35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Linear Dynamic Models for Classification of Single-trial EEG

S. Balqis Samdin, Chee-Ming Ting*, *Member, IEEE,* Sh-Hussain Salleh, A. K. Ariff and A. B. Mohd Noor

Abstract— **This paper investigates the use of linear dynamic models (LDMs) to improve classification of single-trial EEG signals. Existing dynamic classification of EEG uses discrete-state hidden Markov models (HMMs) based on piecewise-stationary assumption, which is inadequate for modeling the highly non-stationary dynamics underlying EEG. The continuous hidden states of LDMs could better describe this continuously changing characteristic of EEG, and thus improve the classification performance. We consider two examples of LDM: a simple local level model (LLM) and a time-varying autoregressive (TVAR) state-space model. AR parameters and band power are used as features. Parameter estimation of the LDMs is performed by using expectation-maximization (EM) algorithm. We also investigate different covariance modeling of Gaussian noises in LDMs for EEG classification. The experimental results on two-class motor-imagery classification show that both types of LDMs outperform the HMM baseline, with the best relative accuracy improvement of 14.8% by LLM with full covariance for Gaussian noises. It may due to that LDMs offer more flexibility in fitting the underlying dynamics of EEG.**

Index Terms- **Linear dynamic model (LDM), hidden Markov model (HMM), brain computer interface (BCI).**

I. INTRODUCTION

Electroencephalogram (EEG) signal contains useful information which can be translated into commands in brain computer interface (BCI) system. BCI systems improve communication among individuals with motor disabilities. The BCI system performance in term of classification rate and computation effect depends on the features and the classification technique chosen. Various classification methods have been used in BCI research. Conventional methods used are static classifiers such as linear discriminant analysis (LDA), k-nearest-neighbor (kNN) and linear support vector machine (SVM) [1]. However, these classifiers are inadequate to capture the temporal information of EEG [2]. Alternative classifiers are dynamic classifiers that can model the temporal changes and classify sequential pattern [3].

The input features for EEG-based BCI are usually varying with time, such as spectral changes of event-related (de)synchronization (ERD/ERS) during motor-imagery. This time course of EEG can provide additional information for classification. To better model the temporal structure of EEG dynamic classifiers are used. Existing dynamic classification studies use hidden Markov models (HMMs) with discrete Markov states [4]-[5], assuming piecewise-stationary of the underlying dynamics. HMM classifier is shown to outperform static classifier like Fisher's linear discriminant [4]. However, this is inappropriate for the continuous and highly non-stationary dynamic underlying EEG.

Linear dynamic model (LDM), a variant of HMMs can better the underlying transient dynamic by allowing the hidden states to be continuous. LDMs formulated in state-space form have been used for time-series of continuous motion trajectory and result in better performance than the conventional HMM [7]-[8]. LDMs have been proposed for speech recognition by [6], [9]-[12]. Use of hidden dynamic state leads to modest accuracy improvement [12]. [5] introduced LDMs for EEG classification using raw signals and shows that use of a simple local level model (LLM) along with PCA transform preprocessing, improve over the HMM. However, using raw signals as input might be too chaotic as representation for modeling. Besides, [5] lacks evaluation on standard dataset using standard features such as autoregressive (AR) parameters.

This paper extends the study [5] by proposing different variants of LDMs for improved modelling of the continuous dynamic of EEG, to enhance the classification performance. We consider two examples of LDM: the LLM in which a simple trend model is used in the observation equation and its extension by allowing the observation to follow a time-varying autoregressive (TVAR) process. The hidden states of trend and TVAR coefficients are modeled as Markov process in state equation. These hidden states of LDM in state-space form are estimated sequentially using Kalman filter (KF). The parameter estimation is performed based on maximum-likelihood (ML) criterion using expectation-maximization (EM) algorithm. Instead of raw signals, short-time AR parameters and band-power are used as input features.

In our earlier work [19], we investigate different covariance modeling of noises in LDMs for denoising of single-trial event-related potentials (ERPs). The covariance of noise in LDMs can be allowed to be of arbitrary structure to capture correlation in observation noise and state evolution. Here, we study the effect of these varying noise covariance structures on classification accuracy. We compare the discrete-state HMM and continuous-state LDMs on two-class motor-imagery EEG classification task using dataset IIIa from BCI Competition III.

