*fetal monitoring, fetal heart rate,
acceleration/deceleration,
expert inconsistency*

Janusz JEŻEWSKI*, Krzysztof HOROBA*,
Janusz WRÓBEL*, Adam GACEK*

INCONSISTENCY IN EVALUATION OF CLINICAL PATTERNS IN FETAL HEART RATE WAVEFORMS

Acceleration and deceleration patterns in fetal heart rate (FHR) are widely used as an assessment of fetal condition. The basis for both visual and automated quantitative analysis is correctly determined fetal heart rate baseline. There are many methods for automated fetal heart rate baseline estimation. Since there are no definitive criteria or "gold standard" to evaluate algorithms for baseline determination, a clinical expert could be considered as a reference. Taking into account that not the shape of baseline itself but the results of quantitative analysis obtained using this baseline are important, we stated that such evaluation should be done only on the basis of comparison of effects – the recognized accelerations and decelerations. However, the first step for such comparison is to check how well expert meets established FIGO definition of accelerations and decelerations patterns in FHR trace. The paper presents a method for evaluation of expert's interpretation of these definitions. Two indices describing the inconsistency among fetal monitoring system and experts have been defined. The results obtained indicate the essential differences in procedure of accelerations/decelerations recognition by the computerized system and human expert.

1. INTRODUCTION

Fetal heart rate (FHR) monitoring in the assessment of fetal condition during high-risk pregnancy is based on the presence or absence of accelerations (A) and decelerations (D) patterns in a limited time period. These episodic events can be identified by a clinician who visually interprets printed FHR trace or the patterns can be recognized automatically by means of computerized fetal monitoring system [8]. A correctly determined FHR baseline is a precondition for correct identification of the acceleration and deceleration patterns in the fetal heart rate waveform [1, 2]. The association of the events with fetal movements and uterine contractions is critical in the assessment of fetal condition. In 1986 the FIGO Subcommittee on Standards in Perinatal Medicine raised the following definition of baseline [9]: "Baseline is the mean level of the FHR when this is stable, accelerations and decelerations being absent. It is determined over a time period of 5 or 10 minutes and expressed in beats per minute (bpm)". Acceleration is defined as a transient increase in heart rate of 15 bpm or more and lasting 15 sec or more. Deceleration is a transient episode of slowing of heart rate bellow the baseline level of more than 15 bpm and lasting 10 sec or more. This baseline definition is only sufficient for visual analysis and it is completely useless for automated signal

* Institute of Medical Technology and Equipment, Roosevelt'a 118, 41-800 Zabrze, Poland

analysis [7]. The computer cannot reject acceleration and deceleration patterns before they are detected so as to draw the baseline in relation to which the A/D patterns are determined. Moreover, baseline estimation is the most difficult although being the most important in these segments of FHR where accelerations and decelerations occur. The definitions of A/D imply that, in the case of automated signal analysis, a strict sequence must be kept, whereby the baseline estimation is the first stage of processing.

Figure 1 presents a segment of FHR signal with its automated interpretation (estimated baseline and graphic markers of recognized accelerations) done by a computer-aided fetal monitoring system The detailed parameters, including duration of accelerations, amplitude and area are calculated and displayed in a numerical form.

