Communications

Hidden Markov Models Based on Symbolic Dynamics for Statistical Modeling of Cardiovascular Control in Hypertensive Pregnancy Disorders

V. Baier, M. Baumert, P. Caminal, M. Vallverdú, R. Faber, and A. Voss*

Abstract—Discrete hidden Markov models (HMMs) were applied to classify pregnancy disorders. The observation sequence was generated by transforming RR and systolic blood pressure time series using symbolic dynamics. Time series were recorded from 15 women with pregnancy-induced hypertension, 34 with preeclampsia and 41 controls beyond 30th gestational week.

HMMs with five to ten hidden states were found to be sufficient to characterize different blood pressure variability, whereas significant classification in RR-based HMMs was found using fifteen hidden states. Pregnancy disorders preeclampsia and pregnancy induced hypertension revealed different patho-physiological autonomous regulation supposing different etiology of both disorders.

Index Terms—Blood pressure variability, cardiovascular control, heart rate variability, hidden Markov model, preeclampsia, pregnancy induced hypertension.

I. INTRODUCTION

Hypertensive pregnancy disorders are a leading cause of fetal and maternal mortality [1]. The etiology of this maladaptation of the cardiovascular system to pregnancy is unknown, but two mechanisms, the immune maladaptation and the genetic imprinting have been discussed [2], [3]. A new report of the Working Group on High Blood Pressure in Pregnancy classified the relevant hypertensive pregnancy disorders in: chronic hypertension, pregnancy-induced hypertension (PIH), and preeclampsia (PE), whereby it is discussed whether preeclampsia is pregnancy-induced hypertension plus proteinuria or is characterized by its own etiology [4].

Heart rate and blood pressure variability (HRV and BPV) are generated by the rhythmic actions of cardiovascular hormones and neuronal pathways on effector organs such as the heart, kidneys, and vessels [5] and are independent predictors for sudden cardiac death after acute myocardial infarction or chronic heart failure [6], [7]. Thus, they also might reflect different cardiovascular patho-physiologies in pregnancy, like PIH and PE. Several studies investigated HRV and BPV in normal pregnancy compared to nonpregnant women and women with

PE [8], [9]. Since those studies were performed with parametric analyzes of HRV and BPV measures, we propose a more complex approach here, based on the modeling of the cardiovascular system with hidden Markov models (HMMs).

The theory of HMMs was introduced in the late 1960s by Baum and colleagues [10]. They are statistical models where it is assumed that the system being modeled generates a Markov process, i.e., a stochastic process with the conditional probability distribution of future states depending only on current state. The observation is a probabilistic function of the state that is not observable (hidden), but can be observed through another set of stochastic processes that produce the sequence of observations (e.g., RR and blood pressure time series). This type of statistical modeling was applied to biomedical data, especially in speech recognition [11] or bioinformatics [12], but only a few applications are published for modeling the hidden dynamics of the cardiovascular system [13]–[15].

Assuming the mentioned pregnancy disorders cause a change in the dynamics of this biological system and consequently the physically measured observation sequences, the statistical (linear and nonlinear) properties have to be different. Therefore, we hypothesize that the mathematical structure of HMMs can be used to describe patho-physiological pregnancies and may be helpful in the discussion about a different etiology of PE compared with PIH.

II. METHODS

A. Patients

We recruited 15 women with pregnancy-induced hypertension (PIH), 34 women with preeclampsia (PE) and 41 pregnant controls at the Department of Obstetrics and Gynecology, University of Leipzig, between June 2000 and December 2002. All diagnoses at admission were confirmed 6 weeks after delivery. The classification of the hypertensive disorders is according to the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy [16]. The investigation conforms to the principles outlined in the Declaration of Helsinki. Local ethics committee approval and informed consent of all subjects have been provided.

B. Data Acquisition and Preprocessing

Continuous blood pressure was recorded noninvasively via finger cuff (100 Hz, Portapres model 2). All measurements were performed over 30 minutes under standardized resting conditions between 8 AM and 12 AM as described before [17]. Time series of beat-to-beat intervals (BBI) and systolic blood pressure were automatically extracted, and thereafter, visually inspected. To exclude ventricular premature beats and artifacts, time series were filtered, using an algorithm based on local variance estimation [18].

