34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

An Expectation-Maximization Algorithm based Kalman Smoother Approach for Single-Trial Estimation of Event-related Potentials

Chee-Ming Ting, S. Balqis Samdin, Sh-Hussain Salleh, M. Hafizi Omar and I. Kamarulafizam

Abstract— **This paper applies an expectation-maximization (EM) based Kalman smoother (KS) approach for single-trial event-related potential (ERP) estimation. Existing studies assume a Markov diffusion process for the dynamics of ERP parameters which is recursively estimated by optimal filtering approaches such as Kalman filter (KF). However, these studies only consider estimation of ERP state parameters while the model parameters are pre-specified using manual tuning, which is time-consuming for practical usage besides giving suboptimal estimates. We extend the KF approach by adding EM based maximum likelihood estimation of the model parameters to obtain more accurate ERP estimates automatically. We also introduce different model variants by allowing flexibility in the covariance structure of model noises. Optimal model selection is performed based on Akaike Information Criterion (AIC). The method is applied to estimation of chirp-evoked auditory brainstem responses (ABRs) for detection of wave V critical for assessment of hearing loss. Results shows that use of more complex covariances are better estimating inter-trial variability.**

Index Terms- **Event-related potentials, Kalman smoother, expectation-maximization algorithm.**

I. Introduction

Event-related potentials (ERPs) are scalp-recorded bioelectrical response of the brain elicited by specific stimulation. The challenge is to extract the underlying ERPs in various noises e.g. background electroencephalogram (EEG) and non-neural artifacts, typically with poor signal-to-noise (SNR) ratio. The conventional ensemble averaging of time-locked single-trials cancels out the assumingly random background noise, however, requires many repeated simulations and implies loss of information related to trial-to-trial variability due to different degree of fatigue, habituation, or levels of attention of subjects [1]. Various approaches have been proposed to solve single-trial ERP estimation problem which is also considered in this paper.

Optimal filtering approaches have been recently introduced for single-trial dynamical estimation of the non-stationary ERPs [1]-[4]. The inter-trial dynamics of ERP parameters are modeled as a Markov process observed in observation noise to give noisy measurements. The model is formulated into state-space form. Then, the single-trial ERP estimation problem is cast into optimal filtering one which involves recursive estimation of filtering densities of the underlying clean ERPs given noisy measurements. The filtering density of ERP can be computed exactly by Kalman filter (KF) [1] or approximated by swarms of weighted samples using particle filter (PF) [3], [4]. Evaluation results show the capability of these methods in estimating the inter-trial ERP parameter changes in low SNR condition. The KF approach by [1] is adopted here, justified by the linear-Gaussian form assumption of the abovementioned ERP dynamic model for which KF yields minimum mean-square error (MMSE) estimators. Besides, KF provides exact solution in simple analytical way while PF suffers approximation error and high computational complexity. The KF based ERP estimation has been extended in [2] to Kalman smoother (KS) algorithm which utilizes both the past and future observations to infer more accurate ERP estimates. Use of KS shows better performance in term of tracking ability and noise reduction of ERPs.

The estimation of the state of ERP parameters in the above-mentioned studies is performed by assuming that the model parameters are *known*, which in fact pre-specified by subjective manual tuning. This is impractical and only yields suboptimal solutions. The aim of this paper is to consider the problem where the model parameters of the ERP dynamic model are *unknown*, and need to be estimated automatically from the data. Expectation-maximization (EM) algorithm in conjunction with KS for maximum likelihood parameter estimation in linear-Gaussian state-space models has been developed [5], [6]. This paper applies the EM method for estimating the parameters of the ERP dynamic model which is a special case of linear-Gaussian model. The EM algorithm has been used for EEG spectral estimation of event-related desynchronization (ERD) by [7] based on time-varying autoregressive model which is distinct from the ERP model considered here. The issue of optimal choice for covariance of model noises is discussed but not studied in [1] who only set diagonal with identical entries, based on uncorrelation assumption of ERP parameter evolution, which is often invalid for real ERP processes. We allow the noise covariance to be of arbitrary structure where the optimal model choice is selected based on Akaike Information Criterion (AIC) which balance the goodness of fit and model complexity. The better modeling of the process is expected to give more accurate ERP estimates and better de-noising. For observation model, we use wavelet coefficient as ERP parameters as suggested in [3] instead of measurement vector

*This work is supported by Universiti Teknologi Malaysia (UTM), under Fundamental Research Grant Scheme, Vot R.J130000.7836.3F504, Minister of Higher Education (MOHE), Malaysia.

