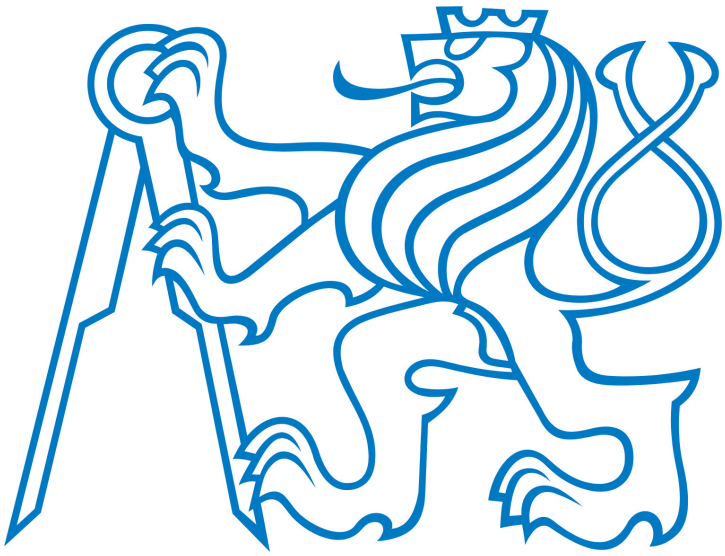
CZECH TECHNICAL UNIVERSITY

IN PRAGUE



DOCTORAL THESIS STATEMENT

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Cybernetics
Gerstner Laboratory

Václav Chudáček

Automatic Analysis of Intrapartum Fetal Heart Rate

Ph.D. Programme: Electrical Engineering and Information Technology
Branch of study: Artificial Intelligence and Biocybernetics

Doctoral thesis statement for obtaining the academic title of "Doctor",

abbreviated to "Ph.D."

Prague, February 2011

The doctoral thesis was produced in full-time manner Ph.D. study at the Department of Cybernetics, Faculty of Electrical Engineering of the Czech Technical University in Prague.

Candidate: Václav Chudáček
Department of Cybernetics
CTU in Prague, Faculty of Electrical Engineering
Technická 2
166 27 Praha 6
email: chudacv@fel.cvut.cz

Supervisor: Doc. Ing. Lenka Lhotská, CSc.
Department of Cybernetics
CTU in Prague, Faculty of Electrical Engineering
Technická 2
166 27 Praha 6
email: lhotska@fel.cvut.cz

Opponents: .....

.....

.....

The doctoral thesis statement was distributed on ..... The defence of the doctoral thesis will be held on ..... at ..... o'clock in room no. ..... at the Faculty of Electrical Engineering, Czech Technical University in Prague.

Those interested may get acquainted with the doctoral thesis concerned at the Dean Office of the Faculty of Electrical Engineering of the CTU in Prague, at the Department for Science and Research, Technická 2, 166 27, Praha 6.

Prof. Ing. Vladimír Mařík, Dr.Sc.
Chairman of the Board for the Defence of the Doctoral Thesis in the branch of study
Artificial Intelligence and Biocybernetics
Czech Technical University in Prague, Fac. of Electrical Engineering
Technická 2
166 27 Praha 6

Contents

1 Problem introduction

Fetal heart activity is the prominent source of information about the fetal well being during delivery. Cardiotocography (CTG) – recording of fetal heart rate (fHR) and uterine contractions (TOCO) enables obstetricians to detect possible ongoing fetal hypoxia, which may occur during delivery even in a previously uncomplicated pregnancy. Figure 1 shows an example of the CTG recording.

Even though fetus has its own natural defense mechanism to tackle the oxygen insufficiency during delivery, in some cases, only timely intervention can prevent adverse consequences [1]. Hypoxia, with prevalence lying in the region of 0.6% [21] to 3.5% [34], is considered to be the third most common cause of newborn death [9] in developed countries.

Instrumental evaluation of the fetal well-being during delivery is more than hundred years old. Auscultation – sensing of the fetal heart rate (fHR) using a fetal stethoscope – introduced by Pinard in 1876 – was replaced in the decades after the second world war by electronic fetal monitoring (EFM) with cardiotocography as the most important representant.

Cardiotocography was introduced in late 1960s and is still the most prevalent method of intrapartum hypoxia detection. It did not, however, bring the expected improvements in the delivery outcomes in comparison to previously used intermittent auscultation [33] and, moreover, continuous CTG is one of the main suspects for increased rate of cesarean sections for objective reasons [33].

To improve the results of cardiotocography, International Federation of Gynecology and Obstetrics (FIGO) introduced general guidelines [13] for CTG assessment. They are based on evaluation of macroscopic morphological fHR features and their relation to the tocographic measurement. Even though the guidelines are available for more than twenty years poor interpretation of CTG still persists [3] with large inter-observer as well as intra-observer assessment variations [6, 4]. Therefore one of this thesis' goals is to contribute to the discussion about the feasibility of the automatic evaluation of the fHR and to the possibility of improving the inter-observer and intra-observer variability.

Recently ST-analysis has been getting much attention as an extension of the classical CTG measurements using additional information from the invasive measurement of the fetal ECG [28]. Although most studies show that ST-analysis is performing slightly better [2], it is important to keep in mind, that the first step to correctly interpret the ST ratio in ST-analysis is to correctly evaluate the CTG itself. Therefore the results of this thesis can be also used to further improve the performance of the ST-analysis.

Attempts to computer evaluation of the CTG are as old as the FIGO guidelines. Features described in the guidelines have become fundamental in most of the clinically oriented systems and automatically extracted morpho-

logical features have been integrated also into automatic systems for CTG analysis [10, 17] mostly for antepartum CTG evaluation.

In majority of the papers only the fHR signal is used. Since it is fHR that contains direct information about fetal state. In this thesis we follow this assumption, also because of inferior quality of the electronically stored TOCO recordings. Extension of this work proposed towards the full CTG (with TOCO included) should be straightforward in the future. To be able to represent the fHR for the computer descriptive parameters of the fHR signal – features – are needed. Different fHR features were investigated in the past, many of them heavily influenced by the research in adult heart rate variability (HRV) analysis.

