

Local Polynomial Modeling of Time-Varying Autoregressive Models With Application to Time–Frequency Analysis of Event-Related EEG

Z. G. Zhang*, *Member, IEEE*, Y. S. Hung, *Senior Member, IEEE*, and S. C. Chan, *Member, IEEE*

Abstract—This paper proposes a new local polynomial modeling (LPM) method for identification of time-varying autoregressive (TVAR) models and applies it to time–frequency analysis (TFA) of event-related electroencephalogram (ER-EEG). The LPM method models the TVAR coefficients locally by polynomials and estimates the polynomial coefficients using weighted least-squares with a window having a certain bandwidth. A data-driven variable bandwidth selection method is developed to determine the optimal bandwidth that minimizes the mean squared error. The resultant time-varying power spectral density estimation of the signal is capable of achieving both high time resolution and high frequency resolution in the time–frequency domain, making it a powerful TFA technique for nonstationary biomedical signals like ER-EEG. Experimental results on synthesized signals and real EEG data show that the LPM method can achieve a more accurate and complete time–frequency representation of the signal.

Index Terms—Electroencephalogram, event-related potential, local polynomial modeling (LPM), time–frequency analysis (TFA), time-varying autoregressive (TVAR) model.

I. INTRODUCTION

MOST biomedical signals are nonstationary in nature and usually contain numerous time-variant and transient components associated with underlying physiological or psychological activities. The time-varying autoregressive (TVAR) model is one of the most commonly used time-series models employed to describe the dynamics of nonstationary signals or characterize the variations associated with such signals, due to its simplicity and effectiveness. For instance, they have been used to classify electroencephalogram (EEG) in brain–computer interfaces [1], [2], to estimate changes of electromyogram during muscle contraction [3], and to track blood pressures [4]. The TVAR coefficients can also be used to estimate time–frequency distributions of biomedical signals. For example, TVAR time–frequency distribution has been employed to identify EEG os-

cillations [5]–[8] and to detect clinical events from intracranial pressure [9].

Generally, methods for estimation or identification of TVAR coefficients can be classified into three categories, namely: 1) the sliding-window approach [3], [10], 2) Kalman filtering (KF) [6], [9], and 3) basis expansion modeling [3], [4]. The sliding-window approach employs a time shifting window to derive a short-time data segment from the original signal at a particular time. The local data samples are assumed to be wide-sense stationarity, and hence, the AR coefficients of each segment can be estimated independently at different time instants using classical AR identification methods [11]. The selection of the window size is critical to the performance of the sliding-window approach and automatic window selection is still a difficult problem, which considerably hinders its practical application. The KF method employs a stochastic state-space model to describe the variation of the TVAR coefficients and it is an optimal estimator in the minimum mean-square error sense if the model is precisely given. Other adaptive algorithms similar to KF include recursive least-squares [3], [5] and the Kalman smoother (KS) [7], [8]. The basis expansion method, on the other hand, uses deterministic basis expansion to model the coefficient variations over a given time windowed signal, and the TVAR coefficients are approximated by a linear combination of known basis functions [3], [12], [13]. In real-world applications, however, the model parameters of the KF or the basis functions in basis expansion modeling are often unknown and they have to be estimated or chosen by trial and error [1], [6], [8], [12]. Therefore, the KF or basis expansion modeling method is highly dependent on these parameters, and their performance will be considerably degraded when the parameters are inappropriately estimated or chosen [12].

In this paper, we propose a new local polynomial modeling (LPM) approach for identification of TVAR models and, in particular, for time–frequency analysis (TFA) of nonstationary EEG signals. The LPM method models the TVAR coefficients locally by a set of polynomials within a data window having variable bandwidth. Given the local bandwidth, the polynomial coefficients can be estimated using the weighted least-squares estimator [14]–[17]. Asymptotic expressions for the bias and variance of the LPM estimator suggest that both of them are functions of the window bandwidth and there exists an optimal local bandwidth that minimizes the mean square error (MSE) at each time instant. For slowly varying TVAR coefficients, we would like the bandwidth to be large to reduce the estimation variance. For fast-varying coefficients, a small bandwidth is

Manuscript received August 4, 2010; revised September 22, 2010; accepted October 15, 2010. Date of publication October 25, 2010; date of current version February 18, 2011. This work was supported by the University of Hong Kong CRCG Small Project Funding. *Asterisk indicates corresponding author.*

*Z. G. Zhang is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam, Hong Kong (e-mail: zgzhang@eee.hku.hk).

Y. S. Hung and S. C. Chan are with the Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam, Hong Kong (e-mail: yshung@eee.hku.hk; scchan@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBME.2010.2089686

desirable in order to reduce the bias error. Since the TVAR coefficients may vary considerably over time, it is crucial to allow variable bandwidths to achieve the best bias-variance tradeoff at every time instant. Based on analytical expressions of the bias and the variance, a data-driven variable bandwidth selection (VBS) scheme is developed for the LPM method.

The proposed LPM-VBS TVAR identification method is general and may find various applications in analysis of biomedical signals. In this paper we will focus on its application to TFA of event-related EEG (ER-EEG). TFA techniques have been extensively and systematically applied in EEG study, and a desirable TFA technique for ER-EEG study should have both good time resolution (i.e., the ability to resolve adjacent events in time) and good frequency resolution (i.e., the ability to resolve adjacent sinusoids). However, the performances of conventional TFA techniques are usually highly dependent on the selection of model parameters (including the window size in the sliding-window approach, the forgetting factor in recursive least-squares algorithms, and model order and covariance of the KF) as well as signal statistics. As a result, higher time resolution is achieved at the expense of undesirable lower frequency resolution and vice versa. The proposed LPM-VBS method addresses the time-frequency resolution tradeoff problem in the TFA of ER-EEG by choosing its bandwidth adaptively to strive for a better compromise between time and frequency resolution. The experimental results on simulated signals and real ER-EEG signals recorded in a visual oddball paradigm show that the LPM-VBS TFA method can reveal various time-frequency components of ER-EEG more clearly and accurately than conventional TFA methods like continuous wavelet transform.

The rest of the paper is organized as follows. In Section II, the LPM for TVAR models is introduced. The VBS scheme for the LPM-based TVAR identification is developed in Section III. The LPM-VBS method is further extended to analyze ER-EEG in Section IV. Experimental results are presented in Section V. Final discussions and conclusions are given in Section VI.

II. LPM FOR IDENTIFICATION OF TVAR MODEL

In the TVAR model, a nonstationary discrete-time signal $z(t)$ is modeled as follows:

$$z(t_n) = \sum_{m=1}^M a_m(t_n) z(t_{n-m}) + \varepsilon(t_n) = \mathbf{a}^T(t_n) \mathbf{z}(t_n) + \varepsilon(t_n), \quad (1)$$

where t_n ($n = 1, 2, \dots, N$, N is the total number of samples) are the sampling time instants, M is the order of the TVAR model, $\mathbf{a}(t_n) = [a_1(t_n), a_2(t_n), \dots, a_M(t_n)]^T$ represents the TVAR coefficient vector, $\mathbf{z}(t_n) = [z(t_{n-1}), \dots, z(t_{n-M})]^T$, and $\varepsilon(t_n)$ is a zero mean white Gaussian process with variance $\sigma^2(t_n)$.

In the proposed LPM of TVAR model, the m th TVAR coefficient $a_m(t)$ is modeled locally at a time point $t = t_o$ as a p th-order polynomial [14]–[17]:

$$a_m(t) \approx \sum_{j=0}^p \frac{1}{j!} a_m^{(j)}(t_o) (t - t_o)^j = \sum_{j=0}^p \beta_m^{(j)}(t_o) (t - t_o)^j \quad (2)$$

where $a_m^{(j)}(t_o)$ are the j th derivatives of $a_m(t_o)$ and $\beta_m^{(j)}(t_o) = \frac{1}{j!} a_m^{(j)}(t_o)$ are the polynomial coefficients. Since the model residual $\varepsilon(t)$ is assumed to be a zero mean Gaussian process, the maximum likelihood estimation of $\beta_m^{(j)}(t_o)$ is equivalent to minimizing a locally weighted least squares criterion as follows:

$$\min_{\beta} \sum_{n=1}^N \left[z(t_n) - \sum_{m=1}^M \sum_{j=0}^p \beta_m^{(j)}(t_o) (t_n - t_o)^j z(t_{n-m}) \right]^2 K_h(t_n - t_o) \quad (3)$$

where $K_h(t_n - t_o) = \frac{1}{h} K(\frac{1}{h}(t_n - t_o))$ is a window used to control the number and weights of neighboring samples around t_o on the estimation of $\beta_m^{(j)}(t_o)$. $K_h(\cdot)$ can be obtained by scaling a basis window $K(\cdot)$ by a bandwidth h .