C. Symbolic Dynamics

HRV and BPV analysis using symbolic dynamics has proven to be a powerful tool to assess cardiovascular control [19]–[23]. The concept of symbolic dynamics goes back to J. S. Hadamard [24] and allows a simplified description of the dynamics of a system with a limited amount of symbols. To obtain coarse-grained values of heart rate and blood pressure, both time series were transformed into a symbol sequence that maps main features of the time series. Several methods for symbol transformation of RR interval sequences have been proposed

[e.g., [20] and [25]–[27]. For the development of Markov models the following transformation by Kurths *et al.* has been applied.

In X (1), x^{BBI} and x^{SP} are n beat-to-beat values of BBI and SP, respectively

$$X = \left\{ \left[x_n^{\mathrm{BBI}}, x_n^{\mathrm{SP}} \right]^T \right\}_{n=0,1,\dots,} \quad x \in R. \tag{1}$$

Each time series in X is transformed in the symbol sequence Z (2) defined as

$$Z = \{z_n\}_{n=0,1,\dots,} s \in 0, 1, 2, 3 \tag{2}$$

using the following transformations:

$$z_n = \begin{cases} 0 : \mu & < x_n \leq (1+a) \cdot \mu \\ 1 : (1+a) \cdot \mu & < x_n < \infty \\ 2 : (1-a) \cdot \mu & < x_n \leq \mu \\ 3 : 0 & < x_n \leq (1-a) \cdot \mu \end{cases} \tag{3}$$

with μ as mean of the time series and threshold values $a^{\mathrm{BBI}} = 0.5$ and $a^{\mathrm{SBP}} = 0.2$. The choice of the threshold values is based on our experimental findings in previous studies. Small changes in the threshold values do not influence the results considerably [21]. Furthermore, the threshold value for systolic blood pressure time series has to be lower than those for beat-to-beat time series because of the lower variance in the blood pressure.

Subsequently, Z is subdivided into short sequences of words O with a length of three symbols to assess cardiovascular short-term regulations. The length of words is limited due to the requirement of a statistically sufficient representation of each single word type.

D. Statistical Modeling of Cardiovascular Control in Pregnancies

The word sequences represent main information of the two physically measured time series as observations of a system with an unknown number of regulatory sources. The theoretical description of such a system can be achieved by statistical modeling of the word sequences. For this purpose discrete HMMs of ergodic topology were developed for both pregnancy disorders PIH, PE and controls, respectively. In this type of HMM every state of the model can be reach in a single step from every other state of the model. The word sequences are given as observation sequences $O = \{O_1, O_2, \cdots, O_T\}$ of the cardiovascular system, where T is the number of words (observations).

The main characteristics of the HMMs are as follows.

N) Number of hidden states in the model. Individual states are denoted as $S = \{S_1, S_2, \cdots, S_N\}$ and the state at position t as qt. For the modeling of the regulatory systems models with $N = \{5, 10, 15\}$ were calculated.

M) Number of distinct observation (emission) symbols per state. The symbol sequence results in an alphabet of size 4. Investigating the word sequence with different word lengths L, the alphabet size increases with $M = 4^L$. Individual observation symbols are stated as $V = \{V_1, V_2, \cdots, V_M\}$.

The initial state distribution $\pi = \{\pi_i\}$ of hidden states where

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N. \tag{4}$$

The probability distribution of state transitions $A = \{a_{ij}\}$ where

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N. \tag{5}$$

The observation (emission) symbol probability distribution $B = \{b_j(k)\}$, in state j where

$$b_j(k) = P[V_k \, at \, t | q_t = S_j], \quad 1 \leq j \leq N, \, 1 \leq k \leq M. \tag{6}$$

Given the observation sequence, the number of hidden states and the number of observation symbols, the probability measures of each