Morphological features were already mentioned and were used by the founder of the automatic CTG evaluation prof. Bernardes [4] and were further employed also in other works [10, 17]. Statistical description of CTG tracings was employed in the work of [26] and in the following study of [16]. Another approach to fHR analysis examined frequency content by spectral analysis and the paper of Laar [24] gives a short overview of most of the works where fHR spectrum was analyzed. The fHR was also analyzed by wavelets with different properties [30, 29]. Other works analyzed nonlinear properties of fHR such as fractal dimension of reconstructed attractor [7] and waveform fractal dimension [11]. Different estimations of fractal dimension were reviewed in [22]. In general the most successful nonlinear methods, used for examination of all kinds of nonlinear systems including the adult HRV were those based on entropy estimation whether on approximate or sample entropy estimation. They have been documented to be applicable also for fHR analysis [15, 14]. Another method that performed well on the fHR recordings is Lemple-Ziv complexity employed in the paper of [12].

Two remarks are necessary to the above mentioned methods of fHR feature extraction. The new features were usually evaluated by themselves and all of the above mentioned works used umbilical artery pH measurement as an annotation of recorded signals. Although pH of the umbilical artery blood is certainly an objective value, there are many studies showing the relation between the CTG/fHR signal recorded during delivery and outcome of delivery to be rather weak [23, 36].

The main motivation for this work was the persistent feeling of disconnection between the technical papers and clinical practice. There is no unified approach to the automatic evaluation of the fHR. Most of the works deal with particular methods (e.g. methods for feature extraction, for classification) and do not pay any attention to the details. Those details are nevertheless very significant for the clinical usability of the results of any paper.

Therefore, in this thesis, we propose an universal, completely new methodology for automatic evaluation of the fHR recordings starting from fHR recording itself to automatic evaluation of fHR. The evaluation of a signal

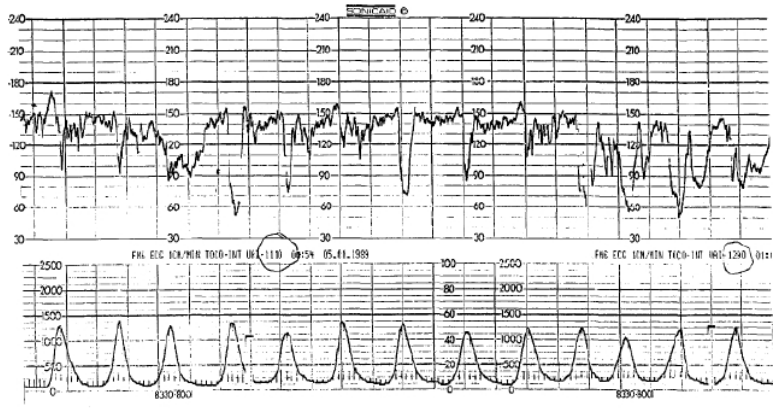


Figure 1: CTG recording. Fetal heart rate – cardiogram – is the upper signal. Tocogram – recording of the uterine pressure – is the lower one. Adapted from [35]

could be used as one of the inputs for the decision support system in the future.

Within the scope of the proposed methodology we mainly focus on the preprocessing steps – often neglected – and their influence on different methods for feature extraction. The assessment of the results against both objective and subjective annotation on one large database is completely novel approach, presented for the first time here for evaluation of intrapartum fHR.

Such an approach enables us to propose novel and useful methodological framework for complex fHR analysis for the future works.

2 Aims of the Doctoral Thesis

The main aim in the background for this work is to bring the results – "that the biomedical engineers are able to generate quite easily" – to the real attention of the clinical practice.

There was no complex methodological approach to the automatic evaluation of the fHR recordings so far. Aim of the majority of published works in this field was to use any method available – usually from the much more researched field of adult heart rate analysis – and apply it on any fHR data. Though the resulting impact on the clinical community is still rather weak. The only clinically used systems were developed by clinicians and use simple macroscopic features introduced in the 1980s (mean frequency, standard deviation, number of accelerations/decelerations).

It seems that it is the inability of the technical research to take into account the surrounding affecters that disqualifies most of the works from

any real clinical use.

Therefore in this work we will develop new robust complex methodology for automatic processing and evaluation of the fHR recordings. Within this thesis we aim to critically review the pros and cons of the various attempts of particular steps of the fHR analysis and put them into one unifying framework. Such an framework – the newly proposed methodology – will enable any future research to be repeatable and therefore possibly clinically usable.

As mentioned in the introduction, one of the main threads of motivation was the persistent feeling of the disconnection between technical and clinical papers on the topic.

Therefore, in this, thesis additionally to the proposal of new methodology for evaluation of the fHR signal during the delivery we evaluate novel (in the clinical domain), although already published (in the technical domain), features for their usability in the clinical environment. The clinical environment is simulated by the annotation of the data set by experts. Together with evaluation of the data using the standard objective pH measurement, we should be able to provide the bridge that could join the results from technical and clinical papers.

2.1 Partial aims

To achieve the main goal – the development of the novel methodology – several partial goals have to be achieved first:

1. Review of the current state of the art
2. New database
3. fHR signal pre-processing
4. Annotation of the signal
5. Feature extraction
6. Feature selection, classification and evaluation

3 Thesis Structure – Methods

The work is structured as follows:

- First, general introduction to the fetal physiology during the labor and concepts related to obstetrics and discussed within this thesis is given.
- Next we provide the reader with review of current methods and approaches for CTG assessment within the clinical and technical settings.

- The data preprocessing, processing and data analysis part of the thesis consists of the following chapters:

  - Data

    Properly assembled database is a keystone on the path to create generally functional evaluation of any medical data. Although this fact is widely accepted, the realization itself is often neglected. One of the reasons might be the fact that the database construction rarely depends on the author of the evaluation algorithms.

    We can state with certainty that the our database is definitely better, in size as well as composition than those used in works of our predecessors, namely [14, 8, 29].

    Final and crucial step in the database creation was the acquisition of an expert evaluation, which necessitated development of the annotation tool. However this step enabled us to provide a totally novel insight into the evaluation of the fHR signal by expert obstetricians.