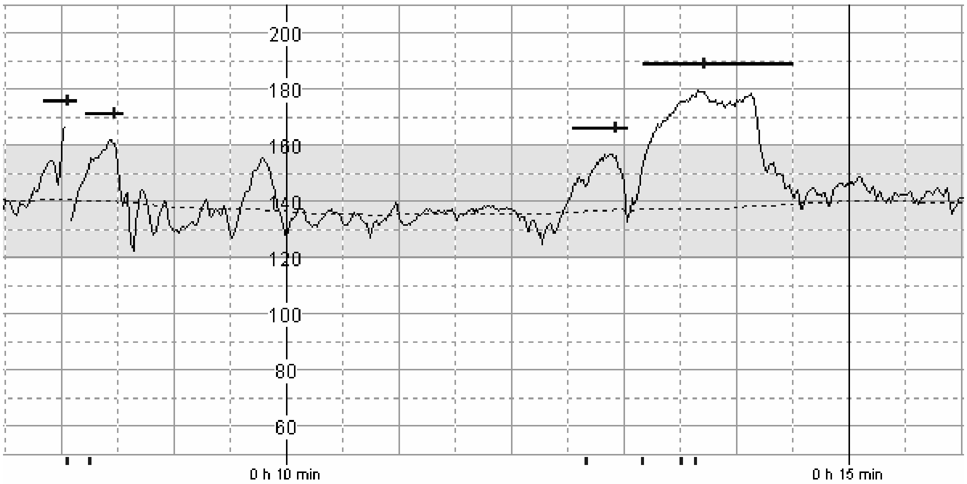

Fig.1. Segment of FHR waveform and its automated interpretation: estimated baseline (dotted line) and recognized accelerations (graphical markers). Horizontal lines above accelerations correspond to the pattern duration, and the vertical bar represents the maximum peak.

There is a common opinion that algorithm for the FHR baseline estimation is responsible for the efficiency of the entire computer-aided fetal monitoring system. Even a small inaccuracy in the FHR baseline curve may significantly distort the detection of A/D, which may subsequently lead to false interpretation of clinical symptoms. Therefore, it is necessary to develop a method and criteria, which would allow for evaluating the baseline estimation algorithms. The method should allow the comparison of results of two different algorithms, as well as the comparison to a reference, which could be a baseline from human expert. Taking into account that not the shape of baseline itself but results of quantitative analysis obtained using this baseline are important, we stated that such evaluation should be done only on the basis of comparison of effects – the recognized accelerations and decelerations. However, the first step for such comparison is to check how well expert meets established FIGO definition of accelerations and decelerations. We have developed a method for evaluation of expert's interpretation of FIGO criteria in reference to the computerized system interpretation. This work has been done using the fetal monitoring system developed by us and described elsewhere [3].

2. METHODS

Block diagram of the method for extraction and processing of expert's interpretation of FHR trace is presented in Fig 2. The expert draws his baseline and indicates start and end points of A/D with vertical bars directly on the FHR trace from the system archive. The printout contains a 30 min recording and is compatible with fetal monitor recorder format. Archive records can be printed as many times as necessary, so they enable experts to repeat evaluations and choose the best interpretation. Moreover, it is possible to enrich printouts with additional elements facilitating processing of expert's interpretation: the calibration line for setting scan area, and two reference lines bounding the full range of baseline variability 50÷210 bpm. Essential elements for conversion are printed in black, while the others in bright colors.

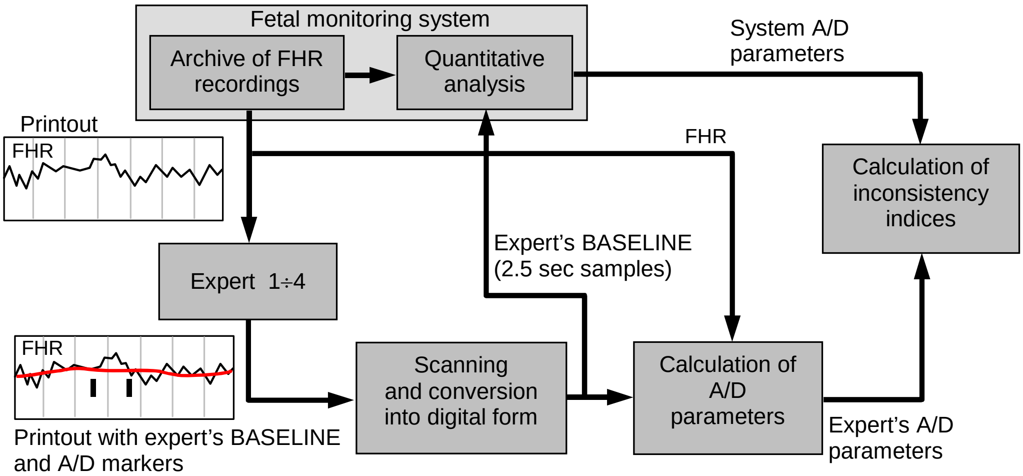

Fig.2. Block diagram of the procedure of extraction of expert's interpretation of cardiotocograms

Figure 3 shows the FHR trace prepared for processing. Printout is scanned and then extraction of baseline curve and A/D markers from bitmap image along with their conversion into digital representation are performed by dedicated software [4, 5]. Result of printout processing is a file containing samples of baseline with accuracy of 1 bmp and averaged over 2.5 sec periods, which corresponds to its representation in the fetal monitoring system. Additionally, there are start and end points of each acceleration and deceleration also expressed as multiple of 2.5 sec periods. These data together with FHR signal are used as inputs for a procedure, which calculates detailed parameters of expert's A and D patterns (maximum amplitude, duration and area). The expert's baseline is also used for automated detection and evaluation of A/D parameters according to the FIGO definition. Finally, two sets of parameters of events, recognized visually and detected automatically, are provided for each FHR trace processed.

