Classification of Fetal Heart Rate during Labour using Hidden Markov Models

George G. Georgoulas
Laboratory for Automation & Robotics
University of Patras
Patras, 26500, Greece
E-mail: georgoul@ee.upatras.gr

George Nokas
Wire Communications Laboratory
University of Patras
Patras, 26500, Greece
E-mail: nokas@george.wcl2.ee.upatras.gr

Chrysostomos D. Stylios
Department of Computer Science
University of Ioannina
Ioannina, 45100, Greece
E-mail: stylios@cs.uoi.gr

Peter P. Groumpos
Laboratory for Automation & Robotics
University of Patras
Patras, 26500, Greece
E-mail: groumpos@ee.upatras.gr

***Abstract*— Intrapartum Electronic Fetal Monitoring (EFM) is *an indispensable means for fetal surveillance. However, the* early enthusiasm was followed by scepticism, since the introduction of EFM in every day practise resulted in an increase in operative deliveries. Nevertheless the drawbacks of EFM relate not so much to the technique itself but more to the difficulties in reading and interpreting the Fetal Heart Rate (FHR). In an attempt to develop more objective means to analyse the FHR recordings and compensate for the different levels of expertise among clinicians, computerized systems have been developed during the last 2 decades. In this work we present an approach to automatic classification of FHR tracings belonging to hypoxic and normal newborns. The classification is performed using a set of parameters extracted form the FHR signal and two Hidden Markov Models (one for each class). The results are satisfactory indicating that the FHR convey much more information than what is conventionally used**

Index Terms— **Cardiotocography, Fetal Heart Rate (FHR), Hidden Markov Models (HMM)**

I. Introduction

Electronic Fetal Monitoring (EFM), also known as cardiotocography, has been widely used for antepartum and intrapartum fetal surveillance. Moreover, by the term cardiotocogram (CTG) we mean the continuous recording and monitoring of the instantaneous Fetal Heart Rate (FHR) (beats/min), which can be either obtained by Doppler ultrasound (the most common method employed during the antepartum period) or directly from the fetal electrocardiogram via scalp electrodes (during the intrapartum period and after the rapture of the membranes), and the Uterine Activity (UA), which is measured using an external tocodynamometer or an intra-uterine pressure catheter (mmHg) [1]. Fig. 1 shows a typical CTG segment with the FHR at the upper part of the figure and UA at the lower part. During the crucial period of labour, CTG is used primarily as a mean to avoid fetal and neonatal compromise, namely metabolic acidosis [2].

The FHR signal, especially, is considered as the subtlest component of the CTG. However, studies of FHR reliability have shown significant inter-observer and intra-observer variation in tracing interpretation [3], indicating that even though specific guidelines have been published for its interpretation [4],[5], the different levels of experience of the various specialists have catalytic influence on the final judgment.

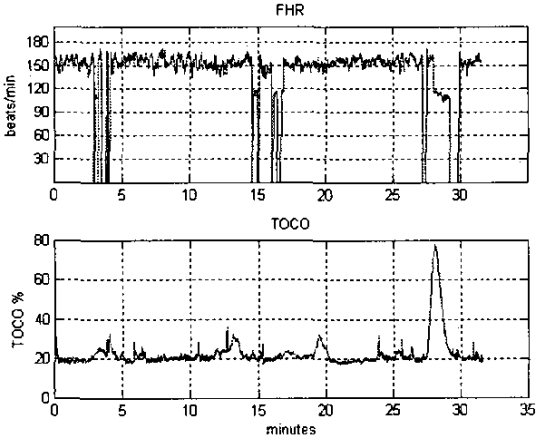


Fig. 1. A typical cardiotocogram (FHR at the upper part and UA at the lower part).

Moreover, the difficulty in distinguishing benign variant patterns from patterns associated with significant fetal acidemia may have arisen because FHR monitoring was introduced into clinical practice before the cause of FHR patterns was well understood [6].

0-7803-8359-1/04/$20.00 ©2004 IEEE

The inconsistency in interpretation and the increase of false positive diagnosis [3], on one hand, and the technological advances in computers along with new signal processing methods, on the other hand, have prompted many researchers to develop computer systems to analyze [7]-[12] or analyze and, additionally, classify the FHR signal [13]-[19].