Next, we rewrite (3) more compactly in matrix form as

$$\min_{\beta} \{ [\mathbf{y} - \mathbf{X}(t_o) \boldsymbol{\beta}(t_o)]^T \mathbf{W}(t_o) [\mathbf{y} - \mathbf{X}(t_o) \boldsymbol{\beta}(t_o)] \} \quad (4)$$

where $\mathbf{y} = [z(t_1), z(t_2), \dots, z(t_N)]^T \in \mathbf{R}^N$,

$$\mathbf{X}(t_o) = \begin{pmatrix} z^T(t_1) & (t_1 - t_o) z^T(t_1) & \cdots & (t_1 - t_o)^p z^T(t_1) \\ z^T(t_2) & (t_2 - t_o) z^T(t_2) & \cdots & (t_2 - t_o)^p z^T(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ z^T(t_N) & (t_N - t_o) z^T(t_N) & \cdots & (t_N - t_o)^p z^T(t_N) \end{pmatrix} \in \mathbf{R}^{N \times (p+1)M}$$

$\boldsymbol{\beta}(t_o) = \{[\boldsymbol{\beta}^{(0)}(t_o)]^T, \dots, [\boldsymbol{\beta}^{(p)}(t_o)]^T\}^T \in \mathbf{R}^{(p+1)M}$ with $\boldsymbol{\beta}^{(j)}(t_o) = [\beta_1^{(j)}(t_o), \dots, \beta_M^{(j)}(t_o)]^T$, and $\mathbf{W}(t_o) = \text{diag}\{K_h(t_n - t_o)\}$. Note that $\boldsymbol{\beta}$, \mathbf{X} , and \mathbf{W} are all functions of time t_o , but the index t_o is omitted in the following discussion for notational simplicity when there is no ambiguity. The weighed least-squares solution to (4) is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y} \quad (5)$$

and $\hat{\mathbf{a}}$ is the first M entries of $\hat{\boldsymbol{\beta}}$.

Unlike the KF and basis expansion modeling, which assume explicit models for the TVAR coefficients, the LPM method relies on the Taylor expansion to construct local data models and to describe the coefficient variations. Therefore, LPM method is a good alternative to conventional KF and basis expansion modeling especially when the underlying coefficient model is unavailable. To investigate in-depth the behaviors of the LPM estimator, asymptotic expressions for the estimation bias and variance of $\hat{\boldsymbol{\beta}}$ are derived. (Due to space restrictions, the detailed asymptotic analysis is not given here, but can be found in Appendix I of the supplementary material available at http://www.eee.hku.hk/~zgzhang/publication/tbme2010_supp.pdf.) The asymptotic expressions show that the bias is an increasing function of h as $\mathbf{b}(\hat{\boldsymbol{\beta}}) \propto h^{p+1}$, while the variance is a decreasing function of h as $\mathbf{V}(\hat{\boldsymbol{\beta}}) \propto h^{-1}$. Using these expressions, the MSE of the TVAR coefficients \mathbf{a} can be

written as

$$\begin{aligned} \text{MSE}(\hat{\mathbf{a}}) &= E\{\|\mathbf{z}^T[\mathbf{a} - \hat{\mathbf{a}}]\|_2^2\} \\ &= \mathbf{b}^T(\hat{\mathbf{a}})\mathbf{R}_{zz}\mathbf{b}(\hat{\mathbf{a}}) + \text{Tr}(\mathbf{R}_{zz}\mathbf{V}(\hat{\mathbf{a}})), \end{aligned} \quad (6)$$

where $\mathbf{b}(\hat{\mathbf{a}})$ is the first M entries of $\mathbf{b}(\hat{\beta})$, $\mathbf{V}(\hat{\mathbf{a}})$ is the matrix consisting of the upper left $M \times M$ subblock of $\mathbf{V}(\hat{\beta})$, $(\mathbf{R}_{zz})_{1 \leq m_1, m_2 \leq M} = E[z(t_o - t_{o-m_1})z(t_o - t_{o-m_2})]$ is the cross-correlation matrix of z , and $\text{Tr}(\cdot)$ denotes the trace operation.

Consequently, it can be seen that, the bias term $\mathbf{b}^T(\hat{\mathbf{a}})\mathbf{R}_{zz}\mathbf{b}(\hat{\mathbf{a}})$ in (6) increases with h , while the variance term $\text{Tr}(\mathbf{R}_{zz}\mathbf{V}(\hat{\mathbf{a}}))$ decreases with h . Hence, there exists a locally optimal bandwidth $h^{\text{opt}}(t_o)$ that minimizes the MSE of \mathbf{a} . However, as quantities such as $\mathbf{b}(\hat{\mathbf{a}})$ and \mathbf{R}_{zz} in (6) are hard to obtain, it is difficult to determine the optimal bandwidth in practice. In the next section, we will introduce an empirical bandwidth selection scheme to approximate this optimal bandwidth.

III. LPM WITH VBS

In the empirical bandwidth selection method, a finite set of bandwidth parameters in an ascending order is given as

$$\mathbf{H} = \{h_1 < h_2 < \dots < h_\Gamma\} \quad (7)$$

where Γ is the number of bandwidths. The bias, covariance, and MSE values for each bandwidth in the set \mathbf{H} will be approximated and the optimal bandwidth $h^{\text{opt}}(t_o)$ will be selected as the one that minimizes the approximated MSE.

A. Approximation of Bias, Covariance, and MSE

The approximation of the bias, covariance, and MSE for the proposed TVAR method are developed based on classical technique of local polynomial regression [14], which have been proved to be effective in numerous applications. More precisely, the bias of $\hat{\beta}$, which is based on a p th-order LPM, is estimated using a Taylor expansion with an order higher than p (say, $p + p_{\text{ex}}$, where p_{ex} is an excess order) as

$$\hat{\mathbf{b}}(\hat{\beta}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \boldsymbol{\tau} \quad (8)$$

where $\boldsymbol{\tau}$ is an $N \times 1$ vector with its n th element given by $\sum_{\vartheta=1}^{p_{\text{ex}}} \sum_{m=1}^M \beta_m^{(p+\vartheta)}(t_o)(t_n - t_o)^{p+\vartheta} z(t_{o-m})$, and the higher order derivatives $\beta_m^{(p+\vartheta)}(t_o)$ are estimated by fitting a polynomial of degree $p + p_{\text{ex}}$ [14]. With this higher order model, the bias of our degree p model can be approximately estimated. However, the $(p + p_{\text{ex}})$ th-order LPM still requires an initial bandwidth, called the pilot bandwidth $h^*(t_o)$, the selection of which will be discussed later.

For the covariance, we consider the case of local homoscedasticity, from which the covariance matrix of $\hat{\beta}$ can be estimated from (5) as

$$\hat{\mathbf{V}}(\hat{\beta}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \sigma^2(t_o). \quad (9)$$

The variance of residual $\sigma^2(t_o)$ can be estimated as the normalized weighted residual sum of squares [14]

$$\hat{\sigma}^2(t_o) = \frac{\sum_{n=1}^N (z_n - \hat{z}_n)^2 K_{h^*}(t_n - t_o)}{\text{tr}\{\mathbf{W}^* - \mathbf{W}^* \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{W}^* \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{W}^*\}} \quad (10)$$

where \mathbf{X}^* and \mathbf{W}^* are, respectively, the design matrix and weighting matrix in the $(p + p_{\text{ex}})$ th-order LPM using the pilot bandwidth $h^*(t_o)$.

