Fetal Heart Rate Complexity Measures to Detect Hypoxia

Óscar Barquero-Pérez, Rebeca Goya-Esteban, Antonio Caamaño, Carlos Martín-Caballero and José Luis Rojo-Álvarez

Abstract

Background Perinatal hypoxia is a severe condition that may harm fetus organs permanently or even cause dead. When the fetus brain is partially deprived from oxygen, the control of the fetal heart rate (FHR) is affected.

Objective. We hypothesize that the complex physiological mechanisms of the FHR are perturbed under perinatal hypoxia. We propose measure entropy and time irreversibility of the FHR to quantify the loss in the complexity.

Materials and Methods. We estimated the complexity of the FHR signal using Sample Entropy (SampEn), Permutation Entropy (PE), and Time Irreversibility (TI). We compared the results with time (Short Time Variability, STV) and frequency domain (High Frequency Power, PHF) methods. We computed every one hour before delivery. FHR traces were preprocessed to remove artifacts. A database of 32 FHR recordings were acquired with cardiotochography, 15 controls and 16 cases. A case was declared whether: 1) the PH of the umbilical artery was $\leq$ 7.05; or 2) the APGAR score 5 minutes after delivery was $\leq$ 7 and a reanimation type III or greater was required. Resampling methods were used to establish the statistical differences.

Results. TI was significantly different for healthy and hypoxia fetuses (-0.38 +- 0.19 vs. -0.21+-0.37, p-value=0.063). Entropy indices were higher for healthy fetuses (SampEn:0.33+-0.12 vs 0.28 +- 0.09, p-value=0.11; PE:0.72+-0.04 vs 0.69+-0.07, p-value= 0.12). STV (3.23+-1.15 vs 3.45+-1.35, p-value=0.30) and PHF (0.40+-0.18 vs 0.43+-0.25, p-value=0.31) indices showed no differences.

Conclusions. Complexity measures of the FHR were different for healthy and hypoxia fetuses. These indices may help to early detect hypoxia with less invasive methods.

1. Introduction

Perinatal hypoxia is a fetus and newborn disease due to the lack of tissues oxygenation. Although it can occur in earlier gestation phases, childbirth and immediate neonatal hours are the fundamental risk periods.

Perinatal hypoxia severity spectrum conveys very mild cases (only requiring neonatal resuscitation with environmental oxygen), more serious cases needing intubation and acidosis correction with bicarbonate (reanimation types V and VI), and critical cases that can cause perinatal death or serious sequels, such as brain or adrenal hemorrhage, necrotizing enterocolitis, delayed neurological development, mental handicap, seizures (West syndrome) or cerebral palsy [**?**, **?**]. Diagnosis is performed at the birth time by evaluating the cardio-respiratory depression and the muscle tone. The severity of the hypoxia is commonly quantified using the Apgar Score [**?**, **?**], with a score lower than 7 at five minutes after delivery being considered as pathological, which is usually confirmed with gas analysis of the umbilical cord, whereas low pH values evidence metabolic acidosis. Typical values considered for diagnosis are pH $\leq$ 7.05 or pH $\leq$ 7, and intrapartum pH values $\leq$ 7.20 are considered pathological in terms of risk of perinatal hypoxia.

Continuous electronic fetal monitoring, also known as Cardiotocography (CTG), was developed around 1960 [**?**, **?**] and consists of the simultaneous evaluation of the Fetal Heart Rate (FHR) and the uterine activity. After CTG generalization, two relevant signs of suspicious fetal hypoxia were recognized, namely, the late decelerations of the FHR in relation to uterine contractions, and the FHR variability decrease [**?**]. Although visual interpretation of CTG has an acceptable sensitivity for risk of hypoxia detection (specially in pathological traces), the specificity still is low (specially for suspicious traces), and it requires the confirmation with invasive pH determination of scalp blood of the fetus, which is technically cumbersome and not always feasible [**?**]. When considering the risk of hypoxia, gynaecologists indicate interventions (cesareans, forceps, and vacuum extraction) more often than necessary [**?**], hence increasing sensitivity at the expense of specificity. In addition, visual assessment of bradycardias and late decelerations is simple, whereas visual assessment of the loss of variability is not and even it varies among observers, representation type (computer display or paper), or cardiotocograph model [**?**, **?**, **?**].

cut the clutter in the previous paragraphs.

