Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Linear models of intrapartum uterine pressure-fetal heart rate interaction for the normal and hypoxic fetus

Philip A. Warrick[1,4], Robert E. Kearney[1] Doina Precup[2] and Emily F. Hamilton[3,4]

***Abstract*— We construct input-output models by linear system-identification methods for uterine pressure - fetal heart rate data collected during labour and delivery. Using standard hypothesis tests, the impulse response model coefficients show statistically significant differences between normal and pathological cases.**

I. INTRODUCTION

The difficulties of visual cardiotocography (CTG) interpretation have been discussed in many previous clinical and technical studies: the sensitivity is clinically useful but the low specificity can increase cesarean section rates [1]. We would like to use automated methods to model the maternal-fetal interaction available via CTG and eventually use these models to improve the differential diagnosis of the fetus during labour.

It is well known that the primary physiological mechanisms for fetal heart rate (FHR) decelerations are: 1) contraction-induced umbilical-cord compression and 2) contraction-related decreases in oxygen delivery through an impaired utero-placental unit. Furthermore there is a general consensus that deceleration frequency and timing with respect to contractions can be an indicator of the ability of the fetus to withstand these types of insults. Hypothesis-driven modeling from these facts would focus on contraction-deceleration detection and gross estimates of timing between these events. This has been the approach in numerous CTG studies [2]–[4].

However, it is also possible to direct attention to the interactions between the CTG signal pair of uterine pressure (UP) and FHR, which can be viewed from the perspective of maternal stimulus and fetal response. It is natural to model this signal arrangement as an input-output system where the fetus is the 'system' that senses the UP and reacts with changes to the FHR. In this way, contractions, decelerations and their temporal relationships (and possibly other phenomena) are implicitly rather than explicitly modelled. This is the approach of our study and to our knowledge it has not yet been applied to the FHR response to contractions.

Corresponding author: Philip Warrick (e-mail: philip.warrick@mcgill.ca).
The authors acknowledge the financial support of this work by LMS Medical Systems, Ltd.
[1]Biomedical Engineering Department, McGill University, Montreal, Quebec, Canada. [2]School of Computer Science, McGill University, Montreal, Quebec, Canada. [3]Department of Obstetrics and Gynecology, McGill University, Montreal Quebec, Canada. [4]LMS Medical Systems, Inc., Montreal, Quebec, Canada.

In hypothesis-driven (or 'deductive') modelling, accepted physiological models drive the search for features correlating to those models. On the other hand, our 'data-driven' (or 'inductive') approach directly models the coupling of the signals and in so doing introduces fewer *a priori* assumptions on the information content of the data. The two approaches should be considered as two poles of a continuous spectrum since the hypotheses can provide reasonable starting points for the inductive analysis (e.g. an appropriate timescale for local analysis in a non-stationary environment). In turn, the data-driven and relatively unbiased knowledge can be used to inform hypothesis-driven research [5].

The investigation of this paper compares input-output models of UP and FHR (see Fig. 1) for normal cases to those of suspected hypoxia. The models are constructed from a database of UP-FHR pairs using linear system-identification techniques to determine the system impulse-response functions (IRFs) over several hours of data collection. We then average the models over time and compare normal and pathological cases.

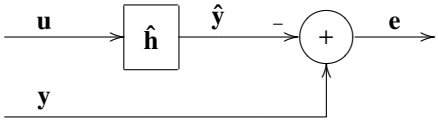

Fig. 1. General linear input-output system modelled by impulse-response estimate $\hat{\mathbf{h}}$ with residual error signal **e**. The inputs, desired and estimated outputs are **u**, **y** and $\hat{\mathbf{y}}$, respectively. In this study **u** and **y** correspond to the UP and FHR signals.