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

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

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

I. Kamarulafizam is with the Center for Biomedical Engineering, UTM, 81310 Skudai, Johor, Malaysia (e-mail:kamarulafizam@yahoo.com).

itself as in [1] to reduce state dimension and hence computational effort. Our method is evaluated on auditory brainstem responses (ABRs) estimation especially the wave V for detection of hearing loss.

The paper is organized as follow. Section II describes the state-space formulation of ERP models with state and model parameter estimation using the EM based KS approach. Section III presents evaluation results on ABR estimation. Conclusion is given in final section.

II. Methods

A. State-Space Modeling of ERPs

We consider the state-space formulation of ERP dynamic model suggested by [1], but use wavelet coefficients as ERP parameters instead of direct measurement samples as in [3] to reduce state dimension. Approximation coefficients of discrete wavelet transform (DWT) are used to represent the low frequency spectral components in ERP which is typically a smooth transient wave [3]. Let denote by $\mathbf{y}_n = [y_{n1}, \cdots, y_{nt}, \cdots, y_{nT}]^T$ sequence of T approximation wavelet coefficients extracted from single-trial ERP measurements at trial n. The state-space model of ERP dynamics consists of observation equation and state equation, respectively as [1], [3]

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

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

The observation process of ERPs $\{\mathbf{y}_n\}$ is typically modeled by linear additive noise model of the form (1) where $\mathbf{x}_n$ is the state vector of clean ERP parameters, here also represented by wavelet coefficients, hidden in background noise $\mathbf{v}_n$ which is $T \times 1$ i.i.d. Gaussian noise with mean zero and static covariance matrix $\mathbf{R}$, $\mathbf{v}_n \sim N(\mathbf{0}, \mathbf{R})$. The hidden state $\mathbf{x}_n$ is assumed to follow first-order Gauss Markov process as in (2) where $\mathbf{w}_t$ is i.i.d. zero mean Gaussian state noise with covariance matrix $\mathbf{Q}$, $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q})$. $\mathbf{R}$ and $\mathbf{Q}$ are $T \times T$ matrices and assumed static i.e. do not vary with time. To better model the ERP process, we allow arbitrary form for $\mathbf{R}$ and $\mathbf{Q}$, which are assumed inappropriately as diagonal with identical entries $\mathbf{R} = \sigma_v^2 \mathbf{I}$ and $\mathbf{Q} = \sigma_w^2 \mathbf{I}$ in [1], [3]. $\mathbf{R}$ and $\mathbf{Q}$ can be set full matrix to model respectively the correlation in observation noise along each dimension and individual ERP parameter evolutions, which is typically present in real ERP process. Besides, the diagonal of $\mathbf{Q}$ can be non-identical to allow different magnitude of changes for each parameter. Different choices of covariance have been investigated for state-space modeling of speech [8]. By denoting $\boldsymbol{\theta} = (\mathbf{R}, \mathbf{Q})$ the model parameters of the fully specified ERP state-space model, the objective is to estimate recursively the unknown trial-varying state vectors of clean ERP wavelet parameters $\mathbf{x}_n$, given $\boldsymbol{\theta}$. This is solved here by KS. Inverse wavelet transform is then applied to the estimated wavelet coefficients to reconstruct the clean ERP waveform. In this paper, $\boldsymbol{\theta}$ is assumed unknown and estimated automatically by maximum likelihood using EM algorithm from the dataset.

