Computers in Biology and Medicine 114 (2019) 103468

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Fetal heart rate baseline computation with a weighted median filter

Samuel Boudet [a,*], Agathe Houzé de l'Aulnoit [a,b], Romain Demailly [a,b], Laurent Peyrodie [c,d], Régis Beuscart [e], Denis Houzé de l'Aulnoit [a,b]

[a] Univ Nord de France, UCLille, Faculté de Médecine et Maïeutique, Biomedical Signal Processing Unit (UTSB), F-59800, Lille, France
[b] Lille Catholic Hospital, Obstetrics Department, F-59020, Lille, France
[c] Yncréa École des hautes études d'ingénieur, Biomedical Signal Processing Unit (UTSB), 59800, Lille, France
[d] I3MTO EA 4708 Orléans, France
[e] Univ Nord de France, CHU Lille, UDSL EA2694, F-59000, Lille, France

ARTICLE INFO

Keywords:
Biomedical signal processing
Nonlinear filter
Fetal heart rate
Baseline
Deceleration
Median filter
Fetal acidosis

ABSTRACT

Background Automated fetal heart rate (FHR) analysis removes inter- and intra-expert variability, and is a promising solution for reducing the occurrence of fetal acidosis and the implementation of unnecessary medical procedures. The first steps in automated FHR analysis are determination of the baseline, and detection of accelerations and decelerations (A/D). We describe a new method in which a weighted median filter baseline (WMFB) is computed and A/Ds are then detected.
Method The filter weightings are based on the prior probability that the sampled FHR is in the baseline state or in an A/D state. This probability is computed by estimating the signal's stability at low frequencies and by progressively trimming the signal. Using a competition dataset of 90 previously annotated FHR recordings, we evaluated the WMFB method and 11 recently published literature methods against the ground truth of an expert consensus. The level of agreement between the WMFB method and the expert consensus was estimated by calculating several indices (primarily the morphological analysis discordance index, MADI). The agreement indices were then compared with the values for eleven other methods. We also compared the level of method-expert agreement with the level of interrater agreement.
Results For the WMFB method, the MADI indicated a disagreement of 4.02% vs. the consensus; this value is significantly lower ($p < 10^{-13}$) than that calculated for the best of the 11 literature methods (7.27%, for Lu and Wei's empirical mode decomposition method). The level of inter-expert agreement (according to the MADI) and the level of WMFB-expert agreement did not differ significantly (p=0.22).
Conclusion The WMFB method reproduced the expert consensus analysis better than 11 other methods. No differences in performance between the WMFB method and individual experts were observed. The method Matlab source code is available under General Public Licence at http://utsb.univ-catholille.fr/fhr-wmfb.

1. Introduction

Since the 1970s, the fetal heart rate (FHR) has been a key parameter for monitoring fetal well-being during pregnancy and labor. An accurate FHR analysis can reduce the frequency of inappropriate obstetric interventions (e.g. instrumental vaginal deliveries and cesarean sections) and decrease the risk of fetal acidosis.

In an FHR analysis, one must first determine the value of several elementary parameters: the baseline (i.e. the mean level of stable FHR segments [1]), the variability (i.e. variations in amplitude during stable periods), accelerations (i.e. a sudden increase in FHR for more than 15 s, with an amplitude of more than 15 beats per minute (bpm) above the baseline), decelerations (i.e. a temporary decrease in the FHR for more than 15 s, with an amplitude of more than 15 bpm below the baseline), and sinusoidal patterns [1]. The FHR recording is considered to be normal if the baseline is between 110 to 160 bpm, the variability is between 5 to 25 bpm, and there are accelerations but no decelerations.

Conversely, abnormal FHR patterns comprise episodes in which (i) the FHR baseline is above 160 bpm (i.e. tachycardia), (ii) the FHR baseline is below 110 bpm (i.e. bradycardia), (iii) the variability is reduced (bandwidth < 5 bpm for more than 50 min), (iv) the variability is high (bandwidth > 25 bpm for more than 30 min), (v) the FHR signal contains decelerations of various types (late, variable, or prolonged), and (vi) a sinusoidal pattern is apparent. The absence of

[*] Corresponding author.
E-mail address: samuel.boudet@gmail.com (S. Boudet).

https://doi.org/10.1016/j.compbiomed.2019.103468
Received 26 March 2019; Received in revised form 14 September 2019; Accepted 23 September 2019
Available online 24 September 2019

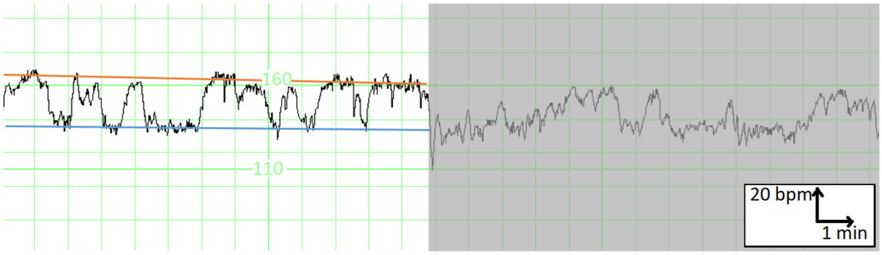


Fig. 1. Example of an FHR recording with ambiguous baseline position. If one is not aware of the end of the recording (the part with a gray background), the red (high) baseline seems more appropriate. However, when the end of the recording is taken into account, the blue (low) baseline seems more probable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

accelerations is not reassuring but cannot be interpreted with certainty. Depending on the combinations of abnormal episodes and the intensity of the abnormal pattern, the criteria in the FIGO consensus guidelines enable an FHR recording to be defined as “normal”, “suspicious” or “pathologic”[1].

As long as one knows exactly where the baseline is located, all these abnormal patterns can be detected easily. However, trying to compute the FHR signal baseline creates a circular definition: FHR accelerations/decelerations (A/D) are defined as periods that are 15 bpm above or below the baseline, whereas the baseline is defined as the mean signal after A/D have been excluded from the data. Fig. 1 shows an example of an FHR recording with an ambiguous baseline position. If the baseline is set at the top (the red line), the FHR will correspond to tachycardia (i.e. a baseline above 160 bpm) and will have many decelerations — which is a worrying clinical sign. If the baseline is set at the bottom (the blue line), the FHR will be normal and will include many accelerations - a perfectly healthy pattern sometimes referred to as "pattern D" [2]. Misinterpretation of “pattern D” can lead to inappropriate clinical management [2]. This example shows that correct placement of the baseline requires examination of the FHR signal over a long time period (> 30 min) - even though the FIGO consensus guidelines suggest that the baseline should be determined in a 10 min window [1].

The interpretation of FHR signals is subject to considerable inter-observer and intra-observer variability [3], as a result of sometimes imprecise assessment criteria and a lack of practical training.

Several researchers have developed automatic analysis methods for determining the baseline and A/D. The first automatic method was developed by Dawes et al. [4], and the concepts behind several other automatic methods are summarized in [5]. In earlier research, we compared 11 automatic analysis methods with an expert consensus [6]. Although the method developed by Lu and Wei [7] gave the best results, the difference with the expert consensus was still marked (i.e. greater than the inter-expert differences). Jezewski et al. [8] also compared several automatic analysis methods, although most of the recently published methods were not included. Lastly, various software packages for automatic FHR analysis are commercially available (for a review, see [9]) but have not been compared with Lu and Wei’s method.

Here, we describe a novel weighted median filter baseline (WMFB) method for baseline determination. The main challenge is positioning the baseline when there are prolonged and/or repeated A/D; the FHR is rarely at the basal level. Our design was prompted by the following seven (informal) hypotheses:

H1. At any given time point, the FHR is either in a baseline state or in an A/D state. After A/D periods have been excluded, the baseline corresponds to the average FHR measured during a time window of a few minutes around the current time point.

H2. One consequence of H1 is that the FHR value during A/D periods should not influence the baseline.

H3. Over an approximately 10 min window, strong baseline fluctuations are monotonic. Indeed, if the FHR falls and then rises during the window, a deceleration is present and conversely, if the FHR rises and then falls during the window, an acceleration is present.

H4. Short (< 10 min) down-up (or up-down) baseline fluctuations may occur but they are slow (low derivative) and have a low amplitude.