II. METHODS

A. Data

The database consisted of 161 intrapartum CTG tracings (762 hrs) for pregnancies having a birth gestational age greater than 36 weeks and having no known genetic malformations. The FHR was acquired at f_S = 4Hz while the UP was acquired at 1Hz and up-sampled to 4Hz. The examples were labelled by outcome according to their arterial umbilical-cord base deficit (BD) and neonatal indications of severe neurological impairment. Base deficit is considered an important marker for hypoxia leading to intrapartum asphyxia with metabolic acidosis [1], [6]. There was an approximately equal distribution of normal cases (56 'D': BD < 8), intermediate cases (56 'C': BD ≥ 8) and severely

compromised cases (49 'A': BD ≥ 12, compromised neurological function). The letter labels 'A', 'C', and 'D' are our own internal labelling scheme.

B. Preprocessing

The FHR and UP signals may be temporarily interrupted by loss of sensor contact, and the FHR may be corrupted by maternal heart-rate interference. We used a Schmitt trigger to detect such segments, and sequential suspect segments in close proximity (4 samples) were merged. If the segment was less than 15s, the segment was bridged by linear interpolation; if it was greater than 5min, the segment was removed from consideration. If the segment length was between these limits, it was retained. We also removed segments having negligible slope or variance which were clearly non-physiological. At this stage the acceptance criteria were rather loose and the main goal was to repair short gaps and to remove the most obvious longer gaps (which can last several hours). More filtering for artifact was done based on the success of the system-identification step, described later.

Evidently, the FHR is influenced by a number of other (unobserved) physiological factors besides UP, making the proposed model only a partial reflection of reality. In fact, to reduce the impact of these other factors on the system-identification process, the signals were detrended by a high-pass filter that would pass a reasonably long contraction or deceleration (cutoff frequency $f_{HP} = \frac{1}{220s} = 4.5 \times 10^{-3}$Hz). Also, our initial studies showed that the UP-FHR interaction manifests itself predominantly at lower frequencies and so the signals were decimated by a factor $D_S = 32$ to improve the signal-to-noise (SNR) ratio of the impulse-response estimate. Modelling the filtered and residual FHR that was not linearly predictable by the UP will be addressed in a subsequent phase of our research.

C. System identification

In the general depiction of system identification in Fig. 1, the estimated system $\hat{\mathbf{h}}$ responds to the input UP signal $\mathbf{u}$ and forms an estimate of the output FHR $\hat{\mathbf{y}}$. When the estimate is imperfect, the residue signal $\mathbf{e}$ is generated, where $\mathbf{e}$ is defined as the difference between the true and estimated FHR signals $\mathbf{y} - \hat{\mathbf{y}}$. A quality measure of the estimated model is the *percent variance accounted for* (%VAF) defined as

$$\%\text{VAF} = 100 \times \left(1 - \frac{\sigma_e^2}{\sigma_y^2}\right) \tag{1}$$

where σ_e^2 and σ_y^2 are the variances of the residue and desired signals, respectively. Lower residual energy thus corresponds to higher %VAF values.

A linear *dynamic* model of the finite-impulse-response (FIR) type is one in which the output estimate $\hat{y}_n$ at sample n, is a function of M previous inputs $\mathbf{u}_n^T = [u_n\, u_{n-1}\, \ldots\, u_{n-M+1}]$. The model is dynamic in the sense that it depends on the recent history of inputs; more generally, a dynamic model can depend on the history of both inputs and outputs. Ideally the desired output y_n can be written as a linear convolution of the inputs with some unknown vector $\mathbf{h} = [h_0\, h_1\, \ldots\, h_{M-1}]^T$:

$$\begin{aligned} y_n &= \mathbf{h} * \mathbf{u}_n \\ &= h_0 u_n + h_1 u_{n-1} + \cdots + h_{M-1} u_{n-M+1} \end{aligned} \tag{2}$$

If it is assumed more realistically that the output is corrupted by the measurement-noise signal $\mathbf{v}$, the measured output can be written as $\mathbf{z} = \mathbf{y} + \mathbf{v}$. Better estimates of $\mathbf{h}$ can be calculated over a time period with multiple output samples, written succinctly in matrix form for an analysis of N input and output samples as $\mathbf{z} = \mathbf{U}\mathbf{h} + \mathbf{v}$ where