B. Estimation of ERP Parameters

The estimation problem involves estimating recursively the filtering density of $\mathbf{x}_n$ conditional on measurement sequence $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \cdots, \mathbf{y}_n\}$, given the model $\boldsymbol{\theta}$, here denoted by $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:n})$. 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. The conditional mean $E(\mathbf{x}_n \mid \mathbf{y}_{1:n})$ is the MMSE estimator of $\mathbf{x}_n$. Let $\hat{\mathbf{x}}_{n|n-1}$ and $\mathbf{P}_{n|n-1}$ denote the mean and covariance of the one-step ahead prediction density $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:n-1})$; $\hat{\mathbf{x}}_{n|n}$ and $\mathbf{P}_{n|n}$ denote the mean and covariance of the filtering density $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:n})$. The Kalman recursions for $1 \le n \le N$ are given as [1]:

$$\hat{\mathbf{x}}_{n|n-1} = \hat{\mathbf{x}}_{n-1|n-1} \tag{3}$$

$$\mathbf{P}_{n|n-1} = \mathbf{P}_{n-1|n-1} + \mathbf{Q} \tag{4}$$

$$\mathbf{K}_n = \mathbf{P}_{n|n-1} + (\mathbf{P}_{n|n-1} + \mathbf{R})^{-1} \tag{5}$$

$$\hat{\mathbf{x}}_{n|n} = \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n(\mathbf{y}_n - \hat{\mathbf{x}}_{n|n-1}) \tag{6}$$

$$\mathbf{P}_{n|n} = (\mathbf{I} - \mathbf{K}_n)\mathbf{P}_{n|n-1} \tag{7}$$

with initial condition $\hat{\mathbf{x}}_{0|0} = \mathbf{x}_0$ and $\mathbf{P}_{0|0} = \boldsymbol{\Sigma}$. $\mathbf{K}_n$ is called Kalman gain. $\mathbf{e}_n = (\mathbf{y}_n - \hat{\mathbf{x}}_{n|n-1})$ and $\mathbf{P}_{\mathbf{e}_n} = (\mathbf{P}_{n|n-1} + \mathbf{R})$ are respectively the prediction error and its covariance.

Future measurements $\mathbf{y}_{n+1:N}$ can be used to correct filtered estimates by performing fixed-lag smoothing of $\mathbf{x}_n$, which involves estimating the smoothing density $p_{\boldsymbol{\theta}}(\mathbf{x}_n \mid \mathbf{y}_{1:N})$ given measurements $\mathbf{y}_{1:N}$. The smoothed estimator $E(\mathbf{x}_n \mid \mathbf{y}_{1:N})$ is more accurate than the filtered one. 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 $n = N-1, N-2, \cdots, 1$ [6], [7].

$$\mathbf{J}_n = \mathbf{P}_{n|n}\mathbf{P}_{n+1|n}^{-1} \tag{8}$$

$$\hat{\mathbf{x}}_{n|N} = \hat{\mathbf{x}}_{n|n} + \mathbf{J}_n(\hat{\mathbf{x}}_{n+1|N} - \hat{\mathbf{x}}_{n+1|n}) \tag{9}$$

$$\mathbf{P}_{n|N} = \mathbf{P}_{n|n} + \mathbf{J}_n(\mathbf{P}_{n+1|N} - \mathbf{P}_{n+1|n})\mathbf{J}_n^T \tag{10}$$

with initial estimates $\hat{\mathbf{x}}_{N|N}$ and $\mathbf{P}_{N|N}$ given by KF.