Based on the belief that the FHR signal may convey much more information than what is usually interpreted by doctors, we propose a new method to discriminate fetal acidemia based on features extracted mathematically from the FHR signal. In this work, we employ the use of a set of features/observations, which have been successfully used for the antepartum case [17], slightly modified, for the classification of fetuses into two categories based on the value of their umbilical cord pH.

For the classification of the aforementioned categories, a new approach based on Hidden Markov Models (HMMs) is presented. More specific, two HMMs, one for each category are estimated using a set of training sequences. The system classifies a test sequence according to the model that best "describes" it.

This paper is structured as follows: section 2 is a brief introduction to HMMs; section 3 describes the processing steps of the FHR for the extraction of the features to be used by the HMM; and in section 4 the experimental results and some conclusions and ideas for future work are presented

II. Hidden Markov Models

HMM [20],[21] is a powerful and popular tool in pattern recognition. It can be viewed as two related stochastic processes that occur at the same time (Fig. 2). The first process produces a sequence of observed symbols (observations) and the second is an underlying process that consists of inter-connected states. Each observation in the first stochastic process relates to each state of the hidden layer by means of a probability distribution. Given a set of observations, it is not possible to determine the exact state sequence that produces these observations. That is, the underlying state sequence that is associated with an observed sequence is hidden and from that stems the name, Hidden Markov Model. HMMs are characterized as follows:

a. The number of states N in the model. The set of states is denoted as $\mathbf{s}=\{s_1, s_2, ..., s_N\}$ and "being in state" s_i at time t is denoted as $q_i(t)$.

b. The state probability transition matrix $\mathbf{A}=\{a_{ij}\}$, where

$$a_{ij}=p\left(q(t+1)\,|\,q(t)\right).$$

From the definition of $\mathbf{A}$ arises that:

$$\sum_{j=1}^{N} a_{ij}=1.$$

c. The output probability distribution $\mathbf{B}=\{b_j\}$, associated with each emitting state, where

$$b_j\left(\mathbf{y}(t)\right)=p\left(\mathbf{y}(t)\,|\,q_j(t)\right)$$

and $\mathbf{y}(t)$ is the feature vector (observation) at time t.

d. The initial state distribution π, where

$$\pi_i=p\left(q_i(1)\right)$$

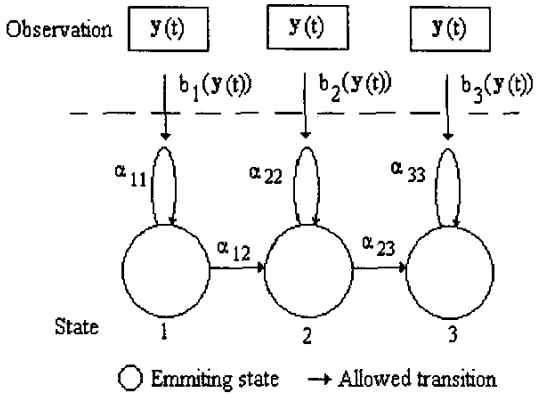


Fig.2. Emitted state left to right HMM

All standard HMMs are characterized by the aforementioned parameters. For the application of HMMs in a classification problem one or more models can be trained with data of a certain class. During classification, the corresponding class to a given set of parameters is identified from the model of the class that best fits the given parameters.

III. Processing Of FHR For Feature Extraction

A. Pre-processing-artifact removal

The fetal heart rate is a noisy signal due to the method that is used for its acquisition and also due to extraneous factors that cannot be isolated. Although the missing or "spiky" data do not create problems to simple eye inspection, they may lead to wrong results when further digital processing is going to take place. Thus, in order to remove "spiky" segments or segments where the signal is zeroed, a pre-processing stage of the FHR signal has to take place. The pre-processing stage introduced in [14], firstly detects a stable FHR segment, which is defined as a segment where the difference (in beats/min) between five adjacent samples is less than 10 beats/min. Whenever a difference between adjacent beats higher than 25 beats/min is found, a linear interpolation is applied between the first of those two signals and the first signal of a new stable FHR segment (Fig. 3).