  - Preprocessing

    The signal pre-processing can influence significantly all further steps in the proposed methodology. Values and further usability of extracted features are especially dependent on the quality of signal preprocessing.

    If the preprocessing is performed carelessly we risk loss of information. On the other hand the preprocessing is necessary, since signal loss and artifacts due to movements of the mother and the baby together with uneven sampling intervals are among the most common distortions of the fHR signal.

    With exception of the artefact removal algorithm, the methods described were never used for evaluation and preprocessing of the fHR recordings. The segment selection algorithm was fully based on the experience with the problematic signal quality just before delivery. The reasoning behind the length where gaps are being interpolated is based on [32] and the methods for resampling of the signal take into account very recent research into the evaluation of re-sampling methods and their impact on the data [25].

    In the end, our preprocessing process consists of four main steps as described further: segment selection, artifacts removal, interpolation and signal detrend.

  - Feature Extraction

    Only features based on fHR signal were evaluated in the context of this work. Only the fHR signal is used in most papers – since the

fHR is considered to be the signal containing direct information about the fetal state. In this thesis we follow this assumption, also because of the inferior quality of the available electronically stored TOCO recordings.

Research of the features for description of the human heart rate was the focal point of electro-cardiological research for decades. The last really new feature for description of the HRV and subsequently fHR was the sample entropy introduced in 2000 by [27].

Even though the features described in this chapter are not new, since the main aim of this thesis was to create a new methodology for automatic fHR processing, we decided to focus on three aspects of feature extraction and evaluation that are indeed novel. Let us mention the main points in here:

* We compute all features under the same conditions and on the same dataset. This enables us to make experiments that are unique in the way that we are able to confront and compare the features directly.
* We compare the features based on their statistical significance and performance using two different approaches to data annotation – where the subjective (expert) annotation, even though it is much closer to the clinical way of evaluation of the signal, was never used for evaluation of the intrapartum fHR before.
* Using both the subjective and objective annotation we are able to arrive at conclusions about the expert evaluation and identify the set of features that shows the greatest promise for real clinical use.

– Feature Evaluation

The extracted descriptive characteristics of the fHR signal – features – represent only the first step in automatic analysis of the fHR signal. Next step is to evaluate, assess the features and distinguish between the useful ones and those that are of no use for any practical deployment.

The classical way how to approach this part of the work would be to do the array: Feature extraction; Feature Selection; Classification. But even though such an approach is used in most of the papers, e.g. [14, 8, 29], we felt that it does not give the correct overall impression of the features. The main reason for this would be most likely the assumption that the better the number in results at the end of the chapter we get, the better results we have achieved. As shown e.g. in great paper of [20] the absolute value of the results is not that important in most of the cases.

In the full chapter we first mention the statistical testing of individual features, followed by correlation evaluation. Further we describe the feature selection algorithms with focus on those that provide ranking of the features according their usefulness for classification. The penultimate part deals with the classification techniques used. The last part of the chapter describes the methods for evaluation of the classification results.

4 Results

4.1 Expert evaluation

To be able to put our results to context of the clinical praxis, we first present the results obtained by the expert clinicians.

Evaluation of the objectively annotated data brings the problem of mapping three class expert evaluation onto the two class objective pH annotation. We decided that since the experts did not have any additional information about the ongoing delivery we will assume that the suspect class is always correct (either pathological, when pH was less then 7.15 or normal otherwise). The results are presented in the Table 1. It was one of the eye openers during the work on this theses - it brought strong doubts on the way the "standard" experiments are carried out.

Evaluation of the subjective GS data is straightforward and bears no need for further explanation. The experts' results are evaluated using their annotation of the data in contrast to Gold standard – derived from their voting.

The table also presents the resulting intra- and inter-observer variabilities. Finally we used kappa statistics to compare expert agreement against an agreement, which might be expected by chance – value of 0.36 corresponds to fair expert agreement. We should mention here that kappa value depends largely on the data used and cannot be used for comparison with performance on different datasets [18].

The results presented in Table 1 are consistent with papers of [4, 6], where, although the name suggests otherwise, the inter-observer variability represents the measure of agreement of the various experts. The results align perfectly also with other works, e.g. [19] and therefore we can say that our experts are very good example of the common clinical decision making.

4.2 Subjective evaluation

We obtained a database of 613 intrapartum fHR recordings with subjective evaluation. Expert annotations of preselected 20 minute long data were collected using the proprietary software. In total 11 cases were excluded

Table 1: Classification results of clinicians when comparing their evaluation of fHR to objective pH annotation. Intra and inter-observer variability and evaluation of the expert agreement.

All in [%]	Expert #1	Expert #2	Expert #3
Sensitivity pH	34.38	48.96	40.63
Specificity pH	14.07	16.30	8.55
Sensitivity GS	71.80	72.45	85.90
Specificity GS	92.72	92.72	67.55
Intra-observer variability	70.83	56.20	76.67
Inter-observer variability		80.61	
Kappa statistics		0.36	

due to total expert disagreement in annotation, hence the database used for feature evaluation consisted of 602 cases.

Results of expert annotation depicting the sensitivity and specificity of each individual and collectively built-up Gold standard, computed using majority voting of three experts, are presented in Table 1. It is necessary to mention that the experts only assessed recordings and no additional information about the patients was provided.

Considering Gold standard annotation as the main one for our work 151 cases were annotated as Normal, 120 as Pathological, and 331 as Suspect. The high prevalence of the pathological class was due to our efforts to collect as much cases with low neonatal pH values (90 records did have pH lower then 7.15). The suspect class resulted from the above mentioned fact that the clinicians were left without any additional clinical information, which is otherwise routinely used in their decision making process – thus any uncertainty usually resulted in the suspect class.

Before proceeding with the statistical testing we have tested correlation in between the features. We have used value of 0.90 as a threshold, above which we considered features correlated enough to include only one of them – the most representative. The intercorrelated groups were as follows, except the last pair, the relationships were not surprising:

- the meanHR correlated with the VLF and A5mean
- LTV; Delta
- ApEn; SampEn; Sevcik
- FD_HigD; FD_HigDs
- PoincareSD2; A5std

Table 2: Classification results for all features and for the subset of features selected based on their statistical significance.