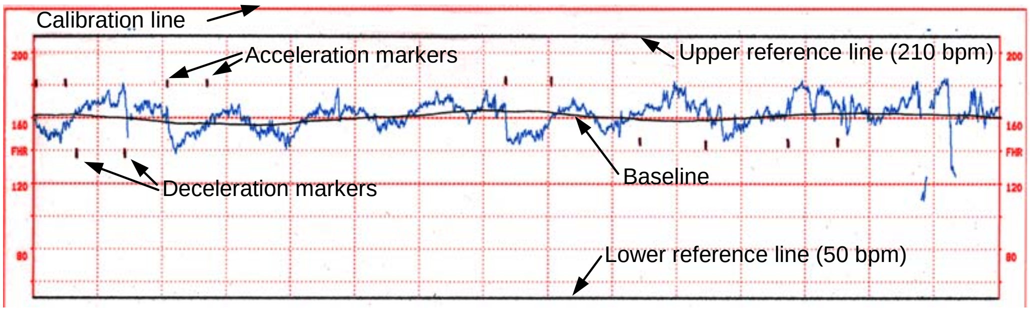

Fig.3. Example of FHR trace printed from system archive and prepared for scanning and conversion

In the next stage, accelerations and decelerations determined by the expert and the computerized system were compared in relation to their compliance with FIGO criteria. Two indices describing inconsistency among system and experts with FIGO have been defined. Indices were calculated separately for acceleration and deceleration patterns. Value of 100 % of a given index means a total inconsistency. Expert's inconsistency index (E_A) as to accelerations expresses percentage of accelerations which were incorrectly interpreted by the expert, e.g. those that were recognized but were to small ($E\overline{C}\overline{F}_A$), or those fulfilling FIGO criteria and detected by the computer but unrecognized by the expert ($\overline{E}CF_A$), in relation to all accelerations that the expert identified both correctly (AE) or incorrectly ($E\overline{C}\overline{F}_A$) as well as should have identified ($\overline{E}CF_A$):

$$E_A = \frac{E\overline{C}\overline{F}_A + \overline{E}CF_A}{\overline{E}CF_A + E\overline{C}\overline{F}_A + AE} \cdot 100\% \qquad (1)$$

Computer inconsistency index (C_D) as to decelerations represents the percentage of decelerations which were improperly detected by the computer, e. g. those recognized by the expert and fulfilling FIGO criteria ($\overline{E}CF_D$) in relation to all decelerations the system detected (DC) and should have detected ($E\overline{C}F_D$):

$$C_D = \frac{E\overline{C}F_D}{E\overline{C}F_D + DC} \cdot 100\% \qquad (2)$$

3. RESULTS

Four experts interpreted 41 FHR traces recorded between 31st and 41st week of pregnancy (duration 31 ÷ 60 min). These traces were selected from the system archive among hundreds of recordings and they represented a various level of interpretation difficulty. As a result we obtained 164 values of inconsistency indices. Summary of results for accelerations are presented in Table 1 and in Table 2 for decelerations.

Expert	Number of accelerations		Number of traces with particular inconsistency			
	AE	AC	NL	$N\overline{L}$	$\overline{N}L$	$\overline{N}\overline{L}$
E1	294	525	9	1	25	6
E2	174	455	8	1	28	4
E3	391	478	14	1	20	6
E4	237	353	11	0	29	1

NL – consistency of number (N) and localization (L)

$N\overline{L}$ – consistency of number and inconsistency of localization

$\overline{N}L$ – inconsistency of number and consistency of localization

$\overline{N}\overline{L}$ – inconsistency of number and inconsistency of localization

Tab.1. Summary of results of experts and system interpretation in reference to accelerations

Expert	Number of decelerations		Number of traces with particular inconsistency			
	DE	DC	NL	$N\overline{L}$	$\overline{N}L$	$\overline{N}\overline{L}$
E1	90	160	21	0	19	1
E2	54	151	17	0*	24	0
E3	140	188	20	2	16	3
E4	112	197	14	0*	27	0