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} \quad \text{and} \quad \mathbf{U} = \begin{pmatrix} u_1 & 0 & \ldots & 0 \\ u_2 & u_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_N & u_{N-1} & \ldots & u_{N-M-1} \end{pmatrix} \tag{3}$$

If it is further assumed that the additive noise $\mathbf{v}$ has zero mean, the least-squares estimate of $\mathbf{h}$ is then given by:

$$\begin{aligned} \hat{\mathbf{h}} &= (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{z} \\ &\approx \boldsymbol{\Phi}_{\boldsymbol{uu}}^{-1}\boldsymbol{\phi}_{\boldsymbol{uz}} \end{aligned} \tag{4}$$

where, for $N \gg M$, $\mathbf{U}^T\mathbf{U}$ (the Hessian $\mathbf{H}$) and $\mathbf{U}^T\mathbf{z}$ are approximated by the input autocorrelation matrix $\boldsymbol{\Phi}_{\boldsymbol{uu}}$ and the input-output cross-correlation $\boldsymbol{\phi}_{\boldsymbol{uz}}$, which are readily calculated [7], [8].

D. Analysis Parameters

The parameter N of the calculation for IRF $\hat{\mathbf{h}}$ defines the length of the analysis window. A larger value of N improves the SNR of the estimate. On the other hand, artifact and non-stationarities in the signal are likely to increase model degradation for larger N. We initially chose a value of N corresponding to a $T_N = 20$min period on the grounds that it was twice the length of a very long 10min deceleration (in fact decelerations are more typically in the 15s - 5min range). It also permitted the analysis of a reasonable number of artifact-free windows for each of the 161 CTGs. Typically there were between 20 and 30 analysis windows per case.

The IRF parameter M corresponds to the length of the input history (or lag) used to estimate the output. To estimate the optimal value of M, the quality of the models were first compared (by their VAFs) using several experimental values of M and a very long N (i.e. the entire CTG tracing). Then the model quality was compared for various values of M using the shorter N given above. Finally the performance of the discrimination (described below) for various values of M were compared. Using all these results, we choose the value of M corresponding to $T_M = 8$min. This value was reasonable given the typical range of deceleration durations. We may return to these selection criteria for T_N and T_M in a more rigorous way in future work.

Finally, to facilitate model comparison, the model for each window was normalized and compressed in order to make the model invariant to scale and reduce its dimensionality. The UP amplitude is not known absolutely in standard CTG measurements due to intra- and inter-subject variability in both the pressure-sensor contact quality and abdominal-tissue thickness [9]: as a result an unknown scaling factor will be introduced into the estimate $\hat{\mathbf{h}}$. To reduce the impact of this scaling, the IRF was normalized to set the total energy (variance) of $\hat{\mathbf{h}}$ to unity. The Euclidean norm $\|\hat{\mathbf{h}}\|$ was used to normalize the coefficients of $\hat{\mathbf{h}}$:

$$\sqrt{\sum_{\tau=0}^{T-1}\left(\frac{\hat{h}(\tau)}{\|\hat{\mathbf{h}}\|}\right)^2} = 1 \quad \text{where} \quad \|\hat{\mathbf{h}}\| = \sqrt{\sum_{\tau=0}^{T-1}\hat{h}^2(\tau)} \qquad (5)$$

The normalized model $\hat{\mathbf{h}}_N = \frac{1}{\|\hat{\mathbf{h}}\|}\hat{\mathbf{h}}$ was decimated by a factor $D_M = 4$ to reduce the number of model coefficients further. The resulting IRF model $\hat{\mathbf{h}}_{ND}$ contained $M = \frac{T_M f_S}{D_S D_M} = 15$ coefficients.

III. RESULTS

We removed intervals where the system identification failed, which occurred for approximately 5% of the 4073 analysis windows. These were generally caused by artifacts that had not been filtered at the preprocessing stage.

Fig. 2 shows a typical result for UP-FHR identification over several 20min analysis windows. For this 'A' case, the normalized impulse responses $\hat{\mathbf{h}}_N$ for each window are similar to each other and to the average model for the entire tracing. The %VAF values for these three windows were 85.0, 88.5 and 86.3. The %VAF values varied considerably across the 161 cases, with pathological cases tending to have higher VAFs (see the distributions of %VAF for each outcome type in Fig. 4).

