_n)}\prod_{i=1}^{n} p(X_i|D) \tag{4.54}$$

Then, to minimaze error classification, we chose the decision rule with maximum a posterior probability referred as MAP:

$$D_{MAP} = \operatorname*{argmax}_d p(D=d)\prod_{i=1}^{n} p(X_i = x_i|D=d) \tag{4.55}$$

Decision tree - C4.5

The C4.5 algorithm was proposed by (Quinlan, 1992) and is used to generate a decision tree. Generally, decision tree divides data into subgroups where it is desired that one class prevails in each subgroup. The C4.5 employs the concept of information entropy for choosing the attribute, see equation 4.39. First, we create root of tree using the attribute with highest information gain (difference in entropy), eq. 4.52. Then we make decision and move to the sublists of the tree until every example is covered. The decision tree is prone to overfitting, hence it has to be pruned in order to ensure the generalization capabilities. The C4.5 utilises error based pruning. The algorithm goes backwards and removes branches that do not help towards the goal by replacing them with leaf nodes.

Support Vector Machine

The main purpose of Support Vector Machina (SVM) is to minimize the structural risk, i.e. the risk of error prediction on unseen data. Let $x_i \in \mathbb{R}^n$ are features and $y_i = \{-1, 1\}$ the resulting class. The SVM algorithm searches the hyperplane (w, b) which maximize the distance (margin) between the hyperplane and instances closest to it (Vapnik, 1995). These instances are called support vectors. To find the optimal hyperplane we have to solve the primal optimization task by minimizing following equation:

$$(w, b) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{4.56}$$

subject to

$$\begin{aligned} \langle w\cdot\varphi(x_i)\rangle + b \geq +1-\xi_i, \quad y_i = +1,\\ \langle w\cdot\varphi(x_i)\rangle + b \leq -1+\xi_i, \quad y_i = -1, \end{aligned} \tag{4.57}$$

where ξ_i are called slack variables that allow the margin constraints to be violated and $\varphi(\cdot)$ is kernel providing nonlinear feature mapping. Constant C is tradeoff between maximization of margin and minimization of error. It is convenient to transform primal optimization task to dual optimization task. The purpose is to find numbers $\alpha_i, i = 1, \ldots, n$ that are solution of dual optimization task:

$$\vec{\alpha} = \underset{\vec{\alpha}}{\operatorname{argmax}} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \varphi(x_i) \cdot \varphi(x_j) \rangle \right), \tag{4.58}$$

subject to

$$\begin{aligned} C \leq \alpha + i \leq 0, \quad i = 1, 2, \ldots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{aligned} \tag{4.59}$$

Points that satisfy condition $\alpha_i > 0$ are called support vectors and determine the hyperplane. By choosing mapping function $\varphi(\cdot)$ properly, the inner products $\varphi(x_i) \cdot \varphi(x_j)$ can be rewritten using a kernel function $k(x_i, x_j)$. Classification of instance x is then obtained as the $sign$ of following function:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i k(x_i, x_j) + b \tag{4.60}$$

This is only determination on which side of hyperplane $\langle w, \varphi(x) \rangle + b = 0$ is an instance situated.

4.6.3 Estimating classification performance

Cross-validation is commonly used to estimate classification performance. The cross-validation has determined number of folds, e.g 10-fold cross-validation. In each step, the data set is divided into training and testing data. Then a learning machine is trained and performance evaluated on testing data. This procedure is repeated in each fold of cross-validation with differently divided data set.

A different performance measures may be used to evaluate classification performance. The most common is represent performance by confusion matrix in figure 4.16. In the confusion matrix, TN (true negative) expresses number of correctly classified negative examples, TP (true positive) is number of correctly classified positive examples, FN (false negative) is number of incorrectly classified negative examples, and FP (false positive) is the number of incorrectly classified positive examples.

	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Figure 4.16: Confusion Matrix.

The overall classification *accuracy* is computed as $a = (TP+TN)/TP+FP+TN+FP$. This could be further divided into accuracy a^+ observed separately on positive examples $a^+ = TP/(TP+FN)$ and accuracy a^- observed only on negative examples $a^- = TN/(FP+TN)$. The a^+ and a^- are called sensitivity and specificity, respectively. Furthermore, when dealing with imbalanced dataset, these separate accuracies are combined to metric called geometric mean $g = \sqrt{a^+ \cdot a^-}$.

Other approach for classifiers comparison is to use receiver operation characteristic (ROC) or simply ROC curve. The ROC curve is shown in figure 4.17. In this graphical plot the sensitivity is plotted as a function of 1 – specificity.

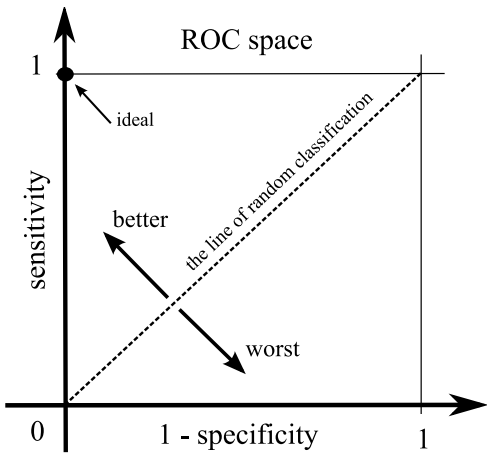


Figure 4.17: The receiver operation characteristic. Sensitivity is plotted as function of 1 – specificity. Ideal classifier is marked in the upper left corner with coordinates (0,1)

The line of random classification (random guess) is a straight line at a 45° diagonal. Successful classifier is placed above this line and tends to upper left corner with coordinates (0,1), that is, all positive examples are classified correctly and no negative example is misclassified. Next usefull value for comparison is area under the ROC curve (AUC). The AUC expresses probability that classifier rank randomly chosen positive instance higher than randomly chosen negative instance.

We shall note here that we mainly focus on sensitivity and specificity. Classification is usually a trade-off between high sensitivity and low specificity and vice versa. In our work, we rather prefer those results with higher specificity, i.e. we are trying to avoid unnecessary interventions.

Chapter 5

Application of nonlinear methods to fetal heart rate

Before we present classification results of fetus status, let us consider and stress practical aspects and limitations of nonlinear methods for fetal heart rate analysis.

As mentioned above, we are dealing with the data that has finite length, finite precision and is sampled either equidistantly or non-equidistantly. Moreover, the data are contaminant with noise which means that more robust tools have to be utilised in order to have unbiased results or, on the other hand, if we are aware of sensitivity to noise we can either modify methods to correct our results or carefully interpret them. We stress several restrictions resulting from properties of data set and propose several possibilities how to deal with them. In some cases, we alleviate demands on the data size and signal to noise ratio as we did for stationarity. Recall that we are focused on minutes preceding the delivery, hence we may consider the data as stationary. The main intention of this section is to show steps to successful nonlinear data analysis.

Before we approach nonlinear analysis we have to establish whether data contain nonlinearity or not. The surrogate data test was developed and is commonly used for this purpose. We introduced this test in section 4.5; here let us mention that we created iteratively refined surrogates as it was described by (Schreiber and Schmitz, 1996). The null hypothesis, we test against time series, is that data were generated by gaussian linear stochastic process. We used higher order statistics and correlation dimension for the null hypothesis testing. In order to estimate correlation dimension of possible nonlinear time series we have to search for optimal embedding parametres.

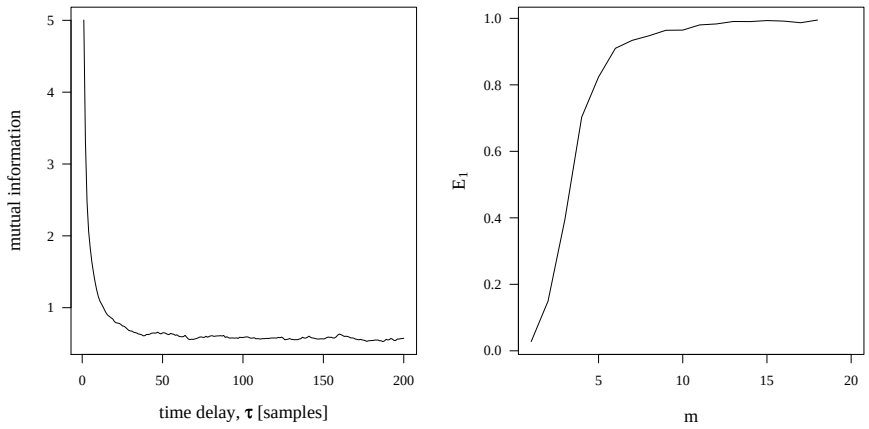
5.1 In search of optimal embedding parametres

The optimal embedding parametres are needed in order to reconstruct unfolded attractor of given time series. The two parametres have to be estimated for proper embedding: time delay τ and embedding dimension m. These parametres are coupled together, nevertheless, for convenience, we evaluate them separately.

To find optimal time delay we used method of auto mutual information. As optimal time we chose the time when the mutual information exhibits first marked minimum. If the mutual information has no minimum, we choose time delay, such that $S(\tau) = S(\tau)/S(0) \sim 1/5$. As (Abarbanel et al., 1993) pointed out, "this is time delay in which some nonlinear decorrelation is at work but not too much". In figure 5.1(a) we show the result of auto mutual information

for sample record from database. It is apparent, the first minimum is at $\tau = 40$ samples. Since the sampling frequency is 4 Hz, the optimal time delay is $\tau = 10$ s. The average value of first minimum of all records was $\tau = 12.3 \pm 9.5$ s.

In order to find the optimal embedding parameter m we applied Cao's method on the whole data set. The result is shown in figure 5.1(b). It is obvious, that E_1 stops changing at value of $m = 6$ when the value of nearest neighbors is constant. We shall note here that for some records sufficient embedding dimension is lower, $m = 5$. Nevertheless, since we will analyse correlation dimension for different embedding, the estimated m rather indicates data dimensionality and may be larger than necessary.



(a) (b)

Figure 5.1: Optimal embedding parametres for sample record. (a) Auto mutual information. The first marked minimum, $\tau = 40$ samples, determines optimal time delay. (b) Results of Cao's method. When the value of E_1 stops changing, $m = 6$, the sufficient embedding dimension has been found.

5.2 Establishing nonlinearity - the surrogate data test

In order to establish nonlinearity of the fetal heart rate we performed surrogate data test. Recall from section 4.5, we test null hypothesis that data originates from gaussian linear stochastic process. To be able to reject this hypothesis on 95% confidence level, we generated 39 different surrogates sequences. As the discrimination method we applied the correlation dimension.