C. Estimation of Model Parameters with EM Algorithm

The ML estimate of $\boldsymbol{\theta}$ is obtained by maximizing the marginal likelihood of $\mathbf{y}_{1:N}$ with respect to $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\mathbf{y}_{1:N}) \tag{11}$$

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

$$\log p_{\theta}(\mathbf{y}_{1:N}) = \sum_{n=1}^{N} \log p_{\theta}(\mathbf{y}_n \mid \mathbf{y}_{1:n-1})$$

$$= \sum_{n=1}^{N} \log N(\mathbf{y}_n ; \hat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{\mathbf{e}_n}) \quad (12)$$

$$= -\frac{1}{2}\sum_{n-1}^{N} \left\{ \log \left| \mathbf{P}_{\mathbf{e}_n} \right| + \mathbf{e}_n^T \mathbf{P}_{\mathbf{e}_n}^{-1} \mathbf{e}_n \right\} - C$$

EM method is used for ML estimation when both model and state parameters are unknown. EM algorithm was first introduced by [9] and has been used for parameter estimation in linear Gaussian state-space models in [5] [6]. We describe the EM algorithm for our model $\boldsymbol{\theta}$ based on the procedure in [6], [7], as following two-steps repeated iteratively:

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:

$$\hat{\mathbf{x}}_{n|N} = E(\mathbf{x}_n \mid \mathbf{y}_{1:N}) \quad (13)$$

$$\mathbf{S}_{n|N} = E(\mathbf{x}_n \mathbf{x}_n^T \mid \mathbf{y}_{1:N}) = \mathbf{P}_{n|N} + \hat{\mathbf{x}}_{n|N} \hat{\mathbf{x}}_{n|N}^T \quad (14)$$

$$\mathbf{S}_{n,n-1|N} = E(\mathbf{x}_n \mathbf{x}_{n-1}^T \mid \mathbf{y}_{1:N}) = \mathbf{P}_{n,n-1|N} + \hat{\mathbf{x}}_{n|N} \hat{\mathbf{x}}_{n-1|N}^T \quad (15)$$

The first two quantities are obtained from KS estimates while for the last through backward recursion [6]:

$$\mathbf{P}_{n,n-1|N} = \mathbf{P}_{n|n}\mathbf{J}_{n-1}^T + \mathbf{J}_n(\mathbf{P}_{n+1,n|N} - \mathbf{P}_{n|n})\mathbf{J}_{n-1}^T \quad (16)$$

which is initialized $\mathbf{P}_{N,N-1|N} = (\mathbf{I} - \mathbf{K}_N)\mathbf{P}_{N-1|N-1}$.

1) M-Step: The model parameters are re-estimated by maximizing the Q function over $\boldsymbol{\theta}$ which is done by taking the corresponding partial derivative of Q and setting zero. Solving it gives the updated parameter estimates as follows:

$$\frac{\partial Q}{\partial \mathbf{Q}^{-1}} = \frac{N-1}{2}\mathbf{Q} - \frac{1}{2}\sum_{n=2}^{N}(\mathbf{S}_{n|N} - \mathbf{S}_{n-1,n|N} - \mathbf{S}_{n,n-1|N} + \mathbf{S}_{n-1|N}) = 0$$

$$\mathbf{Q}^{k+1} = \frac{1}{N-1}\sum_{n=2}^{N}(\mathbf{S}_{n|N} - \mathbf{S}_{n-1,n|N} - \mathbf{S}_{n,n-1|N} + \mathbf{S}_{n-1|N}) \quad (17)$$

and

$$\frac{\partial Q}{\partial \mathbf{R}^{-1}} = \frac{N}{2}\mathbf{R} - \sum_{n=1}^{N}(\frac{1}{2}\mathbf{y}_n\mathbf{y}_n^T - \hat{\mathbf{x}}_{n|N}\mathbf{y}_n^T + \frac{1}{2}\mathbf{S}_{n|N}) = 0$$

$$\mathbf{R}^{k+1} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{y}_n\mathbf{y}_n^T - 2\hat{\mathbf{x}}_{n|N}\mathbf{y}_n^T + \mathbf{S}_{n|N}) \quad (18)$$