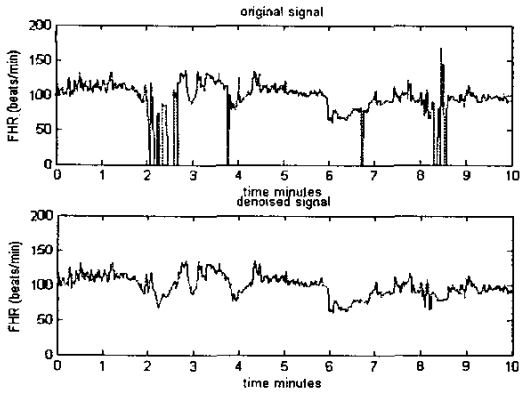


Fig. 3. Data before and after the removal of artefacts.

For this work we use 36 recordings from 36 pregnant women (38–42 weeks of gestation age). The FHR recordings have various lengths, ranging from 20 minutes to more than 1 hour. Thirty of them were acquired using a HP 1350 fetal monitor at a sampling frequency of 4 Hz, and 6 of them were acquired using a Toitu MT810B. In both cases, scalp electrodes were used for the acquisition. The latter cardiotocograms were irregularly sampled and we transform them into "pseudo-regularly" sampled signals. In order to do so, we copy the way HP 1350 operates. To be more specific, if a new beat is detected during the sampling interval (sampling interval=0.25 seconds), the cardiotocograph assigns the computed FHR value to the next output value; otherwise, it assigns the previous FHR value. Thus, from the irregularly sampled FHR signal we reconstruct the sequence of the detected beats (R-peaks) and then we created a regularly sampled FHR sequence the way HP 1350 would have done (Fig. 4).

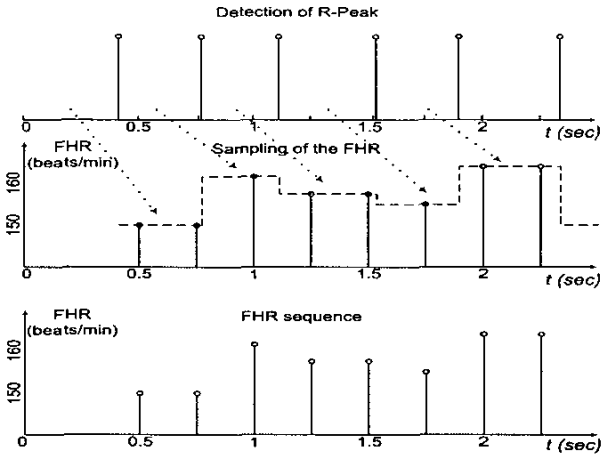


Fig. 4. Transformation of the irregular FHR to a "regularly" sampled sequence.

Due to the different duration of the recordings, we use segments of equal duration from each case and perform the subsequent analysis on these segments only. Therefore we crop, starting from the end of the recording (or as close to the end as possible), segments lasting 20 minutes (maximum duration of some recordings). The segments are chosen as close to delivery time as possible so as to avoid time-bias. The problem encountered is that because of the stress during the last stage of delivery, the last minutes of the recordings are completely "contaminated" by artifacts. Therefore, for some of the recordings, the very spiky 1-3 last minutes are not included in our data set.

B Feature extraction

Each 20-minute segment is divided to overlapping 5-minute segments. For each one of these segments, we employe a feature extraction stage in order to find a group of indexes that can characterize the fetal condition. Those indexes –features- are derived both from the time domain (following [17]) and also the frequency domain.

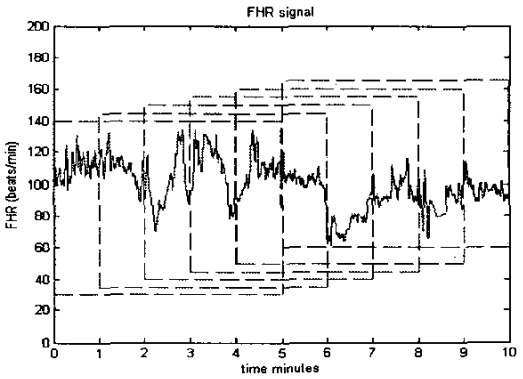


Fig. 5 Overlapping windows for the extraction of sequential features

The set of parameters and their definitions are as follows:

- Mean value of FHR signal
- Standard deviation of FHR signal
- $\text{Delta}=\dfrac{\sum_{i=1}^{m}\left[\max_{i\in m}\left(FHR(i)\right)-\min_{i\in m}\left(FHR(i)\right)\right]}{m}$

  where max and min are computed within each minute of the signal and m is the number of minutes
- $STV=\dfrac{\sum_{i=1}^{24}\left|sFHR(i+1)-sFHR(i)\right|}{24}$ where $sFHR(i)$ is the value of the signal $FHR(i)$ taken each 2.5 sec (i.e. once each 10 samples)