The surrogate data have the same distribution and autocorrelation function as the original data but the nonlinear properties are destroyed. We estimated correlation dimension, D_2, for all records and based on these estimations we rejected the null hypothesis on 95% confidence level, i.e all estimates of D_2 for original and surrogate data were different. In figure 5.2 we present the difference of $\log C(r)$ for original and surrogate time series with different embedding dimension $m = 3, 4, 5$. Notice that for $m = 4, 5$ the lower bound of surrogate data is very close or even touches the solid line representing the $\log C(r)$ of original data. There are several reasons for this behaviour. First, we might have not found optimal embedding parametres (time-delay and embedding dimension). As a consequence, attractor was not properly reconstructed. Second, the nonlinear determinism was partially destroyed by measurement process, such as quantization, filtering, and averaging. Last, and most likely, the measurement noise and

artefacts partially distorted the nonlinear determinism of data. We incline to this hypothesis because the data we have available are of poor quality. We discuss this issue more thoroughly in chapter 6.

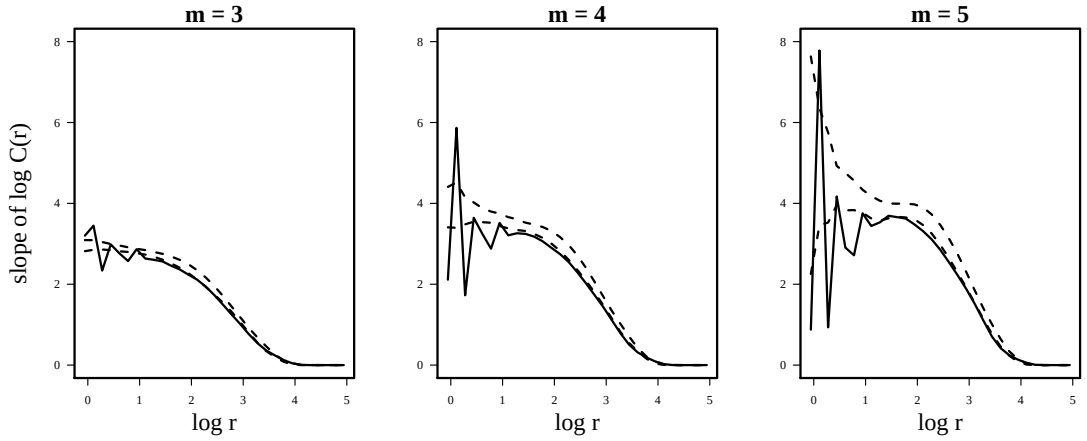


Figure 5.2: Surrogate data testing with correlation dimension. The $d \log C(r)/d \log r$ is plotted as function of $\log r$ for difference embedding dimension $m = 3 - 5$. Dashed lines represent confidence interval of surrogate data and solid line represents original time series.

5.3 Practical consideration for using nonlinear methods

5.3.1 Dimension of attractor

We used estimated optimal embedding parametres for the state space reconstruction and then analysed fractal dimension of attractor. The fractal dimensions of attractor D_0, D_1, and D_2 are numerically so similar that there is no advantage to use them all (Abarbanel et al., 1993). Therefore, we used only correlation dimension for attractor analysis since it is robust to noise. To be able to reliably estimate the fractal dimension a certain data length is necessary. There exists many theories about required data length with general agreement that needed data length increase exponentially with data dimension. We followed data size requirements as were suggested by (Rulle, 1990); for estimating a dimension d, a minimum data length is $N_{min} = 10^{d/2}$ required. Since the embedding dimension is $m = 6$, the maximum dimension is $d = 6$. Therefore, required data length is $N_{min} = 10^{d/2} = 1000$ samples. As was already mentioned, we worked with segments of length 4800 samples, hence we can reliably estimate the correlation dimension.

In addition to embedding parametres, we have to also establish correct scaling region as was described in section 4.4.3. Recall that for Lorenz system we estimated proper scaling region using local slope approach when we searched for plateau on the log-log plot of $d \log C(r)/d \log(r)$ versus $\log r$. In the following figure 5.3, we present estimated local slopes of $\log C(r)$ of normal fetal heart rate record. The attractor was reconstructed with time delay $\tau = 5.5$ s and embedding dimension $m = 2 - 16$. It is apparent that the proper scaling region ranges from $\log r_l = 1$ to $\log r_u = 2$ ($r_l = 2.7$ to $r_u = 7.4$) where $D_2 = 3.5$. Note that embedding dimensions $m = 2$ and $m = 3$ are not sufficient to properly reconstruct the attractor. If we used low-dimensional

embedding, the D_2 would be underestimated. In the following figure 5.4, we also show the $\log C(r)$ versus $\log r$ for different embedding dimension.

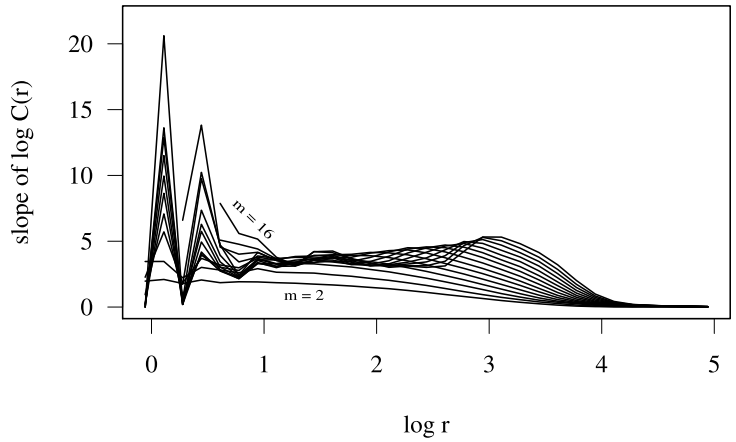


Figure 5.3: Correlation sum of fetal heart rate for increasing embedding dimension $m = 2 - 16$. Estimated correlation dimension by local slope approach. The proper scaling region ranges from $r_l = 2.7$ to $r_u = 7.4$, hence $D_2 = 3.5$.

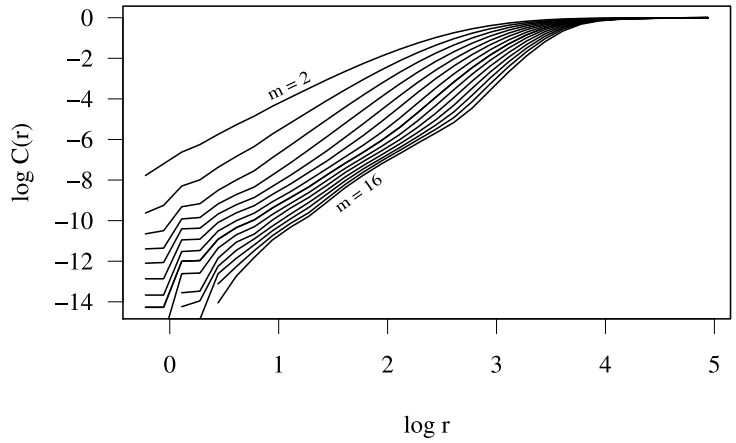


Figure 5.4: Estimated correlation dimension using $\log C(r)$ versus $\log r$ for increasing embedding dimension $m = 2 - 16$.

5.3.2 Wavefrom fractal dimension

Generally, these methods are based on applying measurement function in the selected interval or region. For different size of interval we get different number of points satisfying given condition. The fractal dimension is then estimated as a slope on the log-log plot of the intervals length versus measuring function. We used following methods for waveform dimension estimation: box counting, Higuchi's, and variance. Note that Higuchi's and variance method estimate Hurst exponent, H, that is related to fractal dimension $D = E + 1 - H$, where E stands for Euclidian dimension which is equal to one for time series. We also include detrend fluctuation

analysis in this section although it used for estimation of fractal slope α. As mentioned before, α is related to Hurst exponent $H = \alpha - 1$, hence we apply the same formula for dimension computation as we did for Higuchi's and variance method. We theoretically described all methods above; here let us mention only practical considerations.

Recall that computation of box counting dimension is slow and ineffective but if we consider functions only the whole procedure could be simplified. These simplifications were proposed by Boshoff in his paper (Boshoff, 1992). The main considerations are as follows:

- We assume functions only; this means only boxes in columns are counted.
- Highest and lowest value is considered in each column and every box in between is entered.
- Starting with a number of samples which is a power of two it allows the number of columns to be halved recursively.

Since FHR is self-affine width and height of a box has to be unequal in favor of width or FHR signal has to be normalized. We utilised the later approach and normalized signal by mapping it into the unit square (Sevcik, 1998).

Other interesting property is size of interval where the measuring function is applied. This feature is common for all respective methods. The interval size should not be chosen arbitrarily. On the small scale, where size is very small, the measure may be biased due to noise. Moreover, statistical significance have to be ensured (Kinsner, 1994). The restriction on the interval range results in required data size in order to estimate dimension properly. We examined how the data length affects estimate of fractal dimension for each method. Let $w(t)$ be a Weierstrass deterministic cosine function expressed as follows:

$$w(t) = \sum_{i=0}^{N} \gamma^{-iH} cos(2\pi\gamma^i t), \quad 0 < H < 1, \tag{5.1}$$

where $\gamma > 1$ and N is the data length, The fractal dimension of this function is given by $D = E + 1 - H$. We generated a set of sequences of different lengths N with chosen dimension $D = 1.45$. The length of sequences ranged from 1000 to 10000 samples. Then, for each length, we generated 39 sequences in order to establish 95 % confidence intervals. The results are shown in figure 5.5. The Higuchi method provides good estimates for all data lengths while variance and box counting method offers biased estimate of fractal dimension.The DFA method converges to theoretical value for increasing N. The different algorithms for estimation of fractal dimension were probed by (Esteller et al., 1999). They also concluded that Higuchi's method is the most reliable but, also, the most sensitive to noise. Since FHR certainly contains noise we will also take advantage of the variance method offering robustness to noise (Kinsner, 1994).

The estimated dimension is not dependent only on data length but also on the dimensionality of data. We examined bias of estimated fractal dimension by varying dimension $D = (1, 2)$ of Weierstrass function $w(t)$. For each value of D we generated 39 sequences of length $N = 5000$. The results in figure 5.6 show that dimension estimated by the box counting method is biased. The Higuchi's and variance method provide unbiased estimate of fractal dimension only for dimension $D \geq 1.5$. The DFA method offers biased estimation for $D \leq 1.3$ and $D \geq 1.7$.

