

Comparison of Doppler Ultrasound and Direct Electrocardiography Acquisition Techniques for Quantification of Fetal Heart Rate Variability

Janusz Jezewski*, *Member, IEEE*, Janusz Wrobel, and Krzysztof Horoba

Abstract—A method for comparison of two acquisition techniques that are applied in clinical practice to provide information on fetal condition is presented. The aim of this work was to evaluate the commonly used Doppler ultrasound technique for monitoring of mechanical activity of fetal heart. Accuracy of beat-to-beat interval determination together with its influence on indices describing the fetal heart rate (FHR) variability calculated automatically using computer-aided fetal monitoring system were examined. We considered the direct fetal electrocardiography as a reference technique because it ensures the highest possible accuracy of heart interval measurement, and additionally all the definitions of popular time domain parameters quantifying FHR variability formerly have been created using the fetal electrocardiogram. We evaluated the reliability of various so called short-term and long-term variability indices, when they are calculated automatically using the signal obtained via the Doppler US from a fetal monitor. The results proved that evaluation of the acquisition technique influence on fetal well-being assessment can not be accomplished basing on direct measurements of heartbeats only. The more relevant is the estimation of accuracy of the variability indices, since analysis of their changes can significantly increase predictability of fetal distress.

Index Terms—Doppler ultrasound, fetal electrocardiogram, fetal heart rate, fetal monitoring.

I. INTRODUCTION

PRESENT-DAY medicine is characterized by development of new measurement methods which should be more efficient, less invasive and in case of long lasting monitoring less annoying for a patient. Usually, more invasive method ensures higher measurement accuracy because it records in a direct way signals emitted by the investigated object. Therefore, before applying every new method its measurement accuracy has to be evaluated in relation to the method assumed as a reference one. It is very important because this probable lower accuracy can affect clinical assessment of the process. Such an evolution process of measurement methods from invasive to noninvasive approach can be observed in biophysical fetal monitoring

Manuscript received February 14, 2005; revised October 9, 2005. This work was supported in part by the State Committee for Scientific Research (resources 2004–2006) under Research Project 3 T11E 01726. Asterisk indicates corresponding author.

*J. Jezewski is with the Department of Biomedical Informatics, Institute of Medical Technology and Equipment, Roosevelt Str 118, Zabrze 41800, Poland (e-mail: jezewski@itam.zabrze.pl).

J. Wrobel and K. Horoba are with the Department of Biomedical Informatics, Institute of Medical Technology and Equipment, Zabrze 41800, Poland (e-mail: jwrobel@itam.zabrze.pl; kris@itam.zabrze.pl).

Digital Object Identifier 10.1109/TBME.2005.863945

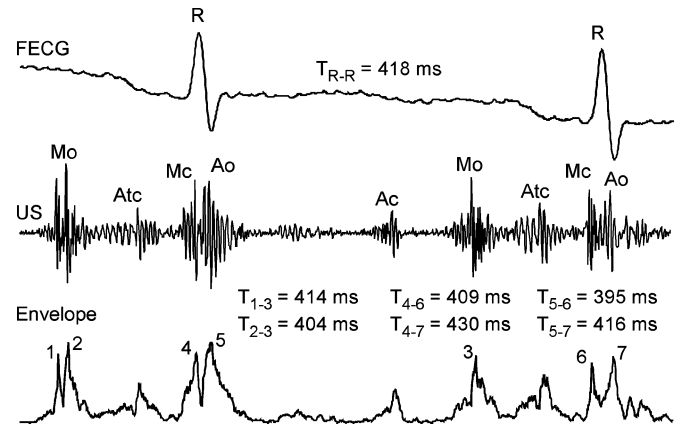


Fig. 1. Determination of TRR interval from the Doppler US envelope using peak detection method. Various durations of the cardiac cycle can be obtained depending on which event is selected as representing a given cycle. Cardiac cycle events: atrial wall contraction—Atc, mitral valve opening and closure—Mo and Mc, and aortic valve opening and closure—Ao and Ac.

which relies on analysis of the fetal heart activity. Recording of fetal electrocardiogram during labor by means of an electrode attached to the fetus head was performed already in 1960s [1]. The noninvasive ultrasound (US) method has become the standard approach since early 1970s because it is possible to use both during pregnancy and labor.

Operation principle of these methods is based on measurement of duration of fetal cardiac cycles. In fetal electrocardiography (FECG), this duration is represented by T_{RR} interval between two consecutive R-waves. The signal of a good quality allows detection of R-waves using quite simple algorithms based on peak detection. In Doppler US technique, heartbeats are detected from the envelope of US wave reflected from moving parts of fetal heart—valves or walls [2]. Peak detection can provide incorrect data due to complex and unstable shape of the envelope signal (Fig. 1). Therefore, for the detection of consecutive heartbeats the correlation techniques considering full shape of the analyzed signal are applied. Autocorrelation with adaptive window selection or cross-correlation with changeable template are mostly used [3], [4]. Distance between two consecutive peaks of autocorrelation function corresponds to the interval between two consecutive R-waves in electrocardiogram. Values of T_{RR} intervals are transformed into instantaneous fetal heart rate (FHR) expressed in beats/min (bpm) accordingly to the equation: $FHR [bpm] = 60000/TRR [ms]$. Such data set creates a signal of fetal heart activity, which as printed waveform is visually analyzed by a clinician. This

classical interpretation of FHR signal comprises recognition of baseline as well as acceleration and deceleration episodes [5]. Since the FHR signal is determined by a simple recalculation of T_{RR} intervals, the terms: instantaneous FHR value and T_{RR} interval are used interchangeably in this paper.

Preliminary evaluation of both measurement techniques lets us to conclude that the US method ensures lower measurement accuracy of cardiac cycle duration than FECG does. Using US we can only approximately determine R-wave locations in time, because heart movements are only responses to the primary electrical excitation of the heart [2]. The simplest approach to the evaluation of US method accuracy can be carried out by direct comparison of corresponding intervals, simultaneously obtained from US signal and fetal electrocardiogram. Only fetal electrocardiogram can be used to deliver reference data. If the differences can not be noticed with the naked eye then we can assume that influence of the US method on classical assessment of FHR is not significant. However when automated analysis of FHR signal is used the question still arises if direct comparison really estimates the influence of the US method on fetal condition assessment.

Complexity of FHR signal causes that a large part of the information still remains hidden. It was found that the evaluation of fetal well-being based on visual interpretation only is characterized by low inter- and intraobserver agreement [6]. One of the most important signs of normal FHR patterns is continuous fluctuation in time of beat-to-beat intervals. This is called short-term variability. These changes are considered to be the most important FHR characteristics reflecting appropriate neurological modulation of the FHR [7]. Due to a certain periodicity in the direction and magnitude of these changes, the values of FHR are distributed around its mean level. These changes are called long-term variability and described by two parameters: amplitude and frequency. There are various time domain parameters (mathematical indices) used for quantitative evaluation of both types of FHR variability (Table II). Definitions of these indices have been created basing on T_{RR} intervals precisely determined from direct fetal electrocardiogram. However, at that time a technology level made impossible the calculation of variability indices by a built-in procedure of bedside monitor, and an external computer was required. In the next years, Doppler US method became a standard approach to fetal monitoring, but even then the monitor functions were limited only to the recording and printing of the FHR signal. Widespread application of computer-aided systems acquiring signals from bedside monitors enabled the on-line automated analysis of FHR [8]. However, terminology for quantitative evaluation of the FHR variability, originally proposed for description of direct FECG signal, has been applied without adaptation for US channel [9].