(Short Term Variability)

- $II = \frac{STV}{std\left[sFHR(i)\right]}$ (Interval Index)
- LTI is defined as the interquartile range $\left[\frac{1}{4}, \frac{3}{4}\right]$ of the distribution $m(j)$ with $m(j) = \sqrt{FHR^2(j) + FHR^2(j+1)}$ (Long Term Irregularity)
- $\text{Delta\_total} = \max(\text{FHR(i)}) - \max(\text{FHR(i)})$

These are the 7 parameters calculated in the time domain. In addition to them we also extracted 4 simple features from the frequency domain.

- Power at the range 0-0.5 Hz
- Power at the range 0.5-1 Hz
- Power at the range 1-1.5 Hz
- Power at the range 1.5-2 Hz

The aforementioned time domain parameters have been successfully used in the antepartum case, so it is reasonable to assume that they may also perform well in the intrapartum case. Regarding the frequency domain, it is common practice to analyze not the FHR signal itself but the beat-to-beat intervals that produce it. For this signal specific bands have been identified, reflecting different activities and fetal states [17]. In the case of the FHR, no standard band selection exists. Therefore, the selection of the frequency bands is done arbitrarily.

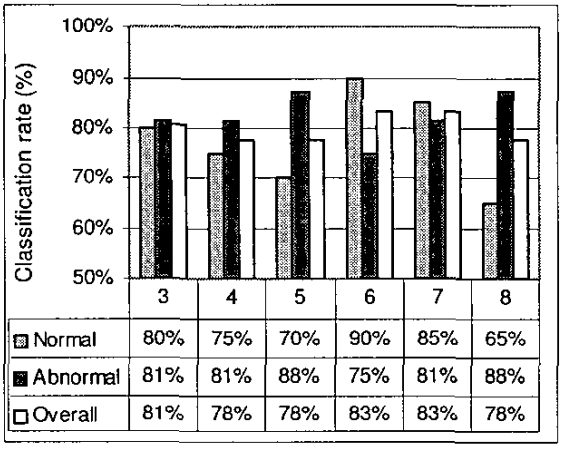
Moreover, conventional interpretation of CTG is based on certain morphological characteristics, according to the guidelines given in [4] or [5]. Of paramount importance is the baseline, which is the mean level of fetal heart rate when this is stable, accelerations and decelerations being absent and it is determined over a time period of 5 or 10 minutes [4]. By applying the algorithm described in [10], we calculate the baseline as the final feature to be used in the HMM. Therefore for each 20-minute FHR signal we extracted sixteen vectors with 12 elements each.

These features are subsequently fed to an HMM classifier in order to test whether we can find a way to discriminate between fetuses with "normal" pH values and those with decreased pH that are suspicious of developing acidemia. For the needs of this work we choose a cut-off value of 7.05 as the borderline/threshold between hypoxic and normal cases. With this convention, the 36 recordings are divided into two groups: the "normal" group containing 20 cases and the "abnormal-hypoxic" group with 16 cases.

Two HMMs, one for the "normal" case and one for the "hypoxic" case are estimated using the segmental k-means training algorithm [22]. Each model is a left to right Continuous Density Hidden Markov Model (CDHMM) with no state skip. The output distribution probabilities are modeled by means of a Gaussian component with diagonal covariance matrix

IV. Results-Conclusions

Because of the restricted number of cases, we use the multifold cross validation scheme [23] in order to evaluate the performance of the proposed methodology. In compliance with that scheme, we divide the 36 cases into 4 non-overlapping groups containing 9 cases each (5 normal and 4 hypoxic). Each time we exclude one of them from the training process and we use it only for testing the performance of the constructed HMM classifier. We repeate this procedure 4 times and we average the classification performances. Various configurations of HMMs (different number of hidden states) are tested and the results are summarized in the following bar-plot (Fig. 6).