With the bias and covariance obtained, respectively, by (8) and (9), the MSE can be computed if an estimate of \mathbf{R}_{zz} is available. Because $z(t)$ is nonstationary, \mathbf{R}_{zz} is also time-varying and should be estimated locally. With the pilot bandwidth $h^*(t_o)$, \mathbf{R}_{zz} can be estimated as

$$\hat{\mathbf{R}}_{zz}(t_o) = \frac{\sum_{n=1}^N [K_{h^*}(t_n - t_o) z(t_n) z^T(t_n)]}{\sum_{n=1}^N K_{h^*}(t_n - t_o)}. \quad (11)$$

Finally, the MSE of $\hat{\beta}$ can be approximated by (6) using the quantities calculated in (8)–(11).

B. Pilot Bandwidth Selection by Intersection of Confidence Interval

An intersection of confidence intervals (ICI) method is employed in this study to determine the pilot bandwidth h^* . The ICI method is an empirical bandwidth selection method, which has been successfully applied to various areas, including local polynomial regression, image processing, and TFA [15] and [16]. The theoretical background of the ICI method is omitted to save space, and the reader is referred to [15] and [16] for details. Here, we only briefly review the algorithm of the ICI method.

Given the bandwidth set \mathbf{H} in (7), the ICI method determines the optimal bandwidth by comparing the confidence intervals of the estimate $\hat{\beta}_m(t_o; h_\gamma)$ with different bandwidths h_γ , $\gamma = 1, 2, \dots, \Gamma$, in \mathbf{H} . Consider a series of confidence intervals $D_\gamma = [L_\gamma, U_\gamma]$ with

$$U_\gamma = \hat{\beta}_m(t_o; h_\gamma) + \kappa \cdot \text{SD}(\hat{\beta}_m(t_o; h_\gamma)) \quad (12)$$

$$L_\gamma = \hat{\beta}_m(t_o; h_\gamma) - \kappa \cdot \text{SD}(\hat{\beta}_m(t_o; h_\gamma)) \quad (13)$$

where $\text{SD}(\hat{\beta}_m(t_o; h_\gamma))$ are the square roots of the diagonal elements of $\hat{\mathbf{V}}(\hat{\beta}(t_o; h_\gamma))$ in (9), κ is a threshold to adjust the width of the confidence intervals and can be chosen by the cross-validation criterion [15]. The ICI method examines the following quantities from the confidence intervals:

$$\bar{L}_\gamma = \max[\bar{L}_{\gamma-1}, L_\gamma], \quad \text{for } \gamma = 2, 3, \dots, \Gamma$$

$$\underline{U}_\gamma = \min[\underline{U}_{\gamma-1}, U_\gamma], \quad \text{for } \gamma = 2, 3, \dots, \Gamma$$

$$\bar{L}_1 = \underline{U}_1 = 0, \quad \text{for } \gamma = 1. \quad (14)$$

\bar{L}_γ is the largest upper bound of the confidence interval for bandwidth up to h_γ , while \underline{U}_γ is the corresponding lower bound. The largest γ for which $\underline{U}_\gamma \geq \bar{L}_\gamma$ gives the ICI-selected optimal bandwidth $\tilde{h}_m(t_o)$. When h is increased beyond $\tilde{h}_m(t_o)$, the bias will suddenly increase, while the variance will decrease gradually. Hence, the confidence intervals will no longer intersect above the bandwidth $\tilde{h}_m(t_o)$.

Since the ICI method actually produces an individual bandwidth for each tap of the TVAR coefficient vector, $\tilde{h}_m(t_o)$ will be further combined (generally, averaged) to produce the pilot bandwidth, i.e., $h^*(t_o) = \frac{1}{M} \sum_{m=1}^M \tilde{h}_m(t_o)$.

C. LPM With VBS

We now summarize the proposed LPM with VBS (LPM-VBS) method for identification of TVAR models as follows.

- Step 1) At each time instant t_o , $\hat{\beta}_m(t_o; h_\gamma)$ is calculated as in (5) by a p th-order LPM with each bandwidth h_γ , for $\gamma = 1, 2, \dots, \Gamma$, in the set \mathbf{H} .
- Step 2) A pilot bandwidth $h^*(t_o)$ is estimated from $\hat{\beta}_m(t_o; h_\gamma)$ and its estimated covariance using the ICI method as in (12)–(14).
- Step 3) The quantities $\beta_m^{(p+\theta)}(t_o)$, $\sigma^2(t_o)$, and $\hat{\mathbf{R}}_{zz}(t_o)$ are estimated with the pilot bandwidth $h^*(t_o)$.
- Step 4) For each bandwidth h_γ , $\text{MSE}(\hat{\beta}(t_o; h_\gamma))$ is approximated using (6)–(11), and the optimal bandwidth $h^{\text{opt}}(t_o)$ is obtained as $h^{\text{opt}}(t_o) = \arg \min_{h_\gamma \in \mathbf{H}} \{\text{MSE}[\hat{\beta}(t_o; h_\gamma)]\}$.
- Step 5) Finally, a p th-order LPM with $h^{\text{opt}}(t_o)$ is performed to obtain $\hat{\beta}(t_o; h^{\text{opt}}(t_o))$.

IV. LPM-VBS TFA FOR ER-EEG

A. TFA of ER-EEG

We now study the application of the proposed LPM-VBS method to the TFA of ER-EEG. ER-EEG is the stimulus-evoked or induced electrical activity originated from cortical neurons and measured on the scalp, and is an effective technique to disclose brain mechanisms [17]–[25]. Different experimental investigations have shown that ER-EEG is a very complex brain activity consisting of a number of components, including event-related potentials (ERPs), event-related synchronization/desynchronization (ERS/ERD), and other oscillations and potentials [18]–[22]. These components provide important information on anatomical sources, brain connectivity, sensory and cognitive processes, etc. Generally, the ERS/ERD and other oscillations are identified by detecting the concentration and change of the signal power in the time–frequency distributions (TFDs). As ER-EEG reflects rapidly changing brain states at a time scale of milliseconds [18]–[22], good time resolution is essential to the accurate identification of the onset and end time of EEG activities. On the other hand, the ER-EEG activities in different frequency bands carry distinct physiological significances [18]–[22]. Hence, both good time and frequency resolution are important to TFA of ER-EEG.

The most popular TFA method in EEG study may be the continuous wavelet transform (CWT) [19], [20]. The superiority of the CWT is that it addresses the time–frequency resolution tradeoff problem by applying a short window at high frequency and a long window at low frequency. However, CWT has degraded frequency resolution for high-frequency components and degraded time resolution for low-frequency components.

Therefore, CWT may not be able to precisely identify EEG components in a wide frequency range and, in particular, cannot acquire accurate temporal and spectral information on rapid changes of ER-EEG. The CWT method is a nonparametric TFA method that estimates the TFD entirely from the signal. Other nonparametric TFA methods, such as short-time Fourier transform [23], matching pursuit [24], windowed Lomb periodogram [25], Hilbert spectrum from empirical mode decomposition [26], [27], were also proposed for analyzing ER-EEG but they did not gain the same popularity as CWT.

Unlike the nonparametric TFA methods, the TVAR-based power spectral density (PSD) estimation is a parametric TFA method. Provided with an appropriate (e.g., TVAR) model, parametric TFA methods can usually achieve higher frequency resolution than nonparametric methods. In EEG studies, the TVAR model and the KF algorithm are amongst the most popular parametric models and identification methods respectively [1], [5]–[8]. As discussed before, the conventional KF is rather sensitive to the selection of model parameters and it often leads to a long tracking lag or a large estimation variance. The reported applications of KF/KS were mainly focused on EEG data with long-lasting and stable time–frequency characteristics, such as eye open/closed EEG [7] and motor-imagery EEG [8]. As far as we are aware, the KF/KS method is still not a well-accepted option for TFA of highly nonstationary ER-EEG data, say, recorded in an ERP experiment.

