System-Identification Noise Suppression for Intra-Partum Cardiotocography to Discriminate Normal and Hypoxic Fetuses

PA Warrick[1,4], RE Kearney[1], D Precup[2], EF Hamilton[3,4]

[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

Abstract

We construct linear system-identification models of cardiotocography (CTG) data collected during labour and delivery. The models are the impulse response functions (IRFs) of the input-output system relating the uterine pressure (UP) stimulus to the fetal heart rate (FHR) response. We compare models obtained with and without applying noise suppression via the pseudo inverse technique. Finally, to determine the ability of the models to discriminate healthy from hypoxic fetuses, we use the average models as feature vectors of a support-vector-machine (SVM) classifier.

Applying the pseudo-inverse resulted in cleaner models with lower variance accounted for (VAF), likely indicative of reduced overfitting. The area under curve of the receiver-operator characteristic (ROC) without applying pseudo-inverse was 0.695 $\pm$ 0.054. Similar results over a useful operating range of false-positive rates were observed with the pseudo-inverse applied.

1. Introduction

The difficulties of visual 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 FHR decelerations are: 1) contraction-induced umbilical-cord compression and 2) contraction-related decreases in oxygen delivery through an impaired uteroplacental 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 directly model the input-output interaction by system identification. We first applied this approach in a previous study [5] by calculating the impulse response functions (IRFs) of the system relating the UP stimulus to the FHR response. The correlation approach used in that study is known to be susceptible to noise [6]. In this study, we attempt to reduce model noise by applying the pseudo inverse technique. Finally, to determine the ability of these models to discriminate healthy from hypoxic fetuses, we use time-averaged models as feature vectors of an SVM classifier.

2. Methods

2.1. Data

As in [5] 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 $f_S = 4$Hz 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. 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.

2.2. Preprocessing

As described in [5], loss of sensor contact and maternal heart-rate interference produce signal artifact that can

degrade modelling. Using a Schmitt trigger to detect these segments, we merge, bridge and exclude them as necessary to obtain cleaner signals. The signals are then detrended by a high-pass filter (cutoff frequency $f_{HP} = \frac{1}{220s} = 4.5 \times 10^{-3}$Hz) in preparation for signal identification. Finally, the signals were decimated by a factor $D_S = 32$ to a sampling rate of $f_{SD} = 0.125$Hz, since most UP-FHR interaction occurs in the lower frequency spectrum.

2.3. System identification

This section will summarize the techniques [6–8] we use to calculate the IRFs by least-squares and by pseudo inverse for noise suppression.

2.3.1. Linear least-squares

An output $\mathbf{y}$ corrupted by the zero-mean measurement-noise signal $\mathbf{v}$ can be written as the linear convolution of the input matrix $\mathbf{U}$ with the unknown IRF $\mathbf{h}$:

$$\begin{aligned} \mathbf{z} &= \mathbf{y} + \mathbf{v} && (1)\\ &= \mathbf{U}\mathbf{h} + \mathbf{v} && (2)\end{aligned}$$

where for N samples and IRF lag M,

$$\mathbf{z} = [z_1, z_2, \dots z_N]^T \quad (3)$$

$$\text{and} \quad \mathbf{U} = \begin{bmatrix} u_1 & 0 & \dots & 0 \\ u_2 & u_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_N & u_{N-1} & \dots & u_{N-M-1} \end{bmatrix} \quad (4)$$

The least-squares estimate of $\mathbf{h}$ is then given by:

$$\begin{aligned} \hat{\mathbf{h}} &= (\mathbf{U^T U})^{-1}\mathbf{U^T z} && (5)\\ &\approx \boldsymbol{\Phi}_{\boldsymbol{uu}}^{-1}\boldsymbol{\phi}_{\boldsymbol{uz}} && (6)\end{aligned}$$

where, for $N \gg M$, $\mathbf{U^T U}$ and $\mathbf{U^T z}$ 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.