II. LINEAR DYNAMICAL MODELING OF EEG

A. Local Level Model

We denote by $\mathbf{y}_n = [y_{n1},...,y_{nd},...,y_{nD}]^T$ the D-dimensional vector of features estimated from n-th short window of a single-trial EEG signal. The observation process $\{\mathbf{y}_n\}$ is

*This work was supported by Universiti Teknologi Malaysia (UTM) under Research University Grant (GUP) –TIER 1 Q.J130000.2545.04H21.

S. B. Samdin is with the Centre for Biomedical Engineering, UTM, 81310 Skudai, Johor, Malaysia (e-mail: balq_qiss@yahoo.com).

C.-M. Ting is with the Center for Biomedical Engineering (CBE), UTM, 81310 Skudai, Johor, Malaysia (e-mail: cmting1818@yahoo.com).

Sh-Hussain Salleh is with the Center for Biomedical Engineering, UTM, 81310 Skudai, Johor, Malaysia (e-mail:hussain@fke.utm.my).

A. K. Ariff is with the Center for Biomedical Engineering, UTM, 81310 Skudai, Johor, Malaysia (e-mail: armouris@gmail.com).

modeled as local level model consisted of observation and state equations respectively as

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{v}_n \tag{1}$$

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \mathbf{w}_n \tag{2}$$

$\mathbf{y}_n$ is assumed to be generated from a trend model (1) with additive Gaussian noise, where $\mathbf{x}_n$ is a hidden slowly varying *trend* component observed in a $D \times 1$ identically independent distributed (*i.i.d*) zero-mean Gaussian noise with covariance matrix **R**, $\mathbf{v}_n \sim N(\mathbf{0}, \mathbf{R})$. The hidden state $\mathbf{x}_n$ is assumed to follow first order Gauss-Markov process as in (3) where $\mathbf{w}_n$ is *i.i.d.* zero-mean Gaussian state noise and covariance matrix **Q**, $\mathbf{w}_n \sim N(\mathbf{0}, \mathbf{Q})$. The $D \times D$ **Q** and **R** are assumed constant with time. Simple model for **Q** and **R** can assume diagonal matrix with identical entries $\mathbf{Q} = \sigma_w^2 \mathbf{I}$ and $\mathbf{R} = \sigma_v^2 \mathbf{I}$, which however, are inappropriate to capture the correlations typically exist in real-world process like EEG. Different choices of covariance have been investigated for state-space modeling of speech [6] and ERPs [19]. **Q** and **R** can be allowed to be of arbitrary form to improve modeling and hence classification performance. **R** and **Q** can be set full matrix to model respectively the correlation between features and evolutions of states, as our previous work [19]. Besides, the diagonals can be non-identical to allow different magnitude of variance. We denote $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{Q})$ the model parameters of a fully specified LLM.

B. TVAR State-space Model

We extend the simple LLM by assuming the observation model (1) as a TVAR process with the evolution of the unknown TVAR coefficients modeled as a random-walk in state equation as (2). Assuming that the elements of feature vector are mutually uncorrelated, we build a separate univariate AR model for each feature dimension as

$$y_n = \sum_{k=1}^{p} a_{n,k} y_{n-k} + v_n \tag{3}$$

where $\{a_{n,k}\}_{k=1}^{p}$ are TVAR coefficients at time *n*, *p* is the model order and v_t is a 1-dimensional *i.i.d.* zero-mean Gaussian noise with variance σ_v^2, $v_t \sim N(0, \sigma_v^2)$. The process is formulated into state-space form as

$$y_n = \mathbf{C}_n \mathbf{x}_n + v_n \tag{4}$$

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{w}_n \tag{5}$$

Denoting $\mathbf{x}_n = \{a_{n,k}\}_{k=1}^{p}$ the state vector of TVAR coefficients, the process (3) is written in a compact form of (4) with a linear mapping of $\mathbf{C}_n = [y_{t-1}, y_{t-2}, ..., y_{t-p}]$.The evolution of $\mathbf{x}_n$ is assumed to follow another multivariate AR(1) process with a constant $p \times p$ coefficient matrix **A**. The *D*-dimensional feature observation is modeled by a composite of these individual TVAR state-space models (SSMs) for each dimension respectively. $\boldsymbol{\theta}_d = (\sigma_{v,d}^2, \mathbf{Q}_d)$ denotes parameters of the TVAR model for dimension *d*. The aim of inference is to estimate the unknown state vectors and model parameter $\boldsymbol{\theta}$.

