Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Evaluating the fetal heart rate baseline estimation algorithms by their influence on detection of clinically important patterns

*Janusz Jezewski**, *Krzysztof Horoba, Dawid Roj, Janusz Wrobel, Tomasz Kupka, Adam Matonia*

Institute of Medical Technology and Equipment ITAM, Zabrze, Poland

ARTICLE INFO

Article history:
Received 18 May 2016
Accepted 16 June 2016
Available online 25 June 2016

Keywords:
Fetal monitoring
Fetal heart rate
Baseline estimation
Acceleration
Deceleration

ABSTRACT

A correctly estimated component of fetal heart rate signal (FHR) – so called baseline – is a precondition for proper recognition of acceleration and deceleration patterns. A number of various algorithms for estimating the FHR baseline was proposed so far. However, there is no reference standard enabling their objective evaluation, and thus no methodology of comparing the different algorithms still exists. In this paper we propose a method for evaluation of automatically determined baseline in reference to a set of experts, based on ten separate groups of signals comprising typical variability patterns observed in the fetal heart rate. As it was proposed earlier [1], the given algorithm is evaluated based on the characteristic patterns detected using the obtained baseline, instead of direct analysis of the baseline shape. For the purpose of quantitative assessment of the estimated baseline a new synthetic inconsistency coefficient was applied. The proposed methodology enabled to evaluate eleven well-known algorithms. We believe that the method will be a valuable tool for assessment of the existing algorithms, as well as for developing the new ones.

1. Introduction

Automated analysis of fetal heart rate signal (FHR) is today one of the most common diagnostic methods used in perinatal medicine [2,3]. It relies on assessment of some FHR variability patterns, such as baseline, acceleration and deceleration episodes, as well as oscillations and others [4–6]. Among these patterns, the one most crucial for reliable interpretation of recording is the baseline – a line indicating the long-term changes of the fetal heart rate in time [7]. There is a common opinion that algorithm for the FHR baseline estimation is responsible for the efficiency of the entire computer-aided fetal monitoring system [8,9]. Therefore, it is necessary to develop a method and criteria, which would allow for evaluating the baseline estimation algorithms.

The baseline definition provided by the FIGO guidelines says: ''Baseline is the mean level of the FHR when this is stable, accelerations and decelerations being absent'' [10]. The impact of the baseline on detection of acceleration and deceleration

* *Corresponding author at*: Institute of Medical Technology and Equipment ITAM, Roosevelta 118, 41-800 Zabrze, Poland.
E-mail addresses: jezewski@itam.zabrze.pl (J. Jezewski), krzysztof.horoba@itam.zabrze.pl (K. Horoba), dawid.roj@itam.zabrze.pl (D. Roj), januszw@itam.zabrze.pl (J. Wrobel), tomekk@itam.zabrze.pl (T. Kupka), adamm@itam.zabrze.pl (A. Matonia).
http://dx.doi.org/10.1016/j.bbe.2016.06.003

(A/D) episodes is evident and results directly from the FIGO guidelines, which define these episodes as temporary deviations of the FHR curve from the baseline. Acceleration is defined as an increase of the FHR with amplitude of at least 15 beats per minute (bpm) and lasting more than 15 s. For deceleration the thresholds are equal to 15 bpm and 10 s.

Presence or absence of an appropriate number and type of A/D episodes recognized in the FHR signal is a basis for evaluation of the fetal state during pregnancy and labor [11–13]. Therefore, reliable detection of these episodes is a crucial issue in automated analysis of FHR recordings. It is, however, infeasible without correct baseline estimation, and even small inaccuracy may lead to misdetection of the fetal distress [14,15].

Automated methods of baseline estimation can be classified into three main groups:

- Filtering – based on filtering approach, most often relying on digital low-pass filtering. Also very common is nonlinear filtering, with filter parameters being modified according to the history of input and/or output signal.
- Statistical – based on statistical measures of central tendencies, being determined in successive segments of signal or for signal as a whole.
- Statistical and filtering mix – utilizes nonlinear filtering technique with parameters being modified according to statistical properties of the signal (as a whole or in overlapping windows).