Feature set	All in [%]	NaiveBayes	SVM	C4.5 Tree
All features	Sensitivity	66.7	64.3	62.4
	Specificity	73.8	72.7	69.9
	Precision	67.4	62.1	66.3
	F-measure	71.2	71.4	66.8
	AUC	0.78	0.65	0.68
Selected features	Sensitivity	67.5	68.5	64.3
	Specificity	78.5	84.1	72.7
	Precision	70.0	64.3	68.3
	F-measure	74.3	72.6	71.1
	AUC	0.83	0.72	0.66

Chi-square test was performed prior to statistical testing of individual features. Most of the features were found having not-normal distribution.

Appropriate statistical tests against the expert annotation were used. The results of the tests are presented in Table 3, where out of 55 features only those having significance level $p < 0.01$ are presented.

Finally we have used three different ranking algorithms to rank the significant features from the classification point of view. The features' ranks are presented in the last column of the Table 3, with number of acceleration and deceleration, interval index, as well as Lempel-Ziv complexity and Higuchi's fractal dimension among the top five features.

For comparison we have also tested performance of the features on objective annotation – as depicted in last but one column of the Table 3.

Table 3: Statistical significance of the features when tested against different types of annotation. Only features where level of confidence $p < 0.01$ was fulfilled are presented. Annotations used are: individual experts; Gold standard (GS); significance of features for record where objective annotation was available (pH (Sel.)). The last column represents averaged rank of the features assessed by three different ranking algorithms.

Domain	Features	Statistical significance of features [p-values]					
		Exp #1	Exp #2	Exp #3	GS	pH (Sel.)	Feature rank
Time	baselineSD	–	✓	–	–	–	10
	# Accel.	✓	✓	✓	✓	✓	1
	# Decel.	–	–	✓	✓	✓	2-3
	II	–	✓	–	✓	✓	5
	LTV	–	–	–	–	✓	6
Frequency	VLF	✓	–	–	–	✓	8-9
	MF	–	–	–	–	✓	11
HRV	Poincaré SD2	–	–	✓	–	–	8-9
Wavelet	D2mean	–	✓	✓	✓	✓	7
Nonlinear	ApEn	–	✓	–	✓	✓	16
	LZc	–	–	✓	–	✓	2-3
	FD_BoxDs	–	–	–	✓	✓	13-14
	FD_BoxDl	✓	✓	✓	✓	✓	13-14
	FD_BoxD	–	✓	–	–	–	12
	FD_HigDl	✓	✓	✓	✓	✓	15
	FD_HigD	✓	✓	–	✓	✓	4
	FD_Var	✓	✓	✓	✓	✓	17

5 Methodology summarized

Although the methodology is based on the sum of the results from previous chapters we feel it is important to recapitulate, what is genuine novel contribution to the research field of automatic fHR analysis.

5.1 Motivation

The reason why we proposed new methodology for the fHR processing was the fact that although the electronic fetal monitoring devices were introduced in 1980s, automatic evaluation of the CTG and fHR signal is virtually nonexistent. The reason for that is not lack of research interest. Since the beginning of the electronic era more then 100 papers were published on automatic processing of the CTG itself (in peer-review journals), not to mention the whole adjoining area of HRV research with thousands of papers.

The papers usually deal with different aspects of signal processing, focusing mainly on methods for feature extraction and classification, without any regards towards their usability in real life situations.

One great drawback when analyzing the published papers clearly stands out. The papers have no common point. The database, the preprocessing methods, even the evaluation process is different in each one of the papers. Therefore it is impossible to compare the results presented in them and draw any general conclusion. Clearly the only way to be able to propose any sustainable solution for the clinical practice is to set some standards. This research work tries to achieve exactly that.

5.2 Initial methodology proposal

Our initial aim was to provide framework that would allow researchers and clinicians to compare the results and enable to draw conclusions about usefulness of each newly proposed method for the clinical practice.

To be able to fulfill such a task we had to go through several bottlenecks. First reasonably large database had to be assembled. Second, critical review of the methods used by other researchers in the field had to be performed and evaluation of the usefulness of the features had to be carried out. Based on those steps we started to create the completely new proposal of comprehensive methodology for fHR processing.

5.3 Critical review of methods used

In general the problem with coming up with any new approach to fHR processing lies in the fact that there is no common ground, no publicly available database of the intrapartum fHR data nor agreement on how to preprocess the data. To crown the dismay the objective annotation used in all papers dealing with intrapartum evaluation is very tricky anchor point

– there are many papers suggesting that the correlation between pH value and the fHR signal is only mild.

Therefore any new method that is proposed starts to exist in its own context, description of which is often suppressed in the papers. It is very hard to connect its results to results of others. The methodology proposed within this thesis should be able to provide the researchers with framework, where such an comparison should start to be possible.

The most problematic parts of the fHR signal assessment are summarized further:

- **Preprocessing** Preprocessing of the fHR signal is largely neglected in current papers. Even though the fHR, especially in the intrapartum period, suffers from all kinds of distortions – movement artifacts, loss of signal and uneven sampling rate are among the most prevalent.

  The artefact reduction algorithm by prof. Bernardes is the only documented method in the field of fHR processing. Gap interpolation and signal resampling is usually either not present or done by any ad-hoc method available.

- **Feature extraction** The last really new feature introduced into the heart rate processing field was sample entropy proposed in [27]. From the point of view of creating the methodology the amount of features available is satisfactory and based on the adult HRV research we are able to describe the fHR signal well enough in all domains.

- **Feature evaluation** The problem with the feature evaluation lies in the fact that the features are evaluated separately – usually only features from one domain are evaluated. In many papers clear bias towards the new features is present. Their novelty is preferred to their usefulness.

  Additionally the influence of the methods that are used for the classification are often exaggerated. As discussed thoroughly in [20] the effect of the slightly better classification vehicle is not comparable to the effect of the whole (possibly wrong) task setting.