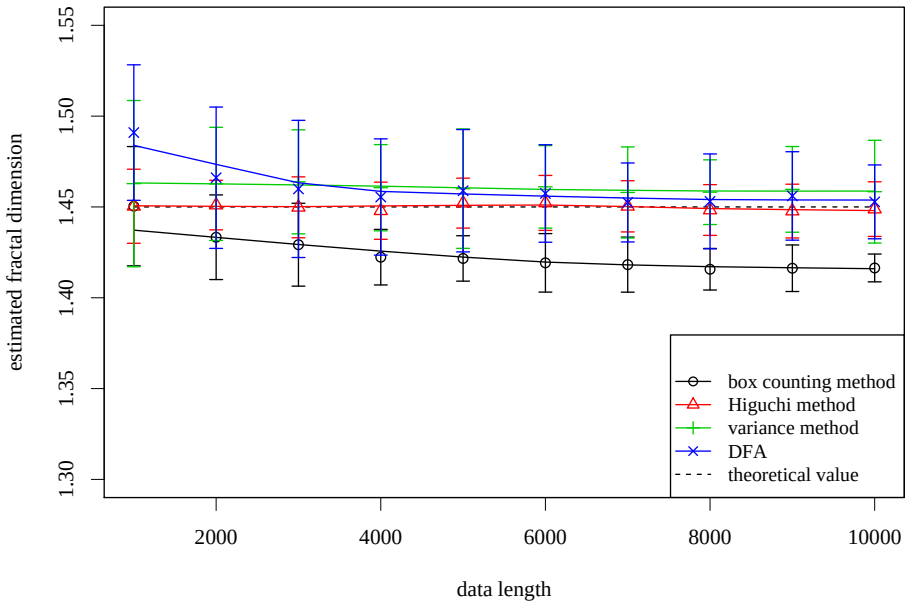


Figure 5.5: Dependence of waveform dimension estimation on the data length. The Weierstrass function was generated with different lengths. Then fractal dimension was estimated by box counting, Higuchi's, variance, and DFA method.

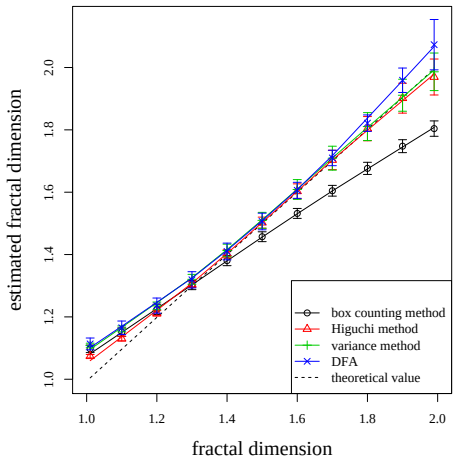


Figure 5.6: Estimated fractal dimension versus theoretical value. The sequences were generated by Weierstrass function where each sequence has length of 5000 samples.

Next, we estimated the two scaling regions as were described by (Higuchi, 1988). He suggested two scaling regions on the log-log plot of some measurement function, e.g. number of boxes, versus size of region, e.g size of box. These two regions are illustrated in figure 5.7. Higuchi named the time where the curve bends as critical time τ_c. This time separates short D_s and long D_l scale waveform fractal dimension for $\leq \tau_c$ and $> \tau_c$, respectively. The region of the short scale reflects the short time variability while the longer scale represents the long time irregularity. To standardize estimated dimension we determined the τ_c for all methods.

Needless to say that τ_c was approximately same for all methods, $\tau_c \approx 3$ s.

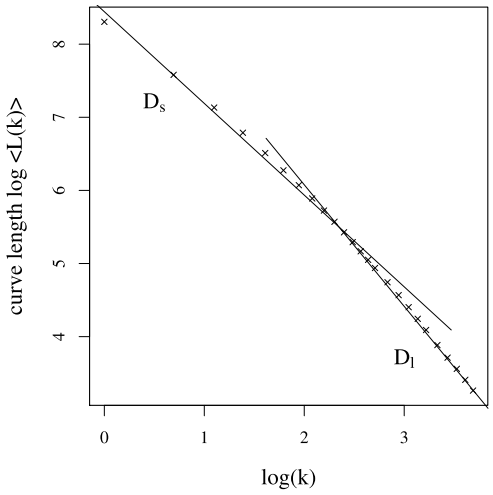


Figure 5.7: Short, D_s, and long, D_l, scale of waveform fractal dimension estimated by Higuchi's method. The curve breaks at $\log(k) = 2.5$. This equals to critical time $\tau_c = e^k/f_s \approx 3$ s, where sampling frequency is $f_s = 4$ Hz.

To conclude this section, the data we use meets demands on required length and reliable estimate of fractal dimension is provided by Higuchi's method. The limitation of waveform fractal analysis lies in biased estimate for low dimension as is shown in figure 5.6. In our work, we mainly focus on 20 minutes directly before delivery which gives 4800 samples for sampling frequency $f_s = 4$ Hz. Performed analysis on synthetic Weierstrass function suggests that we can reliably, by Higuchi's method, estimate fractal dimension. It is necessary to point out, that Weierstrass function does not completely reflects nonlinear and stochastic properties of fetal heart rate. Hence, we have to carefully interpret results of estimated waveform fractal dimensions. For instance (Peng et al., 1995) suggested the minimal data length, for DFA method, to be $N = 8200$ samples. Nevertheless, we relax requirements on data size and use this method anyway but great caution is needed.

5.3.3 Entropy

First of all, let us summarize basic facts about estimation of entropy which were explained above in detail, see sec. 4.4.5. First, KS entropy is useless for a noisy time series of finite length; therefore, ApEn was designed and proved its usefulness for a short and noisy time series. However, it is biased estimation and dependent on data length. Lastly, suggested SampEn eliminated drawbacks of ApEn by reducing bias and data length dependence.

Since FHR records have finite length, the entropy estimation in terms of length is of a major interest. (Pincus, 1995) showed that ApEn is broadly applicable for data series of length $N > 100$. Nevertheless, this was suggested for wide spectrum of applications. In our case, meaningful data length for ApEn is $N \geq 1000$ as it illustrated in figure 5.8(c). ApEn dependence on data length is shown on random numbers with a uniform distribution. Figure 5.8 was proposed by (Richman and Moorman, 2000) and also depicts ApEn and SampEn as a function of r.

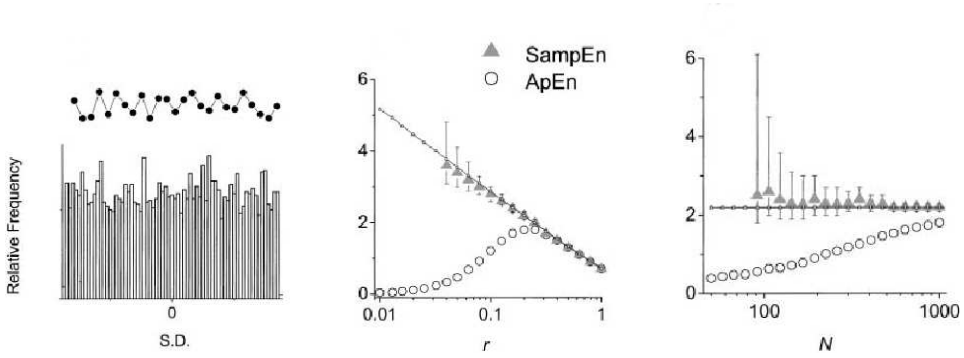


Figure 5.8: SampEn and ApEn of random numbers with uniform distribution. (a) Frequency histogram of 1000 numbers. (b) SampEn and ApEn as function of r,$(m = 2)$. Straight line represents the theoretical value. (c) SampEn and ApEn as function of N. For more details about data generation and theoretical computation see (Richman and Moorman, 2000).

In the previous paragraph, we have shown entropy dependence on data length. Let us now generate signal with variable regularity, Mix(p) process, and estimate entropy. The Mix(p) process, figure 5.9, is defined as $(1 - z) \cdot x + z \cdot y$, where z is a random variable assuming value 1 with probability p and 0 with probability $1 - p$, x is a sequence generated as $x_j = \sqrt{2}sin(2\pi j/12)$ and y is uniformly distributed variable on $[\sqrt{3}, -\sqrt{3}]$.

The results of simulation are presented in figure 5.10. Generated sequence were with $p(0 \div 1)$ and $N = 5000$. The tolerance r is generally considered as $r = (0.15 - 0.2) \cdot SD$, where SD stands for standard deviation of time series. The choice of m was proposed by (Pincus and Viscarello, 1992). They concluded, the best results are achieved when $m = 2$. However, this holds only in cases when a dynamical system is not purely deterministic (Pincus, 1995).

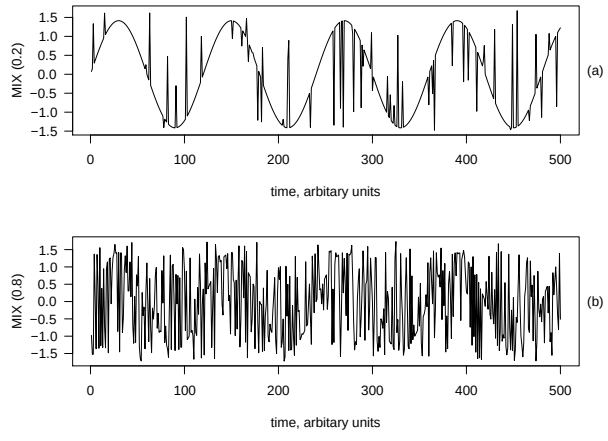


Figure 5.9: The Mix(p) process with $p = 0.2$ and 0.8.

Performed simulation met our expectations and for periodic and regular process the entropy was lower; whereas, for more irregular and random process it was higher. Thus, we verified that entropy is useful for the signal differention.

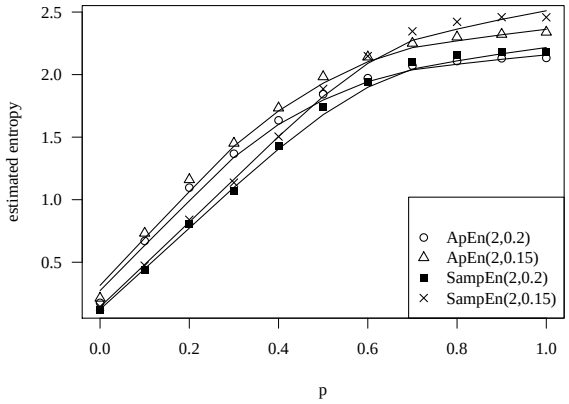


Figure 5.10: Estimated ApEn and SampEn for Mix(p) process. Adjusted values: $m = 2$ and $r = \{0.15, 0.2\}$.