H5. It is unlikely that a period with no FHR variations corresponds to an A/D. It is even more unlikely that the FHR is in baseline state when there is strong variation in the FHR.

H6. Over a long time window (>~ 20 – 40 min), the stable periods (i.e. those without a strong variation) corresponding to the baseline state should account for more than 50% of the total stable periods in the window.

H7. The baseline passes through the points of the FHR if the FHR is a monotonic signal with a slight slope.

The WMFB method’s novelty relates to the combined implementation of these seven hypotheses in a balanced manner. Our analysis is currently performed off-line but could be adapted for real-time analysis — although this would come at the cost of a decrease in precision due to the lack of knowledge of the future FHR.

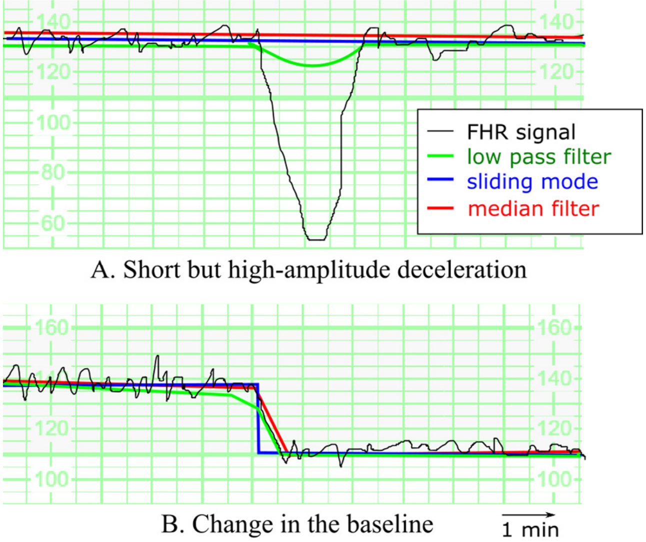
We evaluated the WMFB method by applying a previously described methodology and dataset [6], and using the open source FHR morphological analysis toolbox for MATLAB [10] [http://utsb.univ-catholille.fr/fhr-review].

2. Signal processing method

2.1. Signal acquisition and preprocessing

The FHR signals were recorded using an AVALON FM50 fetal monitor (Philips Healthcare, Amsterdam, The Netherlands). We had previously developed a specific program for capturing the FHR signal package via the RS232 serial interface (available on most fetal monitors) [11]. For both the training and testing datasets, the FHR recordings had been obtained via a Doppler ultrasound probe placed on the women’s abdomen or an electrode placed on the fetus scalp. The FHR and tocometry data are recorded at 4 Hz. All the parameters are set using the open-access training dataset described in our previous publications [5,12]. This dataset contains 66 recordings that have been analyzed by expert consensus. The evaluation dataset will be described in Section 3.1.

Classical preprocessing methods (close to those described in [13] and [14]) were then applied: (i) aberrant samples (> 220 bpm or < 50 bpm) were considered to be missing data. (ii) contiguous periods of less than 30 s and that differed by more than 25 bpm from the previous and following FHR periods were considered to be unreliable, and so were considered to be missing data. (iii) periods with missing data were completed by linear interpolation.



A. Short but high-amplitude deceleration

B. Change in the baseline

Fig. 2. An illustration of the results produced by a low-pass filter, a sliding mode and a sliding median, as applied to two types of signal.

2.2. Application of a median filter

In the literature, most of the baseline calculation methods are based on either a low-pass linear filter or a sliding mode (i.e. determination of the most frequent FHR value within a sliding window). Fig. 2 shows baselines determined in each of those ways for two particular FHR recordings.

In Fig. 2.A, the sliding window mode provides a satisfactory result but the low-pass filter tends to follow the decelerations — especially when the latter have a high amplitude. This prevents accurate estimation of the deceleration start, end and amplitude. From a theoretical viewpoint, the low pass filter does not comply with hypothesis H2 (the FHR value in A/D periods markedly influences the baseline).

In Fig. 2.B, the low-pass filter provides a satisfactory result for a change in baseline but the sliding mode creates an abrupt step-down that has no physiological significance. Depending on the signals before and after the baseline change, there may also be a time delay between the step-down and the true change in the baseline FHR. In principle, the sliding mode does not comply with hypothesis H7 (the baseline of a decreasing curve contains steps, whereas it should be equal to the FHR itself).

Our sliding median approach gave satisfactory results in both situations. In fact, the sliding median does not depend on the deceleration amplitude (and thus complies with hypothesis H2) and tracks monotonic changes in the signal (H7). Moreover, if the window length is more than 10 min, H3 (but not H4) is met. If the window length it is less H4 is respected (but not H3).

Fuentealba et al. also suggested the use of a sliding median but implemented it in a quite different manner [15]. The results appeared to be satisfactory in simple cases but may have been insufficient when prolonged or repeated decelerations were present.

2.3. A weighted median filter

To make the method more accurate, we used a weighted window that gave more weight to the points close to the current point. The working window was set empirically to $T = 40$ min, centered on the current point (sliding step: 1 sample). The window is described in Eq. (1). In Section 2.5, we shall explain how several successive baselines can be calculated using variable windows.

$$W(t) = \begin{cases} 1 - \left|\frac{t}{T/2}\right| & \text{, if } -\frac{T}{2} < t < \frac{T}{2}, \\ 0 & \text{, otherwise.} \end{cases} \qquad (1)$$

However, the sliding median alone is not sufficient; whenever decelerations are prolonged and/or frequent, the baseline deviates — albeit to a lesser extent than with a low-pass filter (thus H2 is not perfectly respected). To solve this problem, we calculated the median by decreasing the weight given to points within a priori A/D period (thus exploiting H1).

For a given FHR time sample i, we assigned a weighting equivalent to the prior probability P(i) with which i is in a baseline period (rather than an A/D period). This weighting is then multiplied by the weighting of the window $W(j)$. Eq. (2) is used to calculate this baseline, where T is the width of the window (40 min x 240 points/min = 9600 points), P is the prior probability, and W is the window weighting (detailed in Section 2.5).

$$BL(i) = \underset{j=-\frac{T}{2}\cdots\frac{T}{2}}{median}\ (FHR(i-j), P(i-j)W(j)) \qquad (2)$$

Here, the median corresponds to a weighted median: if $C(i)$ are the weights and $X(i)$ are the values, $median_i\,(X(i), C(i))$ is defined as the highest value of m fulfilling $\sum_{\{i|X(i)\leq m\}} C(i) < \frac{1}{2}\sum_{\{i\}} C(i)$.

The sections below describe how P and W are obtained. P will be defined as a product of two probabilities P_{stab} and P_{trim}: P_{stab} will be used to eliminate periods with strong variations (thus exploiting H5), and P_{trim} will be used to eliminate periods in which the FHR is far from the long-term average baseline (thus exploiting H3, H4, and H6).

2.4. Estimation of the prior probability, based on the FHR stability

According to H5, the FHR baseline necessarily corresponds to stable periods, i.e. periods during which the low-frequency FHR signal does not fluctuate significantly. Depending on the derivatives of the FHR, we set a high probability for stable periods and a low probability for unstable periods.

On a smoothed signal, the points with a high first derivative have a high likelihood of corresponding to the slope of an A/D. Likewise, the points with a high second derivative have a high likelihood of corresponding to the trough of an A/D. The prior probability P_{stab} for these time points can be set to a very low value — enabling these periods to be ignored. During the trough of an A/D, the first derivatives are almost null; hence, it would not be possible to distinguish between the baseline and the A/D on the basis of the first derivative alone.

An effective way of accounting for both the first and second derivatives is to calculate the signal derivative analytical envelope [16]. Fig. 3. A shows an FHR recording (in black) and the same signal smoothed at 3 bpm (in red). Fig. 3.B shows the derivative of this smoothed signal (in black) and the derivative envelope (the dashed blue line). Whereas the derivative is null in the trough of the deceleration, the envelope has a high value throughout the deceleration. It is therefore possible to set P_{stab} as a function of this envelope.

The accuracy of P_{stab} can still be improved. Ideally, P_{stab} should be as polarized as possible (i.e. close to 1 for stable periods, and close to 0 for unstable periods). More particularly, P_{stab} must change as rapidly as possible near the start and end of A/Ds.