Fig. 3 shows the average normalized and decimated impulse responses $\hat{\mathbf{h}}_{ND}$ across all models of all examples (after decimation by D_M) for each outcome type. The pathological 'A' and 'C' cases exhibit a deeper initial trough compared to the the normal 'D' cases and tend to remain below thereafter. As well, they display longer recovery times before reaching the near-zero state that indicates the time lag of the system. Note that the time $t = 0$ response was removed by the decimation step although we expect that it is also informative.

Fig. 4 shows the distributions of the IRF coefficients for each outcome type. As suggested by Fig. 3, the differences in distributions were most pronounced for the earlier coefficients. Other than the 'A' distribution, the distributions of the final coefficient h_{15} tend to diverge from a Gaussian shape, more so for the 'D' cases than the 'C' cases. This can also be seen in the sharp decrease in the IRF for the last coefficient in Fig. 3, making this coefficient suspect for 'D' cases. We expect that this outlier will be accounted for in future work.

To examine the statistical significance of the differences in the IRF coefficients and overall %VAF, we performed several types of hypothesis tests comparing the pathological and intermediate examples (the union of the 'A' and 'C' groups) to normal examples (those in the 'D' group). The results are tabulated in Table I. The t-test tests the hypothesis that two samples from Gaussian distributions of equal variance have the same mean. The Wilcoxon rank sum and Kruskal-Wallis (K-W) tests assess the hypothesis that two samples from arbitrary distributions have the same median. The Kolmogorov-Smirnov (K-S) test addresses the hypothesis that the distributions of two samples are the same. In the table, asterisks indicate that the null hypothesis (no difference) can be rejected ($p < 0.01$).

All tests were in agreement that four of the first five coefficients h_1, h_2, h_3, h_5 and the %VAF were significantly different across the classes. The tests were substantially in agreement that h_8 and h_9 were significantly different. The later coefficients $h_{10} - h_{15}$ were not found to be significantly different, suggesting that the time lag for most of the discriminating system energy in the sample population for the selected frequency band ($\frac{1}{220s} < f < \frac{1}{32s}$) is approximately 10 IRF samples, or $10 \times 32\text{s} = 320\text{s}$.

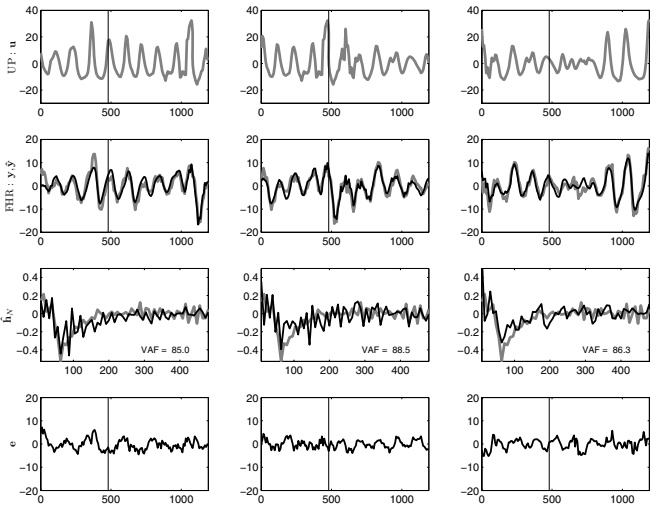

Fig. 2. Typical system identification results for an 'A' case. The three columns are successive 20min analysis windows with 10min overlap. The first and second rows are the UP and FHR, respectively. The true FHR $\mathbf{y}$ and estimated FHR $\hat{\mathbf{y}}$ are shown as thick grey and thin black lines, respectively. In the third row, normalized impulse responses $\hat{\mathbf{h}}_N$ averaged over the entire CTG and for the 20min window are shown in thick grey and thin black lines, respectively. The fourth row shows the residue signal $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. The vertical bars indicates the duration of the IRF lag. The indicated VAFs are calculated over the estimates after the lag, to eliminate filtering end effects.