2.3.2. Pseudo-inverse

To examine the noise characteristics of the the IRF estimate, we look at the factors $(\mathbf{U^T U})^{-1}$ and $\mathbf{U^T z}$ in eqn. 5 separately. The Hessian $\mathbf{H} = \mathbf{U^T U}$ is a positive definite matrix and will therefore have the singular value decomposition (SVD) and inverse

$$\mathbf{H} = \mathbf{V S V^T} \quad \text{and} \quad \mathbf{H^{-1}} = \mathbf{V S^{-1} V^T} \quad (7)$$

respectively. Furthermore, from eqn. 1,

$$\begin{aligned} \mathbf{U^T z} &= \mathbf{U^T U h} + \mathbf{U^T v} \\ &= \mathbf{V S}\zeta + \mathbf{V}\eta && (8)\end{aligned}$$

where $\zeta = \mathbf{V^T h}$ and $\eta = \mathbf{V^T}(\mathbf{U^T v})$ are the projections of the IRF $\mathbf{h}$ and the input-noise cross-correlation $\mathbf{U^T v}$ onto the Hessian eigenvectors $\mathbf{V}$. Substituting eqns. 7 and 8 into eqn. 5,

$$\begin{aligned} \hat{\mathbf{h}} &= \mathbf{V S^{-1} V^T}(\mathbf{V S}\zeta + \mathbf{V}\eta) \\ &= \mathbf{V}\zeta + \mathbf{V S^{-1}}\eta \\ &= \sum_{i=1}^{M}(\zeta_i + \frac{\eta_i}{s_i})\mathbf{v_i} && (9)\end{aligned}$$

Eqn. 9 shows that small eigenvalues s_i can amplify their associated noise terms n_i and corrupt the model term ζ_i when projected onto the i-th eigenvector $\mathbf{v_i}$. The condition number $\frac{s_1}{s_M}$ determines the relative impact of this noise. Thus limiting the summation to the most significant eigenvalues should improve model SNR (signal-to-noise ratio).

We use the minimum description length (MDL) to choose the number of retained terms. The MDL criteria considers both model parsimony and fidelity:

$$\text{MDL}(m) = \left[1 + \frac{m\log(N)}{N}\right]\sum_{i=1}^{N}[y(i) - \hat{y}(i,m)]^2 \quad (10)$$

where m is the number of singular values retained. Since y and the error $e = y - \hat{y}$ are orthogonal in the least-squares solution, the summation in eqn. 10 is the error energy $\xi_e = \xi_y - \xi_{\hat{y}}$. To efficiently calculate the MDL for increasing m, we note that $\xi_y = \sigma_y^2$ and that $\xi_{\hat{y}}$ can be written in terms of the SVD of the Hessian:

$$\begin{aligned} \xi_{\hat{y}} &= \frac{1}{N}\sum_{i=1}^{N}\hat{y}^2(t) \\ &= \hat{\mathbf{h}}^T\boldsymbol{\Phi}_{\boldsymbol{uu}}\hat{\mathbf{h}} \\ &= \hat{\mathbf{h}}^{\mathbf{T}}\mathbf{V S V^T}\hat{\mathbf{h}} \\ &= \sum_{i=1}^{M} s_i\left(\mathbf{v_i^T}\hat{\mathbf{h}}\right)^2 && (11)\end{aligned}$$

Therefore eqn. 10 can be rewritten as

$$\text{MDL}(m) = \left[1 + \frac{m\log(N)}{N}\right]\left[\sigma_y^2 - \sum_{i=1}^{m} s_i\left(\mathbf{v_i^T}\hat{\mathbf{h}}\right)^2\right] \quad (12)$$

2.4. Model processing

The calculation and processing of the IRF models are described in detail in [5]. Briefly, the parameter N of the calculation for IRF $\hat{\mathbf{h}}$ defines the length of the analysis window. We chose N corresponding to $T_N = 20$min, a value as large as possible without incurring undue model degradation due to artifact and non-stationarity. The IRF

	NPI-VAF	Success	PI-VAF	Success
A	55.9 ± 17.1	92.0	40.1 ± 22.3	91.3
C	53.0 ± 16.1	95.5	35.1 ± 20.5	94.2
D	50.9 ± 15.8	96.7	32.4 ± 19.2	94.6
All	52.9 ± 16.4	95.1	35.3 ± 20.7	93.6

Table 1. System identification results without noise suppression (NPI) and with the pseudo inverse applied (PI). For each of the classes as well as the overall results, the mean and standard deviation of the VAFs and the rate of successful identification are shown.

parameter M corresponds to the length of the input history (or lag) used to estimate the output. We chose a value of M corresponding to $T_M = 8$min based on model quality as measured by variance accounted for (VAF) and discrimination performance. Finally, to facilitate model comparison across subjects, the model for each 20-min window was normalized for scale invariance, decimated for model reduction (to 15 coefficients) and then time averaged to obtain an overall model for the entire CTG tracing.