Equation (17) and (18) involve estimation of full matrix, however, estimation for cases when $\mathbf{R}$ and $\mathbf{Q}$ are diagonal and/or with identical entries is not considered in [6], [7]. Here, we adopt the derivations in [10] for estimation of these constrained cases. Details refer to [10]. By denoting $\mathrm{vec}(\mathbf{D}_{n\times m}) = [d_{11} \cdots d_{n1} d_{12} \cdots d_{n2} \cdots d_{1m} \cdots d_{nm}]^T$, the general constrained estimation equation for $\mathbf{Q}$ is given as

$$\mathbf{q}^{k+1} = \frac{1}{N-1}(\mathbf{D}_q^T\mathbf{D}_q)^{-1}\mathbf{D}_q^T \mathrm{vec}(\mathbf{H})$$

$$\mathrm{vec}(\mathbf{Q}^{k+1}) = \mathbf{D}_q\mathbf{q}^{k+1} \quad (19)$$

where $\mathbf{q}$ contains column vector of p free parameters to estimate with $\mathbf{q} = [\sigma_w^2]$ and $\mathbf{q} = diag(\mathbf{Q})$ for diagonal with identical and non-identical entries respectively, we denote $diag(\mathbf{A}_{n\times n}) = [a_{11}, a_{22}, \cdots, a_{nn}]^T$. $\mathbf{D}_q$ is the $T^2 \times p$ design matrix and $\mathbf{S}$ is computed as (17)

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

The constrained estimations for $\mathbf{R}$ are performed similarly with $\mathbf{H}$ computed using (18). We found that estimating diagonal matrix using (19) gives the same estimates as the updated matrices by (17) and (18) with off-diagonals set zeros [8].

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.

III. Experimental Results

This section presents performance evaluation of the proposed methods for single-trial estimation of chirp-evoked ABRs from a subject with normal hearing. ABR comprises the early portion of auditory evoked potentials elicited by acoustic stimulus, and is composed of several waves labeled with roman numerals I-VII, among which waves III and V are the focus of this paper. The ABR waveform characteristics are useful for objective assessment of hearing loss and pathologies affecting auditory brainstem pathways. Refer [11] for details. The ABR has variabilities across trials, different subjects and stimulus intensities [12], [13].

The data are obtained following the procedure in [14]. The ABRs were elicited by the properly calibrated chirps presented at rate 20 Hz through a headphone. The potentials were recorded using electrode placed at positions A1, A2, Cz and the ground Fpz, at sampling frequency of 19.2 kHz. Data of 500 ABR single-trials from passive electrode at 60 dB intensity levels of sound pressure level (SPL) is used for analysis. The data was band-pass filtered with cutoff frequencies of 0.1 and 1.5 kHz and down-sampled to 14.4 kHz. The data is segmented to fixed time frame where the wave III and IV are located. The trials are smoothed with 10-trial moving window for every 10 trials, and parameterized by level-6 approximation wavelet coefficients of bior5.5 wavelet transform, which are used in subsequent estimation.

Selection of the optimal choice for noise covariances is performed using AIC criterion

$$AIC = -2\log p_{\hat{\boldsymbol{\theta}}_{ML}}(\mathbf{y}_{1:N}) + 2k \quad (20)$$

where k is the number of estimated parameters. The optimal

TABLE I. LOG-LIKELIHOODS AND AIC OF FITTED MODELS WITH DIFFERENT NOISE COVARIANCES.

Model	No. of Parameters	Log-like	AIC
(a) $\mathrm{R} = 0.193\mathbf{I}$, $\mathrm{Q} = 0.247\mathbf{I}$	2	-354.58	713.2
(b) R-Identity, Q-Full	196	-483.8	1359.5
(c) R-Identity, Q-Diagonal	14	-486.8	1001.5
(d) R-Diagonal, Q-Diagonal	28	-439.8	935.6
(e) R-Full, Q-Diagonal	210	-74.4	568.7
(f) R-Full, $\mathrm{Q} = 0.073\mathbf{I}$	197	-61.5	517.0