In order to increase the method's time-domain responsiveness, the low-pass filter's cut-off must be high. However, the higher the cut-off frequency, the more the filtered signal will contain high-frequency oscillations of little relevance for calculating the baseline. This compromise was addressed by using a logistic regression model: the objective was to estimate the state of the signal (baseline or A/D) at each instant (the dependent variable) as a function of the following nine parameters (the independent variables):

- $\left|d(FHR_{0-1bpm})\right|$, $\left|d(FHR_{1-3bpm})\right|$, $\left|d(FHR_{3-7bpm})\right|$,
- $\left|d^2(FHR_{0-1bpm})\right|$, $\left|d^2(FHR_{1-3bpm})\right|$, $\left|d^2(FHR_{3-7bpm})\right|$,
- $envelope\left(d(FHR_{0-1bpm})\right)$, $envelope\left(d(FHR_{1-3bpm})\right)$, $envelope\left(d(FHR_{3-7bpm})\right)$.

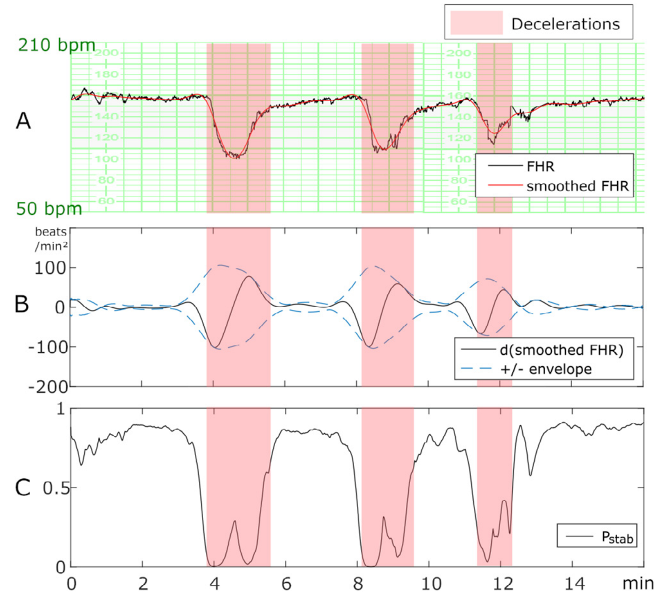


Fig. 3. Illustration of the computation of the *prior* probability signal. - A. An FHR sample with variable decelerations and the smoothed FHR signal (low-pass filter at 1 bpm) - B. The derivative of the smoothed FHR, and the envelope of the derivative - C. The prior probability P_{stab} that the FHR is in baseline state.

where FHR_{a-b} is the bandpass filtered FHR with cut-off frequencies of a and b (Butterworth sixth-order low-pass filter in a forward–backward process), $d(X)$ is the first derivative of X, and $d^2(X)$ is the second derivative of X. They are estimated by finite-difference in beats/min^2 for first derivative and in beats/min^3 for second derivatives.

We constructed a training database comprised 37 recordings (extracted from the training dataset described in Section 2.1) of variable length, with a total duration of 93 h. A total of 1800 A/D were selected by an expert. The selection also included small/short A/D (< 15 bpm and/or < 15 s) that should not be taken into account for baseline calculation.

In order to limit the amount of data and the computation time, we selected one in five of the FHR samples. This data reduction step was performed after the nine features had been computed using the original sampling frequency (4 Hz). To simplify the model, the four variables maximizing the area under the receiver operating characteristic curve (AUC) were selected in a stepwise forward-selection process. Given that the improvement with the 9 variables (instead of 4) was less than 0.01%, we preferred to use the simpler model. The final model of the prior probability is given by Eqs. (3) and (4), where $L(i)$ is the logit from the logistic regression:

$$P_{stab}(i) = \frac{e^{L(i)}}{1 + e^{L(i)}} \tag{3}$$

$$\begin{aligned} L(i) = & -2.4744 + 0.0266 \left| d(FHR_{0-1bpm})(i) \right| \\ & + 0.0413 \; envelope \left(d(FHR_{0-1bpm}) \right)(i) \\ & + 0.0105 \; envelope \left(d(FHR_{1-3bpm}) \right)(i) \\ & + 0.0036 \; envelope \left(d(FHR_{3-7bpm}) \right)(i) \end{aligned} \tag{4}$$

The AUC for this model was 0.87 showing that the classification was even effective when based solely on information from the derivatives. Fig. 3.C shows the values of P_{stab} for a sample FHR recording. As can be seen with the selected variables, it was more effective to use the envelope of the first derivative rather than the absolute value of the second derivative to differentiate the trough of an A/D from a baseline period. We also considered that it was important to use the absolute value of the first derivative in the model, so as to increase the responsiveness of P_{stab}.

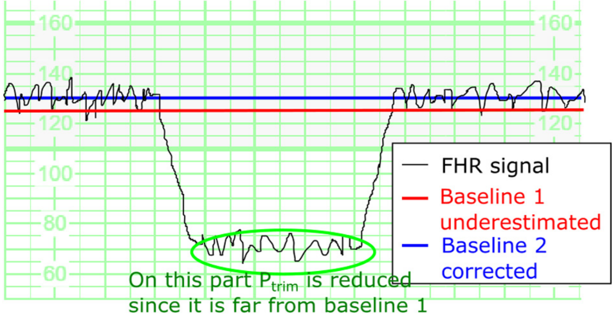


Fig. 4. Illustration of two progressive trimming steps for a deceleration with a "flat trough".

Jimenez et al. had also suggested studying the signal instability [17] but only thresholded the first derivative. A similar approach was also used by Lu and Wei [7].

2.5. Progressive trimming

The methods described above are not, however, sufficient for determining the baseline during a prolonged deceleration with a "flat trough" (Fig. 4).

During these long decelerations, the *prior* probability P_{stab} of a baseline state remains very high over a long period of time; this has a strong influence on the calculated baseline. If the sliding window is narrow, the baseline may even pass through the deceleration's flat trough. In such cases, the only way to find the right baseline is to work in a broad time window (40 min); however, the baseline would then no longer track small variations (H4).

We therefore calculated an initial, approximate baseline with a wide window (40 min) and a very smooth signal (a low cut-off frequency - 1 bpm Butterworth 4-order low-pass filter with a forward–backward process); this yielded baseline 1 in Fig. 4. According to H6, the stable baseline point should account for more than 50% of the total stable periods over the window; since the strong variations are monotonic (H3), the baseline must be a good approximation. Consequently, all the FHR points far from this baseline had a low probability (referred to as P_{trim_1}) of being baseline points.

A new baseline was then calculated with the new weightings $P_1 = P_{stab} * P_{trim_1}$. This baseline is then calculated over a shorter window, to take account of small fluctuations (H4). This process was repeated six times. Depending on the iteration k ($k = 1 \ldots 6$):

- the window W_k becomes narrower;
- we increasingly penalize the points far from the previous baseline by decreasing their probability P_{trim_k};
- we decreasingly smooth the FHR by increasing the cut-off frequency f_k.

Eq. (5) was used to calculate the successive baselines.

$$\begin{aligned} BL_k(i) = \underset{j=-\frac{L}{2} \ldots \frac{L}{2}}{median} \Big(& FHR_{0-f_k}(i-j), \\ & P_{stab}(i-j) P_{trim_k}(i-j) W_k(j) \Big) \end{aligned} \tag{5}$$

To set the window W_k, we decreased the weightings more rapidly as a function of their distance from the current time point (rather than reducing the width of the window, which remained at 40 min). Hence, points far from the current point had low but non-null weightings. If none of the closest points have a high P, the baseline can be positioned by using points further away. Eq. (6) describes the reduction in the

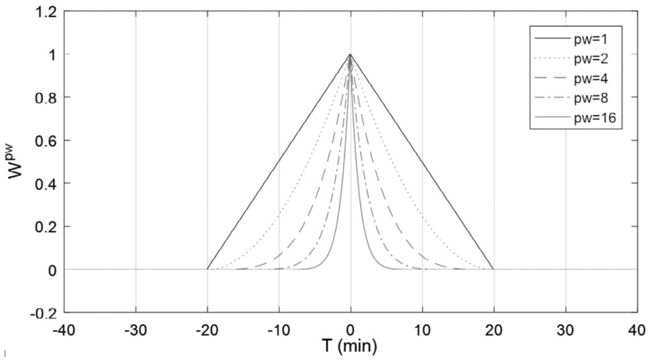


Fig. 5. Window weightings W_k applied with the sliding median, in order to obtain the successive baselines BL_k.