5.3.4 Lempel Ziv Complexity

Lempel Ziv Complexity (LZC) examines reoccurring patterns in time series. The more reoccurring patterns, the less complex signal. The LZC method estimates complexity of encoded signal so that dynamical changes of signal are replaced with particular character. As mentioned in section 4.4.6, we used binary encoding in order to avoid dependence of results on quantification criteria and normalization procedures. The required data length for LZC was examined by (Ferrario et al., 2004). They concluded, the minimum length is 1000 samples for binary encoded data. For the detailed information about used data see referenced paper. Here let us present the estimated complexity of Mix(p) introduced above in section 5.3.3. As expected with growing stochastic component the complexity increased. The encoded Mix(p) process is shown in figure 5.11(a) and the results of estimated LZC are presented in figure 5.11(b).

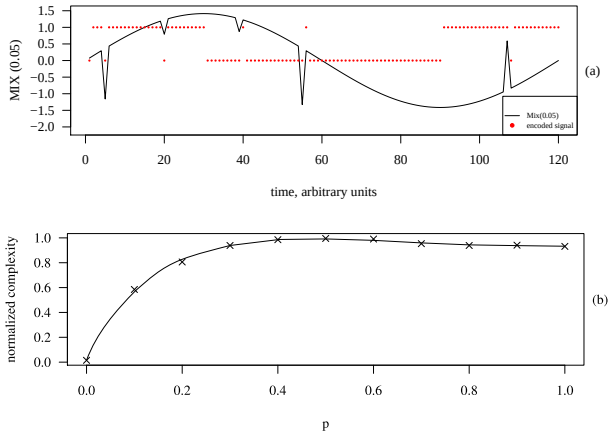


Figure 5.11: Results of estimated complexity for Mix(p) process. (a) encoded sequence of Mix(0.05), (b) normalized complexity of Mix(p) where probability p ranges from 0 to 1. The sequence length is $N = 5000$.

Chapter 6

Results

In this chapter, we present results of fetal heart rate analysis and classification. These results are evaluated in terms of ability to discriminate normal and abnormal fetal heart rate signals. We also perform feature selection in order to find the most suitable features for our domain and, more importantly, reduce dimensionality of the data for further classification. We also emphasize the reasons why some methods are superior to others. At the end of this chapter, a comparison of achieved results with related works is carried out.

6.1 Analysis of fetal heart rate

We analysed all records in database by methods that were described in section 4.4. In table 6.1 we present the abbreviations of respective methods and thus our features.

D_2	correlation dimension estimated from local slopes of correlation sum
D_2 Takens est.	correlation dimension estimated using Takens formula
ApEn	Approximate Entropy, ApEn(2,0.15) - $m = 2$ and $r = 0.15$
SampEn	Sample Entropy
LZC	Lempel Ziv complexity with binary coding
D_σ	Variance fractal dimension
D_0	Box counting fractal dimension
$D_{Higuchi}$	fractal dimension estimated by Higuchi's method
D_{DFA}	fractal dimension estimated by detrend fluctuation analysis method
Ds_{name}	fractal dimension estimated on short scale (below critical time τ_c)
Dl_{name}	fractal dimension estimated on long scale (above critical time τ_c)

Table 6.1: Abbreviations of features

The frequencies of all features for both normal and abnormal records are shown in figure 6.1. The features of normal records are marked by grey color while features of abnormal records are represented by black color. It is apparent that features have different estimated values for different class. The discrimination of two classes is less obvious for features, e.g. Dl_σ and Ds_{DFA} while for features, e.g. $Ds_{Higuchi}$ and LZC the difference between features obtained from normal and abnormal class is evident.

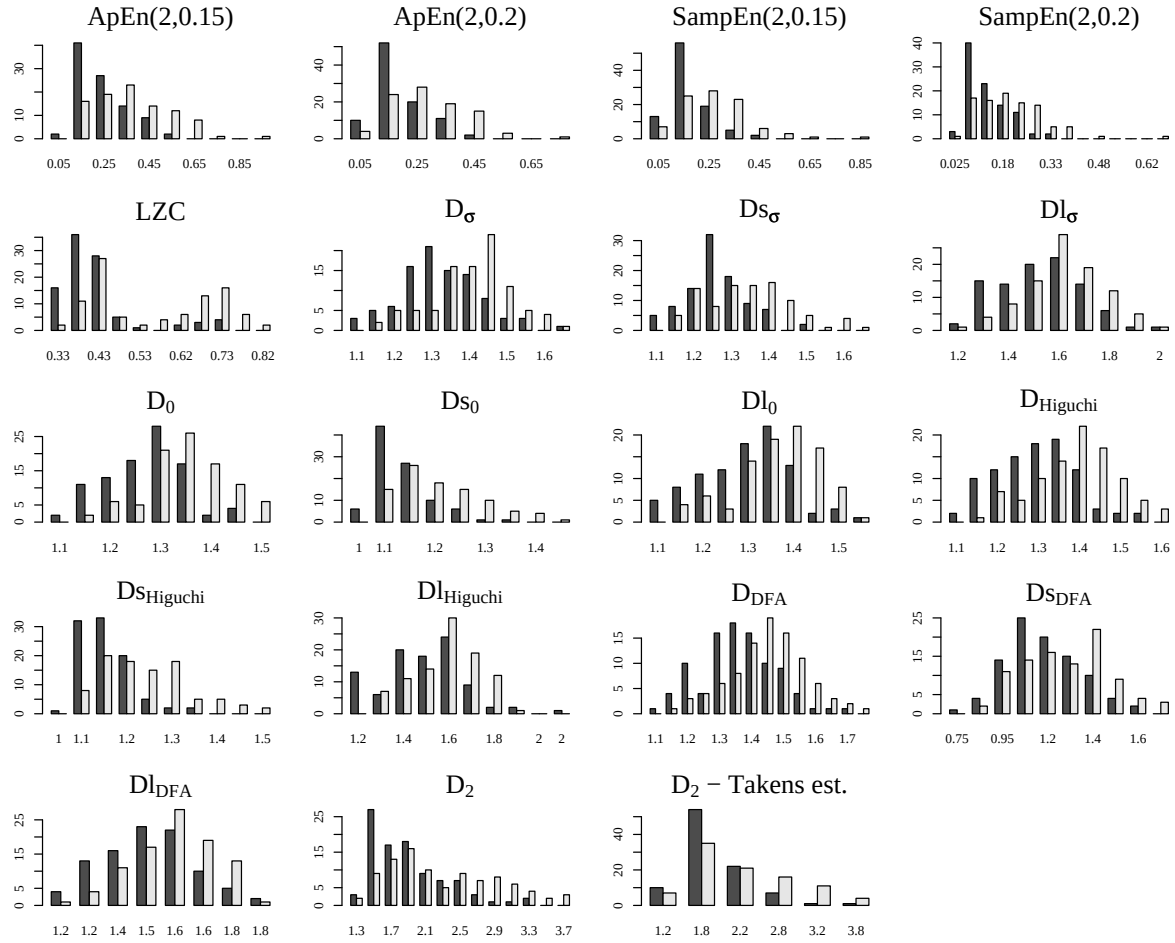


Figure 6.1: The distribution of features' values for respective classes. Grey color represents features of normal FHR tracings and black color represents features of abnormal FHR.

In order to be able to classify significance that these distributions are different, we employed methods of statistical testing. Since the distributions do not follow normal distribution we used Mann-Whitney U test statistics which does not require normal distribution. This test statistics try to reject the null hypothesis that the two distributions differ by the median value on some confidence level, usually $p < 0.01$. In table 6.2 we show mean value μ and standard deviation σ, and Mann-Whitney U test statistics of estimated features.

The null hypothesis can be rejected for all methods. The distribution of normal and abnormal features are better visible on the boxplots shown in figure 6.2. The abnormal class is represented by zero and normal by one. Considering Mann-Whitney U test the best performing features are $Ds_{Higuchi}$, LZC, and Ds_0, D_0, $D_{Higuchi}$, SampEn(2,0.15), and SampEn(2,0.2). This corresponds to observation of results distribution in figure 6.1 and figure 6.2. Note that conclusion about performance are only preliminary. In the following section we will evaluate features and, more importantly, their combination by various methods. It is probable that combination of features will yield to selection of different features.

It is necessary to discuss the values of estimated correlation dimension D_2, see table 6.2. In section 5.1 we searched for optimal embedding dimension m. Cao's method (false nearest

Method	normal ($\mu \pm \sigma$)	abnormal ($\mu \pm \sigma$)	Mann-Whitney
ApEn(2,0.15)	0.373 ± 0.162	0.245 ± 0.108	1.30e-008
ApEn(2,0.2)	0.284 ± 0.132	0.183 ± 0.089	1.50e-008
Sampen(2,0.15)	0.263 ± 0.131	0.169 ± 0.080	2.13e-008
Sampen(2,0.2)	0.198 ± 0.101	0.128 ± 0.063	5.79e-008
LZC	0.556 ± 0.150	0.419 ± 0.095	2.84e-011
D_σ	1.430 ± 0.105	1.348 ± 0.105	1.89e-007
Ds_σ	1.319 ± 0.115	1.235 ± 0.081	2.62e-007
Dl_σ	1.569 ± 0.156	1.476 ± 0.166	1.04e-004
D_0	1.323 ± 0.081	1.251 ± 0.076	4.37e-009
Ds_0	1.183±0.083	1.111±0.053	1.65e-010
Dl_0	1.393±0.092	1.319±0.100	3.20e-007
$D_{Higuchi}$	1.366±0.102	1.274±0.098	2.61e-009
$Ds_{Higuchi}$	1.216±0.096	1.129±0.057	1.51e-011
$Dl_{Higuchi}$	1.536±0.149	1.439±0.181	2.94e-005
D_{DFA}	1.421 ± 0.118	1.338 ± 0.117	1.20e-006
Ds_{DFA}	1.226 ± 0.186	1.129 ± 0.162	3.34e-004
Dl_{DFA}	1.546 ± 0.139	1.462 ± 0.154	5.93e-005
D_2	2.294 ± 0.633	1.902 ± 0.436	8.20e-006
D_2 - Takens est.	2.279 ± 0.629	1.909 ± 0.415	6.64e-006

Table 6.2: Mean, μ, standard deviation, σ, and Mann-Whitney U test statistics for estimated features values.

neighbors algorithm with additional stopping criteria) suggested sufficient m to be in range of $m = 5 - 6$. This implies rather high dimensional underlying dynamics but, as table 6.2 shows, the value of D_2 is small in comparison with found m. There are three possible explanations. First, the algorithm of Cao provides sufficient integer m for proper embedding which is not always necessary. Second, the RR intervals from external (ultrasound) records were smoothed. Last, during preprocessing stage, artefacts were replaced by linear interpolation, i.e. a complex behaviour was replaced by line which yielded to low correlation dimension. We verified the later hypothesis on the Normal Sinus Rhythm RR interval database (Goldberger et al., 2000). The estimated correlation dimension for the signal without artefacts was $D_2 = 4.5$ while for 20% signal contaminated with artefacts the dimension decreased to $D_2 = 3.7$. The additional decrease of D_2 was, most probably, caused by measurement process. To be more specific, the external monitoring uses Doppler ultrasound for fetal heart beat detection. These heart beats are usually determined from periodicity of Doppler envelope using autocorrelation function. This function tends to average slight successive changes of heart beats resulting in lost of fetal heart rate variability (Jezewski et al., 2006).