The proposed LPM-VBS method provides a good alternative for TFA of highly nonstationary ER-EEG data because it can achieve a good tradeoff between time and frequency resolution by employing variable bandwidths. By modeling an ER-EEG signal $z(t)$ by the TVAR model of (1), its time-varying PSD can be calculated from the estimated coefficients and the variance of residual as

$$P(t, f) = \hat{\sigma}^2(t) \left/ \left| 1 - \sum_{m=1}^M \hat{a}_m(t) e^{-j2\pi f m} \right|^2 \right. \quad (15)$$

where the variance $\hat{\sigma}^2(t)$ can be estimated as in (10).

B. Complexity and Parameter Selection

We now discuss several practical issues when applying the LPM method in ER-EEG study.

1) *Computational Complexity*: We first consider the arithmetic complexity of a p -order LPM estimator in (5) at one time point. If $N \gg (p+1)M$, the complexity of LPM is approximately $O(N)$. The complexity can be simplified by using windows with finite support. Denote the number of samples included in a finite-support window as N_K , then the complexity of the LPM estimator will decrease considerably to $O(N_K)$. Since N_K increases with h , a large bandwidth will increase the complexity. Because the approximation of MSE requires a $(p + p_{\text{ex}})$ th fitting and the VBS scheme needs to calculate LPM solutions with a set of bandwidths, the overall complexity of the LPM method is higher.

2) *Selection of Window K* : A window with finite support is desirable because it can reduce the complexity significantly. In this paper, following the recommendation of [14]–[17], the

Epanechnikov window is employed

$$K(u) = \begin{cases} \frac{3}{4}(1 - |u|^2), & |u| < 1 \\ 0, & |u| \geq 1. \end{cases} \quad (16)$$

For an Epanechnikov window with bandwidth h , $K_h(u) = \frac{1}{h}K(u/h)$, only the time points included in the interval $(t_0 - h, t_0 + h)$ are included for local estimation, and thus, the effective window length is $2h$. Other types of windows, such as Gaussian and Hanning windows, can also be used. As their performances are similar, a detailed comparison of different window types is not performed in this paper.

3) *Selection of Bandwidth Set \mathbf{H}* : For the LPM estimator in (5) to be solvable, the minimum bandwidth, h_1 or h_{\min} , should be chosen to ensure that the number of samples in the interval $(t_0 - h_{\min}, t_0 + h_{\min})$ is equal to or larger than the number of variables to be estimated, $(p + p_{\text{ex}} + 1)M$. Suppose the signal is uniformly distributed at a sampling rate of f_s , then we get the following condition for the bandwidth:

$$h \geq [(p + p_{\text{ex}} + 1)M]/(2f_s). \quad (17)$$

On the other hand, there is no theoretical limit on the maximum number of elements in the bandwidth set, h_{Γ} or h_{\max} . However, too large a window will result in high computational complexity. In practice, the maximum bandwidth can be determined from the properties of the signal under study. For example, the maximum window size in our EEG study can be chosen as 120 ms, because previous study on EEG segmentation showed that the duration of most EEG microstates (defined as data segments with stable field topography) is less than 120 ms [28]. As for other bandwidths in \mathbf{H} , more bandwidths may lead to more refined results, but it will also increase the computational load. To achieve a tradeoff between performance and complexity, we generally select 2–3 bandwidths between h_{\min} and h_{\max} . From experimental results, it was found that the proposed bandwidth setting gave satisfactory results in practice.

4) *Selection of Polynomial Order p* : Theoretically, a larger p gives the polynomial more approximation power, and hence, a smaller bias. However, it also increases the variability of the estimates because more variables (namely $(p + 1)M$) have to be estimated. On the other hand, the number of variables should be smaller than the number of measurements N_K in the smallest window so that (8) is solvable. Hence, a large p also requires a larger window and potentially reduces the time resolution of the TFD. More details about the selection of p can be found in the classical text by Fan and Gijbels [14]. Since one of the main advantages of the LPM-VBS method is its fine time-resolution, p and p_{ex} should be selected as small as possible. In this study, we set $p = 0$ and $p_{\text{ex}} = 1$, which gave satisfactory results in the experiments. Although $p = 0$ implies that the LPM method is reduced to conventional sliding-window approach using the Yule–Walker solution, the window size is still variable, and hence, the bias-variance tradeoff problem can still be satisfactorily addressed in the LPM framework.

5) *Selection of TVAR Order M* : The model order M of the TVAR is important to the frequency resolution of the PSD estimation [11]. A small M discriminate different closely spaced

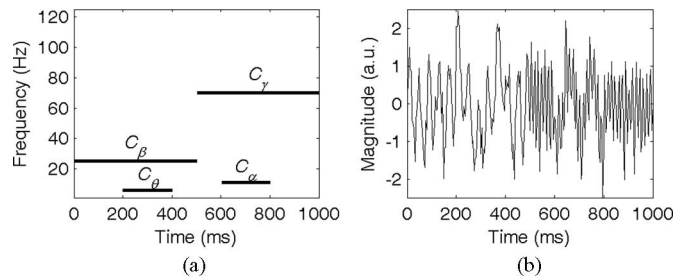


Fig. 1. Simulation model: (a) locations of four sinusoidal components in the time–frequency domain; (b) one realization of the simulated signal.

frequency components, while spurious frequency peaks may appear when M is too large. A number of model-order selection methods, such as the Akaike’s information criterion (AIC) and Bayesian information criterion (BIC), are available for choosing an appropriate TVAR model order M . Note that the order of the TVAR can be made adaptive for each sample or each data segment, which can also be determined from the above criteria. Such time-variant order and bandwidth scheme is potentially a valuable tool for time-variant signal analysis. In this paper, since our focus is on time-variant signals with time-varying bandwidth, the TVAR order will be fixed to simplify the implementation and mathematical analysis.

V. EXPERIMENTAL RESULTS

We first test the LPM-VBS method and compare it with other conventional methods (including random-walk KF/KS, and LPM with fixed bandwidths) for identifying simulated TVAR models. The results are described in [29] and Appendix II of the supplementary material and they show that the LPM-VBS method has better performance for TVAR identification with different extents of coefficient variation.

The performance of the LPM-VBS TFA method is now evaluated and compared using simulated and real EEG data.

A. TFA of Simulated Data

Different TFA methods, including CWT-, KS-, and LPM-based PSD estimation, are compared using simulated EEG signals. This allows us to quantify the performances of the various algorithms. The signal, of duration 1 s and sampling frequency 250 Hz, consists of the sum of four distinct sinusoidal components with different frequencies and durations (C_{θ} : 6 Hz, 200–400 ms; C_{α} : 11 Hz, 600–800 ms; C_{β} : 25 Hz, 0–500 ms; C_{γ} : 70 Hz, 500–1000 ms), which are meant to emulate respectively the theta, alpha, beta, and gamma bands of EEG signals. The four sinusoids have the same amplitude of 1 and random phases between $-\pi$ and π . An additive white Gaussian noise with a SNR of 5 dB is added. The simulated signal model in the time–frequency domain and one realization of the simulated signal are presented in Fig. 1.

The order of TVAR models was determined according to the BIC. The model orders selected by BIC varied slightly in different realizations and for different TVAR identification methods. Since the standard deviation is less than 2 in 100 realizations for each TVAR identification method and the mean values are all

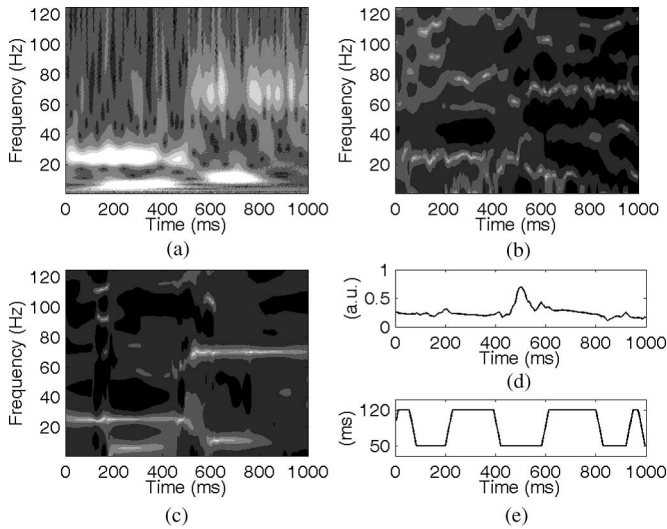


Fig. 2. Time–frequency distributions of one example of simulated signal: (a) CWT, (b) ARWKS, (c) LPM-VBS, (d) the estimated variance of residual in LPM-VBS, and (e) the variable bandwidths used in LPM-VBS. The TFDs of (a)–(c) are shown in a logarithmic scale, and the gray scale is from black (minimum) to white (maximum).

between 11 and 13, we adopted a model order M of 12 in this simulation. In the LPM method, $p = 0$ and $p_{ex} = 1$. The resulting minimum bandwidth is $h_{min} = 48$ ms, and the bandwidth set is $\mathbf{H} = \{50, 75, 100, 120\}$ ms.