IV. DISCUSSION

The impulse responses obtained from the correlation technique contain considerable noise. This may be partially attributable to the low ratio $N/M = \frac{20min}{8min}$; it is normally

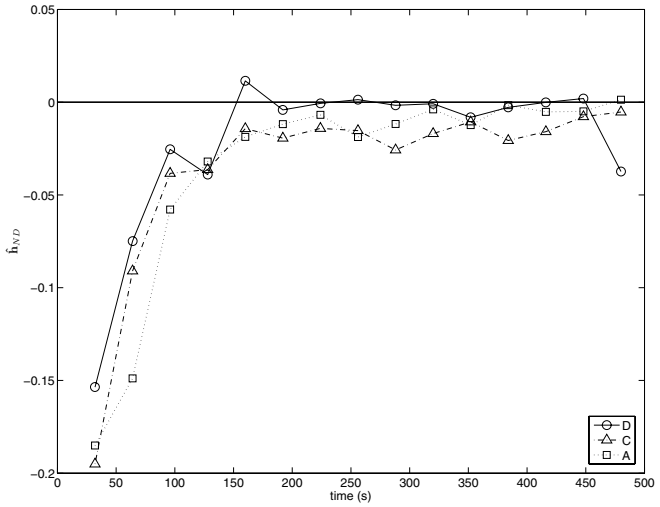

Fig. 3. Average normalized and decimated IRF $\hat{\mathbf{h}}_{ND}$ for each outcome class 'D' (normal), 'C' (intermediate), and 'A' (pathological).

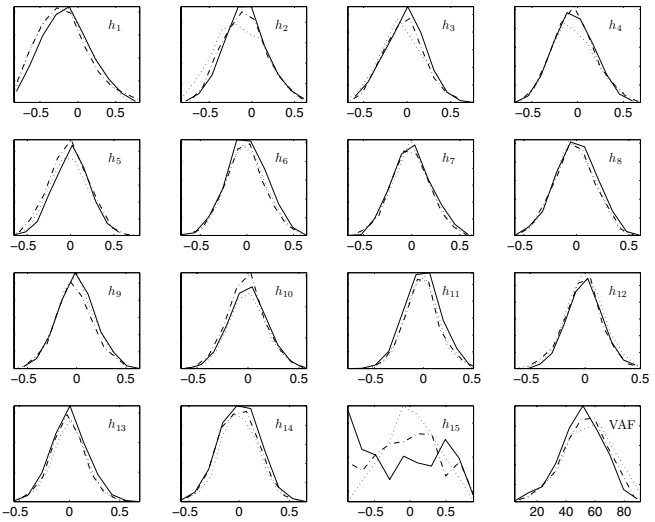

Fig. 4. Histograms of IRF $\hat{\mathbf{h}}_{ND}$ coefficients and overall %VAF for each outcome class (plotted as relative frequency versus value). The legend of Figure 3 is also applicable to this figure.

	t-Test	Rank sum	K-W	K-S
h_1	4.48e-04*	1.03e-04*	1.03e-04*	5.99e-04*
h_2	5.82e-06*	1.81e-05*	1.81e-05*	1.72e-07*
h_3	6.50e-03*	1.34e-03*	1.34e-03*	2.12e-04*
h_4	5.54e-01	8.74e-01	8.74e-01	7.76e-01
h_5	9.79e-05*	3.84e-05*	3.84e-05*	1.28e-04*
h_6	8.02e-02	4.80e-02	4.80e-02	8.05e-02
h_7	1.04e-01	1.95e-01	1.95e-01	4.57e-01
h_8	4.19e-03*	3.91e-03*	3.91e-03*	2.20e-02
h_9	2.87e-03*	1.51e-03*	1.51e-03*	1.16e-03*
h_{10}	6.85e-02	5.26e-02	5.26e-02	1.59e-01
h_{11}	5.86e-01	4.44e-01	4.44e-01	4.15e-01
h_{12}	6.43e-02	1.30e-01	1.30e-01	2.17e-01
h_{13}	4.23e-02	1.08e-01	1.08e-01	2.54e-02
h_{14}	1.68e-01	1.52e-01	1.52e-01	7.42e-02
h_{15}	4.74e-02	5.24e-02	5.24e-02	5.62e-04*
%VAF	1.75e-08*	1.12e-08*	1.12e-08*	7.21e-08*