Autocorrelation technique used in Doppler US method causes averaging of neighboring cardiac cycles durations and, therefore, it can not follow fast changes of FHR signal. Comparing these consecutive values with the reference T_{RR} intervals determined from FECG, we can note that the error of interval determination is not random, but depends on the characteristics of heart rate changes [10]. Therefore, in some epochs when the heart rate raises rapidly, the error takes only positive values, whereas during rapid slowing the negative values only. Distribu-

tion of FHR measurement error has no influence on the results of direct comparison of signals obtained using the descriptive statistics. However, we should assume that this distribution has significant influence on the values of variability indices since they are defined on a basis of absolute difference between neighboring T_{RR} intervals. Consequently, mean value of determination error of a given index calculated within the analyzed trace fragment will not depend directly on mean value of interval error, but rather on instantaneous distribution of interval error in this fragment. It seems necessary to estimate the influence of US method on the correct clinical assessment of fetal condition, assuming that at present this process is computer-aided and accomplished mainly on a basis of instantaneous FHR variability indices calculated automatically [8]. Since definitions of indices are based on T_{RR} intervals, there is a question whether the influence of US method can be evaluated in relation to inaccuracy of directly determined T_{RR} intervals only, or using more complex procedure comprising evaluation of the inaccuracy of popular variability indices.

Several attempts to analysis of the accuracy of the FHR signal acquired by means of Doppler US method have been made. Reference signals were obtained from FECG channel of a bedside monitor [11]–[13]. Modern monitors, which principle of operation is based on US technique, sometimes enable also the recording of electrocardiogram from fetal head. The main task of the applied beat detection procedure is not to achieve maximum accuracy of intervals but to minimize the number of missed or incorrect beats. In this way, the continuity of monitoring is ensured that is very important for visual interpretation of trace. Since FHR values have to be calculated and presented on-line, low sampling frequency of FECG is used to reduce the computational time. Therefore, both the T_{RR} intervals and FHR variability indices calculated using such FECG signal can not be considered as reference values. Evaluation of variability indices errors requires accuracy of measurement of reference T_{RR} intervals not lower than 1 ms, which is ensured by a sampling of the FECG signal with the frequency of 2 kHz [14]. Comparison procedures carried out in previous works did not concern durations of corresponding cycles, but only FHR values averaged over 3.75 s periods [15] or mean values determined for the whole recording [16]. The influence of acquisition method on variability indices was based on mean values determined for segments [16] or for the whole trace [17]. Evaluation of correlation between errors of interval determination and indices errors was not carried out, although it seems to be the most important in our opinion.

The aim of our study was to evaluate the accuracy of the Doppler US method in relation to the FECG by comparison of the primary measurement data— T_{RR} intervals, as well as through comparison of selected parameters of quantitative description of the signal—FHR variability indices. We checked if comparison of T_{RR} intervals is satisfying for determination of the accuracy of the analyzed method and how the error of interval measurements influences the error of FHR variability indices calculated on these intervals. That required solving some essential problems: simultaneous independent monitoring of the same process by means of two different methods, synchronization of measurement values for the same events, elimination of

suspicious values and finally establishing common representation of acquired data in time domain.

II. METHODOLOGY

A. Instrumentation

Measurement station has been based on a laptop PC with the DAQCard-AI-16XE (National Instruments) data acquisition card. This card has eight analog inputs and analog-to-digital (A/D) converter which can operate with the maximum sampling frequency of 200 kHz. Battery power supply and patient's electrical barrier ensures full standard safety for a patient, and minimizes power line interferences. All procedures for acquisition and processing of the signals have been developed in LabView 6.1 environment (National Instruments). Fetal heart rate signals were recorded using the MT-430 (Toitu, Japan) fetal monitor. The monitor is equipped with US transducer which periodically emits (with repetition frequency of 3 kHz) 1 MHz US wave of a very low power <1.5 mW/cm². The wave reflected from moving parts of fetal heart (walls or valves) returns to the transducer, which operates as a receiver between sending successive waves. Frequency shift between emitted and reflected waves is caused by the Doppler effect and provides information on the speed of moving object on which the US beam is focused. Electrical signal from the transducer after amplification and demodulation is used for the detection of heartbeats. At the same time, Doppler signal related to heart movements and contained in the audio frequency range (from 0.2 to 1 kHz) is fed to the speaker, which helps in correct positioning of transducer on maternal abdomen.

Analysis of Doppler envelope is difficult due to a complex structure of the signal comprising components originating from particular events of the cardiac cycle. Additionally, the shape of envelope changes from beat to beat. Amplitude-based detection methods are useless because they may detect events that do not correspond to each other in consecutive cardiac cycles. This consequently leads to incorrect interval determination (Fig. 1). Only correlation techniques are relevant to determine periodicity in Doppler envelope, because they are aimed at recognition of shape similarity. Autocorrelation technique has been used in the MT-430 fetal monitor. Unfortunately, this function tends to average the durations of successive cycles, whereas T_{RR} intervals as a rule are subject to slight changes. Too wide time window, in which autocorrelation function is calculated, causes too strong averaging of T_{RR} intervals and, therefore, a considerable loss of the information on FHR variability. On the other hand, too narrow window decreases the accuracy of interval determination as well as the immunity to interferences. Usually, the window length of 2 or 3 heart cycles is used. The autocorrelation continuously provides information on signal periodicity, e.g., with every window shift a new function value is calculated, whereas in fact the successive interval should be determined only once with every new event—detection of a new heartbeat. Therefore, additional method being able to recognize new heartbeat event has to be used [10], for example, a method based on detailed analysis of the dominant peak of autocorrelation function (Fig. 2).