Several nonlinear indices has been proposed to assess the complexity from the FHR signal []

We aim assess the change in the complexity of the phys-

iological mechanism tha control the FHR due to hypoxia before the partum. We assess the FHR complexity using three differente nonlinea measures, Time Irreversibility, SampEn and Permutation Entropy in 32 foetal recordings (15 controls). The draw of the paper is as follows. Section 2 describes the considered alternatives for the elements of the detection system: signal segmentation, feature extraction, signal similarities computation, feature selection and classification. Then, Section 5 experimentally demonstrates the capability of NCD both for classification of raw signals and for extending the capabilities of conventional analysis in a real FHR dataset. Finally, Section 6 discusses the main advantages of the proposed methodology over other alternatives and concludes the paper.

2. Methods

In this section we present a briefly definition of the nonlinear indices used to assess complexity using the FHR.

2.1. Permutation Entropy

Here your code

2.2. Sample Entropy

Entropy-based methods provide a quantification of the irregularity of a temporal series. Among them, SampEn [**?**] holds some properties which are appropriate for the study of physiological signals, namely it is robust to noise and outliers, and accordingly, it has been widely applied for characterizing the HRV signal. The SampEn, which is a modification of the Approximate Entropy [**?**], is the negative natural logarithm of the conditional probability that two sequences which are similar for m points remain similar for $m+1$ points. Thus, a lower value of SampEn indicates more self-similarity in the time series [**?**]. In order to compute SampEn, the embedded dimension m, i.e., the length of the vectors to be compared, and the noise filter threshold r need to be specified. In this study the values for these parameters are set to $m = 2$ and $r = 0.2$. the standard deviation of the signal, since they are common values used in the literature [**?**].

2.3. Time Irreversibility

Here your code

3. Data description

FHR records[1] were acquired with a Philips cardiotocograph for a total of 32 recordings, 15 controls and 17 cases in the Hospital Universitario Fundación Alcorcón (Madrid, Spain). A case was declared whether: 1) the PH of the umbilical artery was ≤ 7.05; or 2) the APGAR score was ≤ 7 at 5 minutes after delivery and a reanimation type III or greater was required. The institutional Medical Ethics Review Board approved the use of this data.

Records, see Figure 1 for an example, have considerable variability both in start/ending times and pauses as labor duration vary. In addition, the cardiotocograph may be disconnected at any time for a number of reasons. Also, the signal is lost sometimes as the fetus and mother move. The cardiotocograph provides three signal qualities (lost, medium and high). We decided to consider the window between 4 to 1 hours before birth for our analysis, even though not all patients have signal along all this window, e.g. nine patients even start being monitored after 4 hours to delivery (8 cases) or they are removed the cardiotocograph before 1 hour to delivery (one case). When a patient has no signal in the entire interval analyzed in a experiment, it was excluded (see below).

review the real analysis we performed: which interval we used to compute the indices from fhr. We performed some preprocessing?

4. Statistical analysis

To test whether exists statistically significant differences on nonlinear indices between control and cases we performed statistical hypothesis tests based on bootstrap resampling. The null hypothesis (H_0) represents no difference between control and cases, against the alternative hypothesis (H_1) that there exists significant differences. We used the mean difference between each indices computed on control and cases groups as the statistic to summarize our data. Bootstrap hypothesis test is based on the idea of building an empirical distribution of the statistic, under H_0, and then computing the statistic on B different resamplings. Assuming that H_0 is true, bootstrap statistics are computed on resamplings from a pooled population (control union cases). We computed p-value as the fraction of the points on the distribution (probability) that are more extreme than the actual statistic value [**?**, **?**].

5. Results

The results are summarized in the following table:
build a table, is it worthwhile to use some boxplot?