TABLE I
FOURFOLD TABLES WITH CLASSIFICATION RESULTS ACHIEVED BY MODELING SYSTOLIC BLOOD PRESSURE TIME SERIES (SBP) FOR THE PREGNANCY DISORDERS PE AND PIH BASED ON (A) 5, (B) 10, AND (C) 15 HIDDEN STATES, RESPECTIVELY

	$\lambda(SBP, 5)$ PE	$\lambda(SBP, 5)$ PIH	correct	
PE	21	13	62%	
PIH	3	12	80%	A
Exact Fisher test: p = 0.01				

	$\lambda(SBP, 10)$ PE	$\lambda(SBP, 10)$ PIH	correct	
PE	31	3	91%	
PIH	8	7	47%	B
Exact Fisher test: p = 0.01				

	$\lambda(SBP, 15)$ PE	$\lambda(SBP, 15)$ PIH	correct	
PE	34	0	100%	
PIH	15	0	0%	C

Fisher test not applicable, because no word sequence was assigned to model $\lambda(SBP, 15)$ PIH

model $\lambda(\pi, A, B)$ have to be calculated. Therefore, the initial state probabilities were estimated by uniformly distributed random numbers with $\sum_{i=1}^{N} \pi_i = 1$. Initial A as well as initial B were estimated by uniformly distributed random numbers with $\sum_{j=1}^{N} a_{ij} = 1$ and $\sum_{k=1}^{N} b_j(k) = 1$. To find the optimal model parameters, the initial probability parameters are then adjusted to locally maximize $P(O|\lambda)$ from the training sequences (all sequences) applying the iterative Baum-Welch approach using the efficient forward-backward procedure that is described elsewhere [28]. For these calculations the HMM toolbox for Matlab [29] by Murphy was used.

Thus, based on the different number of hidden states $N = \{5, 10, 15\}$ and different time series $X = \{\mathrm{BBI, SBP}\}$ the statistical modeling resulted in six models for both disorders. For convenience, the compact notation $\lambda(X, N)$ is used. Subsequently, each of the 62 observed word sequences (15 PIH and 34 PE) were classified into one of the two disorder models by 1) computing the probability of the observation sequence given the model $P(O|\lambda(X, N))$, and 2) selecting the model with the highest probability.

E. Statistical Modeling of Gaussian White Noise

To evaluate the Markov models of cardiovascular control Gaussian white noise processes were simulated. To meet the 30 minutes recording length of the original data, ten realizations were generated each consisting of 2000 values. Subsequently, symbol sequences were computed as described above and used to 1) develop a separate Markov model $\lambda(\mathrm{Gauss}, N)$ for Gaussian white noise process, and 2) test the models of cardiovascular control developed with measured real data.

III. RESULTS

The results of the classification procedure are presented as fourfold tables (Table I). In order to statistically evaluate these classification tables the exact probability test by Fisher was applied. The null hypothesis is that there is no difference in the probability of an observation sequence being classified into the one or the other model.

As can be seen from Table I models $\lambda(SBP, 5)$ based on the transformed SBP time series with five hidden states revealed a significant classification (p = 0.01; rejection of the null hypothesis). Here, 62% of all PIH and 80% of all PE were classified in the correct model. Classification was also significant for models $\lambda(SBP, 10)$, at which 91% PIH but only 47% PE were correctly identified. In the models with

a higher number of hidden states $\lambda(\text{SBP}, 15)$ the differentiation between the models becomes worse and, therefore, all pregnancies are assigned to one model. The statistical modeling based on BBI reveals no appropriate HMMs using five or ten hidden states, respectively, ($\lambda(\text{BBI}, 5)$ $\lambda(\text{BBI}, 10)$). Solely model $\lambda(\text{BBI}, 15)$ was able to characterize differences in the regulatory systems of heart rate, because 77% PE and 60% PIH could be identified (tables are not presented for BBI-based models). The HMM best distinguishing between normal pregnancies and hypertensive pregnancy disorders was a model based on beat-to-beat interval time series with twelve hidden states. With this settings 57% pregnancy disorders and 74% normal pregnancies were correctly identified.