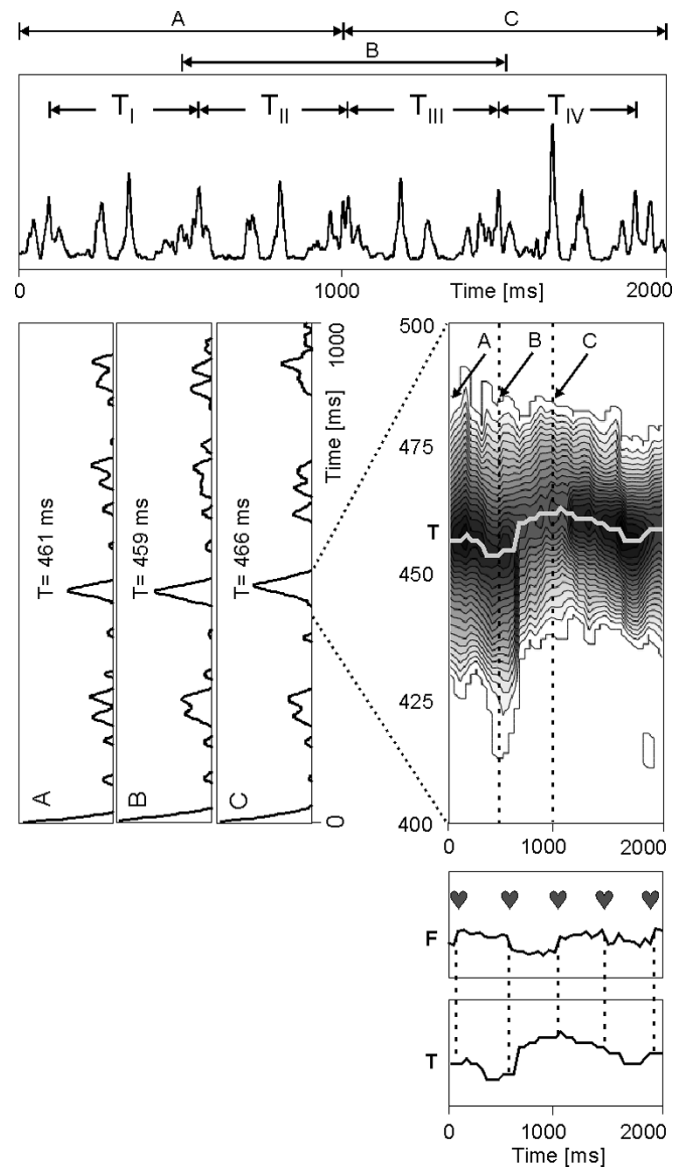


Fig. 2. Simplified procedure of a new heartbeat detection based on autocorrelation function. A fragment of Doppler envelope signal includes four cardiac cycles ($T_I \div T_{IV}$). Below the envelope three examples of autocorrelation function are plotted for successive analyzed windows. Function A includes T_I and T_{II} intervals between heartbeats, function B— T_{II} and T_{III} , whereas function C— T_{III} and T_{IV} . Changes of location and amplitude of autocorrelation function peaks are visualized with a layer graph, which is created by a set of functions determined every 10 ms (the dominant autocorrelation peak located between 400 and 500 ms). Horizontal axis is the time axis, vertical one represents the values of intervals determined in the autocorrelation window, and grey color intensity illustrates a change of amplitude. Below the layer graph there are two plots showing the change of amplitude of the dominant peak—F, as well as the peak movement trajectory—T. Once the abrupt change of peak amplitude is detected, that corresponds to a new heartbeat, a new value of interval is determined as a median value from temporary values which have been determined since the previous beat detection.

Consecutive cardiac cycle lengths expressed in milliseconds are transformed into instantaneous FHR values expressed in bpm. These values are displayed and printed as FHR waveform on a thermal paper. Since they are accessible on the fetal monitor output, they can be used in a computer-aided fetal monitoring system. Usually a serial output is used for sending digital data, however there are various communication protocols. The

FHR samples of 1 bpm resolution can be read out from MT-430 monitor. This corresponds to about 4 ms resolution of T_{RR} determination for a typical value of FHR equaled to 135 bpm, but it is unsatisfactory for analysis of the FHR variability. In addition, the monitor delivers FHR value with resolution of 0.25 bpm in a form of a voltage signal through the analog output. The voltage change occurs with every new heartbeat detected and provides information on current (and more precisely previous) T_{RR} interval. In our measurement station, the voltage signal was fed to the analog input of the data acquisition card, and then after A/D conversion it was used to reconstruct T_{RR} intervals determined in the fetal monitor. Measurement characteristics of the US channel was evaluated by the following experiment. Likewise [18], to simulate fetal heart mechanical action we used a waterproof speaker placed in a container filled with water. US transducer attached to a latex plate was put on the surface of water. Water column height controlling the distance between the transducer and the speaker was set to the mean distance between the fetal heart and maternal abdomen. The speaker was driven by a square-wave generator providing frequency varying in the allowed range—from 50 to 210 bpm (with step of 10 bpm). In that way, we tested the linearity of the measurement characteristics and then calculated its slope and intercept.

Direct fetal electrocardiogram was recorded via spiral electrode (Cetro 15133C, Sweden). Raw signal was fed to the signal conditioning circuit, and then after preliminary amplification and filtration, to the analog input of the data acquisition board. We applied the amplifier with a selectable gain between 500 and 2000 V/V [19]. The use of a band-pass filter at 0.05 Hz ensured suppression of low-frequency noise components and elimination of an isoline drift. At the same time, its upper frequency at 300 Hz overcame the aliasing problem. Electrode attached to patient's hip was used in the active ground circuit to reduce common mode interferences mostly coming from the power line. Acquisition of FECG signal, whose quality depends mainly on correct attachment of the spiral electrode, was on-line controlled on a computer screen. Recorded analog signals were sampled at 2 kHz, which ensured the required accuracy for reference T_{RR} intervals.

B. Reference Data

Determination of the T_{RR} intervals in the reference signal has the essential influence on a whole procedure of the US method evaluation. Therefore, we have decided to use two independent methods in order to eliminate every suspicious interval, when results differs more than 1 ms. We chose one method based on peak detection and the second based on correlation function. Quite different principles of operation let us to assume, that the possible interferences would have different influence on intervals values determined by the chosen methods. Since we were not limited by processing time, we engaged a human expert to control the detection parameters in order to determinate the T_{RR} intervals with the highest accuracy. Thresholds values for peak detection as well as the template QRS complex for correlation method were selected by the expert for every successive 1-min segment. The second-order interpolation used for detection of function peaks in both methods (Peak Detector.vi pro-

cedure from LabView) ensured the accuracy of reference T_{RR} not lower than 0.5 ms.

Peak detection method was based on determination of the first derivative of FECG signal and finding its local maxima corresponding to QR-waves. The zero-crossing of derivative function occurring just after the maximum relates to R-wave. The local maximum of the derivative function was detected on a basis of two conditions: amplitude and distance. The amplitude threshold has been set for every 1-min segment basing on the peak amplitude distribution in the analyzed period. The next peak was expected in the distance from the previous peak corresponding to physiological values (from 300 to 1200 ms). Moreover, a change of distance between consecutive peaks had to be below 10%. If for a given amplitude threshold the number of intervals not matching the distance condition was too high then the distance threshold was increased. If artifact occurred within a given QRS complex, this complex was discarded from further analysis. Very crucial was to check whether the first local maximum really corresponded to the first QRS complex or to an artifact, because only in the first case the distance threshold can be applied.

The second method for QRS detection was based on the cross-correlation function. This function was used to find QRS complexes in FECG signal by comparing the signal with a template. The template was a representative and undistorted QRS complex selected by the expert for every 1-min segment of FECG signal. Since R-wave is the dominating component of the QRS complex, the cross-correlation peak corresponds to the matching of R-waves in two complexes being compared. Therefore, a distance between two consecutive peaks is the T_{RR} interval. The amplitude and distance thresholds were also used for validation of cross-correlation peaks. Because cross-correlation function provides values normalized in the range $\langle -1, 1 \rangle$, the amplitude threshold at 0.7 has appeared to be optimal for most of periods. The distance threshold was applied in the same way like for the peak detection method. When interferences occurred within a given QRS complex that caused cross-correlation peak to be too low, such complex was manually excluded from further analysis, and consequently two neighboring T_{RR} intervals were not calculated. The final stage was the intervals verification. Only this interval whose lengths determined by two detection methods differing by less than 1 ms was included to the reference signal.

C. Signal Synchronization

Reliable comparison requires the synchronization of signals, so as to ensure that just the corresponding cardiac cycles measured by two investigated methods are compared. The Doppler US measurement channel of fetal monitor due to its high complexity introduces a time delay of the FHR signal recorded. Because this delay is important the same fetal heart activity simulator was used for its estimation. Considering quasiperiodicity of the FHR the speaker was driven with a square wave consisting of five cycles which lengths were increasing from 420 to 500 ms, and next five cycles which lengths were decreasing in the same range. The pulses driving the speaker were also sent to an analog input of FECG recording unit. Then the two FHR signals: one measured by the US transducer and the other obtained from the

FECG were visually synchronized on the computer screen and the time shift of 600 ms was determined between them.

D. Final Comparison

For comparison of methods using variability indices, very important was to put markers if at least one of the corresponding intervals has been missed in one of the signals being compared. Thanks to that, for both the FHR signals a given index was calculated using the same number of intervals in every 1-min segment. In order to unify the comparison procedures, the definitions for indices have been a little modified due to the 1-min length of the segments. Indices values for both methods were determined and then their relative errors were calculated for US method where the reference values were obtained from FECG signal. The mean relative error and its standard deviation for each index were calculated over error values from all 1-min segments. The last step was the testing correlation between the error of a given index and the interval error. For this step, the partial errors of interval determination were calculated as mean value of interval differences in 1-min segments.