5.4 Our novel approach

Based on the above presented critical points we came up with several proposals for improvement to the way the fHR signal is processed.

- Preprocessing – We have created a new method for automatic signal segment selection based on the quality of the signal and proximity to the delivery. The decision about the length of a gap, over which it is possible to interpolate was based on [32] as well as in-house experiments. The interpolation of the signal reminds the reader about the

recent paper of [25] and states the possible influence on the extracted features. For the evaluation of the artifacts well established algorithm proposed by prof. Bernardes was used [5].

- Feature extraction – as emerged from the critical review creation of artificial new features would be counterproductive. Therefore in this part of our work we focused on correct implementation of most of the so far published algorithms for feature extraction that were applicable on the signal of 20-minute length. We also researched as many possibilities for each of the feature extraction settings, and selected those that we found out to be the most useful either based on the literature or own experiments.

- Feature evaluation – our approach to this part of the work was completely new when dealing with intrapartum fHR.

  Based on the discovery of the problematic relation between the intrapartum fHR and objective evaluation of measure pH we decided to use both objective and subjective annotation of the data. Automatic evaluation of the intrapartum data subjectively annotated was never done before, but it enabled us to penetrate deeper into the decision making process of the clinicians. That was an important step towards any future working decision support system.

5.5 Methodological conclusion

We can conclude this chapter by stating that after initial research into the topic we have decided to develop a framework for further works, which should enable us to be able to compare the results and evaluate usefulness of individual in the future. The Figure 2 illustrates the steps of the methodology and their interaction. All the steps presented in the Figure 2 are described in detail within the scope of this thesis. From our perspective the development of the methodology was very much needed step, that frees us for the future to explore new features and new methods for classification that would otherwise fall into the same oblivion as the most of the last twenty years of research in the field of automatic fHR analysis.

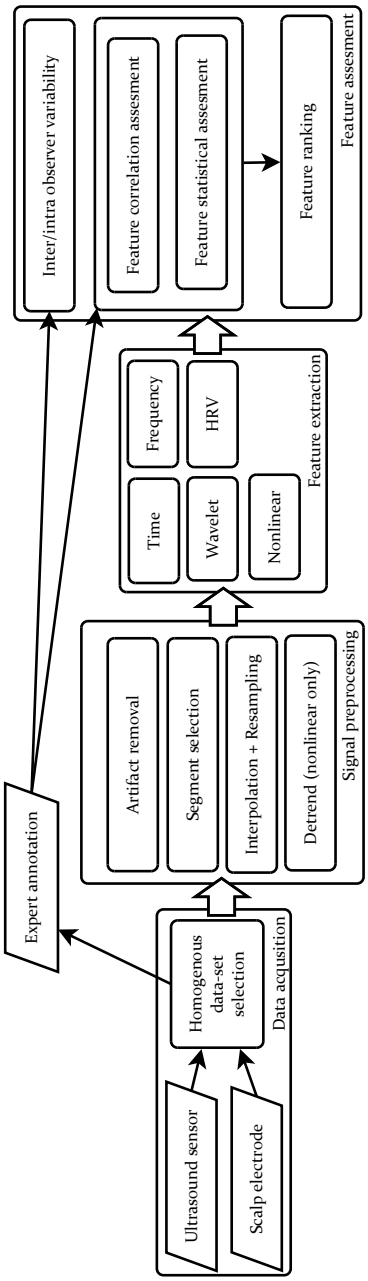


Figure 2: The overall scheme of the presented methodology in the thesis.

6 Conclusion

The main aim of this work was to propose novel methodology that could enable us to compare the results of different methods. Such a complex methodology allows us also to make conclusions about new methods with respect to their usefulness in the clinical setting. We have evaluated all necessary steps from the beginning – the fHR data recording itself – till the end in form of automatic evaluation of the fHR status. We have presented novel approach to the preprocessing of the fHR signal, and discussed widely the limitations of the many features used within this work. Last but not least we have for the first time evaluated the intrapartum fHR against both the expert annotation and the objective annotation.

As a byproduct of the methodology development we found out features that seem to be useful for mimicking the obstetricians behavior when dealing with intrapartum fHR recordings. We based our work on recent research on the extraction of different types of features. We broadened it by comparing directly all the different features on one database using the same preprocessing steps. Additionally, since, based on the clinical experience and as documented in [31], we do not fully agree with the simple and unconditional relation between the pH value and the fHR/CTG recordings, used in majority of the papers, we have used the expert evaluation of the features additionally to the classical, objective pH annotation, approach.

Besides development of the complex methodology for fHR signal processing we have added new conclusions to the findings reported by others – namely:

- We have confirmed that there are other features with information value besides the FIGO guidelines suggested macroscopic features. These features improved performance of the classifiers in both types of annotation.
- We have newly observed that especially the combination of the macroscopic (FIGO) features and non-linear features is worth using.
- We have confirmed that the clinical evaluation of the signals suffers from fairly high inconsistency.
- We can conclude that based on our results the features used by the experts for fHR evaluation suggest high influence of the template matching "intuition" as confirmed by significant non-linear features in expert recording assessment.
- We can confidently conclude based on our results and the close cooperation with obstetricians, that the task of evaluation of the fHR without other clinical data can bring only partial improvements.

- Finally we can conclude that the pH-based classification of the features is insufficient with regards to the simulation of the real-life scenarios. This observation especially is novel – at least from the technical point of view – and should be taken seriously when performing future experiments.

To conclude – in this work we have proposed new methodological framework for evaluation of the intrapartum fHR recordings. Such methodology will be useful in any future evaluation of any method for feature extraction, selection or classification.

Additionally for the first time, according to our knowledge, statistical assessment of the features was performed on large dataset against expert annotation. We warn against ungrounded assumption of automatic large correlation between fHR and umbilical pH. We believe that some relation exists but the type of the relation was never shown in any study, partly due to low numbers of newborns with clearly pathological pH. Expensive follow-up studies would be necessary to link the assumed intrapartum asphyxia and its manifestation into the later stages of newborn's life.