Table 1
The method's coefficients for each iteration.

Iteration (k)	1	2	3	4	5	6
Power weighting (Pw_k)	1	2	4	8	16	16
Cut-off frequency in BPM f_k	1	2	4	8	16	16
C_k	$+\infty$	3.2	2.5	2	1.5	1

weighting as a function of the number of iterations k, and where W is defined by Eq. (1). The values of the coefficients Pw_k are given in Table 1. Fig. 5 shows the curves obtained as a function of Pw_k.

$$W_k(i) = W(i)^{Pw_k} \tag{6}$$

Regarding the tolerable difference between the previous baseline and the FHR, we set the probabilities P_{trim_k} by using a second logistic regression equation (Eq. (7)). Here, the weighting C_k was used to adjust the "tolerance" of the FHR-baseline difference, and the chosen values are given in Table 1.

$$P_{trim_k}(i) = \frac{e^{C_k - 0.19\left|FHR_{0-f_{k-1}}(i) - BL_{k-1}(i)\right|}}{1 + e^{C_k - 0.19\left|FHR_{0-f_{k-1}}(i) - BL_{k-1}(i)\right|}} \tag{7}$$

Table 1 gives the different cut-off frequencies f_k.

We programmed an exception for the case where the average $P_{stab} * P_{trim_k}$ over the window W_k is below 10% of the same average for the iteration $k-1$. This corresponds to cases where there are almost no baseline periods within the majority of the window W_k. In such a case, the baseline value of the previous iteration ($BL_{k-1}(i)$) should be used instead. However to ensure continuity of the baseline signal, we preferred to add ($BL_{k-1}(i)$) to the list of values in Eq. (5) with a weight $P_{BL_{k-1}(i)}$:

$$P_{BL_{k-1}(i)} = 10\% \times \underset{W_{k-1}*\delta_i}{mean}\left(P_{stab} * P_{trim_{k-1}}\right) - \underset{W_k*\delta_i}{mean}\left(P_{stab} * P_{trim_k}\right) \tag{8}$$

$W_k * \delta_i$ corresponds to the window W_k shifted by i samples. The idea of using progressive trimming comes from Taylor et al. [14]. One of the differences here is that we used a less stringent method to define which part of the signal is baseline and which part is an A/D, whereas Taylor used a clear-cut distinction.

2.6. Summary of the method

Our baseline calculation method is summarized in Fig. 6.

2.7. Detecting accelerations and decelerations

We detected A/D by taking account of all the periods during which the FHR was respectively above or below the baseline for more than 15 s and reached an amplitude of 15 bpm.

For the particular case in which (i) the FHR comes within 5 bpm of the baseline without crossing it, and (ii) the amplitude each side of the peak or trough is greater than 15 bpm, the A/D is split into two. Fig. 7 gives an example of a split deceleration.

3. Evaluation

3.1. Evaluation method

The dataset described in the literature [6,12] was used to evaluate our new method and compare it with 11 other published methods. As described in [6], this evaluation dataset contains 90 recordings (lasting between 90 and 120 min) divided into 3 difficulty levels (30 easy-to-interpret recordings, 30 recordings that are neither easy nor difficult to interpret, and 30 difficult-to-interpret recordings). The difficult-to-assess recordings generally included a fluctuating baseline and many often-prolonged A/D. The recording periods were selected by an expert obstetrician from among 12,000 live births having taken place between 2011 and 2016 in the maternity clinic at St Vincent de Paul Hospital (Lille, France). They had been recorded with the process and preprocessing described in Section 2.1. The average signal loss is 2.0%, and the greatest signal loss was 7%). The expert manually excluded periods with unreliable signal (periods which seemed noisy or may have corresponded to the maternal heart rate); this accounted for an additional 0.9% of signal loss. The recordings did not include the second stage of labor because the latter have to be analyzed differently and it is difficult to obtain good-quality signals on this period. All the women concerned had given birth to a live child after 36 to 40 weeks of gestation.

The dataset also contained the morphologic analyses (baselines and A/D) produced by 11 literature methods (as reprogrammed by us [10]; source code available at http://utsb.univ-catholille.fr/fhr-review), namely the methods described by Ayres et al. (referred to hereafter as "A") [13], Cazares (C) [18], Houzé et al. (H) [19], Jimenez et al. (J) [17], Lu and Wei (L) [7], Maeda et al. (MD) [20], Mantel et al. (MT) [21,22], Mongelli et al. (MG) [23], Pardey et al. (P) [24], Taylor et al. (T) [14] and Wrobel et al. (W) [25]. All these researchers defined a baseline computation algorithm but only some defined an A/D thresholding method. For every method, a standard A/D thresholding method (described by A, L, P and W) was applied; these methods are then referred to as A*, C*, H*, J*, etc. This standard thresholding consists in defining A/D by taking all the periods in which the FHR was respectively above or below the baseline for more than 15 s and reached an amplitude of 15 bpm. For the methods incorporating specific A/D thresholding, the thresolding methods were also reprogrammed by us (C, J, MT and T); these methods are then referred to as C, J, MT and T (without asterisk). The pre-processing described in Section 2.1 was applied for all those methods.

This evaluation databank did not contain any recordings used to set the method's parameters described in Section 2.1.

Once the evaluation databank had been created, all the recordings were analyzed independently by three experts, who drew a baseline and positioned the A/D. Next, a fourth (more experienced) expert defined the consensus analysis by choosing the best of the three analyses at each time point. We could therefore be sure that for each time point, at least two of the four experts agreed on a consensus analysis. More details are given in [6].

Disparities between our method's analysis and the consensus analysis were computed for each recording, according to several indices:

- the **morphological analysis discordance index** (MADI) was the present study primary criterion. The MADI was introduced in [6], and was designed to give appropriate weight to differences between two morphologic analyses. It represents a fuzzy evaluation of the proportion of the time during which two baselines are considered to be discordant. Two baselines are said to be