III. MATERIAL

Simultaneous recording of direct FECG and FHR via US could be carried out only during labor. Ethics committee approved all procedures and informed consent was obtained from every woman. Total monitoring time was 242 min. However, during the labor a scalp electrode often lost contact due to the movements of both the mother and fetus, which caused strong interferences or even complete loss of FECG. Finally, 185 min of recording had a satisfactory quality for further processing. In the FHR signal obtained from US channel only temporary signal loss of few beats length was noted. The signal loss together with temporary gaps in FECG were analyzed automatically during signal processing. After the reference FHR determination, we have obtained 24 690 T_{RR} intervals in the recorded FECG signal. In result of its verification, 815 intervals (3.3%) were considered as suspicious, and next 1934 (8.1%) intervals were excluded due to signal loss present in the signal simultaneously recorded via US. Finally, 21 941 pairs of T_{RR} intervals determined for both methods were used for direct comparison. Discarded intervals were replaced by signal-loss markers. For analysis of measurement channel influence on variability indices, the traces have been arbitrary divided into 1-min segments. Limit for signal-loss for the segment has been established at 20% of the number of intervals. That means the segments were excluded whose summary length of all T_{RR} intervals previously marked as lost exceeded 12 s. Assuming that the average length of interval equals from 400 to 500 ms, then 20% threshold corresponds to the range of 20 to 24 lost beats. Because the lost intervals were not regularly distributed in 1-min segments, only 12 of 185 segments were discarded. Finally, 173 1-min segments were used for further processing. Fig. 3 presents successive stages of evaluation of US method accuracy with respect to T_{RR} interval determination and with respect to the calculation of FHR variability indices.

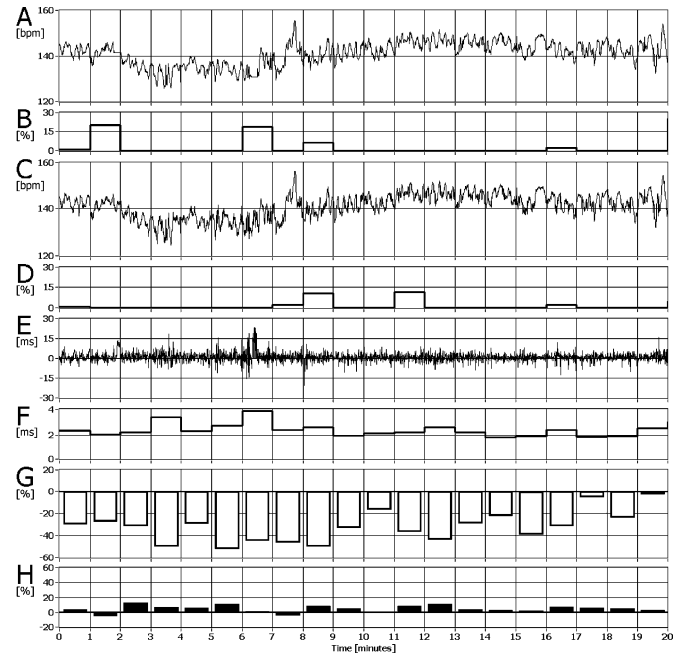


Fig. 3. Successive stages of procedure for evaluation of the US method accuracy with respect to T_{RR} interval determination as well as to the calculation of FHR variability indices. The first trace (A) shows a fragment of FHR signal created from the T_{RR} intervals obtained from electrocardiography as well as the trace C but from the US method (interval values were recalculated into values expressed in bpm). Plots B and D, respectively, let us evaluate a percentage of signal loss in consecutive 1-min segments. Waveform E shows that interval differences are quite equally distributed but at the same time their dispersion is rather large. The partial interval errors calculated over 1-min segments as a mean value of the absolute interval differences in these segments are presented by plot F. The last two plots (G and H) illustrate values of relative errors of de Hann short-term and long-term indices, respectively.

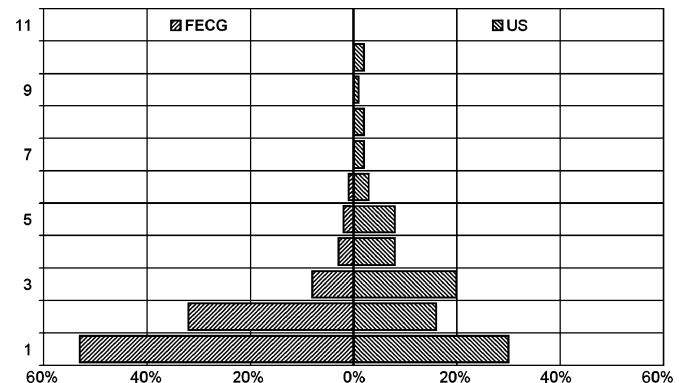


Fig. 4. Relative frequency distributions of signal loss episodes for both methods examined. Classes referring to the duration of episodes are expressed in lost beats and listed along the vertical axis (US method).

IV. RESULTS

Quantity of FHR signal loss episodes and especially their distribution within 1-min segments have essential influence on reliability of the indices describing FHR variability. Fig. 4 presents relative frequency distribution of signal loss episodes for both methods. Very short signal loss episodes were dominating in the reference signal, 85% of episodes did not exceed 2 beats (53% of them—one beat). On the other hand, the duration of these episodes in FHR from the US method differed more, and 85% of episodes lasted from 1 to 5 beats (30% of them—one beat). The influence of lost beats on the quality of the recorded FHR signals can be evaluated with the use of graph showing the percentage of lost beats in episodes of particular durations (Fig. 5).

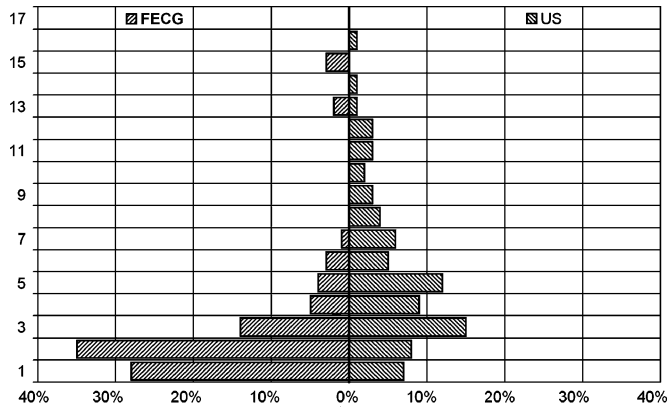


Fig. 5. The influence of the duration of signal loss episode on the overall number of the lost beats in recorded signals (US method). Classes defined as in Fig. 4.

Domination of very short signal loss episodes was confirmed for electrical approach—about 25% of lost beats belonged to the episodes of one-beat duration and over 35% to two-beat duration. Considering the US approach, the largest percentage of lost beats was observed in episodes of the duration from 3 to 5 beats. Obtained results confirmed our previous considerations on the US method reliability.

A lot of very short signal loss episodes in the FHR signal obtained from FECCG means that only short artifacts occur, which distort only one QRS complex. At the same time, the probability of artifacts occurrence in several consecutive QRS complexes, which could cause longer signal loss, is very low. Longer signal loss episodes occurred very rarely, mostly when electrodes lost contact with the fetus as a result of mother's movement. During US monitoring, long signal loss episodes (from 8 to 12 beats) were caused mainly by fetal movements. Shorter signal loss episodes (from 3 to 5 beats) seem to be a result of the algorithms for T_{RR} interval verification built in fetal monitors. For example, the interval verification could be based on testing of three consecutive intervals with respect to the established limit for the length change. This limit can be calculated as a fraction of the average from the last several intervals. This approach is based on the assumption that the fetal heart can not accelerate too fast and then decelerate at once, so every abrupt change must be a result of interference. As a consequence, if one interval is incorrect than the two consecutive intervals will be also discarded. Using this verification procedure the number of signal loss episodes of 3 to 5 beats length increases. However, the level of FHR signal loss achieved during our experiment was satisfactory taking into account the fact that in everyday monitoring it is usually much higher.