* – only inconsistency of number was noted

NL – consistency of number (N) and localization (L)

$N\overline{L}$ – consistency of number and inconsistency of localization

$\overline{N}L$ – inconsistency of number and consistency of localization

$\overline{N}\overline{L}$ – inconsistency of number and inconsistency of localization

Tab.2. Summary of results of experts and system interpretation in reference to decelerations

Experts recognized in all traces 1096 accelerations and 396 decelerations, while the computerized system 1811 and 696 respectively. The difference in numbers of detected accelerations between the best expert (E3) and the computer was 87 (48 for decelerations). This expert has recognized the highest number of accelerations, while the computer has detected the highest number of these events using the baseline provided by expert E1. We can found small number of recordings for which only inconsistency with localization of accelerations was noted. The similar results were obtained in case of decelerations.

Table 3 presents a comparison of accelerations interpreted in various way by the experts and the computer and – for decelerations – Table 4. The largest number of accelerations – 33 that proved to be not fulfilling FIGO has been recognized by expert E4. The best expert E3 has recognized the maximum number (3) of accelerations fulfilling FIGO criteria. According to E_A index, the experts can be lined-up as follows: E3, E4, E1 and E2. C_A index of inconsistency between the computer and FIGO for the best expert is equal to 0.6 and is nearly 40 times lower than corresponding index E_A.

Expert	Number of accelerations not fulfilling FIGO			Mean values of inconsistency indices [%]	
	$E\overline{C}\overline{F}_A$	$E\overline{C}F_A$	$\overline{E}CF_A$	E_A	C_A
E1	11	2	187	40.2	0.4
E2	13	1	266	61.6	0.2
E3	28	3	88	22.9	0.6
E4	33	1	103	36.5	0.3

$E\overline{C}\overline{F}_A$ – accelerations recognized only by expert and not fulfilling FIGO criteria

$E\overline{C}F_A$ – accelerations recognized only by expert and fulfilling FIGO criteria

$\overline{E}CF_A$ – accelerations detected only by the system and fulfilling FIGO criteria

Tab.3. Number of accelerations interpreted in various ways by the experts and the system

Expert	Number of decelerations not fulfilling FIGO			Mean values of inconsistency indices [%]	
	$E\overline{C}\overline{F}_D$	$E\overline{C}F_D$	$\overline{E}CF_D$	E_D	C_D
E1	11	1	65	45.8	0.6
E2	11	1	92	65.6	1.8
E3	39	6	55	40.2	3.1
E4	30	1	77	48.9	0.5

$E\overline{C}\overline{F}_D$ – decelerations recognized only by expert and not fulfilling FIGO criteria

$E\overline{C}F_D$ – decelerations recognized only by expert and fulfilling FIGO criteria

$\overline{E}CF_D$ – decelerations detected only by the system and fulfilling FIGO criteria

Tab.4. Number of decelerations interpreted in various ways by the experts and the system

For the matter of deceleration, although E3 has recognized most of events not fulfiling FIGO (39), he also detected 6 additional correct decelerations. The result is that the lowest value of E_D equals 40.2 being nearly 2 times greater than for acceleration, while C_D index for E3 equals 3.1 being nearly 13 times smaller than E_D.

The Table 5 presents a comparison of average values referenced to a single trace. For the compliance matter of accelerations interpretation with FIGO, even the best expert E3 does not recognize 2.1 accelerations complying with FIGO criteria, yet detecting at the same time 0.7 of accelerations which do not fulfil these criteria. In effect, this becomes nearly 30% of general number (9.5) of the events recognized by him. Expert E2 (mostly inconsistent with the computer) recognizes badly 0.3 and is not detecting the 6.5 accelerations, that together becomes the greater value than the number of his correctly recognized accelerations. In case of deceleration, the corresponding indices are the more unfavourable for the experts.