It is important to notice that results of waveform fractal dimension in table 6.2 resemblance to the results achieved in section 5.3.2 where we analysed the synthetic Weierstrass cosine function. We concluded that Higuchi method provides unbiased estimate of fractal dimension, whereas the box conting method provides underestimated and box counting and DFA method offer overestimated fractal dimension. This conclusion corresponds to the values present in table 6.2.

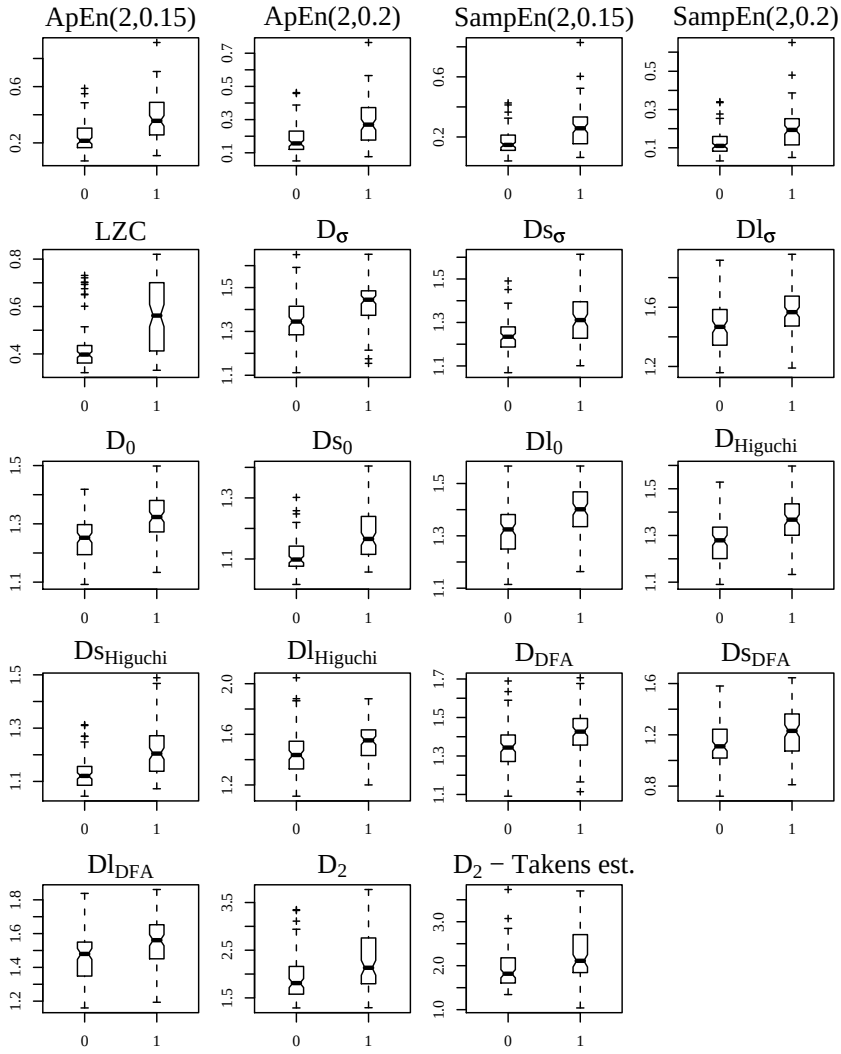


Figure 6.2: Box plots of features for different class. The pathological, abnormal class is represented by the number 0 while normal class by the number 1.

6.2 Feature selection

The statistical analysis performed in the previous section can be considered as simple feature selection where features with the lower p-value are considered to be more significant. In this section we continue with features assessment and rank them according to their power to discriminate normal and pathological fetuses and suitability for the further classification. In order to evaluate the best features we rank features obtained from particular feature selection methods. Those features with the best rank are selected for classification.

First, we examined correlation beetwen particular feature and class. In addition, we estimated the intercorrelations among all features and evaluated the significance of correlation. The results form the correlation matrix where rows and columns are equal to number of features plus class. Since this representation is cumbersome, we present it only in the appendix B. To be able to better distinguish data we picture the correlation matrix as an image where color scales are used to symbolize the values of correlation coefficients. Such representation is illus-

trated in figure 6.3. The corresponding p-values were all $p < 0.01$. Hence, we can conclude that correlations are statistically significant.

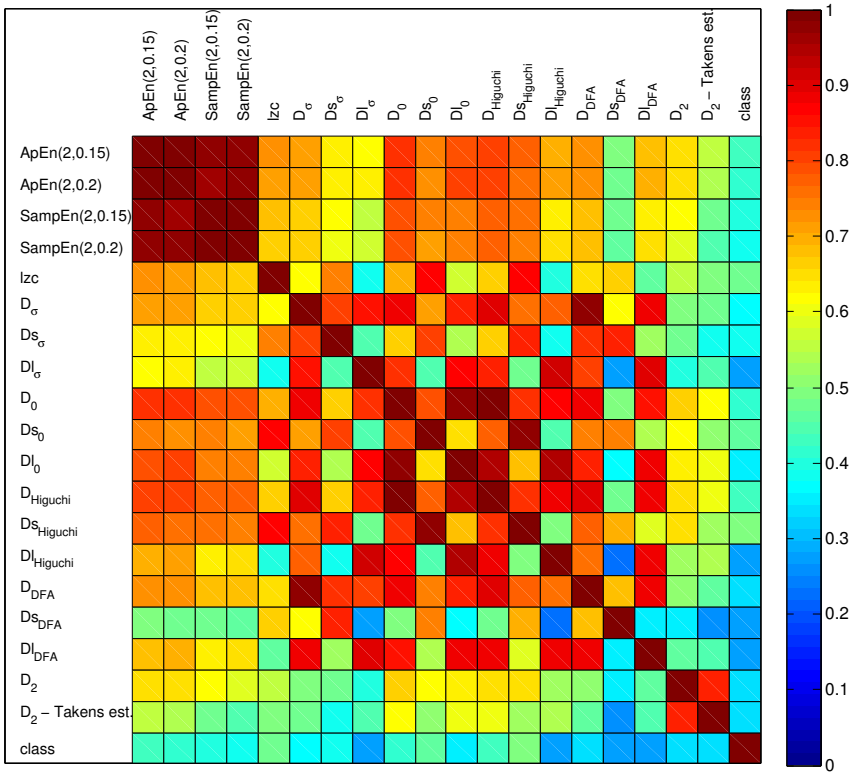


Figure 6.3: Correlation matrix represented as image often referred as correlation map. Scale gives value of R for each variable pair.

In general, we can conclude that fractal dimension estimated by various methods are more or less correlated among themselves. However, some interesting properties can be find by examining the correlation map and these are as follows: First, the fractal dimensions estimated on short scale (Ds_σ, Ds_0, $Ds_{Higuchi}$, Ds_{DFA}) are, as it was expected, in strong correlation. This hold also for long scale dimensions. Second, the long scale dimension Dl is more correlated with dimension D than is the short scale dimension Ds. This is a consequence of imbalanced number of points used for the short and long scale estimation. Due to finite data precision, the size of region can not be decreased towards zero and fractal dimension is estimated only for few lengths. Therefore, dimension D estimated across the whole scale tends to long scale dimension Dl.

The entropy measures ApEn and SampEn are in strong relationship and also moderately correlate with class. The best performance, in terms of correlation, achieved Lempel Ziv complexity and waveform fractal dimension estimated on short scale by Higuchi' and box counting method $Ds_{Higuchi}$ and Ds_0, respectively.

Next, we performed feature selection in Weka data mining software (Witten and Frank, 2005). We used correlation based feature selection (CFS), information gain (InfoGain), and wrapper method using Naive Bayes classifier, for details and theory see section 4.6.1. The 10-fold cross-validation technique was used for evaluation of features quality. Test statistics of Mann-Whitney and correlation criteria assess features individually, i.e. each feature is eval-

uated independently of all others. The rest methods, CFS, InfoGain, and Wrapper of Naive Bayes, assess subset of features. Thus, redundant features are excluded but complementary preserved. We evaluated the results of selection by a hybrid approach. First, the features from Mann-Whitney statistics testing, correlation criteria, and InfoGain were ordered according to performance and non-integer rank between zero and five was assigned. Second, for CFS and wrapper method a feature was considered if appeared at least in one fold of cross-validation. The best performing feature was assigned rank zero and for the worst feature, i.e. feature that not appeared in any fold of cross-validation, rank five.

The results of feature's rank for each method are presented in table 6.3. The best performing features are those with the lowest overall score. It is necessary to point out that this score is rather a suggestion than absolute order.

Method	Mann-Whitney	Corr. criteria	CFS	InfoGain	Wrapper	Overall score
ApEn(2,0.15)	1.58	1.05	3	1.84	4	11.47
ApEn(2,0.2)	1.84	1.84	4	2.89	3	13.58
SampEn(2,0.15)	2.11	2.11	1	2.37	3	10.58
SampEn(2,0.2)	2.37	2.63	3	3.68	4	15.68
LZC	0.53	0.53	0	0.26	0	1.32
D_σ	2.63	2.89	4	3.16	4	16.68
Ds_σ	2.89	2.37	0	1.05	3	9.32
Dl_σ	4.74	4.47	4	5	5	23.21
D_0	1.32	1.58	0	0.53	2	5.42
Ds_0	3.16	0.79	0	0.79	1	5.74
Dl_0	0.79	3.16	4	2.63	3	13.58
$D_{Higuchi}$	1.05	1.32	4	1.58	5	12.95
$Ds_{Higuchi}$	0.26	0.26	4	1.32	4	9.84
$Dl_{Higuchi}$	4.21	4.21	4	3.95	4	20.37
D_{DFA}	3.42	3.68	4	4.21	3	18.32
Ds_{DFA}	5	5	4	4.74	5	23.74
Dl_{DFA}	4.47	4.74	4	4.47	3	20.68
D_2	3.95	3.42	4	3.42	5	19.79
D_2 - Takens est.	3.68	3.95	2	2.11	1	12.74

Table 6.3: Results of feature selection. Feature's rank represents its suitability for our domain. Note that the lower overall score the better feature performs in the domain. For details about ranking, see text.