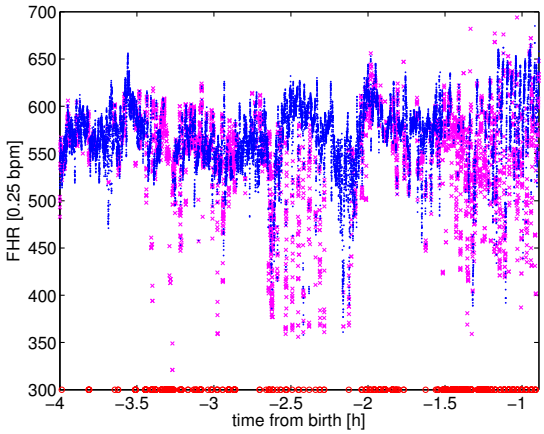
Table1 shows the mean and standard deviation of the nonlinear and linear indices computed on healthy and hypoxia cases.

Entropy indices showed higher complexity (higher values) in healthy fetuses. TI showed values farthest from zero, indicating higher complextiy, in healthy fetuses. Applying the bootstrap hypothesis test we verified a statistical

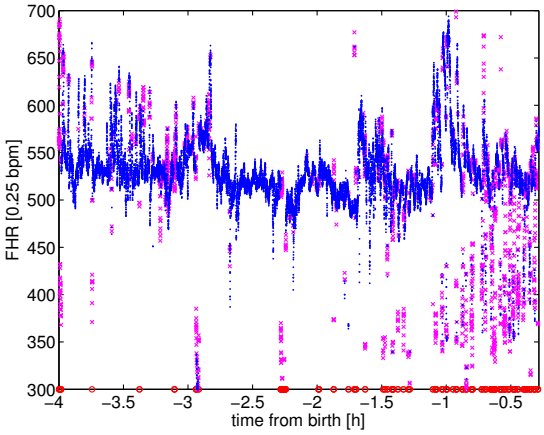
[1] Data is available from the website: http://sites.google.com/site/hufahypoxia.

	PE	SampEn	TI	STV	$\mathbf{P}_{hf}$
Healthy	0.72 ± 0.04	0.33 ± 0.12	$\mathbf{-0.38 \pm 0.19}^*$	3.23 ± 1.15	0.40 ± 0.18
Hypoxia	0.69 ± 0.07	0.28 ± 0.09	$\mathbf{-0.21 \pm 0.37}$	3.45 ± 1.35	0.43 ± 0.25

Table 1. *Mean±standard deviation of the nonlinear and linear indices computed on healthy and hypoxia cases. Symbol * means statistically significant difference (p-value < 0.1) using a bootstrap hypothesis test.*



(a)hypoxic



(b)control

Figure 1. FHR for (a) a hypoxic and (b) a control patient. Signal qualities are 9.9% lost, 19.2% medium and 70.9% high for (a) ; and 1.8% lost, 9.8% medium and 88.3% high for (b). Signal qualities high, medium and lost are respectively represented by the markers: "·", "x" and "o".

significant difference for TI.

Figure**??** show the box plot for nonlinear indices.