2.5. SVM classification

Support-vector classification was chosen for classification because it tends to achieve state-of-the-art classification perfomance, even in high-dimensional input spaces, and has built-in adjustment to avoiding overfitting. It is based on polynomial-time exact optimization rather than approximate methods (e.g. gradient descent). Non-linear classification occurs because of an efficient kernel mapping of the input space to a higher-order space. In this kernel space the best linear classifier is found which maximizes the distance from the boundary to the closest point. Both the optimization step and the classification step are very rapid because they involve only dot products (or similar functions) of the example vector and the support vectors, a small subset of the vector set found close to the class boundaries.

With this feature vector and the outcome label $\mathbf{H_0}$ (normal Ds: the null hypothesis) or $\mathbf{H_1}$ (pathological As and Cs) for each case, an SVM classifier with a radial-basis-function kernel was trained and tested using k-fold cross-validation ($k = 8$).

Multiple cross-validation simulations were done to allow calculation of the confidence bounds of the ROC curve. The entire range of the ROC curve was sampled by modifying the SVM cost function to favour either sensitivity or specificity.

3. Results

Table 1 summarizes the results of the system identification. We removed 20 min intervals where the identification failed, which occurred for approximately 5-6% of the 4073 analysis windows. These were generally caused by artifacts that had not been filtered at the preprocessing stage. The IRF estimates had mean and standard deviation VAF values of 52.9 ± 16.4. Applying the pseudo-inverse resulted in lower VAF (35.3 ± 20.7).

Typical modelling results for a class 'A' case over three 20-min windows with 10min overlap are shown in fig. 1 with and without pseudo inverse. The results without pseudo inverse included significant noise; applying the pseudo inverse resulted in cleaner models. The vertical bars indicates the relative duration of the IRF lag. The indicated VAFs without (85.0, 88.5, 86.3) and with (80.8, 83.6 and 82.8) pseudo inverse are calculated only after the lag to avoid filtering end effects.

Fig. 2 shows the MDL curves and minima for the windows of above case. The number of retained terms of the pseudo inverse in each window was m = 14, 12 and 13 respectively. Without noise reduction, the order was effectively m = 60 (the far right of the MDL curve, or the number of IRF samples before model decimation).

The ROC area under curve of the SVM classifier with and without applying pseudo-inverse was 0.680 ± 0.075 and 0.695 ± 0.054, respectively. However, over a useful operating range of false-positive rates (15-45%), the two models had similar sensitivities (see fig. 3).

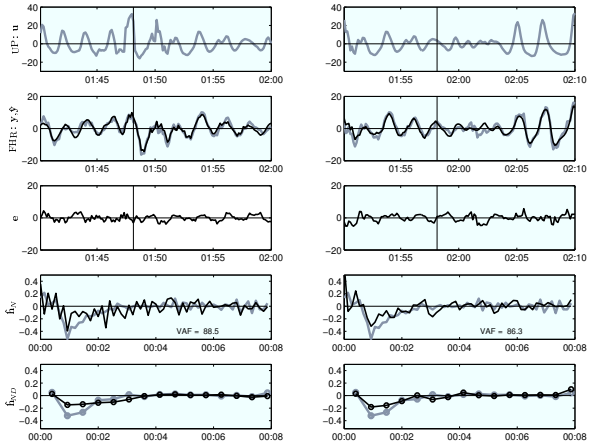
4. Discussion and conclusions

The low VAF values in general are not unexpected since UP is not the sole influence on FHR. The reduction in VAF using pseudo inverse is likely indicative of reduced overfitting: that information was also preserved by this noise suppression is confirmed by the similar discrimination of both models. This similarity can also be explained by the fact that the final model decimation results in IRFs with similar low-pass energy.