We have chosen eight best performing features. These are as follows: LZC, D_0, Ds_0, Ds_σ, $Ds_{Higuchi}$, SampEn(2,0.15), ApEn(2,0.15), and D_2 estimated by Takens. It is surprising that short scale fractal dimension Ds is included three times. Since Ds is strongly correlated for all methods, the inclusion to the eight best performing features is questionable. Considering table 6.3 we can conclude that $Ds_{Higuchi}$ has strong correlation with class and discrimination of two classes as well, whereas Ds_0 and Ds_σ perform best when subset of features is considered. Considering feature subset, only InfoGain method ranks $Ds_{Higuchi}$ as usefull feature, the other methods, CFS and Wrapper, suggest it as insignificant. Based on this evidence we can not exclude any of Ds. The situation is easier for SampEn and ApEn. They are also in strong correlation among themselves but when subset of features is consider they turns out to be almost the same. Therefore, we excluded ApEn(2,0.15) because it performed worst on CFS and Wrapper.

Eventually, we have chosen seven features.

From all of features, the best performing was Lempel Ziv complexity. The complexity of fetal heart rate decreases when repetitive patterns occurs, such as uniform and variable decelerations. However, as we pointed out in section 2.2.1, only uniform late and complicated variable decelerations are connected with marked fetal hypoxia. As a result, those records with early uniform and uncomplicated variable decelerations had also decreased complexity. We shall point out that these records are problematic for all methods and can be properly analysed only if uterine pressure is considered. The short scale fractal dimension estimated by box counting, variance, and Higuchi's method suggests that decreased short time variability is important indicator of metabolic acidemia. The worst results were achieved by DFA. It is a consequence of relaxed requirements on data length. (Peng et al., 1995) proposed the minimal data length of $N = 8200$ samples. However, we used segments of length $N = 4800$ samples.

6.3 Feature classification

A classifier operates in feature space and tries to discriminate features into two groups: normal and pathological. Feature space is formed by the previously selected features (LZC, D_0, Ds_0, Ds_σ, $Ds_{Higuchi}$, SampEn(2,0.15), and D_2 estimated by Takens) with dimension equal to the number of features. As mentioned above, we utilised 10-fold cross-validation and then used following classifiers: Naive Bayes, Support Vector Machine (SVM), and decision tree (C4.5). For the SVM, the polynomial kernel and penalty parameter $C = 1$ were adjusted. The classification results are present in table 6.4. From all performance measures, the specificity is of major importance since a classifier with higher specificity causes lower number of false alarms that leads to lower rate of unnecessary intervention. Regarding the specificity, the SVM performed best. However, statistical tests revealed that difference between individual classifiers is statistically insignificant on $p < 0.01$ confidence level. Only accuracy of Naive Bayes is significant to accuracy of C4.5.

	Naive Bayes	Support Vector Machine	decision tree C4.5
accuracy [%]	73.08	**72.25**	65.2
sensitivity	0.84	**0.78**	0.74
specificity	0.64	**0.7**	0.57
geometric mean	0.73	**0.74**	0.65
AUC	0.79	**0.74**	0.69

Table 6.4: Results of feature classification.

We achieved classification results of 78% of sensitivity and 70% specificity using nonlinear features only. These results are comparable to inter-observer variability (Amer-Wåhlin et al., 2005). Let us, in detail, compare classification performance of nonlinear features with linear and morphological ones (Chung et al., 1995; Salamalekis et al., 2002; Georgoulas et al., 2006, 2007). It is necessary to point out that direct comparison of these works, which also assessed fetal heart rate intrapartum, is not feasible because different methodologies were used. (Chung et al., 1995) reported sensitivity 88% and specificity 75%. They used the same pH threshold 7.15 but worked with small sample size ($n = 73$). (Salamalekis et al., 2002) achieved good performance of 100% sensitivity and 92.5% specificity for pH borderline of 7.15. However,

they worked also with small sample size ($n = 61$) and, in addition, employed combination of fetal heart rate features and features from pulse oximetry signal that are not standard in clinical settings. (Georgoulas et al., 2006) achieved 85% sensitivity and 70% specificity. They used substantially larger data set than the two former works with ($n = 130$) records. However, from the all records only 30 were pathological. Therefore, they used SMOTE technique to balance number of examples in each class. Since this technique operates in feature space it is questionable if the artificially created pathological features reflect the real fetal state. Moreover, they employed the two pH thresholds. First determining pathological fetuses, pH < 7.10, and second determining the normal fetuses, pH > 7.20. Records between these thresholds were excluded. In the later work (Georgoulas et al., 2007), using grammatical evolution, they obtained sensitivity 93 % and specificity 87 %, though the same restriction applies.

Comparing with the results reported by (Chung et al., 1995; Georgoulas et al., 2006), they achieved specificity and sensitivity are comparable to our results. Comparing with the results of (Salamalekis et al., 2002; Georgoulas et al., 2007), they are superior to our. However, caution is needed since it is not direct comparison.

Chapter 7

Conclusion

This work offers a thorough review of nonlinear methods (fractal dimension, entropy, complexity) and examines their applicability to fetal heart rate analysis. The suitability of nonlinear methods for heart rate analysis has already been proven in adults heart rate research. It is generally accepted that human heart is driven by deterministic chaos and fluctuates on different time scales. Because segments of fetal heart rate of short duration were used, the surrogate date test was performed. The null hypothesis that fetal heart rate is linear stochastic process was rejected on $p < 0.05$ confidence level. Hence, the use of nonlinear methods even for fetal heart rate of short duration has been justified.

The fetal heart rate is non-uniformly sampled signal and contains artefacts and noise. In preprocessing stage, artefacts were removed and signal re-sampled in order to obtain evenly spaced time instances. In addition, analysis without resampling was carried out. Based on evidence, it was concluded that non-uniform sampling did not affect analysis results, because sampling was deterministically non-uniform. Segments of 20 minutes length were used for analysis. These segments were chosen as close as possible to delivery. All employed nonlinear methods were theoretically proven only for infinite precision and, more importantly, for infinite data size. Therefore, it was necessary to establish the required data size where minimal data length was not established in literature. Only for the detrend fluctuation analysis the data length was considered insufficient in advance and eventually resulted in poor performance.

In order to analyse fractal dimension of attractor, search for optimal embedding parametres to properly reconstruct state space from fetal heart rate was carried out. The optimal time delay was $\tau = 12.3 \pm 9.5\ s$ and optimal embedding dimension $m = 6$. Afterwards, the correlation dimension of attractor was estimated. The fractal dimension of fetal heart rate was found to be lower than expected. This is a consequence of artefacts interpolation and also of measurement process where subtle variations of successive heart beats were smoothed. Next, several methods for estimation of waveform fractal dimension were utilized. These methods did not require underlying deterministic dynamics but only considered the signal as geometrical representation in two-dimensional space. At last, approximate entropy, sample entropy, and Lempel Ziv complexity were used for analysis.

The several possible approaches to measure outcome of neonates were discused and based on literature, artery pH below 7.15 and Apgar score at five minutes below or equal 7 were used as suitable markers of metabolic acidemia. The performed feature selection showed that most suitable methods for fetal heart rate analysis were: Lempel Ziv complexity, waveform fractal dimension estimated by the box counting method, sample entropy and correlation dimension of

attractor. These methods were used for further classification. Lempel Ziv complexity assesses reoccurring patterns of fetal heart rate and thus reduced variability and repetitive decelerations of the signal resulted in lower complexity. Also the waveform fractal dimension of short scale suggested that short term variability is important marker of fetal distress.

The results showed that nonlinear methods are useful even for classification of fetal heart rate segments of short duration. With arbitrary classifier the achieved results are comparable with inter-observer variability and with related works that used linear and morphological methods, although better results were published. The best performing classifier, in term of specificity, was support vector machine. It achieved sensitivity 0.78 and specificity 0.7. Nevertheless, these results turned as statistically insignificant in comparison with results of other classifiers (Naive Bayes, C4.5).

In this work, using only nonlinear methods that are not commonly used, good performance comparable to inter-observer disagreement was achieved. It is highly probable that inclusion of linear and morphological features will lead to better performance. This is, however, open space for further work.

7.1 Future work

Achieved results are promising for further research. It has been proven that nonlinear approach is useful for fetal heart rate analysis. The performance could be increased by using uterine pressure in connection with fetal heart rate which is not common practice and definitely worth exploring. In addition, reconstructed state space could be used for nonlinear interpolation of artefacts and, moreover, the prediction of fetal heart rate. The future work will also focus on inclusion of linear and morphological features into the existing system.

It should be helpful to include other additional information about the state of the patient such as age and possible risk factors as well as utilization of the information provided by the STAN® system such as T/QRS ratio or width of the QRS complex.

Measures of outcome, artery pH and Apgar score, were found to be not ideal. Also strict division into two groups was not probably appropriate. A better division would be to separate fetuses into three groups (FIGO, 1986): normal, suspicious, and pathological. As another outcome measure assessment of records by experienced obstetricians will be used. Then the two types of measure will be available: objective (pH and Apgar score) and subjective (doctors assessment).

Bibliography

H. Abarbanel, R. Brown, J. Sidorowich, and L. Tsimring. The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, 65 (4):1331–1392, 1993.

A. Alonso-Betanzos, B. Guijarro-Berdiñas, V. Moret-Bonillo, and S. López-Gonźalez. The NST-EXPERT project: the need to evolve. *Artif Intell Med*, 7(4):297–313, Aug 1995.

I. Amer-Wahlin. *Fetal ECG waveform analysis for intrapartum monitoring*. PhD thesis, Dept. Obstetrics and Gynecology, Lund University Hospital, 2003.

I. Amer-Wahlin, C. Hellsten, H. Noren, H. Hagberg, A. Herbst, I. Kjellmer, H. Lilja, C. Lindoff, M. Mansson, L. Martensson, P. Olofsson, A. Sundstrom, and K. Marsal. Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial. *Lancet*, 358(9281):534–538, Aug 2001.

I. Amer-Wåhlin, I. Ingemarsson, K. Marsal, and A. Herbst. Fetal heart rate patterns and ECG ST segment changes preceding metabolic acidaemia at birth. *BJOG*, 112(2):160–165, Feb 2005.

J. Bernardes, C. Moura, J. P. de Sa, and L. P. Leite. The Porto system for automated cardiotocographic signal analysis. *J Perinat Med*, 19(1-2):61–65, 1991.

J. Bernardes, A. Costa-Pereira, D. A. de Campos, H. P. van Geijn, and L. Pereira-Leite. Evaluation of interobserver agreement of cardiotocograms. *Int J Gynaecol Obstet*, 57(1):33–37, Apr 1997.