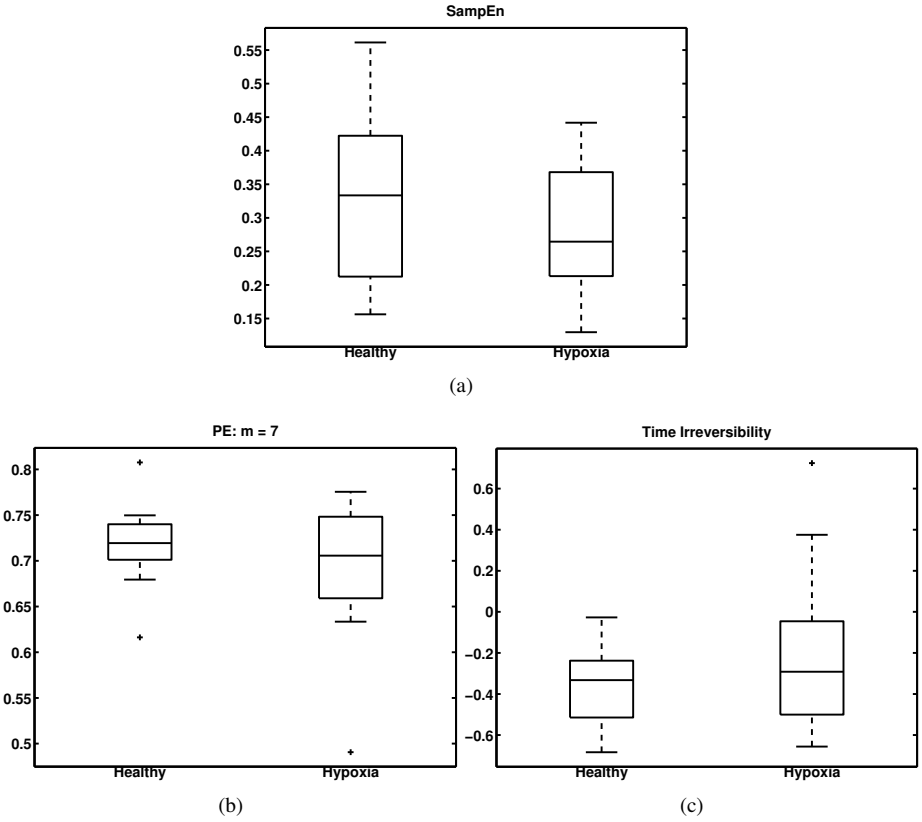
6. Discussion and Conclusions

Several indices have been proposed to analyze FHR. The most common indices are based on time domain and frequency domain methods [**?**, **?**]. Time domain methods aim to assess the long and short term variability of the FHR, whereas frequency domain methods aim to characterize the oscillatory contributions on the FHR. In many cases these indices are reduced to a single number obtained in the entire time series or to a collection of numbers obtained in 5-minute windows slides, which are again reduced to a few numbers like mean or standard deviation.

We have proposed NCD as a similarity measure for FHR registers because it is able to exploit both linear and nonlinear relations among records. We tried it with raw FHR records, time and frequency indices and moments signals extracted from sliding windows. We obtained better performance from the moments than from the raw records, which shows that the compressor is not able to extract all the relations in the data and preprocessing might help. Then, we have shown how to combine several moments (or other type of variable) by using a simple voting scheme and by summing the dissimilarity matrices, which provide overall better results in our case.

Main strengths of using NCD for comparing FHR registers or signals of any statistic applied to signal windows are simplicity and generality. Other commonly used information-theoretical measures, like Approximate entropy [**?**] or Sample entropy [**?**], need to tune the embedding dimension and specially the tolerance, which is a continuous parameter; but there is no parameter to tune for our approach. Indeed, despite we have tried three compressors and two simple approaches to get a symmetric matrix the proposed methodology can be straightforwardly used with a lzma compressor and the min approach to get symmetric NCD matrices. In addition, there is no problem with the common signal losses, which is a problem to apply frequency-related methods as they need signal interpolation, which is not always possible. The similarity can always be computed independently of how the signal losses are addressed.

Visual interpretation values basal FHR, its accelerations



(a)

(b)

(c)

and decelerations in relation to uterine contractions, and beat-to-beat FHR variability [**?**, **?**]. The following signal types are considered clearly pathological (suspicious of hypoxia): late decelerations, whose nadir has a delay of at least 30 seconds with respect to the acme of contractions; maintained bradycardia; low variability (less than 5 beats); and a "sine"-rhythm, named after its wave-like appearance, which is characterized by a long-term variability but almost no variability in the short-term.

First, late decelerations are produced in the following manner. During uterine contraction, when the myometrium pressure exceeds the blood pressure of the intervillous space of the placenta, maternal circulation is interrupted and therefore ceases to carry oxygen to the fetal territory. In a well oxygenated fetus, this is not a problem since the fetus does not consume all the tissue-oxygen before the end of the contraction. In a fetus with poor oxygenation, on the other hand, the oxygen reserves are exhausted before the end of contraction and, particularly in the more sensitive heart cells (which act as a pacemaker), action-potential production mechanisms are delayed, which causes bradycardia.