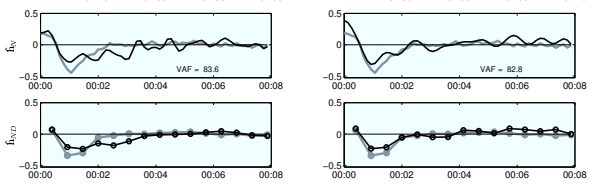
While time averaging also improves the model signal-to-noise ratio, the cleaner pseudo-inverse models may be preferable short-term estimates. In future work, using the pseudo-inverse models to create features with higher time resolution will provide improved dynamic information and should lead to better discrimination.

Acknowledgements

The authors acknowledge the financial support of this work by LMS Medical Systems, Inc.



(a)no pseudo-inverse



(b)with pseudo-inverse

Figure 1. Typical system identification results for an 'A' case without (a) and with (b) pseudo inverse. The columns are successive 20-min windows while the rows are (from top to bottom) the input, output (true in grey, estimate in black) , error, undecimated IRF and decimated IRF (current window in black, time average in grey), respectively. Only the IRFs are shown in (b).

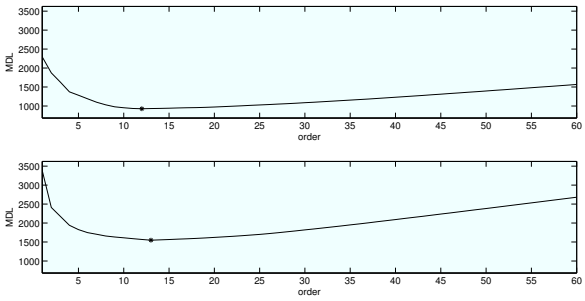


Figure 2. MDL curves and minima for windows of fig. 1.

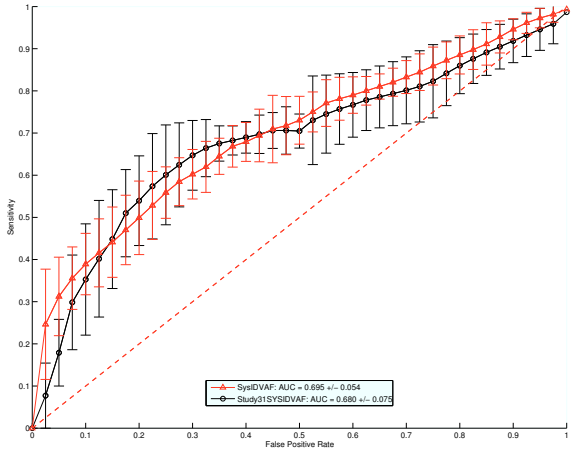


Figure 3. ROC curves of the SVM classifier without (△) and with (○) pseudo inverse.

References

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

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

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

[4] Cao H, Lake D, Ferguson J.E. I, Chisholm C, Griffin M, Moorman J. Toward quantitative fetal heart rate monitoring. Biomedical Engineering IEEE Transactions on 2006; 53(1):111–118. ISSN 0018-9294.

[5] Warrick PA, Kearney RE, Precup D, Hamilton EF. Linear models of intrapartum uterine pressure-fetal heart rate interaction for the normal and hypoxic fetus. In Proceedings of the 2006 IEEE Engineering in Medicine and Biology 28th Annual Conference. 2006; (in press).

[6] Westwick DT, Kearney RE. Identification of nonlinear physiological systems. Hoboken NJ: Wiley-Interscience, 2003.

[7] Hunter IW, Kearney RE. Two-sided linear filter identification. Medical Biological Engineering Computing 1982; 21:203–209.

[8] Haykin S. Adaptive Filtering. Fourth edition. Upper Saddle River, New Jersey: Prentice-Hall, 2002.

Address for correspondence:

Philip A. Warrick
Department of Biomedical Engineering
McGill University
3775 rue University
Montreal, QC H3A 2B4 Canada
philip.warrick@mcgill.ca