	3	4	5	6	7	8
Normal	80%	75%	70%	90%	85%	65%
Abnormal	81%	81%	88%	75%	81%	88%
Overall	81%	78%	78%	83%	83%	78%

Fig. 6. Classification rates for different number of hidden states

As it can be seen we manage to have a maximum overall classification rate of 83% (for seven hidden states) having at the same time high classification rates both for the normal (85%) and the abnormal cases (81%). The same overall classification rate can be achieved with a model of six hidden states but with unbalanced performance between the two classes.

Our results seem to be comparable to those reported in [19], where with a cut-off point for the umbilical arterial blood pH set to 7.15 and in a population of 73 fetuses, (8 fetuses with pH less than 7.15 and 65 fetuses with pH more than 7.15) the developed system manages to classify 7 of the abnormal cases to the right class (87.5% classification rate) and 50 of the normal cases to the right class (76.92% classification rate). However, since the cut off value is different and due to the restricted number of cases, neither direct comparison can be made, nor direct conclusions can be drawn (however according to [24] if a lower value for arterial pH (7.05) is used (in conjunction with a base deficit in the extra-cellular fluid greater than or equal to 12, only one of the 23 patients identified as "abnormal" will fulfill the new criteria). Compared to the results reported in [18],

our results are slightly inferior (for a cut-off value of pH equal to 7.05, the classification rate for the normal cases is 100% and for the "risk" group 84.7%). However, in that work, apart from the FHR signal, measurements of functional oxygen saturation of fetal arterial blood is used, therefore no direct comparison can be made.

Moreover, it must be mentioned that the choice of the threshold for the pH value can be probably chosen lower for the "hypoxic" case. A more justified threshold would be the value of pH at 7, but this would compromise more the classification performance, since only 2 cases would fulfill that criterion, leaving 34 to the normal set. It is obvious that with this partition, overfitting would occur. It is worth mentioning that only very low pH values (6.8) are related to neonatal death or major neurological damage [25].

In conclusion, the results are encouraging, indicating that criteria can be found to discriminate normal from acidemic outcome, something, which was questionable in the early 90s [26]. However, as future work we propose the use the Apgar score as another index component for the formation of the classes. In this work, the data at our disposal prevented us from employing Apgar score as another criterion for the classification, since all Apgar scores but one are higher than 8. By including the Apgar score in the classification process, a more objective categorization may be achieved [27].

V. ACKNOWLEDGEMENTS

The authors would like to thank Prof. Joao Bernardes, Department of Gynecology and Obstetrics, Porto Faculty of Medicine, Porto, Portugal for providing us with the fetal heart data collected within the Research Project POSI/CPS/40153/2001, from Fundacao para a Ciencia e Tecnologia, Portugal.

REFERENCES

[1] M. C. Carter, "Present-day performance qualities of cardiotocographs," *Br. J. Obstet, Gynecol.*, vol. 100, Supplement 9, pp. 10-14, 1993

[2] H. P. V. Geijn, "Developments in CTG analysis," *Bailliers Clin. Obstet. Gynaecol.*, vol. 10, no 2, pp. 185-209, 1996

[3] D. MacDonald, A. Grant, M. Sheridan-Pereira, P. Boylan and I. Chalmers, "The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring," *Am. J. Obstet Gynecol.* vol. 152, pp. 524-39, 1985.

[4] G. Rooth, A. Huch and H. Hough, "Guidelines for the use of fetal monitoring," *Int. J. Gynecol. Obstet.*,vol. 25, pp. 159-67, 1987

[5] National Institute of Child Health and Human Development Research Planning Workshop, "Electronic fetal heart rate monitoring: Research guidelines for interpretation," *Am. J. Obstet Gynecol.*, vol. 177, pp. 1385-90, 1997.

[6] M. Fox, S. Kilpatrick, T. King and J. T. Parer, "Fetal heart rate monitoring: Interpretation and collaborative management," *Journal of Midwifery & Women's Health,* vol. 45, no 6, pp. 498-507, 2000.

[7] D.Arduini, G. Rizzo, G. Piana, A. Bonalumi, P. Brambilla and C. Romanini, "Computerized Analysis of Fetal Heart Rate: I. Description of the System (2CTG)," *J. Maternal Fetal Invest.*, vol. 3, pp. 159-163, 1993.

[8] R. Mantel, H.P. van Geijn, F. J. M Caron, J. M. Swartjes, E. E. van Woerden and H. W. Jongsma, "Computer analysis of antepartum fetal heart rate: 1. Baseline determination," *Int. J. Biomed. Comput.*, vol. 25, no. 2, pp. 261-272, 1990.