H. F. V. Boshoff. A fast box counting algorithm for determining the fractal dimension of sampled continuous functions. In *Communications and Signal Processing, 1992. COMSIG '92., Proceedings of the 1992 South African Symposium on*, pages 43–48, 11 Sept. 1992.

L. Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*, 110:43–50, 1997.

S. Cerutti, S. Civardi, A. Bianchi, M. Signorini, E. Ferrazzi, and G. Pardi. Spectral analysis of antepartum heart rate variability. *Clin Phys Physiol Meas.*, 10:27–31, 1989.

D. G. Chaffin, C. C. Goldberg, and K. L. Reed. The dimension of chaos in the fetal heart rate. *Am J Obstet Gynecol*, 165(5 Pt 1):1425–1429, Nov 1991.

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

T. Chung, M. Mohajer, X. Yang, A. Chang, and D. Sahota. The prediction of fetal acidosis at birth by computerized analysis of intrapartum cardiotocography. *Br. J. Obstet. Gynaecol.*, 102 (6):454–460, 1995.

M. Costa, A. L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of biological signals. *Phys Rev E Stat Nonlin Soft Matter Phys*, 71(2 Pt 1):021906, Feb 2005.

F. Cunningham. *Williams Obsterics*, chapter 18. Intrapartum assesment, pages –. Mc Graw Hill, 2005.

D. A. de Campos, J. Bernardes, A. Garrido, J. M. de Sa, and L. Pereira-Leite. SisPorto 2.0: a program for automated analysis of cardiotocograms. *J Matern Fetal Med*, 9(5):311–318, 2000. doi: 3.0.CO;2-9.

V. Doria, A. T. Papageorghiou, A. Gustafsson, A. Ugwumadu, K. Farrer, and S. Arulkumaran. Review of the first 1502 cases of ECG-ST waveform analysis during labour in a teaching hospital. *BJOG*, 114(10):1202–1207, Oct 2007. doi: 10.1111/j.1471-0528.2007.01480.x.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000. ISBN 0471056693.

J. C. Echeverria, B. R. Hayes-Gill, J. A. Crowe, M. S. Woolfson, and G. D. H. Croaker. Detrended fluctuation analysis: a suitable method for studying fetal heart rate variability? *Physiol Meas*, 25(3):763–774, Jun 2004.

J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57:617–656, 1985.

A. Eke, P. Hermán, J. B. Bassingthwaighte, G. M. Raymond, D. B. Percival, M. Cannon, I. Balla, and C. Ikrényi. Physiological time series: distinguishing fractal noises from motions. *Pflugers Arch*, 439(4):403–415, Feb 2000.

A. Eke, P. Herman, L. Kocsis, and L. R. Kozak. Fractal characterization of complexity in temporal physiological signals. *Physiol Meas*, 23(1):R1–38, Feb 2002.

S. Ellner. Estimating attractor dimensions from limited data: A new method, with error estimates. *Phys. Lett. A*, 133:128–133, 1988.

R. Esteller, G. Vachtsevanos, J. Echauz, and B. Lilt. A comparison of fractal dimension algorithms using synthetic and experimental data. In *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on*, volume 3, pages 199–202vol.3, 30 May-2 June 1999.

C. S. Felgueiras, J. P. de Sá, J. Bernardes, and S. Gama. Classification of foetal heart rate sequences based on fractal features. *Med Biol Eng Comput*, 36(2):197–201, Mar 1998.

M. Ferrario, M. G. Signorini, and S. Cerutti. Complexity analysis of 24 hours heart rate variability time series. *Conf Proc IEEE Eng Med Biol Soc*, 6:3956–3959, 2004.

M. Ferrario, M. G. Signorini, and G. Magenes. Complexity analysis of the fetal heart rate for the identification of pathology in fetuses. In *Computers in Cardiology, 2005*, pages 989–992, Sept. 25-28, 2005.

FIGO. Guidelines for the use of fetal monitoring. *International Journal of Gynecology & Obstetrics*, 25:159–167, 1986.

A. M. Fraser and H. L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2):1134–1140, 1986.

R. Furlan, S. Guzzetti, W. Crivellaro, S. Dassi, M. Tinelli, G. Baselli, S. Cerutti, F. Lombardi, M. Pagani, and A. Malliani. Continuous 24-hour assessment of the neural regulation of systemic arterial pressure and RR variabilities in ambulant subjects. *Circulation*, 81(2):537–547, Feb 1990.

A. Galka. *Topics in nonlinear time series analysis, with implications for EEG analysis, in: Advanced Series in Nonlinear Dynamics*. World Scientific, Singapore., 2000.

G. Georgoulas, C. Stylios, G. Nokas, and P. Groumpos. Classification of fetal heart rate during labour using hidden markov models. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 3, pages 2471–2475vol.3, 25-29 July 2004.

G. Georgoulas, C. D. Stylios, and P. P. Groumpos. Feature extraction and classification of fetal heart rate using wavelet analysis and support vector machines. *International Journal on Artificial Intelligence Tools*, 15:411–432, 2005.

G. Georgoulas, C. D. Stylios, and P. P. Groumpos. Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using svm. *IEEE Trans. Biomed. Eng.*, 53(5):–, 2006.

G. Georgoulas, D. Gavrilis, I. G. Tsoulos, C. D. Stylios, J. Bernardes, and P. P. Groumpos. Novel approach for fetal heart rate classification introducing grammatical evolution. *Biomedical Signal Processing and Control*, 2:69–79, 2007.

A. Goldberg, V. Bhargava, and A. West, BJ. Mandell. On mechanism of cardiac electrical stability: the fractal hypothesis. *Biophys J*, 48:525–528, 1985.

A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):E215–E220, Jun 2000.

A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley. Fractal dynamics in physiology: alterations with disease and aging. *Proc Natl Acad Sci U S A*, 99 Suppl 1:2466–2472, Feb 2002.

H. Goncalves, A. P. Rocha, D. Ayres-de Campos, and J. Bernardes. Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery. *Med Bio Eng Comput*, 44:847–855, 2006.

N. A. Gough. Fractal analysis of foetal heart rate variability. *Physiol Meas*, 14(3):309–315, Aug 1993.

P. Grassberger and I. Procaccia. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev*, 28:2591–2593, 1983.

K. Greene and R. Keith. K2 medical system. http://ww.k2ms.com/, 2002.

B. Guijarro-Berdinas, A. Alonso-Betanzos, and O. Fontenla-Romero. Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system. *Artif. Intell.*, 136:1–27, 2002.

I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

M. A. Hall. Correlation-based feature selection for machine learning. Technical report, The University of Waikato, 1998.

R. Hegger, H. Kantz, and T. Schreiber. Practical implementation of nonlinear time series methods: The tisean package. *Chaos*, 9(2):413–435, Jun 1999.

T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Phys. D*, 31 (2):277–283, 1988. ISSN 0167-2789.

P. Hopkins, N. Outram, N. Löfgren, E. C. Ifeachor, and K. G. Rosén. A comparative study of fetal heart rate variability analysis techniques. *Conf Proc IEEE Eng Med Biol Soc*, 1: 1784–1787, 2006.

J. Jezewski, J. Wrobel, and K. Horoba. Comparison of doppler ultrasound and direct electrocardiography acquisition techniques for quantification of fetal heart rate variability. *IEEE J BME*, 53(5):855–864, 2006. ISSN 0018-9294.

H. Kantz and T. Schreiber. *Nonlinear time series analysis*. Cambridge Univesity Press, UK, 2. edition, 2004.

D. Kaplan and P. Staffin. Software for Heart Rate Variability. [online]. Macalester College, St. Paul, [cited 2009-04-26], April 1998. Available from: http://www.macalester.edu/~kaplan/hrv/doc.

F. Kaspar and H. Schuster. Easily calculable measure of the complexity of spatiotemporal patterns. *Physical Review A*, 36:842–848, 1987.

M. B. Kennel, R. Brown, and H. D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45:3403, 1992.

A. Kikuchi, N. Unno, T. Horikoshi, T. Shimizu, S. Kozuma, and Y. Taketani. Changes in fractal features of fetal heart rate during pregnancy. *Early Hum Dev*, 81(8):655–661, Aug 2005.

W. Kinsner. Batch and real-time computation of a fractal dimension based on variance of a time series. Technical report, Department of Electrical & Computer Engineering, University of Manitoba, Winnipeg, Canada, 1994.

J. O. E. H. V. Laar, M. M. Porath, C. H. L. Peters, and S. G. Oei. Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature. *Acta Obstet Gynecol Scand*, 87 (3):300–306, 2008.

D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol*, 283(3):R789–R797, Sep 2002.

A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, IT-22 (1):75–81, 1976.

G. Magenes, M. G. Signorini, and D. Arduini. Classification of cardiotocographic records by neural networks. In *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, volume 3, pages 637–641vol.3, 24-27 July 2000.

R. Mantel, H. P. van Geijn, F. J. Caron, J. M. Swartjes, E. E. van Woerden, and H. W. Jongsma. Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations. *Int J Biomed Comput*, 25(4):273–286, May 1990a.

R. Mantel, H. P. van Geijn, F. J. Caron, J. M. Swartjes, E. E. van Woerden, and H. W. Jongsma. Computer analysis of antepartum fetal heart rate: 1. Baseline determination. *Int J Biomed Comput*, 25(4):261–272, May 1990b.

C. Merkwirth, U. Parlitz, I. Wedekind, and W. Lauterborn. (2002). Open-TSTOOL. MATLAB toolbox. [cited 2009-03-20]. Accessed at `http://www.phsik3.gwdg.de/tstool`, 2007.

NIH. Electronic fetal heart rate monitoring: research guidelines for interpretation. National Institute of Child Health and Human Development Research Planning Workshop. *Am J Obstet Gynecol*, 177(6):1385–1390, Dec 1997.

NIHCE. Intrapartum care - Care of healthy women and their babies during childbirth. Technical report, National Institute for Health and Clinical Excellence, 2007.

H. Noren, I. Amer-Wahlin, H. Hagberg, A. Herbst, I. Kjellmer, K. Marsal, P. Olofsson, and K. G. Rosen. Fetal electrocardiography in labor and neonatal outcome: data from the Swedish randomized controlled trial on intrapartum fetal monitoring. *Am J Obstet Gynecol*, 188(1): 183–192, Jan 2003.

K. Ojala, M. Vaarasmaki, K. Makikallio, M. Valkama, and A. Tekay. A comparison of intrapartum automated fetal electrocardiography and conventional cardiotocography–a randomised controlled study. *BJOG*, 113(4):419–423, Apr 2006.

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Physical Review Letters*, 45(9):712+, 1980.