model is the one that minimizes the AIC values. We consider the simple ERP model with diagonal noise covariances with identical entries [Table I(a)] and five more complex variants of ERP model [Table I(b)-(f)] formulated by varying the covariance structure. For all models, the covariance parameters are estimated using EM algorithm with stopping criterion $\varepsilon = 0.0001$. The ML estimates of σ_v^2 and σ_w^2 for model (a) are 0.1933and 0.2472 respectively. The ML estimates of σ_w^2 for model (f) is 0.0725. Its estimated full covariance matrix $\mathbf{R}$ is shown in Fig. 1. The obtained log-likelihoods and AICs for different noise covariances are given in Table I. The AIC criterion suggests that the model with full $\mathbf{R}$ and identical diagonal $\mathbf{Q}$ outperforms all other models, with slightly better than use of non-identical diagonal $\mathbf{Q}$ instead. This somehow implies that assumption of each parameter changes differently is unnecessary. Fig. 2 shows that log-likelihood of the AIC best model increase monotonically and converges to -61.5after 10 iterations.

Fig. 2 shows the estimation results of ABRs by different techniques. The single-trial ABR estimates are presented in image (top plots) and epochs (middle plots). Fig. 3(a) shows the noisy ABR measurements with SNR=-18.91dB where the trace of wave V is hardly seen. The SNR for single-trials refer to [15]. The noise is greatly reduced by moving averaging every 10 trials, as shown in Fig. 3(b) which improves the SNR to -5.18dB. However, the wave V trace is barely seen in the epoch plots and still obscured by background noise in the image plot. Reconstruction from level-6 approximation wavelet coefficients which represent the low frequency component reveal the smooth ABR waveforms [Fig. 3(c)], with high frequency noise clearly removed, achieving better SNR of 0.17dB. Besides, the underlying single-trial dynamics are clearly exhibited and wave V is more profound. Fig. 3(d)-(f) show the clean estimates reconstructed from the estimated parameters by EMKS methods. All these estimates shows clearer wave V. As expected, EMKS further reduce random noise of ABR waveform shape in Fig. 3(c) giving SNR=5.13dB, 5.94dB and 7.09dB respectively. The use of full matrix in $\mathbf{R}$, allowing correlation and different noise volatilities along each dimensions in observation noise, gives better reduction in noise, than the setting $\mathbf{R}$ as diagonal with identical entries. Among the full $\mathbf{R}$ model, the estimates are comparable with slightly more changes for using identical $\mathbf{Q}$. From the

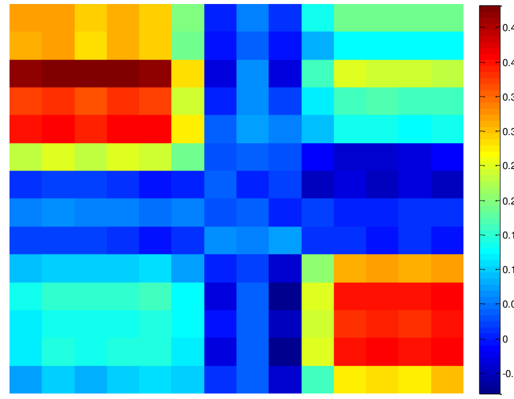

Figure 1. The ML estimates of full covariance matrix $\mathbf{R}$ of Model(f). (The image plot is scaled)

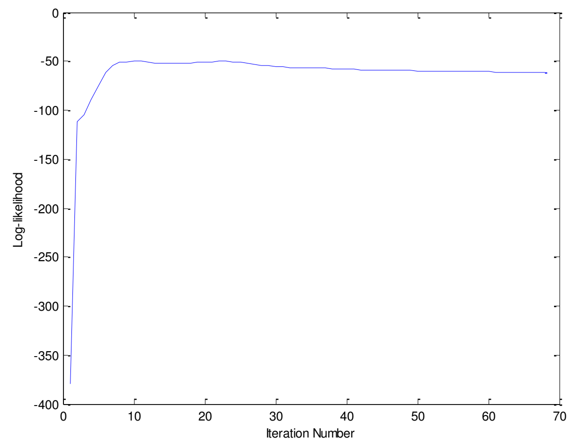

Figure 2. Log-likelihood as a function of interation of EM algorithm.

estimated latencies of wave V at bottom plots, the filtering methods show precise and consistent wave V location.