The descriptive statistics concerning direct comparison of T_{RR} intervals is shown in Table I. The presented values were calculated using T_{RR} intervals expressed in ms and instantaneous FHR values expressed in bpm. The distribution of interval differences was not significantly different from normal distribution (Shapiro-Wilk test, Statistica 6.0 StatSoft). The standard deviation showed that only 5% of differences exceeded 8.37 ms, which is a quite low value in relation to the mean duration of the T_{RR} interval (from 400 to 500 ms). The absolute error of T_{RR} interval measurement is very crucial for

TABLE I
DESCRIPTIVE STATISTICS OF THE T_{RR} INTERVAL
MEASUREMENT ERROR OF THE DOPPLER US

Term	Description	Value in ms	Value in bpm
$\overline{\Delta T_{RR}}$	Mean value of interval differences	0.42	0.13
$SD_{\Delta T_{RR}}$	Standard deviation	4.18	1.25
$2 \cdot SD_{\Delta T_{RR}}$	Double SD	8.37	2.51
$ \overline{\Delta T_{RR}} $	Mean value of absolute interval error	2.98	0.89
Med. $ \Delta T_{RR} $	Median value of absolute interval error	2.17	0.65

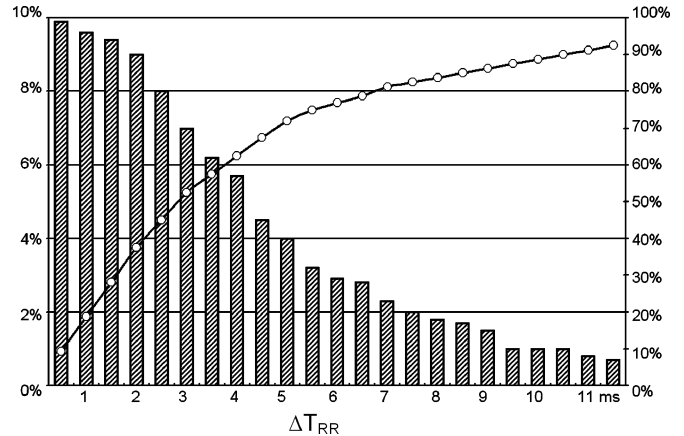


Fig. 6. Absolute error ΔT_{RR} of the measurement of T_{RR} intervals using the US method in relation to the reference FECCG: frequency distribution (bar plot with left vertical axis), cumulative histogram (line plot with right vertical axis).

the visual interpretation of FHR trace. It was defined as the absolute difference between two corresponding T_{RR} intervals transformed into values expressed in bpm. In our study, this mean value reached 0.89 bpm. The median value obtained shows that 50% of errors remains below 0.65 bpm.

It can be noticed from histograms (Fig. 6) that for 60% of intervals the absolute error is below 1 bpm, and only for 25% it exceeds 2 bpm. Taking into account that the line thickness on a strip chart paper corresponds to 0.5 bpm, such differences are difficult to observe with the naked eye. The mean value of relative error of the T_{RR} intervals was equal to 0.8%, which is below 1%—the limit recommended in [14]. The results obtained reveal that accuracy provided by Doppler US technique is satisfactory for visual analysis of the FHR trace.

Detailed evaluation of the influence of the US method on accuracy of the FHR variability indices were carried out by the comparison of their values with the reference data obtained from the analysis of fetal electrocardiogram. For both methods, 13 different indices describing instantaneous FHR variability were determined for all the 173 1-min segments. Mean values as well as their standard errors for all the indices are presented in Table II.

Considering the indices of short-term variability (the first letter is S) we can notice much lower mean values of indices determined with US approach in relation to the reference ones. In case of long-term variability indices this decrease was insignificant. Differences between the corresponding mean values were statistically significant for all indices ($p < 0.01$

TABLE II
SUMMARY VALUES OF THE FHR VARIABILITY INDICES
DETERMINED SIMULTANEOUSLY FROM US AND ECG

Term	Origin	FHR from US	FHR from FECG
Long-term variability indices			
L_HAA	de Haan [25]	25.00 ± 0.86	25.20 ± 0.79
L_YEH	Yeh [26]	2.84 ± 0.09	3.09 ± 0.11
L_HUE	Huey [27]	91.70 ± 2.30	139.40 ± 3.40
L_DAL	Dalton [28]	12.70 ± 0.43	13.80 ± 0.50
L_ZUG	Zugaib [29]	43.20 ± 3.10	51.80 ± 3.90
L_ORG	Organ [30]	7.71 ± 0.26	8.39 ± 0.29
L_OSC	Oscillation [8]	16.30 ± 0.53	18.30 ± 0.64
Short-term variability indices			
S_HAA · 10 ³	de Haan [25]	3.58 ± 0.08	6.02 ± 0.13
S_YEH · 10 ³	Yeh [26]	2.89 ± 0.07	3.20 ± 0.08
S_HUE	Huey [27]	108.00 ± 1.90	152.40 ± 3.00
S_DAL	Dalton [28]	1.50 ± 0.03	1.89 ± 0.05
S_ZUG · 10 ³	Zugaib [29]	1.91 ± 0.05	2.31 ± 0.06
S_GEI	van Geijn [7]	5.37 ± 0.15	7.81 ± 0.20

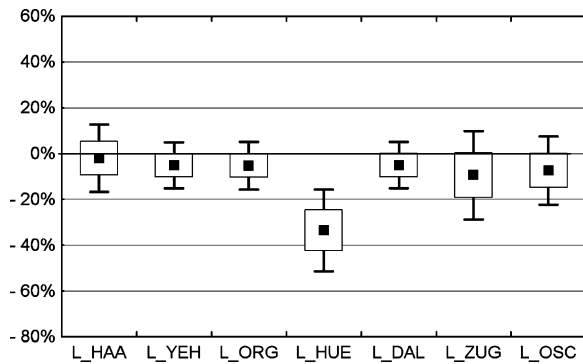


Fig. 7. Descriptive statistics of the relative errors of all the long-term FHR variability indices. Rectangle represents mean ± SD, whereas horizontal lines define the range of ± 2 SD.

for Wilcoxon test) and the distributions for indices significantly differ from the normal distribution ($p < 0.001$ for Shapiro-Wilk test). Obtained results proved that the influence of US method on values of variability indices is significant. In order to evaluate the strength of this influence the relative errors of all indices were examined. Figs. 7 and 8 are the graphical representation of their descriptive statistics. It is worth to note that the errors of indices determined using US always take negative values, which means that these indices are lower than the reference ones calculated for FECG signal. This is mainly the effect of correlation techniques applied in the US method which causes averaging of T_{RR} values. Therefore, differences between consecutive T_{RR} intervals and in consequence values of variability indices decrease.

The best long-term indices (having the lowest sensitivity to the measurement method) are those proposed by: de Haan with error of -2% ($SD = \pm 7\%$) as well as by Yeha, Organ and Dalton with error about -5% ($SD = \pm 5\%$). The next group are: Zugaib and Oscillation index with error reaching -10% ($SD = \pm 10\%$). The Huey index whose error equals -33% ($SD = \pm 17\%$) is outside of any expectable range. Short-term indices can be ordered from the worst to the best: de Haan, van Geijn, Huey, Dalton, Zugaib and Yeh. Their error varies from

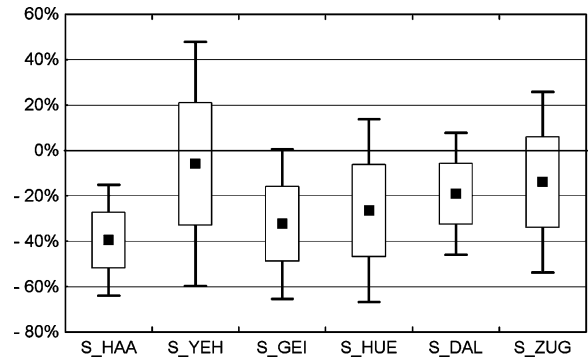


Fig. 8. Summary statistics of the relative errors of all the short-term FHR variability indices. Rectangle represents mean ± SD, whereas horizontal lines define the range of ± 2 SD.