Expert	Accelerations				Decelerations			
	AE	AC	$E\overline{C}\overline{F}_A$	$\overline{E}CF_A$	DE	DC	$E\overline{C}\overline{F}_D$	$\overline{E}CF_D$
E1	7.2	12.8	0.3	4.6	2.2	3.9	0.3	1.6
E2	4.2	11.1	0.3	6.5	1.3	3.7	0.3	2.2
E3	9.5	11.7	0.7	2.1	3.4	4.6	0.9	1.3
E4	5.8	8.6	0.8	2.5	2.7	4.8	0.7	1.9

$E\overline{C}\overline{F}_A$, $\overline{E}CF_A$ – A/D patterns recognized only by the expert and not fulfilling FIGO criteria

$E\overline{C}\overline{F}_D$, $\overline{E}CF_D$ – A/D patterns detected only by the system and fulfilling FIGO criteria

Tab.5. Comparison of average values in reference to a single trace

Figure 4 presents interpretation of three selected FHR traces. Four experts and computer performed recognition of A/D patterns using the same baseline, which was provided by the best expert e.g. the most compliant with FIGO criteria.

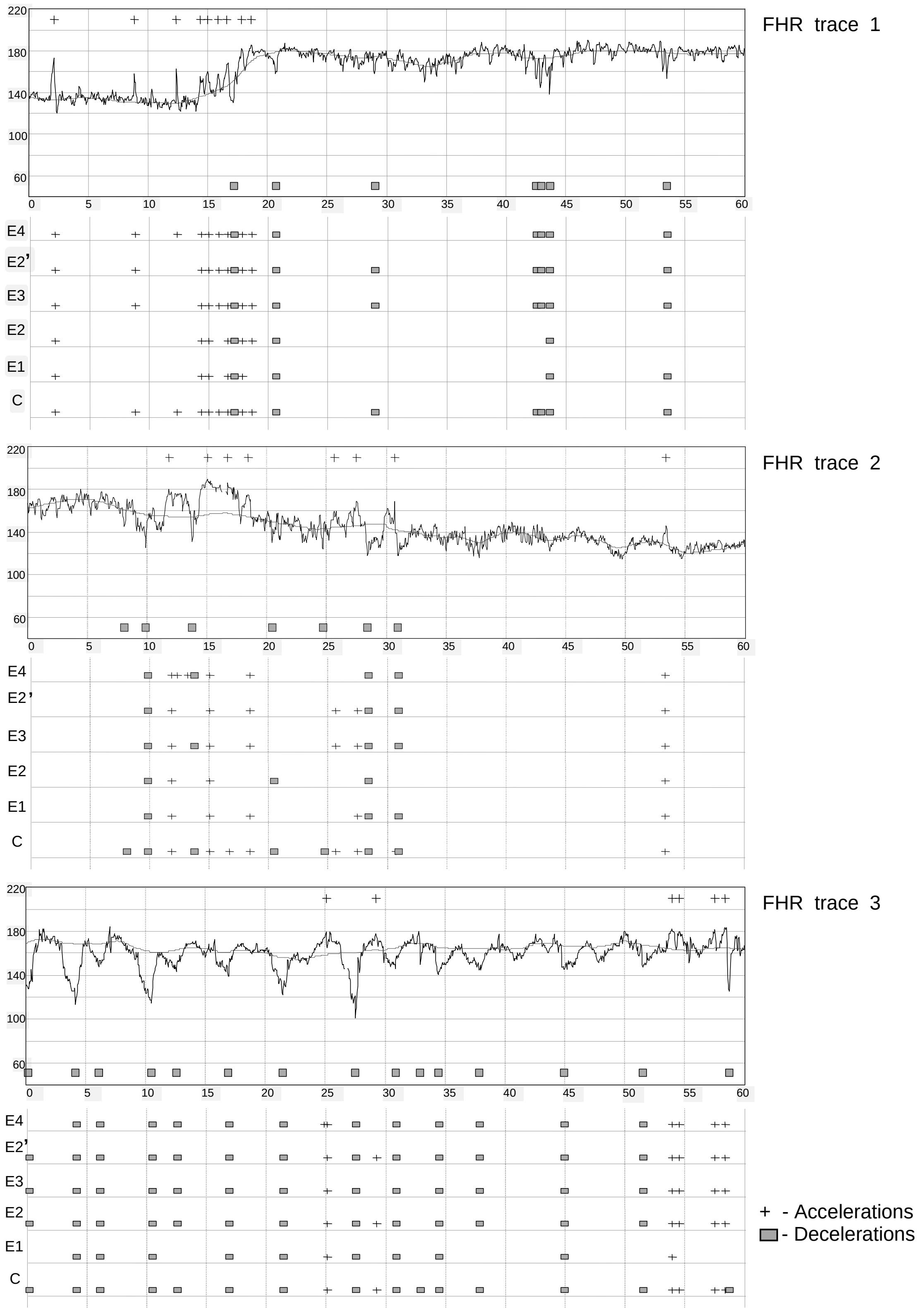

Fig.4. Three various FHR traces with baselines fitted by the best expert and with markers of accelerations and decelerations detected by four experts ($E_1 \div E_4$) and computer (C). FHR trace 1 – inconsistency between experts and computer (E_i-C) is higher than inconsistency among experts (E_i-E_j). FHR trace 2 – inconsistency among experts is higher than among experts and computer. FHR trace 3 – inconsistency (E_i-E_j) for accelerations is higher, while for decelerations is lower.

4. CONCLUSIONS

The results obtained indicate the essential differences in procedure of accelerations/decelerations recognition by a fetal monitoring system and human expert. In practice, there were occurred only few acceleration and deceleration patterns which while being compliant with FIGO criteria, would become recognized by an expert and simultaneously would not be recognized by the system. On the contrary, very often the system recognized such patterns, whereas an expert did not. This indicates unanimously that although being instructed the expert was not using the established FIGO definitions consistently.

The next complication is caused by a fact that an ideal clinical expert and in connection with this, the set of reference accelerations (decelerations) for a given record do not exist. In this situation, it is necessary to use a "quasi-reference" set that is constituted by the team of experts. The sets of accelerations of this team can be acknowledged as almost ideal, however, the subjective comprehension of the phenomenon causes that these sets can differ with each other [6].