IV. CONCLUSION

This paper has applies an expectation-maximization (EM) based Kalman smoother (KS) approach for single-trial event-related potential (ERP) estimation. The existing studies only considers state estimation of ERP parameters by assuming the ERP model parameters known, which is in fact pre-specified by impractical manual tuning. The EM algorithm with KS solves both state and model estimation in a fully automatic way, and gives maximum likelihood estimates of the parameters. We also allow the noise covariance to be of arbitrary form to better models the individual parameter evolution and correlation in observation noise along each dimension. Selection using AIC criterion confirm the superiority of more complex models i.e. full $\mathbf{R}$ covariance over the use of diagonal with fixed entries. Evaluation on single-trial ABR estimation shows that use of full $\mathbf{R}$ covariance gives better performance in noise reduction. The model selected is based on one dataset and might not be robust against other dataset. The optimal model generally for ABR signals could be determined based on large dataset. The proposed method could be further evaluated on ABRs from subjects with hearing loss.

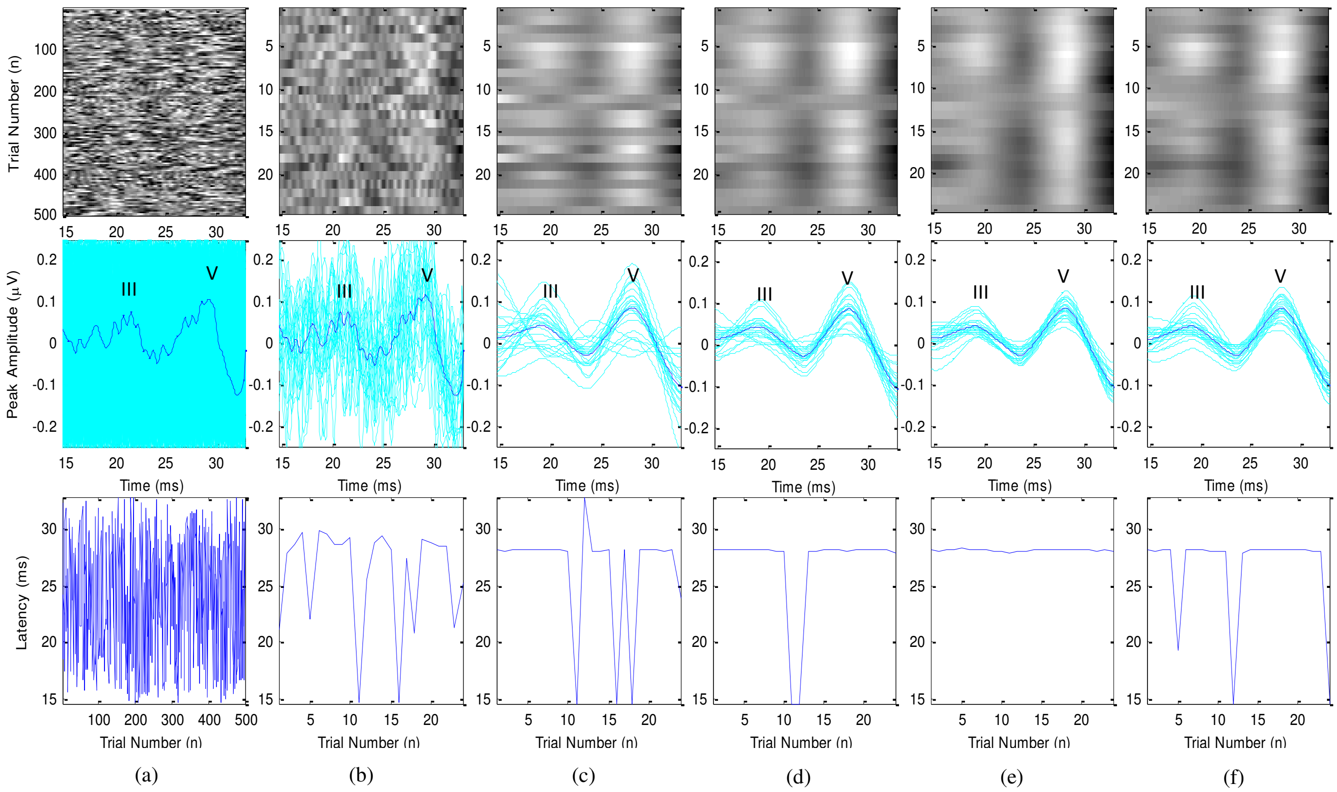

Figure 3. Estimation of single-trial ABR dynamics by different techniques. (a) Raw ABR measurements. (b) Moving -averaged ABRs. (c) Wavelet-denoied ABRs. (d)-(f) ABR estimates by EMKS methods. (d) $\mathbf{R} = 0.193\mathbf{I}$, $\mathbf{Q} = 0.247\mathbf{I}$ (e) R-Full, Q- Diagonal. (f) R-Full, $\mathbf{Q} = 0.073\mathbf{I}$. Top plots: ABR image; middle plots: epochs and their mean; bottom plots: latencies.

REFERENCES

[1] S. D. Georgiadis, P. O. Ranta-aho M. K. Tarvainen and P. A. Karjalainen, "Single-trial dynamical estimation of event-related potentials: A Kalman filter-based approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1397–1406, Aug. 2005.

[2] S. D. Georgiadis, P. O. Ranta-aho M. K. Tarvainen and P. A. Karjalainen, "A subspace method for dynamical estimation of evoked potentials," *Computational Intelligence and Neuroscience*, vol. 2007, 2007.