TABLE I

HYPOTHESIS TEST RESULTS FOR THE SYSTEM IDENTIFICATION PARAMETERS AND VAF VALUE. ASTERISKS INDICATE SIGNIFICANCE LEVEL $p < 0.01$

desirable that this ratio be greater than 10. We intend to improve the IRF signal-to-noise ratio in future work using known noise-reduction techniques such as the pseudo-inverse method [8].

The finding that models of pathological cases tend to have higher %VAF is consistent with clinical expectation: a fetus in distress will tend to be less resistent to the insults of labour, especially the highly compressive contraction events. As a result, their neural compensatory mechanisms may be compromised, causing them to closely follow the onslaught of the stimulus rather than to resist it. In other words, their response is a more predictable phenomenon that can be modelled more precisely. In addition, the deeper initial IRF trough of pathological cases is consistent with the known clinical fact that pathological cases tend to have deeper decelerations. The plot of Fig. 3 highlights the difficulty of the discrimination problem in that healthy fetuses also react initially to this stimulus, but to a slightly lesser degree on average. Similarly, the longer time lag of the pathological cases is consistent with the fact that a compromised neural system may require more time to recover to steady state (recognized by clinicians during 'late' decelerations).

The statistically significant differences between outcome class models indicate their good potential as discriminators. Furthermore, in this study these differences were diminished by time averaging: the closer to delivery that deteriorations in fetal health occur, the less they are reflected in the average model. Thus the model may be less discriminating for these cases. If instead the models reflect conditions with higher time resolution while at the same time maintaining sufficient SNR–often conflicting requirements–the potential for discrimination is even greater than this study suggests. These research questions will be pursued in future work.

REFERENCES

[1] J. Low, R. Victory, and E. Derrick, "Predictive value of electronic fetal monitoring for intrapartum fetal asphyxia with metabolic acidosis," *Obstet Gynecol*, vol. 93, pp. 285–291, 1999.

[2] J. Skinner, J. Garibaldi, J. Curnow, and E. Ifeachor, "Intelligent fetal heart rate analysis," in *Advances in Medical Signal and Information Processing, 2000. First International Conference on (IEE Conf. Publ. No. 476)*, 2000, pp. 14–21.

[3] F. Lunghi, G. Magenes, L. Pedrinazzi, and M. Signorini, "Detection of fetal distress though a support vector machine based on fetal heart rate parameters," in *Computers in Cardiology, 2005*, 2005, pp. 247–250.

[4] H. Cao, D. Lake, I. Ferguson, J.E., C. Chisholm, M. Griffin, and J. Moorman, "Toward quantitative fetal heart rate monitoring," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 1, pp. 111–118, 2006.

[5] V. Z. Marmarelis, *Nonlinear Dynamic Modeling of Physiological Systems*. Hoboken, NJ: Wiley Interscience, 2004.

[6] A. MacLennan, "A template for defining a causal relation between acute intrapartum events and cerebral palsy: international consensus statement," *BMJ*, vol. 319, no. 7216, pp. 1054–1059, 1999.

[7] I. W. Hunter and R. E. Kearney, "Two-sided linear filter identification," *Medical & Biological Engineering & Computing*, vol. 21, pp. 203–209, 1982.

[8] D. T. Westwick and R. E. Kearney, *Identification of nonlinear physiological systems*. Hoboken NJ: Wiley-Interscience, 2003.

[9] T. Vanner and J. Gardosi, "Intrapartum assessment of uterine activity," *Bailliere's Clinical Obstetrics And Gynaecology*, vol. 10, no. 2, pp. 243–257, June 1996.