III. Parameter Inference & EEG Classification

A. State Estimation by KF

The state estimation involve estimating sequentially the filtering density of $\mathbf{x}_n$, $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:n})$ given on observation sequence $\mathbf{y}_{1:n} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ and the model $\boldsymbol{\theta}$. For the linear Gaussian model considered here, the mean and covariance of $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:n})$ can be obtained analytically by KF [15]. The conditional mean $E(\mathbf{x}_n \mid \mathbf{y}_{1:n})$ is the MMSE estimator of $\mathbf{x}_n$. Alternating between state mean and covariance prediction is recursively carried out in KF. The Kalman forward recursion is performed to compute the mean and covariance of $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:n})$ for $1 \le n \le N$ as given in [6], [19]. Future observations $\mathbf{y}_{n+1:N}$ available can be used to correct the filtered estimates by performing fixed-lag smoothing of $\mathbf{x}_n$, which computes the smoothing density $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:N})$. We denote by $\hat{\mathbf{x}}_{n|N}$ and $\mathbf{P}_{n|N}$ the mean and covariance of $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:N})$. Based on the estimates by the forward filtering recursion, the smoothed estimates can be obtained by backward recursion for $n = N-1, N-2, ..., 1$ [5], [19].

B. Model Parameter Estimation & Classification

The ML estimates of model parameters $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{Q})$ are obtained by maximizing the marginal likelihood of $\mathbf{y}_{1:N}$ with respect to $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\theta} \log p_{\theta}(\mathbf{y}_{1:N}) \tag{6}$$

where $\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N})$ for linear Gaussian model here can be computed analytically using KF as follows

$$\begin{aligned} \log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}) &= \sum_{n=1}^{N} \log p_{\boldsymbol{\theta}}(\mathbf{y}_n \mid \mathbf{y}_{1:n-1}) \\ &= -\frac{1}{2} \sum_{n-1}^{N} \left\{ \log\left|\mathbf{P}_{\mathbf{e}_n}\right| + \mathbf{e}_n^T \mathbf{P}_{\mathbf{e}_n}^{-1} \mathbf{e}_n \right\} - C \end{aligned} \tag{7}$$

where $\mathbf{e}_n$ and $\mathbf{P}_{\mathbf{e}_n}$ are respectively the prediction error and its covariance computed from the forward Kalman recursion. For the TVAR state-space model, the composite likelihood is computed by summing log-likelihoods evaluated on each dimension using respective individual models

$$\log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}) = \sum_{d=1}^{D} \log p_{\boldsymbol{\theta}_d}(y_{d,1:d,N}) \tag{8}$$

EM method is used for ML estimation when both model and state parameters are unknown. EM algorithm was first introduced by [16] and has been used for parameter estimation in linear Gaussian state-space models in [17] [18]. The EM algorithm for our models is described based on the procedure in [18], as two-step iteration:

1) E-Step: Involve computing the expected log likelihood $Q = E[\log p_{\boldsymbol{\theta}_k}(\mathbf{x}_{1:N}\mathbf{y}_{1:N} \mid \mathbf{y}_{1:N})]$ given the model estimates at k^{th} iteration $\boldsymbol{\theta}_k$. This quantity depends on expectations:

$$\hat{\mathbf{x}}_{n|N} = E(\mathbf{x}_n \mid \mathbf{y}_{1:N}) \quad (9)$$

$$\mathbf{S}_{n|N} = E(\mathbf{x}_n \mathbf{x}_n^T \mid \mathbf{y}_{1:N}) = \mathbf{P}_{n|N} + \hat{\mathbf{x}}_{n|N}\hat{\mathbf{x}}_{n|N}^T \quad (10)$$

$$\mathbf{S}_{n,n-1|N} = E(\mathbf{x}_n \mathbf{x}_{n-1}^T \mid \mathbf{y}_{1:N}) = \mathbf{P}_{n,n-1|N} + \hat{\mathbf{x}}_{n|N}\hat{\mathbf{x}}_{n-1|N}^T \quad (11)$$