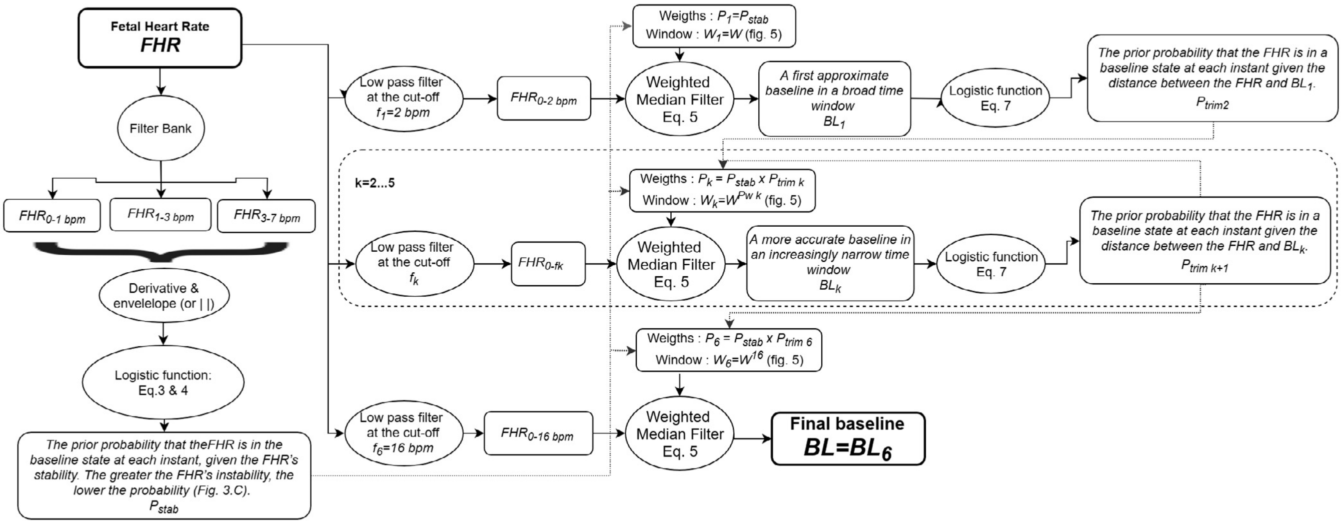

Fetal Heart Rate
FHR
Filter Bank
FHR0-1 bpm
FHR1-3 bpm
FHR3-7 bpm
Derivative & envelelope (or | |)
Logistic function: Eq.3 & 4
The prior probability that theFHR is in the baseline state at each instant, given the FHR's stability. The greater the FHR's instability, the lower the probability (Fig. 3.C).
Pstab
Low pass filter at the cut-off f1=2 bpm
FHR0-2 bpm
Weigths : P1=Pstab
Window : W1=W (fig. 5)
Weighted Median Filter Eq. 5
A first approximate baseline in a broad time window BL1
Logistic function Eq. 7
The prior probability that the FHR is in a baseline state at each instant given the distance between the FHR and BL1.
Ptrim2
k=2...5
Low pass filter at the cut-off fk
FHR0-fk
Weigths : Pk = Pstab x Ptrim k
Window : Wk=W^Pw k (fig. 5)
Weighted Median Filter Eq. 5
A more accurate baseline in an increasingly narrow time window BLk
Logistic function Eq. 7
The prior probability that the FHR is in a baseline state at each instant given the distance between the FHR and BLk.
Ptrim k+1
Weigths : P6 = Pstab x Ptrim 6
Window : W6=W^16 (fig. 5)
Low pass filter at the cut-off f6=16 bpm
FHR0-16 bpm
Weighted Median Filter Eq. 5
Final baseline
BL=BL6

Fig. 6. The flow scheme for baseline determination using the WMFB method.

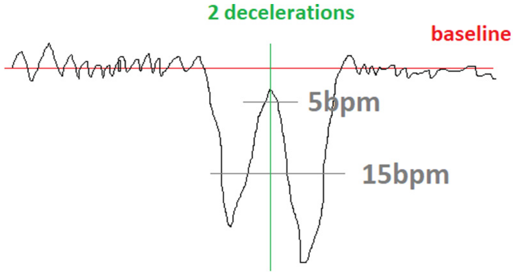


Fig. 7. Illustration of the subdivision of a double deceleration that does not cross the baseline in the middle.

discordant if one suggests that the FHR is in the baseline state and the other suggests that an A/D episode is present. Importantly, a 10 bpm difference between two baselines will be significant when the FHR is in the baseline state (with low variability) but not when the FHR is decelerating or when the variability is high.

- the **synthetic inconsistency (SI) coefficient** (introduced by Jezewski et al. [8]) is a guide to the overall quality of A/D event detection. It takes account of the number, location, and area of A/D, and corresponds to a percentage of the A/D area difference. Although the SI coefficient ranks methods in a reasonable order, we previously found that its values are counter-intuitive: disagreements of around 50% can correspond to minor differences in the analysis [6]. For example, a method that detected 90% of the decelerations and had no false detections would have an SI coefficient indicating a disagreement of 31.6%.
- the **root mean square difference** (RMSD) between baselines [6, 7] is commonly used to evaluate the mean baseline difference in bpm:

$$\mathrm{RMSD}(A,B)=\sqrt{\frac{\sum_{i=1}^{n}(A_i-B_i)^2}{n}} \tag{9}$$

where A and B are the two baselines and n is the number of samples.

- The **deceleration F-measure** is the harmonic mean of the sensitivity and the positive predictive value. Two analyses are said to agree when the respectively detected decelerations overlap by at least 5 s.
- The **acceleration F-measure** is the same but for accelerations.

These indices are clearly measures of agreement, and the present study complies with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [26]. The median for the whole set of recordings was computed for each index, along with its confidence interval (corresponding to the 36th and 55th ranked values over the 90 recordings). A two-sided Wilcoxon signed rank test was then used to compare the WMFB method with Lu and Wei's method (the best literature method among the 11 tested on [7]).

To establish whether or not our method performed as well as an expert, we also computed the differences between the WMFB method's analysis and those performed by independent experts (namely WMFB-Ei for i=1,2 or 3). For each recording, we defined WMFB-E as the median of the three WMFB-Ei differences. These values were compared with the median of the three inter-expert differences (E1–E2, E2–E3, and E3–E1), referred to as E–E. For each index, a Wilcoxon test was used to compare WMFB-E with E–E. We could not use the consensus analysis to say whether a given expert was better or worse than WMFB because the three experts participated in this consensus; hence, the level of expert-consensus agreement would have been biased and markedly overestimated.

3.2. Results

3.2.1. Examples

Fig. 8 shows an example of an FHR recording analyzed with the WMFB method and Lu and Wei's method [7], together with the expert consensus. This example included several variable decelerations and a few accelerations, all of which were detected by the two automated methods. The most difficult-to-interpret feature in this recording is the prolonged deceleration followed by a change in the baseline. Lu and Wei's method underestimates the baseline, which in turn leads to the incorrect detection of several small decelerations (rather than a single prolonged deceleration). In contrast, the WMFB method gave a satisfactory result in this situation.

3.2.2. Comparison with other methods

Fig. 9 illustrates the performance of the WMFB method vs. Lu and Wei's method, using the expert consensus as the reference. For all indices (the MADI, the SI coefficient, the baseline RMSD, and the deceleration F-measure) except the acceleration F-measure, the WMFB method agreed significantly more with the consensus than Lu and Wei's method did (MADI: $p = 7.1 * 10^{-14}$, SI: $p = 3.5 * 10^{-6}$, baseline RMSD: $p = 6.4 * 10^{-8}$, deceleration F-measure: $p = 1.5 * 10^{-5}$ and acceleration F-measure: $p = 0.9$). It must be borne in mind that