Inter- and intra-observer variability presented in previous papers e.g. [6] was confirmed. Additionally in our case we can report that our experts based their decision on the most-easy-to-assess macroscopic features (number of acceleration, deceleration, variability) and the rest of their decision making seems to be based on the "intuition" – possibly correlated with the nonlinear features in Table 3.

Our main goal for the future is to try to integrate additional knowledge into the system that would take into account the clinical context when evaluating the fHR/CTG recording. An attempt to provide a working practical decision support system will be then the next logical step.

6.1 Mapping achievements to goals

We can confidently say that the ultimate goal of this thesis was fulfilled. We have proposed completely new methodology and evaluated it using objective and expert annotation. We believe that the methodology can be used as a solid stepping stone on the way to future endeavors such as creation of fHR/CTG-based decision system.

1. *Review of the current state of the art*

   Comprehensive review of the state of the art of CTG/fHR processing and evaluation was carried out showing two very distinct approaches depending on the field of the researcher.

   The main observation from the technical part of the review: most of the published papers applied known techniques onto new type of signal.

They did it usually only with small data-set (about 50 recordings) and with no attempt on result interpretation.

2. *New database*

   The initial database was cleaned and homogenized into the form of the final data-set containing 613 patients out of which 218 had their umbilical artery pH included. 95 cases from this data-set had pH of less then 7.15, in scope of this thesis considered pathological. This amount of pathological samples was unequivocally big number when compared to other works. The detailed description of the final data-set used within this thesis was presented in Chapter **??**.

3. *fHR signal pre-processing*

   We especially focused on the limitations of each of the preprocessing methods – to ensure that all the possible adverse effects were averted. We have developed new method for segment selection and adopted general approaches of [32] to the fHR environment to decide which gaps can be interpolated over. Well established algorithm was used for artefact removal.

4. *Annotation of the signal*

   To be able to record the annotation we have created an annotation software for online collection of the data. This software is envisioned to be used in further projects exceeding the scope of the thesis.

5. *Feature extraction*

   Based on the thorough study of the literature – feature extraction methods for extraction of more then 50 features were implemented in Matlab. We have evaluated different settings for the features and have chosen either those most recommended in the literature or those performing best in our initial tests.

6. *Feature selection, classification and evaluation*

   The goal of this thesis was not to find the best results possible in terms of sensitivity and specificity. We were well aware of the limitations of the data-set and the whole experimental setting, so often overlooked in other works. Therefore we have focused fully on finding the best methodological approach. We have employed different evaluation techniques that enabled us on one hand comparison with work of others, on other hand we acquired novel views on the data that enabled us to create new and for the further research interesting conclusions.

7. *Methodology*

We have summarized in detail the proposed methodology while uncovering the motivations for each particular step. We sincerely believe that by using the proposed methodology, the field of automatic fHR processing can move forward more quickly and clinicians should be able to benefit from its results.

6.2 Future challenges

There are several challenges that the investigators will have to overcome in order to achieve any reasonably working automatic decision support system. Some of the envisioned goals for the future are summarized bellow:

- It would be well worth to extend the experiments presented in this work. Larger database and more experts for evaluation should be probably one of the first steps. Direct comparison of the results and evaluation of the generality of our approach should be the very next step.
- Lowering the variability of the clinician's decision by means of automatic evaluation of the selected features and their presentation to the doctor is among the goals that should be achievable in short term.
- Using additional information about the actual status of the delivery within the evaluation process is a necessary step to really get the clinicians on board.
- Creating case-based reasoning system to further homogenize the decision making process within the hospital and to learn the invisible paths of the decision making process.
- Concentration of all of the above mentioned processes into one decision support system that will be able to evaluate in real time the actual fHR in the context of available clinical information should be probably for now considered as a holy grail of the automatic fHR/CTG analysis field.

References

[1] Z. Alfirevic, D. Devane, and G. M. L. Gyte. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database Syst Rev*, 3:CD006066, 2006. 1

[2] I. Amer-Wåhlin and K. Maršál. ST analysis of fetal electrocardiography in labor. *Seminars in Fetal and Neonatal Medicine*, 16(1):29–35, 2011. 1

[3] J. Bernardes and D. Ayres-De-Campos. The persistent challenge of foetal heart rate monitoring. *Current Opinion in Obstetrics and Gynecology*, 22(2):104–109, 2010. 1

[4] 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. 1, 4.1

[5] J. Bernardes, D. A. de Campos, A. Costa-Pereira, L. Pereira-Leite, and A. Garrido. Objective computerized fetal heart rate analysis. *Int J Gynaecol Obstet*, 62(2):141–147, Aug 1998. 5.4

[6] E. Blix, O. Sviggum, K. S. Koss, and P. Oian. Inter-observer variation in assessment of 845 labour admission tests: comparison between midwives and obstetricians in the clinical setting and two experts. *BJOG*, 110(1):1–5, Jan 2003. 1, 4.1, 6

[7] 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. 1

[8] T. K. Chung, M. P. Mohajer, Z. J. Yang, A. M. Chang, and D. S. Sahota. The prediction of fetal acidosis at birth by computerised analysis of intrapartum cardiotocography. *Br J Obstet Gynaecol*, 102(6):454–460, Jun 1995. 3

[9] E. d'Aloja, M. Müller, F. Paribello, R. Demontis, and A. Faa. Neonatal asphyxia and forensic medicine. *J Matern Fetal Neonatal Med*, 22 Suppl 3:54–56, 2009. 1

[10] D. A. de Campos, P. Sousa, A. Costa, and J. Bernardes. Omniview-SisPorto 3.5 - A central fetal monitoring station with online alerts based on computerized cardiotocogram+st event analysis. *Journal of Perinatal Medicine*, 36(3):260–264, 2008. 1

[11] 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. 1

[12] M. Ferrario, M. Signorini, and G. Magenes. Complexity analysis of the fetal heart rate variability: Early identification of severe intrauterine growth-restricted fetuses. *Medical and Biological Engineering and Computing*, 47(9):911–919, 2009. 1

[13] FIGO. Guidelines for the Use of Fetal Monitoring. *International Journal of Gynecology & Obstetrics*, 25:159–167, 1986. 1