The first two quantities are obtained from Kalman smoothing estimates while for the last through backward recursion [18]:

$$\mathbf{P}_{n,n-1|N} = \mathbf{P}_{n|n}\mathbf{J}_{n-1}^T + \mathbf{J}_n(\mathbf{P}_{n+1,n|N} - \mathbf{P}_{n|n})\mathbf{J}_{n-1}^T$$

2) M-Step: The model parameters are updated by maximizing the Q function over $\boldsymbol{\theta}$ which is done by setting the corresponding partial derivative of Q zero [19]:

$$\mathbf{Q}^{k+1} = \frac{1}{N-1}\sum_{n=2}^{N}(\mathbf{S}_{n|N} - \mathbf{S}_{n-1,n|N} - \mathbf{S}_{n,n-1|N} + \mathbf{S}_{n-1|N}) \quad (12)$$

and

$$\mathbf{R}^{k+1} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{y}_n\mathbf{y}_n^T - 2\hat{\mathbf{x}}_{n|N}\mathbf{y}_n^T + \mathbf{S}_{n|N}) \quad (13)$$

Equations (12) and (13) provide estimation of full matrix. We investigate a method for estimation for cases when $\mathbf{R}$ and $\mathbf{Q}$ are diagonal entries in [19]. The EM steps increase the likelihood monotonically with guaranteed convergence to a local maximum. The iteration is stopped when $p_{\boldsymbol{\theta}_{k+1}}(\mathbf{y}_{1:N}) - p_{\boldsymbol{\theta}_k}(\mathbf{y}_{1:N}) < \varepsilon$ where ε is a small threshold, and the ML estimates of $\boldsymbol{\theta}$ is obtained.

For the training of classifier, a LDM model is estimated for each class to be classified, using the EM algorithm. The EM algorithm can be easily extended to perform batch training on multiple examples. In the E-step, Kalman smoother is run through each observation sequence, and then the generated expectations (9)-(11) are accumulated over all observations. The re-estimation in M-step proceeds as before by dividing by the total number of observation frames over all examples [6]. Classification requires calculation of likelihood of a given model generating the test sample using (7) by performing one run of KF. In classification, a test EEG signal represented by a sequence of feature vectors $\bar{\mathbf{y}}_{1:N}$ is evaluated over estimated models of each class and assigned to the one with the highest likelihood

$$c^* = \arg\max_{1 \le c \le C} \log p_{\boldsymbol{\theta}_c}(\bar{\mathbf{y}}_{1:N}) \quad (14)$$

where $c \in C$ is set of classes to classify.

IV. Experimental Results

We evaluate the performance of the proposed LDMs on single-trial EEG-based motor imagery classification using dataset IIIa, a subset of BCI Competition III dataset [20]. The task is four-class classification of cued motor imagery EEG (left hand, right hand, foot or tongue movements). The database consists of three subjects each recorded 60-channel EEG data with sampling frequency 250 Hz, with 60 trials per class. We focus on two-class classification (left and right hand movement) using only two unipolar channels i.e. the C3 and C4. The numbers of training and test trials for each subject are shown in Table I. The task is challenging due to very few channels along with small amount of training data.

TABLE I. Number of Training and Test Trials for Each Subject.

Subject	# of training trials		# of test trials	
	Left hand	Right hand	Left hand	Right hand
k3	36	37	38	38
k6	21	26	22	16
l1	20	20	23	19

Segments of single-trial EEGs during motor imagery from 3s to 8s are used for analysis. For feature extraction, short-time autoregressive (AR) parameters and band-power (BP) estimates are estimated from each short segment of 250ms without overlapping. 12-dimensional AR coefficients are extracted with 6 parameters from each channel C3 and C4. To obtain the BP features, the signals of each channels are bandpass-filtered at frequency10-12Hz (alpha band) and 14-18Hz (beta band), then normalized by mean subtraction and squared before averaging over each short-segment to give 4 BP features each per channel and frequency range. Log-transform is then applied and the relative BP is calculated using reference power over window 3-4.25s. Both set of features are concatenated together to form a feature vector of 16 dimension.