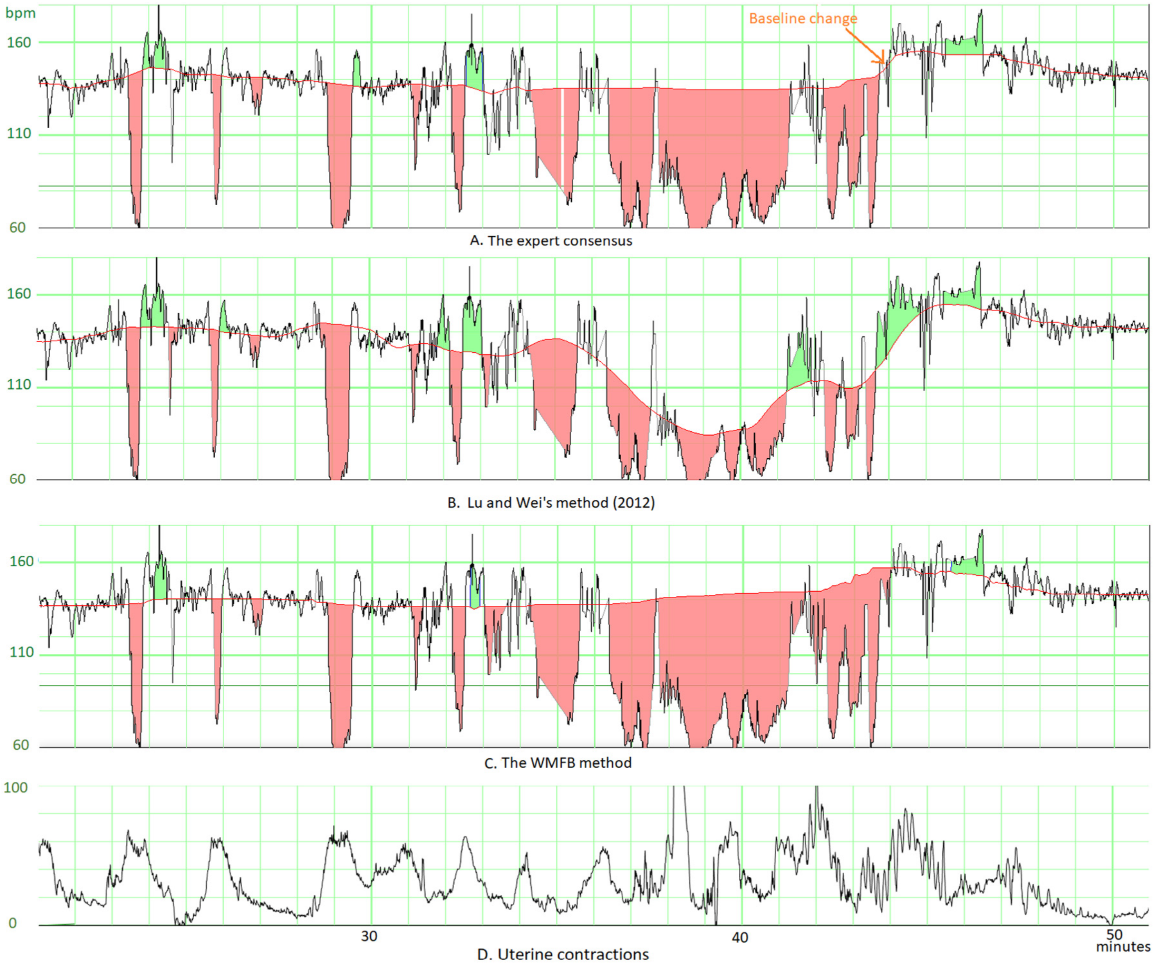


Fig. 8. Examples of FHR recordings analyzed with the WMFB method and Lu and Wei's method, together with the expert consensus analysis. The FHR includes several decelerations (red areas) and accelerations (green areas). One of the decelerations is prolonged, which is due to a hyperkinetic phase. The recordings also include a change in baseline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Lu and Wei's method was the best of 11 automated analyses studied previously [6]. According to the primary endpoint (the MADI), the level of disagreement was 7.3% for Lu and Wei's method and 4.0% for the WMFB — suggesting that the latter method constitutes a notable improvement.

Fig. 10 shows the performance of the WMFB method and all 11 literature methods with regard to three indices evaluating independent morphological features of the FHR (baseline, accelerations, and decelerations).

3.2.3. *Comparison with inter-expert levels of agreement*

A comparison of the WMFB-expert agreement with the inter-expert agreement is shown in Fig. 11. For four of the five criteria (the MADI, the SI coefficient, the baseline RMSD, and the acceleration F-measure), the differences were not statistically significant. For the deceleration F-measure, the level of inter-expert agreement was higher than the level of WMFB-expert agreement (p=0.035).

4. Discussion

4.1. Discussion of our results

Our WMFB method appeared to perform markedly better than 11 literature methods [6], and even seemed to perform as well as three experts did. In fact, the level of WMFB-expert agreement did not differ significantly from the level of inter-expert agreement for all indices other than the deceleration F-measure.

Some of the methods compared in the present study correspond to old versions of commercialized methods [13,18,24]. Although we know that these methods have changed since the original publication, the details have not been published. Hence, we did not evaluate the current versions of the commercialized methods.

The improvement over other automated methods is notably due to the longer analytical time window (40 min for the WMFB method, compared with the usual 10 min – or even less – for the other methods). Given that a prolonged deceleration can last for up to 10 min, appropriate positioning of the baseline requires an analysis over longer time durations. Another contributory factor is perhaps the application of a stability criterion - a concept also used in the methods developed by Jimenez and by Lu and Wei. Relative to the latter methods, the WMFB advantageously weights this stability — giving a better compromise for FHR recordings with few stable periods.

It is probable that the inter-expert variability measured in our study is much lower than inter-expert variability in routine practice. Firstly, the experts annotated the recordings in detail by precisely measuring A/D; this is not the case in routine practice. Secondly, the experts in the present single-center study were all specialists in FHR analysis; it is highly probable that a newly qualified midwife would disagree more with the three experts [27]. It would be interesting to measure the performance levels of various practitioners according to these criteria.

When considering the recordings retrospectively, it appears that there were a few rare cases in which the baseline provided by the WMFB method was not meaningful — notably for abrupt changes in baseline or when the FHR was almost not in the baseline state for

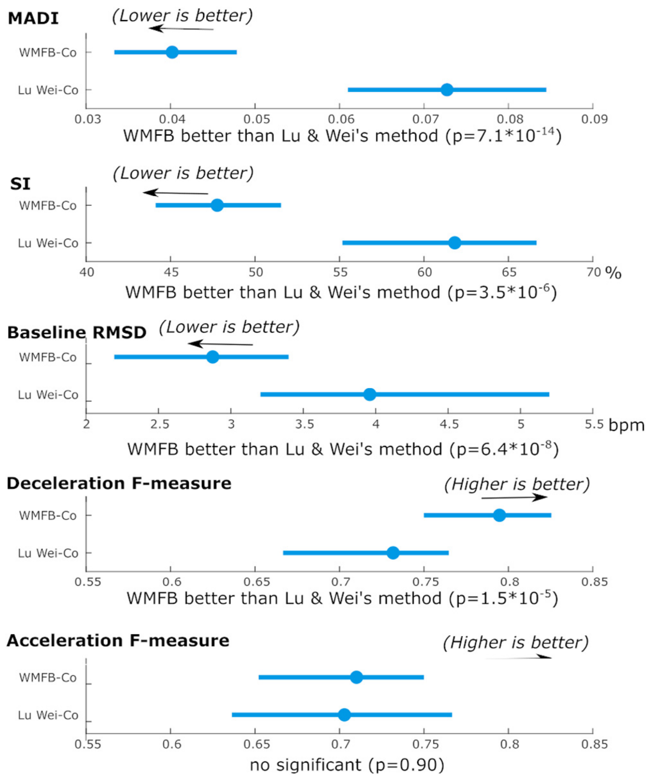


Fig. 9. Indices of the disparity between the two automated methods (the WMFB method and Lu and Wei's method) and the expert consensus analysis. The dot represents the median for the 90 recordings, and the line corresponds to the confidence interval. The p-values were calculated in a Wilcoxon test.

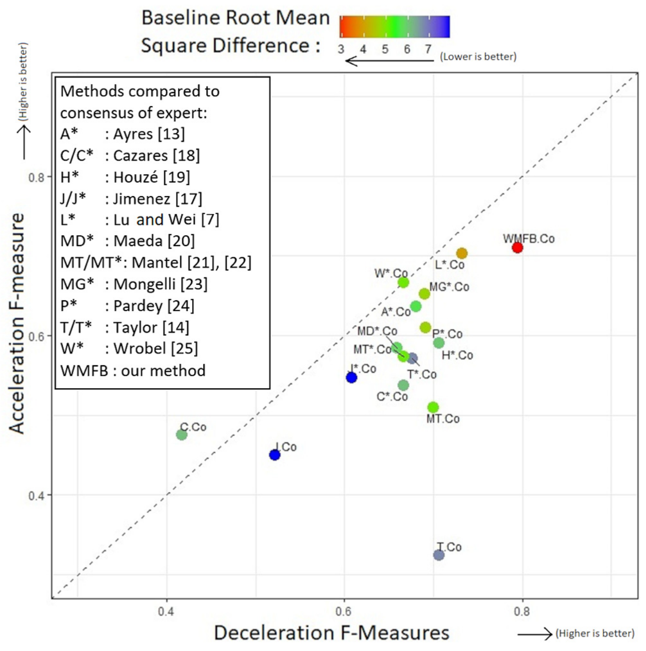


Fig. 10. A comparison of the WMFB method with 11 literature methods (from [6]) for three independent criteria (the deceleration F-measure, the acceleration F-measure, and the baseline RMSD). * indicates that the A/D are detected with a standard thresholding [6] method not directly described in the respective publications. The absence of * means that the A/D thresholding was performed using the method described in the corresponding publication.