Considering our further research on evaluation of various baseline estimation algorithms the final conclusion about reference data can be drawn. We should not use in a direct way accelerations/decelerations recognized by the expert. We should use the baseline curve from expert, and in the next step we should perform automated identification of A/D patterns using dedicated software procedure with the expert's baseline as the input data.

BIBLIOGRAPHY

[1] DALTON K.J. DAWSON A.J. Baseline: A computer method of calculating baseline in fetal heart rate recordings, International Journal of Biomedical Computing, 15, pp. 311-317, 1984.
[2] DAWES G.S. HOUGHTON C.R. Baseline in human fetal heart rate rocords, British Journal of Obstetrics and Gynaecology, 89, pp. 270-275, 1982.
[3] JEŻEWSKI J. WRÓBEL J. HOROBA K. Monitorowanie zagrożeń płodu wspomagane komputerem, in "Biocybernetyka i Inżynieria Biomedyczna 2000 – Tom 7. Systemy komputerowe i teleinformatyczne w służbie zdrowia", ed. M. Nałęcz, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001, /in press/.
[4] JEŻEWSKI J. WRÓBEL J. HOROBA K. GACEK A. GRACZYK S. Cardiotocographic image processing for extraction of FHR baseline drawn by clinical expert, Medical & Biological Engineering & Computing, Vol.34, Suppl.1/1, pp. 110-113, 1996.
[5] JEŻEWSKI J. HOROBA K. WRÓBEL J. GRACZYK S. Extraction of expert's baseline curve from graphical FHR records: Evaluation of the method accuracy. Proc. of the 4th International Conference on Computers in Medicine, Vol.1, pp. 176-180, Zakopane V 1997.
[6] JEŻEWSKI J. WRÓBEL J. HOROBA K. GACEK A. SIKORA J. Fetal heart variability: clinical experts versus computerized system interpretation. 24th International Conference of the IEEE Engineering in Medicine and Biology Society, /in press/, Houston X 2002.
[7] MANTEL R. VERVERS I. COLENBRANDER G.J. VAN GEIJN H.P. Automated antepartum baseline FHR determination and detection of accelerations and decelerations, in "A critical appraisal of fetal surveillance", van Geijn H.P. and Copray F.J.A. editors, Elsevier Science B.V., pp. 333-348, 1994.
[8] MOCZKO J.A. A comparative study of usefulness of quantitative parameters in the description of cardiotocogram records, Medical Informatics, 12, pp. 231-238, 1987.
[9] ROOTH G. Guidelines for the use of fetal monitoring, International Journal of Gynaecology and Obstetrics, 25, pp. 159-167, 1987.

This study was supported by the State Committee for Scientific Research, Warsaw, Poland (KBN Grant No. 4 T11E 006 22).