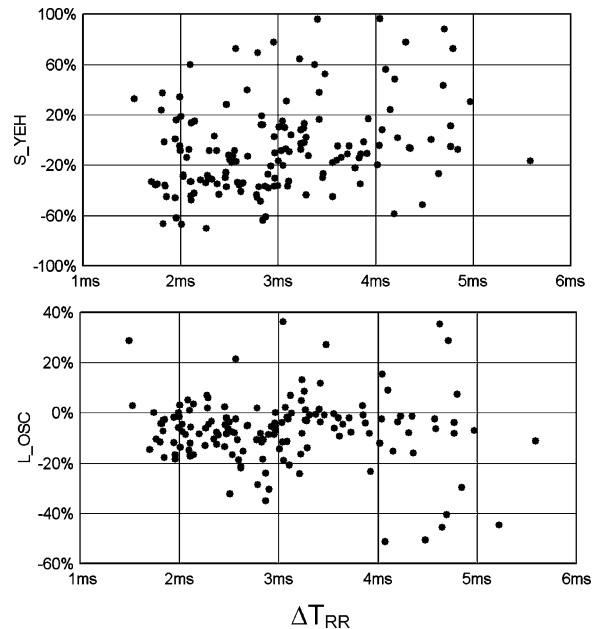


Fig. 9. Diagrams of correlation between partial values of T_{RR} interval errors and errors of two selected FHR variability indices determined within the same 1-min segments.

about -40% , with 7% step for successive indices, to the value of -5% , dispersion of errors is very large ($\pm 20\%$).

Results of correlation analysis between mean values of intervals and selected indices errors are presented in Fig. 9. The observed lack of correlation confirms previous conclusion from the analysis of indices definitions. The T_{RR} determination error itself for a given fragment of the trace is not crucial for indices of instantaneous FHR variability and, therefore, for clinically important signs of FHR trace. This relation is much more complex because just the distribution in time of the T_{RR} interval error is important. Obtained results confirm that the approach to testing the FHR from Doppler US channel based on clinically significant trace parameters is correct.

V. DISCUSSION

The evaluation of the US method was carried out with respect to the accuracy of the FHR measurement and with respect to the influence of this accuracy on quantitative FHR variability analysis. Taking into account that FHR variability analysis

plays the main role in fetal distress assessment, the evaluation of conditions which may have influence on the efficiency of this diagnosis is extremely crucial. Due to the limited resolution of human eye, the accuracy of the US method has not a critical impact on the visual interpretation of FHR trace. However, today's standard in perinatology is the automated FHR analysis comprising calculation of variability indices, whose predictive values for fetal well-being assessment are considered as very high. Analysis of processing techniques of Doppler envelope and indices definitions have led to the conclusion that the algorithms for determination of FHR variability indices are the most sensitive to the signal recording method. Although this problem has been widely discussed, the question: "What is the real influence of T_{RR} measurement accuracy on indices describing FHR variability?" has not been answered yet.

FECG was chosen as a reference method, because it ensured the most accurate measurement of T_{RR} intervals basing on precise detection of QRS complexes. Comparison process was carried out in two stages. The first one was a direct comparison of corresponding T_{RR} intervals determined with the use of two investigated acquisition methods. However, this comparison of primary data is not sufficient for the evaluation of influence of US method on the indices describing the FHR variability. Therefore, we decided to carry out the second stage—comparison of indices calculated by both methods. This stage was more complex because it included the determination of various indices within 1-min segments of signals as well as establishing of measures for the indices comparison.

The dedicated measurement station for simultaneous recording of direct fetal electrocardiogram and FHR signal via US channel was developed. The FECG recorded from fetal head was used to obtain the reference FHR signal. The algorithm for precise detection of R-waves, thanks to the off-line mode, made possible an interactive control of all the suspicious intervals, and the final material comprised 21 941 pairs of T_{RR} intervals. Differences between corresponding intervals provided by the US and the FECG were calculated. Mean value of the absolute error of the US method was equal to 2.98 ms (standard deviation 4.18 ms). These values appeared as lower than those reported elsewhere [13], [20], which could be caused by the fact that very precise synchronization of recorded signals was ensured, and an efficient algorithm for determination of reference data as well as an advanced fetal monitor were used.

The standard deviation of 1.18 bpm has been reported in [21], that is very close to our result (1.25 bpm). However, reported value was calculated from only 267 instantaneous FHR values obtained simultaneously from the US method and the direct FECG. That work was focused mainly at the evaluation of a new processing technique of raw envelope signal from the HP 8040 fetal monitor. In [22], while testing the developed method for fetal abdominal electrocardiogram recording, authors referred to FHR signal simultaneously acquired using very basic Doppler US fetal monitor. However, it is only an estimation of agreement between these methods because none of them can be considered as a reference. The authors showed that 79% of all differences between corresponding values were in the range ± 3 bpm, whereas we noted 87% (Fig. 6). Our result describes the inaccuracy

of Doppler US method since the comparison was carried out using the reference method—direct FECG.

Measurement error of 0.89 bpm related to visual analysis confirms the high consistency of FHR waveforms in global aspect. Similar consistency was reported by Dawes [16], and Lawson [11], [12]. Such differences are difficult to notice because the thickness of line used to plot the FHR waveforms on a thermal paper corresponds to 0.5 bpm. However, very large differences (even exceeding 35 bpm) can occur locally—when due to interferences the US method is not able to follow the changes of FHR. Despite that, occasional doubling or halving of FHR value for no apparent reason are very characteristic drawbacks of the US channel [4].

From a clinical point of view, it was very important to compare the acquisition methods indirectly—by the use of indices describing long- and short-term FHR variability. Mean relative error of the US method with respect to long-term indices was negative, and it varied from -2.0% to -9.3% for particular indices, excluding Huey index. So, the US method decreases the magnitude of FHR variability in the worst case to 90.7% of its true value. This is a very important feature because it makes the assessment of fetal distress slightly more pessimistic. The decrease is an effect of the averaging characteristic of correlation-based techniques used in the US channel. The influence of US acquisition method on short-term variability indices appeared to be more significant. The mean relative error for particular indices varied in a wide range (from -5.9% for Yeh index to -39.5% for de Haan index). Like in the case of long-term FHR variability, the US method decreases values of the short-term indices.

Obtained results can not be simply related to results reported in other works, because very simple comparison procedures were usually used. Long-term variability indices were calculated using samples averaged over 3.75 s periods, or even over one-hour trace. Whereas in our study, indices were calculated in 1-min segments—just as the definitions require. Dawes [16] and Lawson [11], [12] compared FECG and US channels on a basis of FHR values averaged for one-hour traces. The error of long-term variability index (defined simply as a standard deviation of T_{RR} intervals) was equal to 3.3%, whereas for short-term variability (determined as beat-to-beat changes of intervals) was much higher and reached 100% [16]. Lawson [11] determined long-term index error for other type of fetal monitor and obtained 12.5%. Short-term index error was also higher and reached 200%. Finally, when testing another model of fetal monitor (with autocorrelation technique) Lawson [12] obtained lower values: 2.5% and -35% , respectively. Murrills [23] analyzed 5-min segments of signal and obtained results closer to ours: -3% for long- and -15% for short-term variability. However, the use of strongly interfered FECG signal recorded from maternal abdomen seems to be rather controversial [24].