We compare the performance of the proposed LDMs with conventional HMMs in term of classification accuracy. The baseline uses discrete HMM with single Gaussian observation density, trained using Viterbi training algorithm with stopping criterion $\varepsilon = 0.001$. HMMs with 2 and 3 states are used. The likelihood for HMM is calculated based on most likely state sequence obtained by Viterbi algorithm. LLMs with different covariance models of $\mathbf{Q}$ and $\mathbf{R}$ are trained. The performance of LLMs is sensitive to parameter initialization for the model training. We choose the initial parameters based on the highest converged likelihood using EM algorithm with ε=0.1on the training data. For the TVAR SMMs, AR order is varied from 2 to 4. For simplicity, $\mathbf{Q}$ with identical diagonals is used for state noise of individual models. All models are trained on a subject-dependent basis.

The results for each subject and average over all subjects are shown in Table II. The best performing variant of each type of models are based on the highest averaged accuracy, indicated in bold. It is observed that both the proposed LDMs outperform the best HMM baseline, giving relative accuracy improvement of 14.8% and 2.8% for the best LLM and TVAR model. All the variants of LLMs offer significant gains over the baseline, possibly due to improved modeling of the underlying dynamics of EEG by the continuous-state compared to the discrete one. The performance improvement is more significant for subject l1 with very few training data available. Among the LLMs, use of more complex models with full covariance to capture correlations provide better performance with full matrix for both $\mathbf{Q}$ and $\mathbf{R}$ achieving the best. Despite being superior over the baseline, extension by describing the observation as autoregression in TVAR model, however, perform worse than the LLMs. This may be due to the use of independent AR process for each dimension is unable to capture the correlation between features as LLMs

TABLE II. CLASSIFICATION ACCURACY (%) USING HMMS, LLMS AND TVAR SSMS FOR EACH SUBJECT.

Model	k3	k6	l1	Mean
HMM state-2	52.63	73.58	41.42	**55.88**
HMM state-3	52.63	61.08	37.99	50.57
LLM, Full **Q**-Full **R**	57.89	68.18	66.37	**64.15**
LLM, Diagonal **Q**-Diagonal **R**	55.26	61.65	64.30	60.40
LLM, Full **Q**-Diagonal **R**	56.58	53.98	62.82	57.79
LLM, Diagonal **Q**-Full **R**	56.58	67.90	56.30	60.26
TVAR order-2	56.58	34.94	50.00	47.17
TVAR order-3	52.63	56.53	63.15	**57.44**
TVAR order-4	50.00	50.00	52.17	50.72

TABLE III. LOG-LIKELIHOODS (AVERAGE OVER ALL SUBJECTS) AND AIC VALUES FOR DIFFERENT LDMS FITTED TO THE TRANING DATA.

Model	No. of Parameters	Log-likelihood	AIC
HMM state-2	64	-1080.02	2288.04
HMM state-3	96	-905.28	2002.56
LLM, Full **Q**-Full **R**	512	**182.23**	659.54
LLM, Diagonal **Q**-Diagonal **R**	32	-214.27	**492.54**
LLM, Full **Q**-Diagonal **R**	272	-79.86	703.72
LLM, Diagonal **Q**-Full **R**	272	-482.72	1509.44
TVAR order-2	112	-425.59	865.18
TVAR order-3	208	-540.85	1497.70
TVAR order-4	336	-805.01	2282.02

with full covariance do. Future work will extend it by using multivariate TVAR model with full covariance modeling as the LLMs. From Table III, more complex LDMs (more parameters) give better fit to the data than the HMMs, indicated by higher likelihoods. Among LDMs, LLMs are superior. Generally, use of better representing models leads to improved accuracy, e.g. the best fitting LLM, full **Q** and **R** gives highest performance. Akaike information criterion (AIC) suggests LLM, diagonal **Q** and **R** the most optimal model balanced between simplicity and goodness-of-fit.