[9] Mantel, R., H.P. van Geijn, F.J.M Caron, J. M. Swartjes, E. E. van Woerden and H. W. Jongsma, "Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations," *Int. J. Biomed. Comput.*, vol. 25, no. 2, pp. 273-286, 1990.

[10] G. M. Taylor, G. J. Mires, E. W. Abel, S. Tsantis, T. Farrell, P. F. W. Chien and Y. Liu, "The development and validation of an algorithm for real time computerized fetal heart rate monitoring in labour," *Br. J. Obstet. Gynaecol.*, vol, 107, pp. 1130-1137, 2000.

[11] J. Jezewski, and J Wrobel, "Foetal monitoring with automated analysis of cardiotocogram: The KOMPOR system," in Proc. *15th Ann. Card. IEEE/EMBS*, San Diego, CA, pp. 638-639, 1993.

[12] K. Maeda, "Computerized analysis of cardiotocograms and fetal movements," *Bailliers Clin. Obstet. Gynaecol.*, vol. 4, no. 4, pp. 1797-813, 1990.

[13] B. G. Berdinas, A. A. Betanzos and O. F. Romero, "Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system," *Artificial Intelligence*, vol. 136, pp. 1-27, 2002.

[14] J. Bernardes, C. Moura, J.P.M. de Sa and L.P. Leita, "The Porto system for automated cardiotocographic signal analysis," *J. Perinat. Med.*, vol. 19, pp. 61-65, 1991.

[15] G. S. Dawes, M. Moulden, and C. W. Redman, "Computerized analysis of antepartum fetal heart rate," *Amer. J. Obstet. Gynecol.*, vol. 173, no. 4, pp. 1353-1354, 1995.

[16] J. F. Skinner, J. M. Garibaldi and E. C. Ifeachor, "A Fuzzy System for Fetal Heart Rate Assessment," *In Proc of the 6th Fuzzy Days Conference*, Dortmund, Germany, pp. 20-29, 1999

[17] G. Magenes, M.G Signorini and D. Arduini, "Classification of cardiotocographic records by neural networks. *D.* Neural Networks," In Proceedings of the IEEE-INNS-ENNS International Joint Conference, vol. 3 pp. 637-641, 2000

[18] E. Salamalekis, P. Thomopoulos, D. Giannaris, I. Salloum, G. Vasios, A. Prentza and D. Koutsouris, "Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks," *Br. J. Obstet. Gynaeco.*, vol. 109, no. 10, pp. 1137-1142, 2002.

[19] T. K. H. Chung, M. P. Mohajer, X. J. Yang, A. M. Z. Chang and D.S. Sahota, The prediction of fetal acidosis at birth by computerized analysis of intrapartum cardiotocography. *Br. J. Obstet. Gynaecol.*, vol. 102, pp. 454-460, 1995.

[20] S. W. Dharmadhikari, "Functions of finite Markov Chains," *Annals of Mathematical Statistics*, vol. 34 pp. 1022-1032, 1963.

[21] C. J. Burke, M. Rosenblatt, "A Markovian function of a Markov Chain", Annals of Mathematical Statistics, 1958.

[22] M J Hunt and C Lefebre. A comparison of several acoustic representations for speech recognition with degraded and undegraded speech. In Proceedings ICASSP, pages 262-265, 1989.

[23] S. Haykin, *Neural Networks: A Comprehensive Foundation.* 2nd ed. Englewoo Cliffs, NJ: Prentice Hall, 1999.

[24] K. R. Greene, "Intelligent fetal heart rete computer systems in intrapartum surveillance," in Curr. Opin. Obstet. Gynecol, vol. 8, pp. 23-127, 1996.

[25] J. T. Parer, *Handbook of fetal heart rate monitoring*, Philadelphia, Pennsylvania: W. B. Saunders Company, 1997.

[26] G. S. Dawes, "Computerized fetal heart rate analysis" in A critical appraisal of fetal surveillance eds. E. P. van Geijn and F. J. A, Copray: Elsevier Science, 1994.

[27] J. Bernardes, D. A. de Campos, A. C. Pereira, L.P. Leita and A. Garrido, "Objective computerized fetal heart rate analysis," *Int. J. Gynecol. Obst.*,vol. 62 pp. 141-147, 1998.