We proved that variability index error was not correlated with T_{RR} interval measurement error. The correlation coefficient varied from 0.28 to 0.32. These results can be confirmed by a visual evaluation of plot from Fig. 9, showing that a given index error takes random values from its variation range when T_{RR} interval error varies from 1.8 to 4.8 ms. This implies that

indices errors are quite independent from FHR measurement accuracy. Consequently, this confirms our approach to evaluation of the accuracy of FHR measurement technique with the use of the output results of signal analysis—indices describing the instantaneous variability of the fetal hear rate.

VI. CONCLUSION

Modern fetal monitors using the Doppler US technique are not able to provide the signal of the accuracy required for reliable quantitative evaluation of FHR variability—particularly short-term variability—based on the indices calculated automatically. Good news is that this limitation causes a decrease of the variability indices, which prevents fetal distress signs from being unnoticed. Both among the long-term and short-term FHR variability indices, there were two whose sensitivity to the acquisition method was considerably lower: de Haan and Dalton indices.

As we proved the FHR variability index error is not correlated with T_{RR} interval measurement error. Therefore, measurement accuracy from the fetal monitor specification can not be directly related to the results of the computer-aided analysis of FHR variability. Generally, we can say that the influence of measurement method on the description of biophysical process can not be accomplished basing only on errors calculated for direct measurement of the given process parameters. This conclusion results from the fact that usually a given process is evaluated not only using primary data provided by direct measurements, but more frequently, using the results of final analysis of these data.

Above conclusions can be applied particularly when the process is monitored with the use of two independent methods: passive and active. In passive acquisition method, the instrumentation records an energy emitted during the process, for example, thermal (thermvisual imaging), acoustic (phonocardiography), electrical (electrocardiography, electroencephalography) or magnetic (magnetocardiography). Active method uses the energy delivered by the measurement instrument that underwent absorption, dispersion or reflection effect from some objects taking part in the process observed. Examples of such approach are: x-ray, ultrasonography or magnetic resonance imaging. In the analyzed case—biophysical fetal monitoring—passive method determines FHR parameters by recording of electrical heart activity—fetal electrocardiogram, whereas active method relies upon US monitoring of mechanical activity of fetal heart. As it was proved in this work, if one of the measurement methods is considered as a reference, then the evaluation of the influence of inaccuracy of the second one on the diagnosis of fetal distress can not be accomplished basing only on the analysis of error determined for directly recorded signal. This evaluation should be based on errors calculated for parameters having essential predictive value for clinical assessment.

APPENDIX

MEASURES OF FETAL HRV BASED ON DE HAAN INDICES

Instantaneous variability of FHR is divided into two types. Changes concerning the durations of consecutive R-R intervals are called short-term variability or beat-to-beat variability. Due

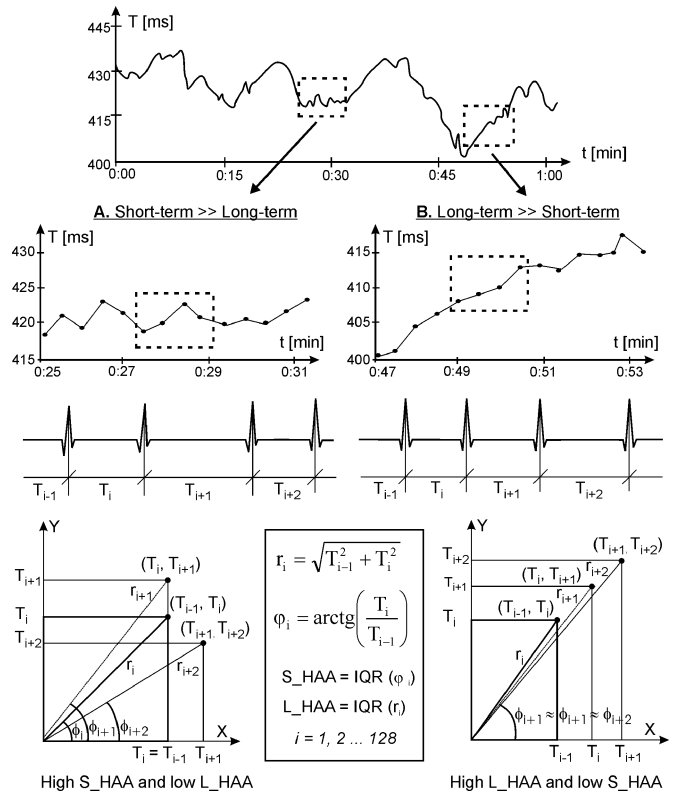


Fig. 10. Graphical representation of instantaneous FHR variability: with dominating short-term variability—A, and with dominating long-term variability—B. The examples are illustrated using the definitions of de Haan variability indices (S_HAA—short-term, L_HAA—long-term variability).

to a certain periodicity in the direction and magnitude of these changes, they result in fluctuations of the FHR around its mean level. These fluctuations are called long-term variability. Indices proposed by de Haan are very frequently used for description of FHR variability [25]. Points corresponding to consecutive pairs of intervals (T_{i-1}, T_i) are put on two-dimensional plot in Cartesian coordinates expressed in milliseconds (Fig. 10). Their polar coordinates: the radial coordinate r_i and the angular coordinate φ_i are used to construct the definition of FHR variability indices. If short-term variability is high the angular coordinate changes from beat to beat. The domination of long-term variability is accompanied by significant changes of radial coordinate during consecutive heartbeats, whereas the angular coordinate has almost the same value. The short-term variability index (S_HAA) is defined as the interquartile range of the angular coordinates, whereas long-term variability index (L_HAA) as the interquartile range of the radial coordinates.

REFERENCES

- [1] E. H. Hon, "Instrumentation of fetal heart rate and fetal electrocardiography. II. A vaginal electrode," *Am. J. Obstet. Gynecol.*, vol. 86, pp. 772–778, 1963.
- [2] T. Kupka, J. Jezewski, A. Matonia, K. Horoba, and J. Wrobel, "Timing events in Doppler ultrasound signal of fetal heart activity," in *Proc. 26th IEEE EMBS Int. Conf.*, San Francisco, CA, 2004, pp. 337–340.
- [3] M. Y. Divon, "Autocorrelation techniques in fetal monitoring," *Am. J. Obstet. Gynecol.*, vol. 151, pp. 2–6, 1985.
- [4] J. Jezewski, J. Wrobel, K. Horoba, J. Moczko, G. Breborowicz, and S. Graczyk, "Advances in Doppler ultrasound FHR monitoring," *Klin. Perinat. Ginekol.*, vol. 9, pp. 241–251, 1995.