In the KS method, the state transition matrix is chosen as an identity matrix to obtain a random-walk model [12]. The covariance matrix of the state noise or the observation noise in the random-walk KS is recursively estimated as a weighted sum of the previous covariance estimate and the outer product of the current residual vector [6]. Such adaptive random-walk KS (ARWKS) coupled with covariance estimation is a commonly used simplification of the standard KS in practical implementations where the true parameters of the state-space model are unknown. The ARWKS framework has been adopted in various applications including TFA study of EEG [1], [6], [7]. In this paper, the parameters for estimating the covariance matrices follow the setting in [6]. In addition, the KF method is also tested but its results are not as good as those of KS, which is consistent with the conclusions in [7] and [8]. Thus, the results of KF are not presented in this paper.

In the CWT algorithm, complex Morlet wavelet basis

$$\psi(t) = \frac{1}{\sqrt{\pi f_b}} e^{2\pi i f_c t} e^{-t^2 / f_b}$$

with $f_b = f_c = 1$ is adopted. Although other choices of wavelet basis may lead to better time resolution or frequency resolution in specific time–frequency regions, they cannot simultaneously improve the time and frequency resolution in the whole time–frequency domain. As shown in the simulation results, the selected wavelet gives a satisfactory tradeoff between time resolution and frequency resolution in a wide frequency range. All the TFDs are evaluated at each sampling time instant and from 0.5 Hz to 125 Hz (the Nyquist frequency) with a frequency step of 0.5 Hz.

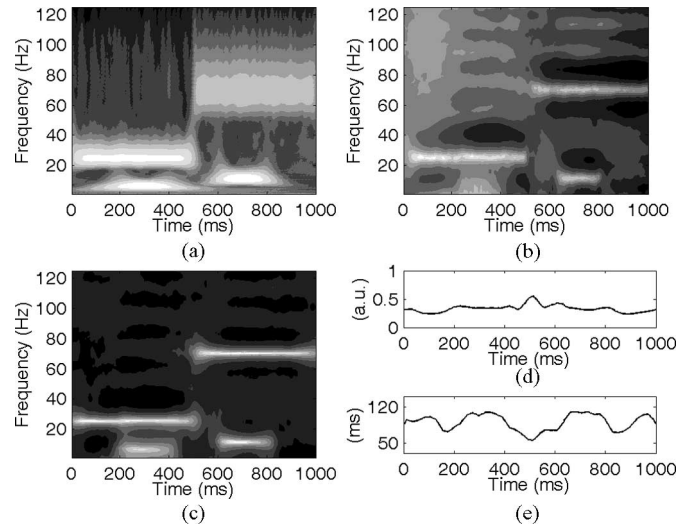


Fig. 3. Averages of time–frequency distributions of 100 trials of simulated signals: (a) CWT, (b) ARWKS, (c) LPM-VBS, (d) the average of estimated variance of residual in LPM-VBS, and (e) the average of variable bandwidths used in LPM-VBS. The TFDs of (a)–(c) are shown in a logarithmic scale, and the gray-scale is from black (minimum) to white (maximum).

The TFDs of one representative simulation are presented in Fig. 2. It can be seen that the CWT has low time resolution for low-frequency components (C_θ and C_α), and low frequency resolution for high-frequency components (C_γ). The ARWKS method exhibits considerable variability in estimating the sinusoidal components and it also produces a number of artificial time–frequency components. The poor performance of ARWKS is due to its inability to accurately estimate the model parameters. In contrast, the LPM-VBS method gives a TFD with good time resolution and few artificial time–frequency components.

In TFA of signals with low SNR, such as ER-EEG, if a number of signal trials repeatedly measured for one specific event are available, it is common to calculate the average of TFDs of all the trials so that the components having stable time–frequency characteristics can be enhanced, while the artificial components can be smoothed out [19], [20]. In this simulation, the averaged TFDs across 100 simulated trials are presented in Fig. 3. It can be seen that, the LPM-VBS method has better representations for the four sinusoidal components than the other two methods. CWT is unable to attain good time resolution and frequency resolution simultaneously, and ARWKS has large variability when estimating stable time–frequency components. Fig. 3(e) showed the average of variable bandwidths used in 100 LPM-VBS simulations and it can be seen that LPM-VBS employs a small bandwidth when the time–frequency characteristics are stable (i.e., the TVAR coefficients are slowly varying in 200–400 and 600–800 ms) and a large bandwidth when the time–frequency characteristics change rapidly (i.e., the TVAR coefficients are fast-varying around 200, 400, 500, 600, and 800 ms).

We next compare the three TFA methods in a more quantitative manner. The instantaneous PSD values at 6, 11, 25, and 70 Hz were extracted from TFDs of 100 realizations for comparison. Because the PSD values obtained from CWT are calculated as the squared magnitude of the CWT coefficients

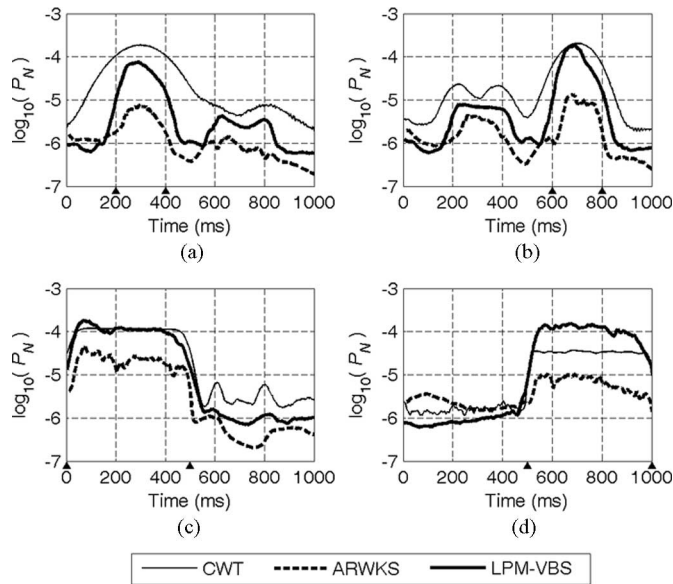


Fig. 4. Instantaneous normalized PSD extracted from time–frequency distributions of 100 trials of simulated signals using CWT, ARWKS, and LPM-VBS at frequencies of: (a) 6 Hz, (b) 11 Hz, (c) 25 Hz, and (d) 70 Hz. The black triangles under the time-axis mark the onset and end time of the sinusoidal component at the frequency.

and they are not in the same scale as the ARWKS- and LPM-based PSDs calculated using (15), the instantaneous PSD values from different TFA methods are normalized for the purpose of comparison. The normalized instantaneous PSD $P_N(t, f)$ is obtained by dividing by the sum of all the PSD values in one TFD, i.e., $P_N(t, f) = P(t, f) / \sum_t \sum_f P(t, f)$. The mean values of $P_N(t, f)$ at $f = 6, 11, 25,$ and 70 Hz calculated from 100 realizations are plotted in Fig. 4.

The time resolution of the three TFA methods was evaluated by checking whether the onset and end times of the four components were accurately identified. It can be seen from Fig. 4 that, the CWT has poor time resolution for the low-frequency component C_θ . But the time resolution of CWT is improved with the increase of frequency, so that the onset time for C_γ can be accurately identified. On the other hand, the time resolution of ARWKS or LPM-VBS is very consistent in a wide frequency range and it is better than that of CWT for low-frequency components. In addition, as we can see from Fig. 3(b) and Fig. 4(d), the ARWKS-based TFD shows an artificial high-frequency (>70 Hz) component before 200 ms, because ARWKS needs some time to reach convergence, and thus, it has a large estimation error shortly after the onset of event. Such error is a big disadvantage of ARWKS, because it will considerably contaminate meaningful short-latency high-frequency components in ER-EEG.