[3] H. R. Mohseni, K. Nazarpour, E. L. Wilding and S. Sanei, "The application of particle filters in single trial event-related potential estimation," *Physiological Measurement*, vol. 30, no. 10, pp. 1101-1116, 2009.

[4] D. Jarchi, S. Sanei, H. R. Mohseni and M. M. Lorist, "Coupled particle filtering: A new approach for P300-based analysis of mental fatigue," *Biomedical Signal Processing and Control*, vol. 6, pp. 175-185, 2011.

[5] 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.

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

[7] M. E. Khan and D. N. Dutt. "An expectation-maximization algorithm based Kalman smoother approach for event-related desynchronization (ERD) estimation from EEG," *IEEE Trans. on Biomed. Eng.,* vol. 52. No. 7, pp. 1191-1198, 2007.

[8] J. Frankel and S. King, "Speech recognition using linear dynamic models," *IEEE Trans. Audio, Speech*, *and Language Process*. vol. 15, no. 1, pp. 246–256, 2007.

[9] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. R. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.

[10] E. E. Holmes, "Derivation of the EM algorithm for constrained and unconstrained multivariate autoregressive state-space (MARSS) models," Northwet Fisheries Science Center, 2010.

[11] R. E. Delgado and O. Ozdamar, "Automated auditory brainstem response interpretation," *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, no. 2, pp. 227-237, 1994.

[12] M. J. Rowe, III, "Normal variability of the brain-stem auditory evoked response in young and old adult subjects," *Electroencephalography and Clinical Neurophysiology*, vol. 44, no. 4, pp. 459-470, 1978.

[13] R. M. Edwards, J. S. Buchwald, P. E. Tanguay and J. A. Schwafel, "Sources of variability in auditory brain stem evoked potential measures over time," *Electroencephalography and Clinical Neurophysiology*, vol. 53, no. 2, pp. 125-132, 1982.

[14] F. I. Corona-Strauss, W. Delb, B. Schick, D. J. Strauss, "Phase stability analysis of chirp evoked auditory brainstem responses by Gabor frame operators," *IEEE Trans Neural Syst. Rehabil. Eng.,* vol. 17. no. 6, pp. 530-536, 2009.

[15] J. Raz, B. Turetsky and G. Fein, "Confidence intervals for the signal-to-noise ratio when a signal embedded in noise is observed over repeated trials," *IEEE Trans. on Biomed. Eng.,* vol. 35, no. 8, pp. 646-649, 1988.