[14] 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 support vector machines. *IEEE Trans Biomed Eng*, 53(5):875–884, May 2006. 1, 3

[15] H. Gonçalves, J. Bernardes, A. P. Rocha, and D. A. de Campos. Linear and nonlinear analysis of heart rate patterns associated with fetal behavioral states in the antepartum period. *Early Hum Dev*, 83(9):585–591, Sep 2007. 1

[16] H. Gonçalves, A. P. Rocha, D. A. de Campos, and J. Bernardes. Linear and nonlinear fetal heart rate analysis of normal and acidemic fetuses in the minutes preceding delivery. *Med Biol Eng Comput*, 44(10):847–855, Oct 2006. 1

[17] B. Guijarro-Berdinas and A. Alonso-Betanzos. Empirical evaluation of a hybrid intelligent monitoring system using different measures of effectiveness. *Artif Intell Med*, 24(1):71–96, Jan 2002. 1

[18] K. L. Gwet. *Handbook of Inter-Rater Reliability*. Advanced Analytics, LLC, 2010. 4.1

[19] L. A. Haggerty. Continuous electronic fetal monitoring: contradictions between practice and research. *J Obstet Gynecol Neonatal Nurs*, 28(4):409–416, 1999. 4.1

[20] D. J. Hand. Classifier Technology and the Illusion of Progress. *Statistical Science*, 21, No. 1:1–15, Jun 2006. 3, 5.3

[21] E. Heintz, T.-H. Brodtkorb, N. Nelson, and L.-A. Levin. The long-term cost-effectiveness of fetal monitoring during labour: a comparison of cardiotocography complemented with st analysis versus cardiotocography alone. *BJOG*, 115(13):1676–1687, Dec 2008. 1

[22] P. Hopkins, N. Outram, N. Lofgren, E. C. Ifeachor, and K. G. Rosen. A Comparative Study of Fetal Heart Rate Variability Analysis Techniques. In *Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS '06*, pages 1784–1787, 2006. 1

[23] J. Jezewski, T. Kupka, and K. Horoba. Extraction of fetal heart-rate signal as the time event series from evenly sampled data acquired using doppler ultrasound technique. *IEEE Trans Biomed Eng*, 55(2 Pt 1):805–810, Feb 2008. 1

[24] J. 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. 1

[25] C. Liu, C. Liu, P. Shao, L. Li, X. Sun, X. Wang, and F. Liu. Comparison of different threshold values r for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups. *Physiological Measurement*, 32(2):167, 2011. 3, 5.4

[26] G. Magenes, M. G. Signorini, and D. Arduini. Classification of cardiotocographic records by neural networks. In *Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN 2000*, volume 3, pages 637–641 vol.3, 2000. 1

[27] 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. 3, 5.3

[28] K. G. Rosén. Fetal electrocardiogram waveform analysis in labour. *Curr Opin Obstet Gynecol*, 17(2):147–150, Apr 2005. 1

[29] E. Salamalekis, E. Hintipas, I. Salloum, G. Vasios, C. Loghis, N. Vitoratos, C. Chrelias, and G. Creatsas. Computerized analysis of fetal heart rate variability using the matching pursuit technique as an indicator of fetal hypoxia during labor. *J Matern Fetal Neonatal Med*, 19(3):165–169, Mar 2006. 1, 3

[30] 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. *BJOG*, 109(10):1137–1142, Oct 2002. 1

[31] S. Schiermeier, S. P. von Steinburg, A. Thieme, J. Reinhard, M. Daumer, M. Scholz, W. Hatzmann, and K. T. M. Schneider. Sensitivity and specificity of intrapartum computerised FIGO criteria for cardiotocography and fetal scalp ph during labour: multicentre, observational study. *BJOG*, 115(12):1557–1563, Nov 2008. 6

[32] J. C. Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, 2003. 3, 5.4, 3

[33] P. J. Steer. Has electronic fetal heart rate monitoring made a difference. *Semin Fetal Neonatal Med*, 13(1):2–7, Feb 2008. 1

[34] B. K. Strachan, W. J. van Wijngaarden, D. Sahota, A. Chang, and D. K. James. Cardiotocography only versus cardiotocography plus PR-interval analysis in intrapartum surveillance: a randomised, multicentre trial. fecg study group. *Lancet*, 355(9202):456–459, Feb 2000. 1

[35] T. Vanner and J. Gardosi. Intrapartum assessment of uterine activity. *Baillieres Clin Obstet Gynaecol*, 10(2):243–257, Jun 1996. 1

[36] P. Warrick, E. Hamilton, D. Precup, and R. Kearney. Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography. *IEEE Transactions on Biomedical Engineering*, 57(4):771–779, 2010. 1

List of Candidate's Works Related to the Doctoral Thesis

This thesis is based also on the following previously published material:

6.3 Journal papers

- Chudáček, V. - Spilka, J. - Janků, P. - Koucký, M. - Lhotská, L. - Huptych, M. Evaluation of intrapartum fetal heart rate: An analysis of useful features *Physiological Measurement* (submitted)
- Spilka, J. - Chudáček, V. - Janků, P. - Koucký, M. - Lhotská, L. - Huptych, M. - et al. Using nonlinear features for fetal heart rate classification *Biomedical Signal Processing and Control* (submitted)

6.4 Selected conference papers

- Chudáček, V. - Spilka, J. - Huptych, M. - Lhotská, L. Comparison of Linear and Non-linear Features for Intrapartum Cardiotocography Evaluation; Clinical Usability vs. Contribution to Classification In: *Proceedings of Biosignal 2010: Analysis of Biomedical Signals and Images* [CD-ROM]. Brno: Brno University of Technology, 2010, p. 369-372. Contribution: 50%
- Chudáček, V. - Spilka, J. - Huptych, M. - Lhotská, L. Automatic Classification of Intrapartal Fetal Heart Rate Recordings; can it compete with experts? In: *Information Technology in Bio-and Medical Informatics*. Berlin: Springer, 2010, p. 57-59. ISBN 978-3-642-15019-7. Contribution: 50%
- Chudáček, V. - Spilka, J. - Huptych, M. - Lhotská, L. Linear and Non-Linear Features for Intrapartum Cardiotocography Evaluation In: *Computing in Cardiology 2010* Preprints [CD-ROM]. New Jersey: IEEE, 2010, Contribution: 50%
- Spilka, J. - Chudáček, V. - Kužílek, J. - Lhotská, L. - Hanuliak, M. Detection of Inferior Myocardial Infarction: A Comparison of Various Decision Systems and Learning Algorithms In: *Computing in Cardiology 2010* Preprints [CD-ROM]. New Jersey: IEEE, 2010, Contribution: 25%