As there are no strict requirements for an algorithm for the FHR baseline estimation, the aim of all existing methods is to simulate the clinical expert's behavior [16]. Therefore, the correctness of estimation can only be verified involving clinical experts of considerable experience in this matter [17,18]. The problem in this approach is that a perfect expert, and thus a reference baseline, does not exist. It was proven that significant interobserver disagreement in baseline interpretation is noted [19]. What is more, those differences occur even between baselines estimated by the same expert after some time (intraobserver disagreement) [20]. For this reason it is necessary to use a pseudo-reference of baseline interpretation, coming from a group of experts [21]. In this approach the final rating of an algorithm is determined by averaging the results obtained with respect to each individual expert's interpretation.

The baseline automatically estimated by a given algorithm should be as close as possible to the baseline assumed as a reference. However, direct assessment of the difference between two baselines does not provide any information on whether differences occur in a segment of FHR signal which is significant for further A/D episodes detection or not [22,23]. The relation between the inconsistency of baselines and inconsistency of A/D episodes is not a one-to-one mapping. There may be high baseline inconsistency and at the same time the inconsistency of the A/D identified using them may be low. Conversely, the baseline inconsistency may be low and it may be accompanied by a high inconsistency in the recognized episodes. It depends on whether the baseline differences occur in the segments with A/D episodes. Keeping in mind that these episodes are the most important in clinical practice [24], the direct evaluation of baseline inconsistency is not a relevant measure of its usefulness for the automated analysis [1]. Hence, the automatically estimated baseline must be evaluated solely on the basis of the A/D episodes recognized on its basis.

The right level of reference standard for baseline algorithm assessment should be provided by the baseline obtained from clinical experts, which already at the process of comparison will be represented exclusively by the A/D episodes detected automatically on its basis. The reason behind using automatically detected A/D episodes, instead of episodes directly indicated by experts, was presented in [25], where essential differences were noted in procedure of A/D recognition by a fetal monitoring system and human expert. It was clearly proven that the expert, although being instructed, was not using the established FIGO A/D definitions consistently. Therefore, it is advantageous to use only the baseline interpretations provided by the experts, while the A/D episodes should be detected automatically, with a rather simple detection algorithm directly following the FIGO guidelines.

The aim of this work is to propose a new, versatile methodology for evaluation of baseline estimation algorithms. This problem has already been raised in [19], where a method was described for quantification of inconsistency between two baselines. It relied on a set of coefficients, separately describing different aspects of inconsistency. However, during the research works it was observed, that for a specific class of FHR signals that inconsistency was incorrectly determined. To overcome this issue we proposed a modified approach, where all aspects of inconsistency are evaluated together, enclosed in the same formula. The proposed measure is called "synthetic baseline inconsistency index" – SI.

2. Methods

Block diagram of the procedure for comparison of automatically estimated baseline with the expert's interpretation of FHR trace is presented in Fig. 1. The expert draws his baseline directly on the FHR trace printed from the fetal monitoring system archive. The printout contains up to 60 min of recording and is consistent with the fetal monitor recorder format. Elements essential for conversion (e.g. calibration line, lines marking the full range of baseline variability) are printed in black, while the others in bright colors. Printout is scanned, and then the baseline curve is extracted from the bitmap image and converted into digital representation by dedicated software [25]. The result of printout processing is a file containing samples of baseline with accuracy of 1 bmp and averaged over 2.5 s periods. These data together with FHR signal are used as inputs for a procedure, which detects A/D episodes and calculates their detailed parameters (maximum amplitude, duration and area extended between FHR trace and the baseline). The same procedure is repeated for the baseline estimated automatically. Finally, two sets of A/D parameters are provided for the processed FHR trace, which in turn are the inputs to the final inconsistency determination procedure.

Method of inconsistency evaluation between two baselines has already been proposed in [19] and relies on a set of