O. Palomaki, T. Luukkaala, R. Luoto, and R. Tuimala. Intrapartum cardiotocography – the dilemma of interpretational variation. *J Perinat Med*, 34(4):298–302, 2006.

C. K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5:82–87, 1995.

V. Pierrat, N. Haouari, A. Liska, D. Thomas, D. Subtil, P. Truffert, and G. d'Etudes en Epidémiologie Périnatale. Prevalence, causes, and outcome at 2 years of age of newborn encephalopathy: population based study. *Arch Dis Child Fetal Neonatal Ed*, 90(3):F257–F261, May 2005.

S. Pincus. Approximate entropy (ApEn) as a complexity measure. *Chaos*, 5 (1):110–117, 1995.

S. M. Pincus and R. R. Viscarello. Approximate entropy: a regularity measure for fetal heart rate analysis. *Obstet Gynecol*, 79(2):249–255, Feb 1992.

J. Quinlan. *C4.5: Programs for machine learning*. Morgan Kaufmann, 1992.

J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*, 278(6):H2039–H2049, Jun 2000.

D. Rulle. Deterministic chaos: The science and the fiction. *Proc. R. Soc. London*, 427:241–248, 1990.

E. Salamalekis, P. Thomopoulos, D. Giannaris, I. Salloum, G. Vasios, A. Prentza, and D. Koutsouris. Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks. *BR. J. Obstet. Gynaecol.*, 10:1137–1142, 2002.

T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.*, 77 (4):635–638, 1996.

C. Sevcik. A procedure to estimate the fractal dimension of waveforms. *Complexity International*, 5:–, 1998.

C. E. Shannon. A mathematical theory of communication. *Bells Systems Tech J.*, 27:379–423, 623–656, 1948.

O. Sibony, J. Fouillot, M. Benaudia, A. Benhalla, P. Blot, and C. Sureau. Spectral analysis: a method for quantitating fetal heart rate variability. In H. van Geijn and F. Copray, editors, *A Critical Appraisal of Fetal Surveillance*, pages 325–332. New Your, Elsevier Science, 1994.

M. G. Signorini, G. Magenes, S. Cerutti, and D. Arduini. Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. *IEEE Trans Biomed Eng*, 50(3):365–374, Mar 2003.

J. C. Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, 2003. ISBN 0198508409.

P. J. Steer. Has electronic fetal heart rate monitoring made a difference. *Seminars in Fetal and Neonatal Medicine*, 13:2–7, 2008.

J. Store and R. Bulirsch. *Introduction to numerical analysis*. Springer-Verlag, 2 edition, 1993.

A. Sundstrom, D. Rosen, and K. Rosen. Fetal surveillance - textbook. [online] Gothenburg, Sweden: Neoventa Medical AB, [cited 2009-02-28], January 2000. Available from: `http://www.neoventa.com/`.

F. Takens. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence*, 4: 366–381, 1981.

F. Takens. On the numerical determination of the dimension of an attractor. *Dynamical Systems and Bifurcations*, 1125:99–106, 1984.

Task-Force. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Eur Heart J*, 17(3):354–381, Mar 1996.

G. M. Taylor, G. J. Mires, E. W. Abel, S. Tsantis, T. Farrell, P. F. Chien, and Y. Liu. The development and validation of an algorithm for real-time computerised fetal heart rate monitoring in labour. *BJOG*, 107(9):1130–1137, Sep 2000.

J. Theiler. *Quantifying chaos:practical estimation of the correlation dimension*. PhD thesis, California Inst. of Technology, Pasadena, 1988.

J. Theiler. Estimating fractal dimension. *J. Opt. Soc. Am. A*, 7 (6):1055–1073, 1990.

P. Van Leeuwen and H. Bettermann. The status of nonlinear dynamics in the analysis of heart rate variability. *Herzschrittmachertherapie und Elektrophysiologie*, 11 (3):127–130, 2000.

V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995. ISBN 0387945598.

M. Westerhuis, A. Kwee, A. van Ginkel, A. Drogtrop, W. Gyselaers, and G. Vissera. Limitations of ST analysis in clinical practice: three cases of intrapartum metabolic acidosis. *BJOG: an International Journal of Obstetrics and Gynaecology*, 114:1194–1201, 2007a.

M. E. M. H. Westerhuis, K. G. M. Moons, E. van Beek, S. M. Bijvoet, A. P. Drogtrop, H. P. van Geijn, and J. M. M. van Lith. A randomised clinical trial on cardiotocography plus fetal blood sampling versus cardiotocography plus ST-analysis of the fetal electrocardiogram (STAN) for intrapartum monitoring. *BMC Pregnancy Childbirth*, 7:13, 2007b.

I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2005.

Appendix A

Used software

This section serves as a review of software and packages that were used for analysis, feature selection and classification. The whole program was implemented in MATLAB® version 7.06 Natick, Massachusetts: The MathWorks Inc., 2008. For the nonlinear time series analysis two packages were used. First, the TISEAN package (Hegger et al., 1999) written in C language that offers a broad area of useful tools. This package is specially powerful in combination with graphical software Gnuplot. The second is MATLAB® toolbox, OpenTSTOOL (Merkwirth et al., 2007). The source can be compiled to the mex functions and computationally heavy methods can be run more effectively. Note that the mex functions were compiled on Linux running kernel 2.6.24-23 by gcc version 4.2.4. We used this package in implementation for estimation of time delay, embedding dimension, and correlation dimension. We also used implementation of (Kaplan and Staffin, 1998) and (Goldberger et al., 2002) for approximate and sample entropy, respectively.

Feature subset selection and classification were performed in the data mining software Weka (Witten and Frank, 2005). All graphs and figures of time series were printed using R-project: A language and environment for statistical computing Vienna, Austria.

Appendix B

Correlation matrix of features

Estimated correlations between individual features and class are represented by correlation matrix shown in table B.1. The corresponding p-values were all $p < 0.01$. Hence, we can conclude that correlations are statistically significant. The corresponding methods for the numbers in the fist and row and column in table B.1 are as follows: 1. ApEn(2,0.15), 2. ApEn(2,0.2), 3. SampEn(2,0.15), 4. SampEn(2,0.2), 5. LZC, 6. D_σ, 7. Ds_σ, 8. Dl_σ, 9. D_0, 10. Ds_0, 11. Dl_0, 11. $D_{Higuchi}$, 12. $Ds_{Higuchi}$, 13. $Dl_{Higuchi}$, 14. D_{DFA}, 15. Ds_{DFA}, 16. Dl_{DFA}, 17. D_2, 18. D_2 Takens est., and number 19 represents the class.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1.00	0.97	0.98	0.73	0.71	0.64	0.62	0.83	0.74	0.80	0.81	0.77	0.69	0.72	0.49	0.69	0.65	0.55	0.42
2	1.00	1	0.97	0.97	0.72	0.71	0.63	0.63	0.83	0.73	0.80	0.81	0.75	0.70	0.73	0.48	0.70	0.64	0.55	0.41
3	0.97	0.97	1	1.00	0.69	0.66	0.61	0.55	0.79	0.74	0.74	0.77	0.76	0.63	0.67	0.47	0.63	0.61	0.47	0.40
4	0.98	0.97	1.00	1	0.67	0.66	0.60	0.57	0.78	0.72	0.75	0.77	0.74	0.64	0.68	0.47	0.64	0.59	0.45	0.38
5	0.73	0.72	0.69	0.67	1	0.61	0.74	0.39	0.69	0.87	0.57	0.67	0.87	0.40	0.64	0.67	0.46	0.55	0.49	0.48
6	0.71	0.71	0.66	0.66	0.61	1	0.81	0.86	0.89	0.71	0.84	0.90	0.75	0.78	0.98	0.61	0.88	0.49	0.47	0.37
7	0.64	0.63	0.61	0.60	0.74	0.81	1	0.45	0.67	0.81	0.55	0.67	0.83	0.39	0.82	0.84	0.53	0.48	0.39	0.39
8	0.62	0.63	0.55	0.57	0.39	0.86	0.45	1	0.83	0.44	0.87	0.84	0.48	0.92	0.81	0.27	0.90	0.40	0.45	0.28
9	0.83	0.83	0.79	0.78	0.69	0.89	0.67	0.83	1	0.79	0.97	0.99	0.82	0.87	0.88	0.50	0.86	0.66	0.61	0.42
10	0.74	0.73	0.74	0.72	0.87	0.71	0.81	0.44	0.79	1	0.64	0.78	0.97	0.45	0.74	0.74	0.54	0.62	0.51	0.46
11	0.80	0.80	0.74	0.75	0.57	0.84	0.55	0.87	0.97	0.64	1	0.95	0.67	0.95	0.83	0.37	0.88	0.63	0.60	0.36
12	0.81	0.81	0.77	0.77	0.67	0.90	0.67	0.84	0.99	0.78	0.95	1	0.82	0.88	0.90	0.48	0.88	0.64	0.60	0.42
13	0.77	0.75	0.76	0.74	0.87	0.75	0.83	0.48	0.82	0.97	0.67	0.82	1	0.49	0.78	0.69	0.58	0.64	0.53	0.48
14	0.69	0.70	0.63	0.64	0.40	0.78	0.39	0.92	0.87	0.45	0.95	0.88	0.49	1	0.76	0.23	0.89	0.52	0.55	0.28
15	0.72	0.73	0.67	0.68	0.64	0.98	0.82	0.81	0.88	0.74	0.83	0.90	0.78	0.76	1	0.68	0.88	0.50	0.46	0.34
16	0.49	0.48	0.47	0.47	0.67	0.61	0.84	0.27	0.50	0.74	0.37	0.48	0.69	0.23	0.68	1	0.34	0.35	0.26	0.27
17	0.69	0.70	0.63	0.64	0.46	0.88	0.53	0.90	0.86	0.54	0.88	0.88	0.58	0.89	0.88	0.34	1	0.46	0.45	0.28
18	0.65	0.64	0.61	0.59	0.55	0.49	0.48	0.40	0.66	0.62	0.63	0.64	0.64	0.52	0.50	0.35	0.46	1	0.84	0.34
19	0.55	0.55	0.47	0.45	0.49	0.47	0.39	0.45	0.61	0.51	0.60	0.60	0.53	0.55	0.46	0.26	0.45	0.84	1	0.33
20	0.42	0.41	0.40	0.38	0.48	0.37	0.39	0.28	0.42	0.46	0.36	0.42	0.48	0.28	0.34	0.27	0.28	0.34	0.33	1

Table B.1: Correlation matrix for all features. The strongest relationship between feature and class appears for Lempel Ziv complexity, waveform fractal dimension estimated on short scale by Higuchi's, $Ds_{Higuchi}$, and box counting, Ds_0, method.