Linear and Non-Linear Features for Intrapartum Cardiotocography Evaluation

Václav Chudáček[1], Jiří Spilka[1], Michal Huptych[1], George Georgoulas[2], Lenka Lhotská[1], Chrysostomos Stylios[2], Michal Koucký[3], Petr Janků[4]

[1]Department of Cybernetics, Czech Technical University, Prague, Czech Republic
[2]Laboratory of Knowledge and Intelligent Computing, TEI of Epirus, Artas, Greece
[3]Obstetrics and Gynaecology clinic, Medical Faculty of Charles University, Prague, Czech Republic
[4]Obstetrics and Gynaecology clinic, University Hospital in Brno, Czech Republic

Abstract

Since 1960's obstetricians have been using cardiotocography (CTG) to detect possible ongoing hypoxia of the fetus. CTG consists of fetal heart rate (fHR) and uterine contraction (TOCO) monitoring. The evaluation of the fHR in clinical settings is ruled by FIGO guidelines, which are based on evaluation of macroscopic morphological features derived from the fHR, such as baseline variability. Although upgrades were proposed to the guidelines – none of them is taking into account results achieved by the adult heart rate variability research.

In this work, almost complete set of features previously used for fHR description is investigated and the features are assessed based on their statistical significance in the task of distinguishing the records into three FIGO classes. Inter-correlation of the features is also discussed. We assess the features on a large data set and use expert signal evaluation instead of pH values with the aim to give an overall view of the potential usefulness of the features in the clinical settings.

We conclude the paper by presenting the best uncorrelated feature subset according to the meta-analysis of three different ranking methods.

1. Introduction

Evaluation of the fetal status during delivery is necessary to enable discovery of possible ongoing fetal hypoxia which might occur even in a previously uncomplicated pregnancy. Hypoxia, severe oxygen deprivation of the fetus, is considered to be the third most common cause of newborn death. Cardiotocography as a diagnostic method was introduced in late 1960s and consists of continuous recording of fetal heart rate (fHR) and pressure of uterine contractions (TOCO). It has not fully delivered the expected improvements in the delivery outcomes in comparison to previously used intermittent auscultation [1] and, moreover, continuous CTG is the main suspect for increased rate of cesarean sections for objective reasons [1].

To improve the interpretation of CTG, and thus to improve the overall – unsatisfactory so far – results, guidelines were introduced [2] based on evaluation of macroscopic morphological fHR features and their relation to the TOCO measurement. Even though the guidelines are available for more than twenty years poor interpretation of CTG still persists with large inter-observer as well as intra-observer assessment variations [3, 4].

In many papers only the fHR signal from the whole CTG recording is used since fHR is the signal containing direct information about the fetal state. Our paper follows this assumption, also because of the inferior quality of the available electronically stored TOCO recordings.

For fHR description different features were investigated in the past, many of them heavily influenced by the research in adult heart rate variability (HRV) analysis. Statistical description of CTG tracings was employed e.g. in the work of Magenes [5]. Another approach to fHR analysis examined frequency content by spectral analysis and Laar [6] gives a short overview of most works where fHR spectrum was analyzed. The fHR was also analyzed by various wavelets with different properties [7, 8]. Other works analyzed nonlinear properties of fHR such as fractal dimension of reconstructed attractor [9] and waveform fractal dimension [10].

2. Data description

Data for this work were obtained at the Dept. of Obstetrics and Gynaecology of Charles University hospital in Prague from 2007 to 2009. The fHR signals were measured on a Neoventa's STAN S21 system using external as well as internal scalp electrodes.

All recordings were checked for patient anamnesis and only one fold pregnancies delivered during 38th – 42nd week of pregnancy were chosen for the final database, which consisted of 613 delivery recordings altogether.

For the evaluation expert annotation was used. It has its drawbacks – it is much more subjective, and suffers from inter- and intra-observer variations. But it gives better insight into the real clinical decision making than the post-delivery scores based on pH or overall newborn behavior, as described by Apgar score.

3. Signal preprocessing

Values of extracted features and their further usability are highly dependent on the quality of signal preprocessing. Our preprocessing process consisted of four main steps: segment selection, artifacts removal, interpolation and signal detrend.