The frequency resolution was evaluated by comparing the instantaneous PSD values of the four sinusoidal components. Because the four components have identical amplitudes of 1, they should have identical P_N . Moreover, the better the frequency resolution, the smaller will be the frequency leakage, and vice versa. It can be seen from Fig. 4 that the P_N 's of the four components estimated by CWT decrease with the fre-

quency ($\log_{10}(P_N)$ is over -4.0 for C_θ from 200 to 400 ms, and is around -4.5 for C_γ from 500 to 1000 ms). This verifies that the CWT has decreased frequency resolution for high-frequency components. In contrast, the ARWKS and LPM-VBS methods have relatively stable P_N for the four components ($\log_{10}(P_N)$ is around -0.4 in LPM-VBS and around -0.5 in ARWKS). The higher PSD value of LPM-VBS over ARWKS suggests that the LPM-VBS-based TFDs have better frequency resolution, which is also observed from Fig. 3.

We further investigate the phenomenon of frequency leakage in different TFA methods. It is often necessary to calculate the power in one given frequency range in EEG study, and thus, small frequency leakage (i.e., high frequency resolution) is highly desirable. As $C_\theta, C_\alpha,$ and C_β are closely spaced in the frequency domain, it is possible that the PSD value leaks out to other frequencies. In Fig. 4(b), for example, P_N is evidently high between 200 ms and 400 ms at 11 Hz for all the TFA methods, which is actually due to the leaking PSD from C_θ at 6 Hz. In CWT of Fig. 4(b), $\log_{10}(P_N)$ from C_α between 600 and 800 ms is slightly larger than -0.4 and $\log_{10}(P_N)$ leaking from C_θ between 200 ms and 400 ms is slightly larger than -0.5 , and hence, the difference is around 0.1. Such differences of P_N at 11 Hz in Fig. 4(b) are around 1.5 and 0.5 for LPM-VBS and ARWKS, respectively, and similar observations can be made in Fig. 4(a) and (c). These results indicate that the LPM-VBS method has less frequency leakage and better frequency resolution than CWT and ARWKS.

B. Real ER-EEG

Next, the LPM-VBS method is evaluated using real ER-EEG data. The ER-EEG data in a visual oddball paradigm are used because this experiment has been extensively studied and the time–frequency characteristics of such ER-EEG data have been revealed by various means. Therefore, we can validate our LPM-based results by comparison with previous findings.

The ER-EEG data were recorded using a 128-channel EEG/ERP system manufactured by Electrical Geodesic, Inc., and referenced to vertex. The standard stimulus is a character ‘‘O’’ with 80% occurrence, while the target visual stimulus is a character ‘‘X’’ with 20% occurrence. A total of 250 stimuli were presented, and thus, 200 were standard stimuli and 50 were target stimuli. The duration of each stimulus was 500 ms, and the interstimulus interval was 1500 ms. The stimuli were presented on a CRT monitor with a refresh rate of 60 Hz.

Ten subjects (eight males and two females) aged 19–32 years (mean \pm SD: 25 ± 4) were recruited in this study. The subject was asked to press the correct button immediately after he/she recognized the stimulus. The sampling rate was 250 Hz. The data were high-pass filtered at 0.1 Hz. Automatic artifact correction for EEG by blind source separation was further performed to attenuate the artifacts [30].

The LPM-VBS method will be compared to the CWT, which contributed most of findings of time–frequency features in such paradigm. The ARWKS method is not used for comparison because it is seldom used in study of ER-EEG due to its limitations such as slow convergence. The orders of TVAR models

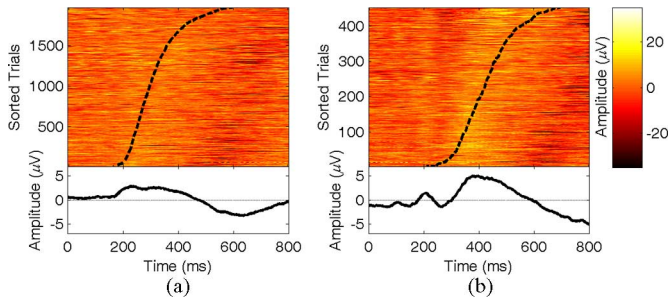


Fig. 5. Grand averages of ER-EEG trials (recorded at Pz): (a) trials with correct responses to standard stimuli, and (b) trials with correct responses to target stimuli. The bold dashed lines denote the response time.

are selected by the BIC. The BIC-selected orders are different from trial to trial. Since it is time consuming to determine an individual order for each trial in a total of 250×10 trials recorded at one electrode, we randomly select 100 trials and obtain their optimal orders using BIC. The 100 BIC-based orders have a mean value of 14.7 and a SD of 1.2. Thus, we employ a TVAR model order of $M = 15$. In the LPM method, $p = 0$ and $p_{ex} = 1$. The minimum bandwidth is $h_{min} = 60$ ms, and the bandwidth set is $H = \{60, 80, 100, 120\}$ ms. Other parameters for the LPM and CWT methods are the same as those in previous simulations.

The performances of CWT and LPM-VBS methods are compared using EEG activities produced by the two types of stimuli: standard or target. Single-trial ER-EEG signals recorded at Pz in correct responses to standard and target stimuli are presented in Fig. 5, and they are sorted by the response time. The mean response time is respectively 315 ms and 425 ms for standard and target stimuli. The grand averages across trials and subjects are also given in Fig. 5 and we can see that P300 and other early ERP components (such as N100 and P200 [17]) are apparent in the grand average of correct responses to target stimuli.

Figs. 6 and 7 show the grand averages of stimulus-locked and response-locked TFDs across trials and subjects, respectively. To observe the changes in power of the TF components caused by the events, the log-transformed LPM- or CWT-based TFDs are baseline-corrected by subtracting the average power of the signal in a pre-stimulus interval (-50 to -300 ms), which is similar to the usual approaches in [19]–[21]. Furthermore, as suggested in [20], all these TFDs are expressed as Z -scores to make the comparison between different TFA methods possible. For each time–frequency point, a t -test is used to determine whether the power spectra of responses to standard and target stimuli are significantly different, resulting in a p -value map in the time–frequency domain. It has been shown in [31] that the log-transformed TFD can be regarded as normally distributed, and thus, the t -test is applicable. The significance threshold α^+ is corrected by the false discovery rate (FDR) procedure [32] to address the problem of multiple comparisons.

We searched all meaningful time–frequency components in TFDs or statistical contrast maps of Figs. 6 and 7, and their locations in the time–frequency domain are indicated by a series of regions of interest (ROIs). Some of the ROIs (A, C, D, E, F, G, I, and J, shown as solid boxes in Figs. 6 and 7) are identified

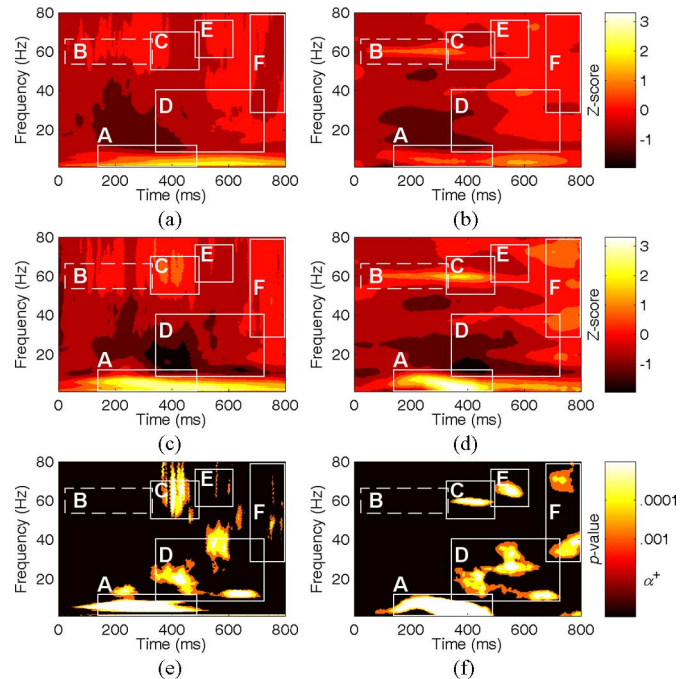


Fig. 6. Grand averages of TFDs of ER-EEG (time-locked to the stimulus, recorded at Pz) with correct responses to standard or target stimuli: (a) CWT of standard stimuli, (b) LPM-VBS of standard stimuli, (c) CWT of target stimuli, (d) LPM-VBS of target stimuli, (e) CWT-based contrast between standard and target stimuli (FDR-corrected $\alpha^+ = 0.008$), and (f) LPM-VBS-based contrast between standard and target stimuli (FDR-corrected $\alpha^+ = 0.008$).