The proposed model $\lambda(\text{SBP}, 5)$ was also tested using surrogate data of PIH and PE, i.e., with the same power spectral densities of the original ones but with completely destroyed phase patterns. A significant difference in the probability of surrogate data from PE being classified into the one or the other model exists no longer, whereas the classification of surrogate data from PIH remains the same as with the original data. Therefore, it is suggested that nonlinear properties contribute to the differentiation between the pregnancy disorders.

The HMMs modeled with the Gaussian white noise processes differed significantly from those of both pregnancy disorders. The ten simulated time series could be fully assigned to the respective models ($\lambda(\text{Gauss}, 5)$ $\lambda(\text{Gauss}, 10)$ and $\lambda(\text{Gauss}, 15)$). Neither were the Gaussian sequences classified in one of the two pregnancy disorder models, nor were a single word sequence from the pregnancy disorder patients assigned to the Gauss model.

IV. Discussion

In order to evaluate the cardiovascular control in women with PIH in comparison to that of women with PE, we investigated HMMs of heart rate and blood pressure time series. Ergodic HMMs of transformed SBP time series with five to ten hidden states were found to be sufficient to characterize the different blood pressure variability of PE and PIH patients. The significant classification of these patients into different HMMs for blood pressure variability suggests a differently altered cardiovascular control, and therefore, a different patho-physiology. Models based on the transformation of BBI revealed sufficient classification using a higher number of hidden states, which seems to point at more complex variations in heart rate than in blood pressure signals. This agrees with the reported findings of different complexity between both time series in normal subjects [30] as well as in heart failure [31]. The tests with the models of Gaussian white noise processes proved that the physiological models of the pregnancy disorders have a deterministic structure, which is preserved through the transformation process (symbol coding).

Earlier analyzes by our group found a significantly increased peripheral blood pressure pulse [32] but no differences in HRV and BPV parameters except of mean systolic blood pressure that was increased in PE [33]. It could be shown that the vascular system in PE is altered due to the increased release of vasoconstrictive substances and an insensitivity to vasodilative hormones, leading to a smaller blood volume and increased peripheral resistance [34]. Since the HRV-based models of PIH and PE also resulted in a significantly different classification (higher number of hidden states), we suggest that the different pathophysiology mainly affects the vascular regulation (i.e., blood pressure control) and in consequence heart rate control. The classification results discriminating CON and combination of PIH + PE are less impressive but still significant. However, this phenomenon is in congruence with our findings [34] showing that HRV differs significantly between PIH and CON but not between PE and CON. Therefore, the global differentiation between CON and hypertensive disorders in pregnancy is of

lower interest because of the different patho-physiological, regulatory, and compensatory mechanisms in PE and PIH.

Our approach is based on the assumption that BBI and SBP time series are quasi-stationary ergodic processes (ensured through measurement under standardized resting conditions) that can be used for statistical modeling using ergodic HMMs. We used words of three successive symbols to characterize the short-term regulation in the cardiovascular system, which results in 64 possible words (output symbols). For this approach the data length of about 2000 sample points is assumed to be sufficient for modeling purposes. Nevertheless, several regulatory mechanism, e.g., vasomotoric activity, last over a longer period of heart cycles. Therefore, further work requires either a longer period of data recording or a different approach for symbol coding. However, it might be problematic to increase the recording length because it is often not reasonable for pregnant woman beyond 30th gestational week to be measured for more then 30 min under resting conditions. It would also be conceivable to use a different type of HMM topology e.g., a left-right model in which the state index increases with time. But this should perform better if one is interested in modeling different stages of a disease.

In conclusion, the pregnancy disorders PE and PIH seem to have a different pathophysiological blood pressure regulation that can be characterized by HMMs of BPV based on symbolic dynamics. Hence, the etiology of both disorders is assumed to be dissimilar.

References

[1] A. H. Shennan, "Recent developments in obstetrics," *BMJ*, vol. 327, no. 7415, pp. 604–608, Sep. 13, 2003.