Segments were selected from the complete recordings, some of them up to 12 hours long, as close as possible to the actual delivery. Signal quality was evaluated in relation to the segment position and the segment with the best score was selected. When available information allowed, we tried to set the end of the segment onto the beginning of the second stage of labor, where the quality of signal sharply decreases. Segments were maximally 24 minutes long and due to further preprocessing (gap interpolation, noisy segments removal) were truncated to 20 minutes long segments – 4800 samples when using 4 Hz sampling frequency.

The algorithm proposed in [11] was utilized for artifact removal and cubic hermite spline interpolation [12] was employed to interpolate over the gaps.

4. Features

Features used for purposes of this paper are almost complete collection of features used for evaluation of intrapartal/antepartal fHR in recently published papers and are presented in following groups. Their exact description can be found in the references mentioned in the introduction section.

4.1. Morphological features

Morphological features proposed in the FIGO guidelines are the features used in the obstetricians wards. These features describe the macroscopic – "visible" – properties of the fHR. A well known algorithm for fHR extraction described in [11] was used in this study. The features extracted were: Mean of the fHR baseline without influence of accelerations and decelerations; Number of accelerations – transient increase in heart rate above the baseline by 15 bpm or more, lasting 15 seconds or more; Number of decelerations, where deceleration is defined as the transient episode of slowing fetal heart rate below the baseline level by more than 15 bpm and lasting 10 seconds or more. Decelerations can be distinguished further according to their length [13]; Percentage of time occupied by prolonged decelerations.

4.2. Time and frequency domain features

Two types of time domain features were computed. First type deals with macroscopic but time demanding features. Second type assesses more subtle changes in fHR behavior, that are impossible to spot by naked eye. Median of the fHR baseline; Standard deviation of the fHR baseline, describing variations of the fHR; Long term irregularity (LTI); Short term variability assesses the variability of the fHR in smaller segments; Interval index; Delta and Delta total values

Various spectral methods have been used for the analysis of adult heart rate [14]. In case of fHR analysis no standardized use of frequency bands exists. Therefore we used two slightly different partitioning to three and four frequency bands and its ratios.

4.3. HRV based statistical features

Fetuses suffering from any possible heart condition were excluded from the database, therefore all beats were considered as normal (N) – thus the distance between two beats is depicted as NN. Based on commonly used features in adult HRV we computed several statistical measures [14]: Standard deviation of the NN intervals (SDNN); root of the mean squared differences (RMSSD) of successive NN intervals; $NN50$ and $pNN50$; Lengths of axis in Poincaré plot.

4.4. Wavelet features

We decomposed the signal into five levels of decomposition using the Malat algorithm with Daubechies order 4 (db4) mother wavelet. Based on the decomposition of the signal we computed the mean and standard deviation, and other statistical parameters in all details and the last – 5^{th} approximation.

4.5. Nonlinear features

Almost all methods used for fHR analysis have their roots in adult HRV research. We have computed: Fractal dimension; Correlation dimension using different algorithms; Approximate Entropy (ApEn); Sample Entropy (SampEn); Lempel Ziv Complexity (LZC).

5. Feature evaluation

For testing of statistical significance of the features for distinguishing between the three classes, Anova test was used for normally distributed features and for the rest Kruskal-Wallis test was used where normality of distribution is not required.

We have used three different feature selection techniques that enabled us to rank the features based on their performance in the potential classification process using 10-fold cross-validation. Based on our previous experience we have used following techniques – each one based on slightly different principle – that are described in larger detail in e.g. [15]:

- **Information Gain Evaluation** (InfoGain) evaluates attributes by measuring their information gain with respect to the class.
- **One Rule Evaluation** uses the simple minimum-error measure adopted by the One Rule classifier.
- **SVM Feature Evaluation** evaluates attributes using recursive feature elimination with a linear support vector machine. Attributes are selected one by one based on the size of their coefficients.

6. Results

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, the most representative. The inter-correlated groups were as follows: the meanHR correlated with the VLF and A5mean; LTV with Delta; ApEn with SampEn and Sevcik; FD_HigD with FD_HigDs and finally PoincareSD2 correlated with A5std. Except the last pair, the relationships were not surprising.

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 1, 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 1, with number of acceleration and deceleration, interval index, as well as Lempl-Ziv complexity and Higuchi's fractal dimension among the top five features.