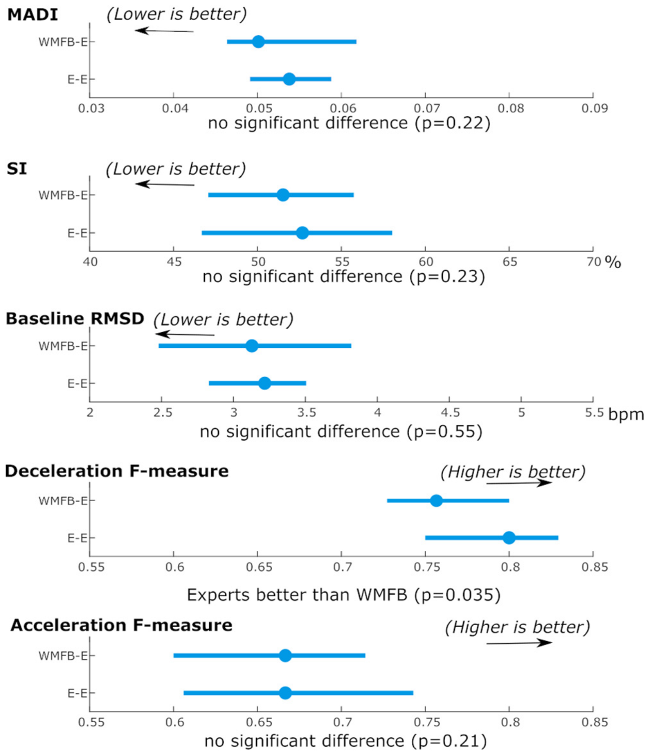


Fig. 11. Comparison of levels of WMFB-expert agreement (the median for each expert) and the inter-expert agreement E–E (the median for each pair of experts Ei–Ej), with regard to five criteria. The dot represents the median for the 90 recordings, and the line corresponds to the confidence interval. The p-values were calculated in a Wilcoxon test.

long periods (> 30 min). Although these cases were too rare to affect the statistical analysis, it appears that the WMFB method can still be slightly improved.

4.2. Discussion of the methodology used for evaluation

All the training and evaluation material used here has been available online (at http://utsb.univ-catholille.fr/fhr-review) since October 2018, so that researchers can try to further improve the performance of their automated morphologic analyses.

The evaluation did not incorporate the neonatal outcome as an efficacy criterion, since this was not a study objective. Firstly, calibrating an automated analysis method with regard to the neonatal outcome is very difficult because other influential factors are present during the delivery. Secondly, a trial probing the relationship with rare outcomes such as perinatal death and cerebral palsy (both with incidences of ~ 2 in 1000) would have to include around 80,000 patients [28, p. 59].

Before further research on predicting the neonatal outcome is performed, it would be also interesting to test the level of agreement with experts in terms of the FHR features used to classify the risk of acidosis: the type of deceleration (early, late, variable and prolonged), FHR variability over baseline periods, and the presence of episodes of tachycardia or bradycardia. Although it is usually assumed that these features can be directly deduced from the morphologic analysis, the use of fuzzy methods is needed for consistency with the experts.

4.3. Possible improvements

4.3.1. Real-time analysis

Just like a practitioner, the WMFB method considers information before and after the current time point. The WMFB method is capable of positioning a temporary baseline instantaneously (using the previous

20 min of signal only) but may then modify it for up to 20 min after the time point has first been analyzed. Although most of these changes will be irrelevant, a few may be major. Future research will have to assess differences between the analysis performed at the current time point and that performed 20 min later. The other literature methods generally use a shorter window time (often < 10 min) but this may be insufficient in some cases (e.g. Fig. 1).

In routine practice, practitioners are probably faced with the same restriction; they sometimes wait for 30 min to be sure of the baseline position after hesitating between pattern D and multiple decelerations (such as those shown in Fig. 1) [29].

4.3.2. Determination of the baseline's reliability

Some cases are ambiguous because a highly fluctuating FHR prevents the definition of a baseline. In some cases, a high number of A/D means that FHR is rarely in a baseline state; the baseline is then situated somewhere but its location cannot be determined. In other cases, there may not even be a baseline when damage to the autonomic nervous system (which regulates the FHR) has induced fetal acidosis — thus resulting in an unassignable baseline and an absence of variability [30,31].

To avoid erroneous interpretations, it is important to determine the baseline level of reliability. Georgieva et al. [32] suggested an index that measures the difficulty of baseline assignment. A non-assignable baseline is then considered to be pathologic. Nevertheless, one limitation of Georgieva et al.'s approach relates to its inability to distinguish between cases without a baseline and cases in which a baseline is present but difficult to determine (as can be observed in a non-pathologic FHR recording, such as that shown in Fig. 1).

4.3.3. Better preprocessing

The FHR recordings used for training and evaluation were selected for their low levels of signal loss, the low proportion of periods with an unreliable signal, and the low risk of confusion with the maternal heart rate signal. Under these conditions, the standard pre-processing used in the present study was sufficient. In routine practice, however, the recording quality can be poor, and preprocessing steps might not remove abnormal points. This can represent a major bias for classifying an FHR recording as normal, suspicious or pathologic [33]. In such a case, taking account of the maternal heart rate recording might improve the preprocessing [34].

4.3.4. Detection of contractions

The WMFB method can also be used to detect uterine contractions. The initial results are promising but require additional investigation. The same parameters can be used by applying a correction factor of 2 mmHg, instead of 1 bpm. It would be possible to improve the results by forcing the baseline to pass through the low values of the tocography signal and thus considering that any increases over this baseline must be uterine contractions.

5. Conclusion

Under the experimental conditions described here and in comparison with an expert consensus, the WMFB method appeared to perform as well as an expert. The WMFB method was evaluated as part of an open competition that we launched in an earlier publication [6,12] (see http://utsb.univ-catholille.fr/fhr-review). The method performed significantly better than 11 literature methods. The method Matlab source code is available under General Public Licence at [http://utsb.univ-catholille.fr/fhr-wmfb] to encourage independent evaluation and to ease its use for research purpose.

Our future research will address the systematic analysis of the FHR in the delivery suite, and the prediction of neonatal outcomes. Automated FHR analyses can quantify anomalies that are not easily visible (the deceleration area, short-term variability, etc.) and can be used to build predictive models [35,36]. We hope that better morphologic analysis will lead to better prediction of acidosis.

6. Ethical approval

All procedures involving human participants were performed in accordance with the ethical standards of the institutional and/or national review boards and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

This research complied with generally accepted scientific principles and medical research ethical standards, and was approved by the local institutional review board (*Comité Interne d'Ethique de la Recherche médicale*; Lille Catholic Hospitals, Lille, France; reference number: 2016-08-06). The study databases were registered with the French National Data Protection Commission. All study participants provided their informed consent.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Ayres-de Campos, S. Arulkumaran, FIGO consensus guidelines on intrapartum fetal monitoring: Introduction, Int. J. Gynecol. Obstet. 131 (1) (2015) 3–4, http://dx.doi.org/10.1016/j.ijgo.2015.06.017.

[2] A. Matias, P. Xavier, J. Bernardes, B. Patrıcio, Fetal heart-rate monitoring during maternal hypoglycaemic coma: A case report, Eur. J. Obstet. Gynecol. Reproduct. Biol. 79 (2) (1998) 223–225, http://dx.doi.org/10.1016/S0301-2115(98)00051-7.

[3] L. Sabiani, R. Le Dû, A. Loundou, C. d'Ercole, F. Bretelle, L. Boubli, X. Carcopino, Intra-and interobserver agreement among obstetric experts in court regarding the review of abnormal fetal heart rate tracings and obstetrical management, Am. J. Obstet. Gynecol. 213 (6) (2015) 856.e1–856.e8, http://dx.doi.org/10.1016/j.ajog.2015.08.066.

[4] G.S. Dawes, C.R. Houghton, C.W. Redman, Baseline in human fetal heart-rate records, Br. J. Obstet. Gynaecol. 89 (4) (1982) 270–275.

[5] A. Houzé de l'Aulnoit, S. Boudet, R. Demailly, L. Peyrodie, R. Beuscart, D. Houzé de l'Aulnoit, Baseline fetal heart rate analysis: eleven automatic methods versus expert consensus, IEEE, 2016, pp. 3576–3581, http://dx.doi.org/10.1109/EMBC.2016.7591501.

[6] A. Houzé de l'Aulnoit, S. Boudet, R. Demailly, A. Delgranche, M. Génin, L. Peyrodie, R. Beuscart, D. Houzé de l'Aulnoit, Automated fetal heart rate analysis for baseline determination and acceleration/deceleration detection: A comparison of 11 methods versus expert consensus, Biomed. Signal Process. Control 49 (2019) 113–123, http://dx.doi.org/10.1016/j.bspc.2018.10.002.