[2] G. A. Dekker and B. M. Sibai, "Etiology and pathogenesis of preeclampsia: current concepts," *Am. J. Obstet. Gynecol.*, vol. 179, pp. 1359–1375, 1998.

[3] M. S. Esplin, M. B. Faussett, A. Fraser, R. Kerber, G. Mineau, J. Carrillo, and M. W. Varner, "Paternal and maternal components of the predisposition to preeclampsia," *N. Engl. J. Med.*, vol. 344, pp. 867–872, 2001.

[4] I. Caniggia, J. Winter, S. J. Lye, and M. Post, "Oxygen and placental development during the first trimester: implications for the pathophysiology of preeclampsia," *Placenta*, vol. 21, pp. 25–30, 2000.

[5] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability—standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, pp. 1043–1065, 1996.

[6] R. E. Kleiger, J. P. Miller, J. T. Bigger, and A. J. Moss, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *Am. J. Cardiol.*, vol. 59, pp. 256–262, 1987.

[7] H. Tsuji, M. G. Larson, F. J. Venditti, E. S. Manders, J. C. Evans, C. L. Feldman, and D. Levy, "Impact on reduced heart rate variability on risk for cardiac events," *Circulation*, vol. 94, pp. 2850–2855, 1996.

[8] A. Voss, H. Malberg, A. Schumann, N. Wessel, T. Walther, H. Stepan, and R. Faber, "Baroreflex sensitivity, heart rate, and blood pressure variability in normal pregnancy," *Am. J. Hypertens.*, vol. 13, pp. 1218–1225, 2000.

[9] S. Rang, H. Wolf, G. A. Montfrans, and J. M. Karemaker, "Non-invasive assessment of autonomic cardiovascular control in normal human pregnancy and pregnancy-associated hypertensive disorders: a review," *J. Hypertens.*, vol. 20, no. 11, pp. 2111–2119, Nov. 2002.

[10] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, pp. 1554–1563, 1966.

[11] R. C. Rose and B. H. Juang, "Hidden Markov models for speech and signal recognition," *Electroencephalogr. Clin. Neurophysiol., Suppl.*, vol. 45, pp. 137–152, 1996.

[12] S. R. Eddy, "Profile hidden Markov models," *Bioinformatics*, vol. 14, no. 9, pp. 755–763, 1998.

[13] R. Silipo, G. Deco, R. Vergassola, and H. Bartsch, "Dynamics extraction in multivariate biomedical time series," *Biol, Cybern.*, vol. 79, no. 1, pp. 15–27, Jul. 1998.

[14] R. Silipo, G. Deco, R. Vergassola, and C. Gremigni, "A characterization of HRV's nonlinear hidden dynamics by means of Markov models," *IEEE Trans, Biomed, Eng.*, vol. 46, no. 8, pp. 978–986, Aug. 1999.

[15] S. J. Merrill and J. R. Cochran, "Markov chain methods in the analysis of heart rate variability," in *Nonlinear Dynamics and Time Series*, C. D. Cutler and D. T. Kaplan, Eds. Toronto, ON, Canada: Fields Institute Communications, 1997, pp. 241–252.

[16] "national high blood pressure education program working group on high blood pressure in pregnancy: report of the national high blood pressure education program working group on high blood pressure in pregnancy," *Am. J. Obstet. Gynecol.*, vol. 183, pp. 1–22, 2000.

[17] A. Voss, H. Malberg, N. Wessel, A. Schumann, T. Walther, H. Stepan, and R. Faber, "Baroreflex sensitivity, heart rate and blood pressure variability in normal pregnancy," *Am. J. Hypertens.*, vol. 13, pp. 1218–1225, 2000.

[18] N. Wessel, A. Voss, H. Malberg, C. Ziehmann, H. U. Voss, A. Schirdewan, U. Meyerfeldt, and J. Kurths, "Non-linear analysis of complex phenomena in cardiological data," *Herzschrittmachertherapie und Elektrophysiologie*, vol. 11, no. 3, pp. 159–173, 2000.