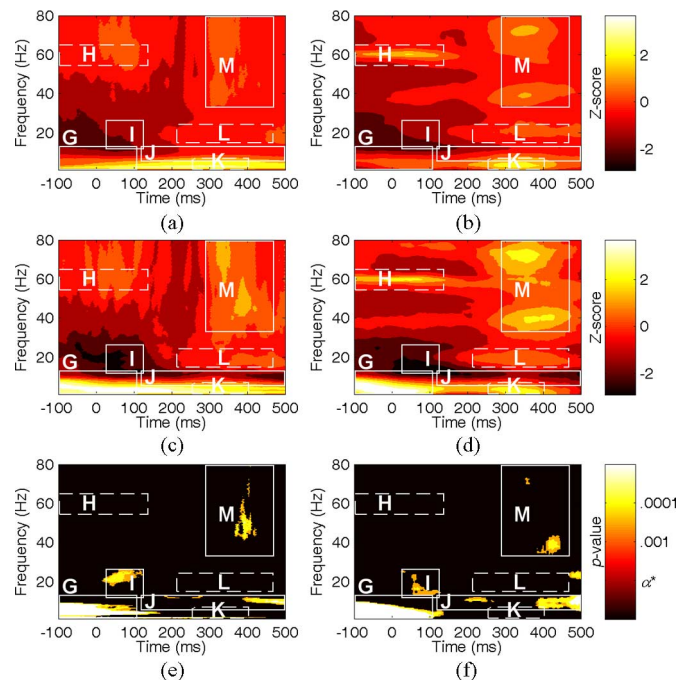


Fig. 7. Grand averages of TFDs of ER-EEG (time-locked to the response, recorded at Pz) with correct responses to standard or target stimuli: (a) CWT of standard stimuli, (b) LPM-VBS of standard stimuli, (c) CWT of target stimuli, (d) LPM-VBS of target stimuli, (e) CWT-based contrast between standard and target stimuli (FDR-corrected $\alpha^+ = 0.004$), and (f) LPM-VBS-based contrast between standard and target stimuli (FDR-corrected $\alpha^+ = 0.004$). The time instant 0 of the time axis denotes the response time.

by significant differences found in the statistical contrast maps, while others (B, H, K, and L, shown as dashed boxes in Figs. 6 and 7) are identified by significant power increases or decreases (t -test, $p < 0.05$; results not shown here) found in the LPM- or CWT-based TFDs.

In Fig. 6, the ROI A reaches its maximum from 100 to 450 ms in the theta and alpha bands for target stimuli. This predominated maximum should be contributed by the P300 and other early ERP components elicited or enhanced by the target stimuli. In Fig. 7(d), the maximal value in G, which should be caused by the P300 component, can be easily distinguished from the later theta-band component in K, which should be caused by eye artifacts or error detection [22]. However, the power peaks in G and K can hardly be separated in the CWT result of Fig. 7(c) due to the coarse time resolution of CWT in low-frequency range. It can also be observed that a long-duration of high-frequency (around 60 Hz) component in B exists in the LPM-based TFDs for both standard and target stimuli, while this 60-Hz component looks like some intermittent responses in the gamma band from the CWT results. In addition, this 60-Hz component does not show significant difference for different stimuli. All these observations (stable frequency around 60 Hz, long-lasting duration, similar responses to different stimuli, duration around 500 ms) suggest that the 60-Hz component in B is very likely to be the steady-state visual evoked potential (SSVEP) in response to the refresh rate (60 Hz) of the computer monitor [33]. We further observed from the LPM-based TFDs that, a gamma-band activity in C is significantly increased in response to target stimuli, which should be the induced gamma-band response (iGBR) reported in [10], [34], [35]. However, the frequency resolution of the high-frequency component in C is rather poor in CWT. Similar differences between the CWT and LPM results can be observed for other high-frequency (especially gamma-band) components. A summary of time–frequency components discovered in the experiments and their properties are given in Table I.

Overall, our LPM-based results (in theta, alpha, and beta bands) are in high accordance with previous findings in the visual oddball paradigm [22]. However, the CWT-based findings in [22] were observed only up to around 40 Hz. Gamma-band activities revealed by CWT or other TFA methods were reported in many other studies, but their observations were generally limited in the high-frequency range. It should be noted that all the previous studies focus only on specific and rather limited time–frequency areas due to the limited TF resolution of conventional TFA methods. On the other hand, the proposed LPM-VBS based TFD can reveal a complete (from theta to gamma bands) and clear (with good time and frequency resolution) time–frequency characteristics of ER-EEG, which is consistent with the findings obtained in the somewhat scattered literature. Therefore, it offers a convenient framework for analyzing EEG and other biomedical signals without the need for user intervention through trial and error selection of model parameters. The proposed method may facilitate further analysis, such as statistical analysis, source location, etc., so as to provide more thorough interpretation of the disclosed time–frequency components, and hence, reveal the underlying physiological and psychological mechanisms.

TABLE I
TIME–FREQUENCY COMPONENTS IDENTIFIED BY THE LPM-VBS METHOD
AND THEIR CHARACTERISTICS IN THE VISUAL ODDBALL PARADIGM

ROI	Identified components, interpretation, and references	Time (ms)	Frequency (Hz)	p -value *
A, G	P300 and early ERP components, elicited by infrequent stimuli; [22]	100 - 450	1-10	<0.0001
B, H	SSVEP, elicited by monitor flicker; [33]	50-500	60	$>\alpha^*$
C, E, F, M	iGBR, related to awareness, attention, memory, etc.; [10], [34], [35]	400-800	30-80	$<\alpha^*$
D, I, J	ERD or alpha/beta blocking; [21], [22]	300-800	10-30	$<\alpha^*$
K	Eye artifacts, error detection, etc.; [22]	after 600	1-5	$>\alpha^*$
L	ERS or beta increase; [21], [22]	after 600	15-25	$>\alpha^*$

The dominant p -values in the ROIs, indicating the significance level of difference between the identified components in response to standard stimuli and target stimuli. α^ is the FDR-corrected significance threshold.

VI. DISCUSSION AND CONCLUSION

A novel LPM approach and a data-driven VBS scheme for identification of TVAR models were presented, with emphasis on its application to TFA of ER-EEG data. The experimental results showed that the LPM-VBS method can reveal a complete and clear picture of time–frequency components in ER-EEG. The proposed LPM-based TFA method is expected to find various clinical and research applications in EEG study of cognitive neuroscience, neurophysiology, and psychology.

Although one major application of LPM method is to investigate the time–frequency distribution of biomedical data (in particular, ER-EEG data), the potential applications of the LPM method, as an effective TVAR identification method, has yet to be fully explored. For instance, since another important application of TVAR modeling is the tracking and prediction of time-varying biomedical signals [3], the LPM method may also find its usage here. Furthermore, it is shown in [17] that the LPM-VBS method can be further extended for identification of time-varying systems with mild nonlinearity by means of appropriate linearization. As shown in [36] and [37], nonlinear dynamics are commonly exhibited in biomedical signals or systems, and therefore the effectiveness of LPM for identification of nonlinear biomedical systems is a potentially fruitful area for future work.