And second, the decrease in variability is more difficult to explain. On the one hand, the regulation of FHR, usually controlled by the vagus nerve, is stopped by cardiac and nervous system hypoxia. On the other hand, it seems that there are less functionally active cells in the pacemaker as the hypoxia progresses. The regulatory system loses "degrees of freedom", which therefore becomes increasingly uniform and deterministic [**?**].

FHR signal can be measured in two ways, namely, external, by using a ultrasonic sensor that observes the Doppler effect; and internal, in which fetal electrocardiogram (ECG) is measured with an electrode in the fetus scalp. Uterine activity can also be measured in two ways: by using a non-invasive pressure transducer in the mother's abdomen, or with an invasive intrauterine catheter pressure sensor.

Automatic analysis of CTG has also been proposed. In [**?**], automatic ST analysis (the ST segment connects the QRS complex and the T wave) combined with CTG was shown to increase the ability of obstetricians to identify hypoxia and to improve the perinatal outcome. In [**?**], a clinical trial was proposed to assess whether computer analysis and alerts improves visual CTG monitoring. A system-identification approach was proposed [**?**] to model FHR and uterine activity as an input-output system, reporting around 50% sensitivity with 7.5% of false positives at 1h and 40 minutes before delivery. Time-frequency analysis and features from time-frequency space decomposition were sucessfully used in an animal study yielding 93% sensitivity and 98% specificity[**?**]. Other non-invasive ap-

proaches have been proposed to complement CTG, such as Doppler velocimetry and pulse oximetry [**?**, **?**, **?**], or near-infrared spectroscopy to measure cerebral metabolic rate of oxygen [**?**]. The performance of automatic approaches currently applied to hypoxia detection has still room for improvement both in accuracy (sensitivity and specificity) and in the time before birth where hypoxia is detected.

NCD technique was successfully used for clustering the fetuses of a multicentric study with the aim of identifying the abnormal ones [**?**].

It is remarkable that, using sliding windows and NCD, both frequency indices and moments obtain the best accuracies in the $4 \leftrightarrow 3$ hours interval whereas time indices obtain the best results in the $3 \leftrightarrow 2$ hours interval. Our comparisons show that the commonly used Time and Frequency indices can be complemented by the moments, which are always applicable and do not suffer from signal loses. In addition, fetus movement might provide valuable information as we noted when analyzing raw signals (Table **??**), and when we observed the performance of $P_{LF}/(P_{MF}+P_{HF})$ index, which depends on fetus movement (Table **??**). Finally, by adding similarity matrices selected by FS a promising 0.88 accuracy is reached in the $4 \leftrightarrow 1$ hours interval, which compares entire records and mimics the processing of a fetus during labor.

Practical implementation of this approach as an plugin of the available CTG systems is straightforward. We recommend to perform a careful selection and labeling of FHR records. Then, the number of cases in the knowledge database and processing capabilities must be balanced. For instance, the analysis of an FHR record every minute against a knowledge database of 1000 patients is easily done in a normal PC using gzip as compressor.

The decision making during the labor is a difficult task for the gynecologist. It always should be intended to be as less invasive as possible, but, of course, ensuring fetal well-being and acting as soon as possible in case of suspicion of fetal hypoxia. Our main contribution shows how the NCD analysis of the readily available FHR traces may help the gynecologists to make the correct decisions. We reach 88% accuracy, which is a remarkable result if we take into account that we are actually identifying stressed fetuses 3 hours before delivery that were not detected by the gynecologist until a later stage. This general methodology is also applicable to other time series classification problems and it is both simple to understand and simple to apply.

The results obtained in this study indicate that a further study with more patients should be performed to open the application of this type of FHR analysis of the fetus condition to the industry.

Acknowledgements

The author Óscar Barquero Pérez has the support of a FPU Grant (AP2009-1726) from the Ministerio de Educación, Spanish Government. Authors want to thanks professor Antonio García Marques for his suggestions and thoughtful comments on LASSO models. Address for correspondence:

O Barquero-Pérez
Department of Signal Theory and Communications
University Rey Juan Carlos. B104, Camino del Molino s/n
28943 - Fuenlabrada (Madrid), Spain
Phone: +34 91 488 84 62
oscar.barquero@urjc.es