[19] A. Voss, K. Hnatkova, N. Wessel, J. Kurths, A. Sander, A. Schirdewan, A. J. Camm, and M. Malik, "Multiparametric analysis of heart rate variability used for risk stratification among survivors of acute myocardial infarction," *Pacing Clin. Electrophysiol.*, pt. 2, vol. 21, no. 1, pp. 186–192, 1998.

[20] J. Kurths, A. Voss, P. Saparin, A. Witt, H. J. Kleiner, and N. Wessel, "Quantitative analysis of heart rate variability," *Chaos*, vol. 5, no. 1, pp. 88–94, 1995.

[21] A. Voss, J. Kurths, H. J. Kleiner, A. Witt, N. Wessel, P. Saparin, K. J. Osterziel, R. Schurath, and R. Dietz, "The application of methods of nonlinear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death," *Cardiovasc. Res.*, vol. 31, no. 3, pp. 419–433, Mar. 1996.

[22] A. C. Yang, S. S. Hseu, H. W. Yien, A. L. Goldberger, and C. K. Peng, "Linguistic analysis of the human heartbeat using frequency and rank order statistics," *Phys. Rev. Lett.*, vol. 90, no. 10, p. 108 103, Mar. 14, 2003.

[23] M. Baumert, T. Walther, J. Hopfe, H. Stepan, R. Faber, and A. Voss, "Joint symbolic dynamic analysis of beat-to-beat interactions of heart rate and systolic blood pressure in normal pregnancy," *Med. Biol. Eng. Comput.*, vol. 40, no. 2, pp. 241–245, Mar. 2002.

[24] J. Hadamard, "Les surfaces a courbures opposees et leurslignes geodesiques," *J. Math. Pures. Appl.*, 1898.

[25] J. A. Palazzolo, F. G. Estafanous, and P. A. Murray, "Entropy measures of heart rate variation in conscious dogs," *Am. J. Physiol.*, pt. 2, vol. 274, no. 4, pp. H1099–H1105, Apr. 1998.

[26] H. Bettermann, D. Amponsah, D. Cysarz, and P. van Leeuwen, "Musical rhythms in heart period dynamics: a cross-cultural and interdisciplinary approach to cardiac rhythms," *Am. J. Physiol.*, pt. 2, vol. 277, no. 5, pp. H1762–H1770, Nov. 1999.

[27] A. Porta, S. Guzzetti, N. Montano, R. Furlan, M. Pagani, A. Malliani, and S. Cerutti, "Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp. 1282–1291, Nov. 2001.

[28] L. R. Rabiner, "A tutorial of hidden Markov Models and selected applications in speech recognition," *Proc. IEEE.*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[29] K. Murphy. Hidden Markov Model (HMM) Toolbox for Matlab. [Online]. Available: http://www.ai.mit.edu/~murphyk/Software/HMM/hmm.html

[30] A. Porta, G. Baselli, S. Guzzetti, M. Pagani, A. Malliani, and S. Cerutti, "Prediction of short cardiovascular variability signals based on conditional distribution," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 12, pp. 1555–1564, Dec. 2000.

[31] G. C. Butler, S. Ando, and J. S. Floras, "Fractal component of variability of heart rate and systolic blood pressure in congestive heart failure," *Clin. Sci. (Lond.)*, vol. 92, no. 6, pp. 543–550, Jun. 1997.

[32] R. Faber, H. Stepan, M. Baumert, A. Voss, and T. Walther, "Analysis of blood pressure waveform—a new method for the classification of hypertensive pregnancy disorders," *J. Hum. Hypertens.*, vol. 18, no. 2, pp. 135–137, 2004.

[33] R. Faber, M. Baumert, H. Stepan, N. Wessel, A. Voss, and T. Walther, "Baroreflex sensitivity, heart rate and blood pressure variability in hypertensive pregnancy disorders," *J. Hum. Hypertens.*, vol. 18, no. 10, pp. 707–712, 2004.

[34] G. A. Dekker and B. M. Sibai, "Etiology and pathogenesis of preeclampsia: current concepts," *Am. J. Obstet. Gynecol.*, vol. 179, pp. 1359–1375, 1998.