V. CONCLUSION

In this study, we examine the use of LDMs to improve the dynamic classification of single-trial EEG. The continuous state of LDMs can better describe the non-stationary underlying dynamics of EEG compared to the piece-wise stationary discrete-state of conventional HMMs. We propose two examples of LDMs: LLMs with covariance modeling of noises to capture feature correlations, and TVAR models using a autoregressive process for observations. Results on two-class motor-imagery EEG classification show that both models outperform the discrete-state HMM, with substantial improvement of 14.8% by the best variant of LLM. Future work will consider extending the single-Gaussian noises of the LDMs to non-Gaussian or mixture of Gaussians, and LDMs to more general state-space models. Besides, the LD modeling of EEG can be extended to multivariate case for multiple channels. To obtain more reliable comparison results, the evaluation would be extended to more complex multiple-class classification task using more subjects. Successful modeling and classification using LDMs for EEG motivates application to other biomedical signals generally.

REFERENCES

[1] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of Neural Engineering*, vol. II, no. 4, pp. L14-L22, August 2005.

[2] F Lotte, M Congedo, A Lecuyer, F Lamarche, and B Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. IV, no. 4, pp. R1–R13, January 2007.

[3] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. LXXVII, no. 2, pp. 257-286, February 1989.

[4] H. Lee and S. Choi, "PCA-based linear dynamical systems for multichannel EEG classification," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, 2002, pp. 745-749.

[5] H. Lee and S. Choi, "PCA-based linear dynamical systems for multichannel EEG classification," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, 2002, pp. 745-749.

[6] J. Frankel and S. King, "Speech Recognition Using Linear Dynamic Models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. XV, no. 1, pp. 246-256, January 2007.

[7] V. Pavlovi´cy, B. J. Freyz, and T. S. Huang, "Time-Series Classification Using Mixed-State Dynamic Bayesian Networks," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[8] V. Pavlovic and J.M. Rehg, "Impact of dynamic model learning on classification of human motion," in *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition, 2000*, 2000, pp. 788-795.

[9] J.Z. Ma and L. Deng, "Target-directed mixture dynamic models for spontaneous speech recognition," *IEEE Transactions on Speech and Audio Processing,* vol. XII, no. 1, pp. 47- 58, January 2004.

[10] J. Ma and L. Deng, "A mixed-level switching dynamic system for continuous speech recognition," *Computer Speech & Language*, vol. XVIII, no. 1, pp. 49-65, January 2004.

[11] A.-V. I. Rosti, *Linear Gaussian Models for Speech Recognition*, Ph.D. Thesis, Eng. Department, Cambridge Univ., Cambridge, U.K., 2004.

[12] J.Frankel, *Linear dynamic models for automatic speech recognition*, Ph.D. thesis, School of Informatics, Univ. of Edinburgh, U.K., 2003.

[13] S. Roweis and Z. Ghahramani, "A unifying review of linear gaussian models," *Neural Comput*, vol. XI, no. 2, pp. 305-45, 1999.

[14] A.V.I. Rosti and Gales M., *Generalised linear Gaussian models*.: Cambridge University Engineering Department, 2001.

[15] R. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. LXXXII, pp. 35–44, March 1960.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. R. Statist. Soc. B*, vol. XXXIX, pp. 1-38, 1977.

[17] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, vol. 3, no. 4, pp. 253–264, 1982.

[18] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Technical Report, University of Toronto, 1996.

[19] C.-M. Ting, S. B. Samdin, S. Salleh, M. H. Omar, and I. Kamarulafizam, "An expectation-maximization algorithm based Kalman smoother approach for single-trial estimation of event-related potentials," in *Proc Annu. Int. Conf. IEEE Eng. Med. Bio. Soc*, pp. 6534-6538, 2012.

[20] B. Blankertz, K. R M¨uller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schl¨ogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schr¨oder, and N. Birbaumer, "The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1044–1051, Jun. 2004.