Finally, it is worth noting that, the polynomial order p is chosen to be zero in the current TFA study but a higher order can be employed to further explore the advantages of LPM in other applications. A higher polynomial order implies a decreased estimation bias, but it will also lead to increased computational complexity and larger variability. Therefore, proper selection and comparison of polynomial order is a potential research area for future applications of the LPM method.

ACKNOWLEDGMENT

The authors would like to thank Dr. K. H. Ting for providing the EEG data and M. Chen for her comments.

REFERENCES

- [1] A. Schlögl, *The Electroencephalogram and the Adaptive Autoregressive Model: Theory and Applications*. Aachen, Germany: Shaker Verlag, 2000.
- [2] Q. G. Wei, Y. J. Wang, X. R. Gao, and S. K. Gao, "Amplitude and phase coupling measures for feature extraction in an EEG-based brain-computer interface," *J. Neural Eng.*, vol. 4, no. 2, pp. 120–129, 2007.
- [3] Z. G. Zhang, H. T. Liu, S. C. Chan, K. D. K. Luk, and Y. Hu, "Time-dependent power spectral density estimation of surface electromyography during isometric muscle contraction: Methods and comparisons," *J. Electromyogr. Kinesiol.*, vol. 20, no. 1, pp. 89–101, Feb. 2010.
- [4] K. H. Chon, H. Zhao, and R. Zou, "Multiple time-varying dynamic analysis using multiple sets of basis functions," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 956–960, May 2005.
- [5] J. P. Kaipio and P. A. Karjalainen, "Estimation of event-related synchronization changes by a new TVAR method," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 8, pp. 649–656, Aug. 1997.
- [6] M. Arnold, H. Miltner, H. Witte, R. Bauer, and C. Braun, "Adaptive AR modelling of nonstationary time series by means of Kalman filtering," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 553–562, May 1998.
- [7] M. P. Tarvainen, J. K. Hiltunen, P. O. Ranta-aho, and P. A. Karjalainen, "Estimation of nonstationary EEG with Kalman smoother approach: An application to event-related synchronization (ERS)," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, pp. 516–524, Mar. 2004.
- [8] 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. Biomed. Eng.*, vol. 54, no. 7, pp. 1191–1198, Jul. 2007.
- [9] M. Aboy, O. W. Márquez, J. McNames, R. Hornero, T. Trong, and B. Goldstein, "Adaptive modelling and spectral estimation of nonstationary biomedical signals based on Kalman filtering," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 8, pp. 1485–1489, Aug. 2005.
- [10] G. G. Supp, A. Schlögl, N. Trujillo-Barreto, M. M. Müller, and T. Gruber, "Directed cortical information flow during human object recognition: Analyzing induced EEG gamma-band responses in brain's source space," *PLoS One*, vol. 2, p. e684, Aug. 2007. DOI: 10.1371/journal.pone.0000684.
- [11] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [12] M. Niedzwiecki, *Identification of Time-varying Processes*. Chichester, U.K.: Wiley, 2000.
- [13] T. S. Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *J. Roy. Stat. Soc. Ser. B*, vol. 32, pp. 312–322, 1970.
- [14] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman and Hall, 1996.
- [15] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567–2571, Sep. 1999.
- [16] Z. G. Zhang, S. C. Chan, K. L. Ho, and K. C. Ho, "On bandwidth selection in local polynomial regression analysis and its application to multi-resolution analysis of non-uniform data," *J. Signal Process. Syst.*, vol. 52, no. 3, pp. 263–280, Sep. 2008.
- [17] S. C. Chan and Z. G. Zhang, "Local polynomial modeling and variable bandwidth selection for time-varying linear systems," *IEEE Trans. Instrum. Meas.*, to be published. DOI: 10.1109/TIM.2010.2064850.
- [18] E. Niedermeyer and F. H. Lopes da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed., Philadelphia, PA: Lippincott Williams & Wilkins, 2005.
- [19] B. J. Roach and D. H. Mathalon, "Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia," *Schizophr. Bull.*, vol. 34, no. 5, pp. 907–926, Sep. 2008.
- [20] A. Mouraux and G. D. Iannetti, "Across-trial averaging of event-related EEG responses and beyond," *Magn. Reson. Imag.*, vol. 26, pp. 1041–1054, 2008.
- [21] G. Pfurtscheller and F. H. Lopez da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, 1999.
- [22] S. Makeig, A. Delorme, M. Westerfield, T. P. Jung, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Electroencephalographic brain dynamics following manually responded visual targets," *PLoS Biol.*, vol. 2, no. 6, pp. 747–762, Jun. 2004.
- [23] S. Makeig, "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones," *Electromyogr. Clin. Neurophysiol.*, vol. 86, pp. 283–293, 1993.
- [24] P. J. Durka, *Matching Pursuit and Unification in EEG Analysis*. Boston, MA: Artech House, 2007.
- [25] Z. G. Zhang, X. L. Cai, S. C. Chan, Y. Hu, L. Hu, and C. Q. Chang, "Time-frequency coherence analysis of multi-channel event-related potential using adaptive windowed Lomb periodogram," in *Proc. IEEE/EMBS Conf. Neural Eng. (NER 2009)*, Antalya, Turkey, 29 Apr. 29–May 2, 2009, pp. 657–660.
- [26] C. M. Sweeney-Reed and S. J. Nasuto, "A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition," *J. Comput. Neurosci.*, vol. 23, pp. 79–111, 2007.
- [27] L. Causa, C. M. Held, J. Causa, P. A. Estévez, C. A. Perez, R. Chamorro, M. Garrido, C. Algarín, and P. Peirano, "Automated sleep-spindle detection in healthy children polysomnograms," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2135–2146, Sep. 2010.
- [28] T. Koenig, L. S. Prichep, D. Lehmann, P. Valdes-Sosa, E. Braeker, H. Kleinlogel, R. Isenhardt, and E. R. John, "Millisecond by millisecond, year by year: Normative EEG microstates and developmental stages," *NeuroImage*, vol. 16, pp. 41–48, 2002.
- [29] Z. G. Zhang, S. C. Chan, and Y. S. Hung, "Local polynomial modelling of time-varying autoregressive processes and its application to the analysis of event-related electroencephalogram," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS 2010)*, Paris, France, May 30–Jun. 2, pp. 3124–3127.
- [30] K. H. Ting, P. C. W. Fung, C. Q. Chang, and F. H. Y. Chan, "Automatic correction of artifact from single-trial event-related potentials by blind source separation using second order statistics only," *Med. Eng. Phys.*, vol. 28, pp. 780–794, 2006.
- [31] S. J. Kiebel, C. Tallon-Baudry, and K. J. Friston, "Parametric analysis of oscillatory activity as measured with EEG/MEG," *Hum. Brain Mapp.*, vol. 6, no. 3, pp. 170–177, 2005.
- [32] P. J. Durka, J. Zygierevicz, H. Klekowicz, J. Ginter, and K. J. Blinowska, "On the statistical significance of event-related EEG desynchronization and synchronization in the time-frequency plane," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1167–1175, Jul. 2004.
- [33] E. Lyskov, V. Ponomarev, M. Sandström, K. H. Mild, and S. Medvedev, "Steady-state visual evoked potentials to computer monitor flicker," *Int. J. Psychophysiol.*, vol. 28, no. 3, pp. 285–290, May 1998.
- [34] V. Wyart and C. Tallon-Baudry, "Neural dissociation between visual awareness and spatial attention," *J. Neurosci.*, vol. 28, no. 10, pp. 2667–2679, Mar. 2008.
- [35] C. S. Herrmann, I. Fründ, and D. Lenz, "Human gamma-band activity: A review on cognitive and behavioral correlates and network models," *Neurosci. Biobehav. Rev.*, vol. 34, no. 7, pp. 981–992, Jun. 2010.
- [36] L. Faes, K. H. Chon, and G. Nollo, "A method for the time-varying nonlinear prediction of complex nonstationary biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, pp. 205–209, Feb. 2009.
- [37] B. Yang and K. H. Chon, "Estimating time-varying nonlinear autoregressive model parameters by minimizing hypersurface distance," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 1937–1944, Aug. 2010.

Authors' photographs and biographies not available at the time of publication.