- Novák, D. - Křemen, V. - Cuesta-Frau, D. - Chudáček, V. - Lhotská, L. Discrimination of Endocardial Electrogram Disorganization Using a Signal Regularity Analysis In: *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society* [CD-ROM]. Piscataway: IEEE, 2009, p. 1812-1815. ISBN 978-1-4244-3296-7. Contribution: 5%

- Chudáček, V. - Huptych, M. - Koucký, M. - Bauer, L. - Spilka, J. - et al. Fetal Heart Rate Data Pre-Processing and Annotation In: *Proceedings of 9th International Conference on Information Technology and Applications in Biomedicine* [CD-ROM]. Piscataway: IEEE, 2009, ISBN 978-1-4244-5378-8. Contribution: 35%

6.5 Other: Related projects and other journal papers

- Grant IGA 2010-2015 - **Cardiotocography evaluation by means of artificial intelligence** – project based on the research described in this thesis.

- Eniac-MAS project 2010-2013 - One part of the project: **mCTG - Mobile CardiTocoGraphy** – telemedicine application for risk pregnancies at home.

- Chudáček, V. - Georgoulas, G. - Lhotská, L. - Stylios, C. - Petrík, M. - et al. Examining Cross-Database Global Training to Evaluate Five Different Methods for Ventricular Beat Classification In: *Physiological Measurement* 2009, vol. 30, no. 7, p. 661-677. ISSN 0967-3334. Contribution: 65%

- Lhotská, L. - Chudáček, V. - Huptych, M. ECG Processing In: *Data Mining and Medical Knowledge Management: Cases and Applications.* Hershey: IGI Publishing, 2009, p. 137-160. ISBN 978-1-60566-218-3. Contribution: 33%

- Kos, P. - Varga, F. - Handl, M. - Kautzner, J. - Chudáček, V. - et al. Correlation of dynamic impact testing, histopathology and visual macroscopic assessment in human osteoarthritic cartilage In: *International Orthopaedics* 2011, vol. 35, no. 1, p. 1-7. ISSN 0341-2695. Contribution: 10%

Summary

Cardiotocography (CTG) – monitoring of fetal heart rate (fHR) and uterine contractions – is routinely used by obstetricians to detect fetal hypoxia. Recent clinical fHR evaluation is based solely on macroscopic morphological features even though automatic adult heart rate analysis research field

is flourishing. One of the causes may be lack of unified approach for the automatic evaluation of the fHR. Most of the works deal with particular methods (e.g. separate methods for feature extraction, for classification) and omit crucial details. Therefore within this thesis we propose completely new methodology for automatic evaluation of the fHR recordings. It covers all the steps between the fHR data recording itself and automatic evaluation of the fHR status. As a byproduct of the developed methodology we have found features useful for mimicking the obstetricians behavior, when dealing with intrapartum fHR recordings. We have used expert evaluation additionally to the objective pH annotation. This approach originated from our disagreement with the assumption of simple and unconditional relation between the pH value and the fHR/CTG recordings, used in majority of the papers.

This thesis provides the reader with description of novel comprehensive methodology for automatic fHR analysis. Using the results obtained on the reasonably large database we were able to make novel conclusions about usefulness of different types of features and provide new insights into the expert decision making process. For the future one of the most challenging research goals remains – to try to integrate the automatically evaluated fHR/CTG into a system that would take clinical context into account. Nevertheless we believe that our proposed methodology will be useful as an stepping stone for such a future researcher.

Shrnutí

Hodnocení srdeční frekvence plodu (fHR) je v klinickém prostředí založeno na hodnocení makroskopických morfologických znaků. Nové přístupy z oblasti výzkumu variability srdeční frekvence se v porodnici zatím neprosadily. Důvodem může být fakt, že neexistuje standardizovaný přístup k automatickému vyhodnocení fHR. Většina prací se zabývá jednotlivými metodami – např. samostatné metody pro extrakci, nebo pro klasifikaci příznaků – a nevěnují dostatečnou pozornost detailům jejich použití. V této práci navrhujeme zcela novou metodiku pro automatické vyhodnocení záznamů intrapartální fHR. Navrhovaná metodika se vztahuje na všechny kroky zpracování fHR – vlastním nahráním signálu počínaje a automatickým vyhodnocení stavu plodu konče. Vedlejším produktem návrhu metodiky je objevení příznaků, které jsou schopny reprezentovat způsob rozhodování porodníka při hodnocení intrapartálního fHR. Mimo klasického způsobu hodnocení záznamu – pomocí objektivní hodnoty pH – jsme v práci poprvé použili i subjektivní hodnocení intrapartálního fHR na základě expertního hodnocení. Motivací pro tento nový přístup byl náš nesouhlas s předpokladem jednoduchého a bezpodmínečného vztah mezi pH a fHR/CTG záznamem, který je používán ve většině publikovaných prací.

Tato práce poskytuje čtenáři popis nové komplexní metodiky pro automatickou analýzu fHR. Výsledky dosažené na dostatečně velké databázi nám umožnily přijít s novými závěry o užitečnosti různých typů příznaků a poskytnout nový náhled na rozhodovací proces expertů v oboru. Do budoucna zůstává aktuální jeden z nejnáročnějších cílů výzkumu v této oblasti – pokus o integraci automaticky vyhodnocených záznamů fHR/CTG do systému pracujícímu v rámci klinického kontextu. Domníváme se, že naše navrhovaná metodika umožní dosáhnout tohoto cíle rychleji.