[7] Y. Lu, S. Wei, Nonlinear baseline estimation of FHR signal using empirical mode decomposition, in: Proc. of the IEEE 11th International Conference on Signal Processing, ICSP, vol. 3, IEEE, 2012, pp. 1645–1649, http://dx.doi.org/10.1109/ICoSP.2012.6491896.

[8] J. Jezewski, K. Horoba, D. Roj, J. Wrobel, T. Kupka, A. Matonia, Evaluating the fetal heart rate baseline estimation algorithms by their influence on detection of clinically important patterns, Biocybern. Biomed. Eng. 36 (4) (2016) 562–573, http://dx.doi.org/10.1016/j.bbe.2016.06.003.

[9] I. Nunes, D. Ayres-de Campos, C. Figueiredo, J. Bernardes, An overview of central fetal monitoring systems in labour, J. Perinat. Med. 41 (1) (2013) http://dx.doi.org/10.1515/jpm-2012-0067.

[10] S. Boudet, A. Houzé de l'Aulnoit, R. Demailly, A. Delgranche, L. Peyrodie, R. Beuscart, D. Houzé de l'Aulnoit, A fetal heart rate morphological analysis toolbox for MATLAB, 2019, http://dx.doi.org/10.20944/preprints201906.0139.v1, Preprints (2019060139).

[11] A. Houzé de l'Aulnoit, S. Boudet, M. Génin, P.-F. Gautier, J. Schiro, D. Houzé de l'Aulnoit, R. Beuscart, Development of a smart mobile data module for fetal monitoring in e-healthcare, J. Med. Syst. 42 (5) (2018) 83, http://dx.doi.org/10.1007/s10916-018-0938-1.

[12] S. Boudet, A. Houzé de l'Aulnoit, R. Demailly, A. Delgranche, L. Peyrodie, R. Beuscart, D. Houzé de l'Aulnoit, Fetal heart rate signal dataset for training morphological analysis methods and evaluating them against an expert consensus, 2019, Preprints (2019060139), http://dx.doi.org/10.20944/preprints201907.0039.v1.

[13] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de Sá, L. Pereira-Leite, SisPorto 2.0: a program for automated analysis of cardiotocograms, J. Matern. Fetal Med. 9 (5) (2000) 311–318, http://dx.doi.org/10.1002/1520-6661(200009/10)9:5<311::AID-MFM12>3.0.CO;2-9.

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

[15] P. Fuentealba, A. Illanes, F. Ortmeier, Analysis of the foetal heart rate in cardiotocographic recordings through a progressive characterization of decelerations, Curr. Direct. Biomed. Eng. 3 (2) (2017) http://dx.doi.org/10.1515/cdbme-2017-0089.

[16] J.O. Smith, Mathematics of the Discrete Fourier Transform (DFT): With Audio Applications, BookSurge, North Charleston, 2010, OCLC: 837679883.

[17] L. Jimenez, R. Gonzalez, M. Gaitan, S. Carrasco, C. Vargas, Computerized algorithm for baseline estimation of fetal heart rate, in: Computers in Cardiology, IEEE, 2002, pp. 477–480, http://dx.doi.org/10.1109/CIC.2002.1166813.

[18] S.M. Cazares, Automated Identification of Abnormal Patterns in the Intrapartum Cardiotocogram, (Ph.D. thesis), University of Oxford, Oxford, 2002.

[19] D. Houzé de l'Aulnoit, R. Beuscart, G. Brabant, L. Corette, M. Delcroix, Real-time analysis of the fetal heart rate, in: Engineering in Medicine and Biology Society, 1990. Proceedings of the Twelfth Annual International Conference of the IEEE, IEEE, 1990, pp. 1994–1995, http://dx.doi.org/10.1109/IEMBS.1990.692125.

[20] K. Maeda, M. Utsu, Y. Noguchi, F. Matsumoto, T. Nagasawa, Central computerized automatic fetal heart rate diagnosis with a rapid and direct alarm system, Open Med. Dev. J. 4 (2012) 28–33, http://dx.doi.org/10.2174/1875181401204010028.

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

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

[23] M. Mongelli, R. Dawkins, T. Chung, D. Sahota, J.A. Spencer, A.M. Chang, Computerised estimation of the baseline fetal heart rate in labour: the low frequency line, Br. J. Obstet. Gynaecol. 104 (10) (1997) 1128–1133.

[24] J. Pardey, M. Moulden, C.W.G. Redman, A computer system for the numerical analysis of nonstress tests, Am. J. Obstet. Gynecol. 186 (5) (2002) 1095–1103, http://dx.doi.org/10.1067/mob.2002.122447.

[25] J. Wróbel, K. Horoba, T. Pander, J. Jeżewski, R. Czabański, Improving fetal heart rate signal interpretation by application of myriad filtering, Biocybern. Biomed. Eng. 33 (4) (2013) 211–221, http://dx.doi.org/10.1016/j.bbe.2013.09.004.

[26] J. Kottner, L. Audigé, S. Brorson, A. Donner, B.J. Gajewski, A. Hróbjartsson, C. Roberts, M. Shoukri, D.L. Streiner, Guidelines for reporting reliability and agreement studies (GRRAS) were proposed, J. Clin. Epidemiol. 64 (1) (2011) 96–106, http://dx.doi.org/10.1016/j.jclinepi.2010.03.002.

[27] R.D. Keith, K.R. Greene, Development, evaluation and validation of an intelligent system for the management of labour, Baillieres Clin. Obstet. Gynaecol. 8 (3) (1994) 583–605.

[28] E. Chandraharan (Ed.), Handbook of CTG Interpretation: From Patterns to Physiology, Cambridge University Press, Cambridge, United Kingdom, New York, 2017.

[29] E.J. Mulder, G.H. Visser, D.J. Bekedam, H.F. Prechtl, Emergence of behavioural states in fetuses of type-1-diabetic women, Early Hum. Dev. 15 (4) (1987) 231–251.

[30] A. Ugwumadu, Are we (mis)guided by current guidelines on intrapartum fetal heart rate monitoring? Case for a more physiological approach to interpretation, BJOG Int. J. Obstet. Gynaecol. 121 (9) (2014) 1063–1070, http://dx.doi.org/10.1111/1471-0528.12900.

[31] A. Ugwumadu, Understanding cardiotocographic patterns associated with intrapartum fetal hypoxia and neurologic injury, Best Pract. Res. Clin. Obstet. Gynaecol. 27 (4) (2013) 509–536, http://dx.doi.org/10.1016/j.bpobgyn.2013.04.002.

[32] A. Georgieva, S.J. Payne, M. Moulden, C.W.G. Redman, Relation of fetal heart rate signals with unassignable baseline to poor neonatal state at birth, Med. Biol. Eng. Comput. 50 (7) (2012) 717–725, http://dx.doi.org/10.1007/s11517-012-0923-7.

[33] P. Pinto, C. Costa-Santos, H. Gonçalves, D. Ayres-De-Campos, J. Bernardes, Improvements in fetal heart rate analysis by the removal of maternal-fetal heart rate ambiguities, BMC Pregnancy Childbirth 15 (1) (2015) http://dx.doi.org/10.1186/s12884-015-0739-1.

[34] R. Nurani, E. Chandraharan, V. Lowe, A. Ugwumadu, S. Arulkumaran, Misidentification of maternal heart rate as fetal on cardiotocography during the second stage of labor: the role of the fetal electrocardiograph: Erroneous recording of maternal heart rate, Acta Obstet. Gynecol. Scand. 91 (12) (2012) 1428–1432, http://dx.doi.org/10.1111/j.1600-0412.2012.01511.x.

[35] J. Spilka, V. Chudáček, M. Koucký, L. Lhotská, M. Huptych, P. Janků, G. Georgoulas, C. Stylios, Using nonlinear features for fetal heart rate classification, Biomed. Signal Process. Control 7 (4) (2012) 350–357, http://dx.doi.org/10.1016/j.bspc.2011.06.008, URL http://www.sciencedirect.com/science/article/pii/S1746809411000619.

[36] M. Lichman, UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Science, 2013, https://archive.ics.uci.edu/ml/datasets/cardiotocography.