7. Discussion and conclusion

The work presented in this paper is novel in the way we perceive the problem area of fHR evaluation. Many of the building pieces of our work were used by others before [16, 17] but our contribution is distinct.

We decided to examine set of features primarily against expert evaluation. This approach enabled us to examine the features from the point of view of clinical experts that are unaware of the final outcome when assessing the ongoing fHR during delivery. More importantly, they should act against adverse outcome of the delivery when pathological CTG occurs. Thus fHR might clearly be pathological (by expert judgement) but the final outcome after e.g. caesarean section can be normal (by pH assessment).

When discussing the feature average ranking as presented in the last column of the Table 1 we can see that from the point of view of automatic serial assessment of the features, the classical – and very distinctive ones – such as number of acceleration and deceleration and LTV are ranked in the top half. The fact that most of the non-linear features are ranked to the bottom half can be justified by their correlation, where the additional features after using LZC and FD_HigD do not contribute significantly to improvement of the final score. The inter-correlation of the nonlinear features that are presented in the Table 1 was in the range of 0.53-0.86 – therefore it did not fulfilled our condition for "high" correlation (above 0.90) but the effect is pronounced in the ranking method results.

To conclude – for the first time statistical assessment of the features was performed on large dataset against expert annotation.

Goal for the future work is to try to verify our findings using different data sets. We will also try to integrate additional knowledge into the system that would take into account the clinical context of the test in an attempt to provide a practical decision support system.

Acknowledgements

This work was supported by the research programs No. NT11124-6/2010 of the Ministry of Health Care, MSM 6840770012 of the Ministry of Education of the Czech Republic and by the CVUT Grant SGS10/279/OHK3/3T/13.

References

[1] Steer PJ. Has electronic fetal heart rate monitoring made a difference. Semin Fetal Neonatal Med Feb 2008;13(1):2–7.

[2] FIGO. Guidelines for the use of fetal monitoring. International Journal of Gynecology Obstetrics 1986;25:159–167.

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

[4] Bernardes J, Costa-Pereira A, de Campos DA, van Geijn HP, Pereira-Leite L. Evaluation of interobserver agreement of cardiotocograms. Int J Gynaecol Obstet Apr 1997; 57(1):33–37.

Table 1. Statistical significance of the features when tested against different types of annotation – individual experts; Gold standard (GS); objective annotation when available (pH (Sel.)). The last column represents averaged rank of the features.

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] Magenes G, Signorini MG, Arduini D. Classification of cardiotocographic records by neural networks. In Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN 2000, volume 3. 2000; 637–641 vol.3.

[6] Laar JOEHV, Porath MM, Peters CHL, Oei SG. Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature. Acta Obstet Gynecol Scand 2008; 87(3):300–306.

[7] Salamalekis E, Thomopoulos P, Giannaris D, Salloum I, Vasios G, Prentza A, Koutsouris D. Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks. BJOG Oct 2002; 109(10):1137–1142.

[8] Salamalekis E, Hintipas E, Salloum I, Vasios G, Loghis C, Vitoratos N, Chrelias C, Creatsas G. 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 Mar 2006;19(3):165–169.

[9] Chaffin DG, Goldberg CC, Reed KL. The dimension of chaos in the fetal heart rate. Am J Obstet Gynecol Nov 1991;165(5 Pt 1):1425–1429.

[10] Felgueiras CS, de S JP, Bernardes J, Gama S. Classification of foetal heart rate sequences based on fractal features. Med Biol Eng Comput Mar 1998;36(2):197–201.

[11] Bernardes J, Moura C, de Sa JP, Leite LP. The Porto system for automated cardiotocographic signal analysis. J Perinat Med 1991;19(1-2):61–65.

[12] Schumaker LL. Spline Functions: Basic Theory. Cambridge University Press, 2007.

[13] de Campos DA, Bernardes J. Early, variable and late decelerations: can a consensus be reached in their identification? Int J Gynaecol Obstet Jun 1999;65(3):305–306.

[14] 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 Mar 1996;17(3):354–381.

[15] Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, 2005.

[16] Georgoulas G, Stylios CD, Groumpos PP. Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines. IEEE Trans Biomed Eng May 2006;53(5):875–884.

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

Address for correspondence:

Václav Chudáček

Department of Cybernetics, Czech Technical University in Prague, Technická 2, Praha 6, 166 27, Czech Republic

chudacv@fel.cvut.cz