- [5] J. Jezewski, K. Horoba, A. Gacek, J. Wrobel, A. Matonia, and T. Kupka, "Analysis of nonstationarities in fetal heart rate signal: inconsistency measures of baselines using acceleration/deceleration patterns," in *Proc. 7th ISSPA*, Paris, France, 2003, pp. 34–38.
- [6] J. Jezewski, J. Wrobel, K. Horoba, A. Gacek, and J. Sikora, "Fetal heart rate variability: clinical experts versus computerized system interpretation," in *Proc. 24th IEEE EMBS Int. Conf.*, Huston, TX, 2002, pp. 1617–1618.
- [7] H. P. van Geijn, "Analysis of heart rate and beat-to-beat variability: interval difference index," *Am. J. Obstet. Gynecol.*, vol. 138, pp. 246–252, 1980.
- [8] J. Jezewski and J. Wrobel, "Fetal monitoring with automated analysis of cardiocotogram: the KOMPOR system," in *Proc. 15th IEEE EMBS Int. Conf.*, San Diego, CA, 1993, pp. 638–639.
- [9] T. Kubo, J. Inaba, S. Shigemitsu, and T. Akatsuka, "Fetal heart variability indices and the accuracy of variability measurements," *Am. J. Perinat.*, vol. 4, pp. 179–186, 1987.
- [10] J. Jezewski *et al.*, "Monitoring of mechanical and electrical activity of fetal heart: determination of the FHR," *Arch. Perinat. Med.*, vol. 8, pp. 33–39, 2002.
- [11] G. W. Lawson, G. S. Dawes, and C. W. G. Redman, "A comparison of two fetal heart rate ultrasound detector systems," *Am. J. Obstet. Gynecol.*, vol. 143, pp. 840–842, 1982.
- [12] G. W. Lawson, R. Belcher, G. S. Dawes, and C. W. G. Redman, "A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems," *Am. J. Obstet. Gynecol.*, vol. 147, pp. 721–722, 1983.
- [13] F. H. Boehm and L. M. Fields, "The indirectly obtained fetal heart rate: comparison of first- and second-generation electronic fetal monitors," *Am. J. Obstet. Gynecol.*, vol. 155, pp. 10–14, 1986.
- [14] J. Jezewski, J. Wrobel, K. Horoba, and A. Gacek, "Instrumentation for fetal monitoring—improvement in Doppler ultrasound technology," *J. Med. Inf. Technol.*, vol. 2, pp. 17–26, 2001.
- [15] D. L. Tuck, "Improvement in Doppler ultrasound human foetal heart rate records by signal correlation," *Med. Biol. Eng. Comput.*, vol. 20, pp. 357–360, 1982.
- [16] G. S. Dawes, G. H. A. Visser, J. D. S. Goodman, and C. W. G. Redman, "Numerical analysis of the human fetal heart rate: the quality of ultrasound records," *Am. J. Obstet. Gynecol.*, vol. 141, pp. 43–52, 1981.
- [17] J. A. D. Spencer, R. D. Belcher, and R. D. Dawes, "The influence of signal loss on the comparison between computer analysis of the fetal heart rate in labor using pulsed Doppler ultrasound (with autocorrelation) and simultaneous scalp electrocardiogram," *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 25, pp. 29–34, 1987.
- [18] S. A. Shakespeare, J. A. Crowe, B. R. Hayes-Gill, K. Bhogal, and D. K. James, "The information content of Doppler ultrasound signals from the fetal heart," *Med. Biol. Eng. Comput.*, vol. 39, pp. 619–626, 2001.
- [19] J. Jezewski, D. Cholewa, K. Kaminski, A. Matonia, T. Kupka, and K. Horoba, "Progress in fetal monitoring—direct or indirect electrocardiography," *Arch. Perinat. Med.*, vol. 9, pp. 15–19, 2003.
- [20] T. Koyanagi, T. Yoshizato, N. Horimoto, T. Takashima, S. Satoh, and H. Maeda, "Fetal heart rate variation described using a probability distribution matrix," *Int. J. Biomed. Comput.*, vol. 35, pp. 25–37, 1994.
- [21] C. H. L. Peters, E. D. M. ten Broeke, P. Andriessen, B. Vermeulen, R. C. M. Berendsen, P. F. F. Wijn, and S. G. Oei, "Beat-to-beat detection of fetal heart rate: Doppler ultrasound cardiocotography compared to direct ECG cardiocotography in time and frequency domain," *Physiol. Measur.*, vol. 25, pp. 585–593, 2004.
- [22] M. I. Ibrahimy, F. Ahmed, M. A. Mohd Ali, and E. Zahedi, "Real-time signal processing for fetal heart rate monitoring," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 2, pp. 258–262, Feb. 2003.
- [23] A. J. Murrills, T. H. Wilmschurst, and T. Wheeler, "Antenatal measurement of beat-to-beat fetal heart rate variation: accuracy of the Hewlett-Packard ultrasound autocorrelation technique," in *Proc. Fetal Physiological Measurement*, Memphis, TN, 1986, pp. 36–44.
- [24] J. Jezewski *et al.*, "Monitoring of mechanical and electrical activity of fetal heart: the nature of signals," *Arch. Perinat. Med.*, vol. 8, pp. 40–46, 2002.
- [25] J. de Haan, J. H. van Bommel, B. Versteeg, A. F. L. Veth, L. A. M. Stolte, J. Janssens, and T. K. A. B. Eskes, "Quantitative evaluation of fetal heart rate patterns I. Processing methods," *Eur. J. Obstet. Gynecol.*, vol. 3, pp. 95–102, 1971.
- [26] S. Y. Yeh, A. Forsythe, and E. H. Hon, "Quantification of fetal heart rate beat-to-beat interval differences," *J. Obstet. Gynecol.*, vol. 41, pp. 355–363, 1973.
- [27] J. R. Huey, R. H. Paul, A. A. Hadjiev, J. Jilek, and E. H. Hon, "Fetal heart rate variability: An approach to automated assessment," *Am. J. Obstet. Gynecol.*, vol. 134, pp. 691–695, 1979.
- [28] K. J. Dalton, G. S. Dawes, and J. E. Patrick, "Diurnal respiratory and other rhythms of fetal heart rate in lambs," *Am. J. Obstet. Gynecol.*, vol. 127, pp. 414–424, 1977.
- [29] M. Zugaib, A. B. Forsythe, B. Nuwayhid, S. M. Lieb, K. Tabsh, R. Erkkola, E. Ushioda, C. R. Brinkman, III, and N. S. Assali, "Mechanisms of beat-to-beat variability in the heart rate of the neonatal lamb I. Influence of the autonomic nervous system," *Am. J. Obstet. Gynecol.*, vol. 138, pp. 444–452, 1980.
- [30] L. W. Organ, P. A. Hawrylyshyn, J. W. Goodwin, J. E. Milligan, and A. Bernstein, "Quantitative indices of short- and long-term heart rate variability," *Am. J. Obstet. Gynecol.*, vol. 130, pp. 20–27, 1978.



Janusz Jezewski (M'93) was born in Zabrze, Poland. He received the M.S. degree in electronic engineering from the Silesian University of Technology, Gliwice, Poland, in 1979 and the Ph.D. degree in biological science from the University of Medical Sciences, Poznan, Poland, in 1997.

From 1979 to 1989, he worked at the Institute of Electronics of the Silesian University of Technology. In 1989, he joined the Institute of Medical Technology and Equipment, Zabrze, Poland where he is currently a Director of Science and a Head of the

Biomedical Informatics Department. His research interests are in biomedical instrumentation and digital signal processing, especially the detection and analysis of fetal heart activity signals for the extraction of clinically relevant information.

Dr. Jezewski is a member of the Polish Society of Biomedical Engineering, the Polish Society of the Perinatal Medicine, the Institute of Physics and Engineering in Medicine (IPEM), the European Society for Engineering and Medicine (ESEM), and the International Federation for Medical & Biological Engineering (IFMBE).



Janusz Wrobel was born in Poland, in 1965. He received the M.S. degree in electronic engineering from the Silesian University of Technology, Gliwice, Poland, in 1990 and the Ph.D. degree in medical science from the University of Medical Sciences, Poznan, Poland, in 2001.

Currently, he is a System Research Engineer in Biomedical Informatics Department of the Institute of Medical Technology and Equipment, Zabrze, Poland. His research interests include clinical applications of digital processing methods and

instrumentation for fetal and maternal biosignals acquisition.

Dr. Wrobel is a member of the Polish Society of Biomedical Engineering.



Krzysztof Horoba was born in Poland in 1968. He received the M.S. degree in electronic engineering from the Silesian University of Technology, Gliwice, Poland, in 1993 and the Ph.D. degree in medical science from University of Medical Sciences, Poznan, Poland, in 2001.

He is currently a Project Leader in Biomedical Informatics Department of the Institute of Medical Technology and Equipment, Zabrze, Poland. His research interests include fetal electrocardiography, electrohysterography as well as the software development of the computerized fetal monitoring systems.

Dr. Horoba was a fellowship holder of the Foundation for Polish Science in 1997. He is a member of the Polish Society of Biomedical Engineering.