Modeling and estimation of single-trial event-related potentials using partially observed diffusion processes

Chee-Ming Ting, PhD
Sh-Hussain Salleh
Z. M. Zainuddin
Arifah Bahar
Modeling and Estimation of Single-Trial Event-Related Potentials using Partially Observed Diffusion Processes

Chee-Ming Ting¹*, Sh-Hussain Salleh¹, Z. M. Zainuddin² and Arifah Bahar²

¹Center for Biomedical Engineering, Universiti Teknologi Malaysia, Malaysia
²Centre for Industrial and Applied Mathematics, Universiti Teknologi Malaysia, Malaysia
cmting1818@yahoo.com, hussain@fke.utm.my, {zmarlizawati, arifah}@utm.my

Abstract

This paper proposes a new modeling framework for estimating single-trial event-related potentials (ERPs). Existing studies based on state-space approach use discrete-time random-walk models. We propose to use continuous-time partially observed diffusion process which is more natural and appropriate to describe the continuous dynamics underlying ERPs, discretely observed in noise as single-trials. Moreover, the flexibility of the continuous-time model being specified and analyzed independently of observation intervals, enable a more efficient handling of irregularly or variably sampled ERPs than its discrete-time counterpart which is fixed to a particular interval. We consider the Ornstein-Uhlenbeck (OU) process for the inter-trial parameter dynamics and further propose a nonlinear process of Cox, Ingersoll & Ross (CIR) with a heavy-tailed density to better capture the abrupt changes. We also incorporate a single-trial trend component using the mean-reversion variant, and a stochastic volatility noise process. The proposed method is applied to analysis of auditory brainstem responses (ABRs). Simulation shows that both diffusions give satisfactory tracking performance, particularly of the abrupt ERP parameter variations by the CIR process. Evaluation on real ABR data across different subjects, stimulus intensities and hearing conditions demonstrates the superiority of our method in extracting the latent single-trial dynamics with significantly improved SNR, and in detecting the wave V which is critical for diagnosis of hearing loss. Estimation results on data with variable sampling frequencies and missing single-trials show that the continuous-time diffusion model can capture more accurately the inter-trial dynamics between varying observation intervals, compared to the discrete-time model.

Keywords—Diffusion models, non-linear state-space models, particle filters, event-related potentials
1. Introduction

Event-related potentials (ERPs) are bioelectrical responses of the brain to specific stimulus, which provide useful information about neurological processing. The problem is to extract the ERPs hidden behind various noise sources e.g. background electroencephalogram (EEG) and non-neural artifacts, typically in poor signal-to-noise ratio (SNR) condition. Conventional ensemble averaging of time-locked single-trials obtained by repeated stimulations can cancel out random background noise, but implies loss of information related to trial-to-trial variability due to different degree of fatigue, habituation, or attention levels of subjects [1]. Auditory brainstem response (or evoked potentials) (ABRs or ABEPs), a particular type of ERPs is the early portion of the auditory evoked potentials. The major waves of ABR are important for hearing and neurological assessment and exhibit variability across trials. Estimation of inter-trial variability and common trend of major waves for these small-amplitude potentials obscured by various noise and artifacts, pose a challenge to this study.

Various approaches were proposed to solve single-trial based estimation of ERPs which is also the focus of this study. State-space approach was recently proposed for single-trial dynamical ERP estimation in [1]-[4] and our work [5]. Within the state-space framework, the underlying dynamics of clean ERPs are modeled as a discrete-time first-order Markov process i.e. a simple Gaussian random-walk process in the state equation, which is observed in additive Gaussian noise as single-trial measurements in the observation equation. The latent ERP process is then estimated recursively using Kalman filter (KF) [1], which is optimal in the mean-square sense. Extension using Kalman smoother [2] further improves the performance in tracking of ERP changes. Similar study [3] proposed a particle filter (PF) for the estimation and use wavelet coefficients, instead of measurement samples, as ERP parameters [1] to reduce the state dimension and hence the computational effort. This study was extended in [4], who models the ERP subcomponent waveforms using shifted Gaussian functions, which can directly characterize important ERP morphological features such as the peak amplitudes and latencies. The Gaussians are also less inclined to capture the abruptly changing random noises, except the usually smooth ERP waves. The parameter estimation is done by Rao-Blackwellised PF (RBPF). However, the
above-mentioned studies [1]-[5] assume discrete-time modeling which is inappropriate to describe the single-trial ERPs, the underlying physiological processes of which are typically of continuous nature. Besides, the discrete-time models are unable to handle the irregularly sampled data problem which is common in practice, where the ERP measurements are unequally spaced with some single-trials occasionally missing, probably due to sensor failures. Moreover, these studies use a Gaussian noise model with constant variance, which fails to describe adequately the actual observational noise process whose volatility typically changes over time.

We develop a new framework for modeling the inter-trial dynamics of ERPs based on continuous-time partially-observed (PO) diffusion process, with applications to estimation of single-trials corrupted by noise. Our approach is rooted in the state-space framework of the aforementioned studies, but offer several enhancements over the existing models, by relaxing the inadequate discrete-time and constant noise-volatility modeling assumptions and incorporating additional, important features of ERPs which have not so far been accounted for, such as the trend of the single-trials. In particular, we decompose the single-trial observations into three components: (1) latent clean ERP component, (2) observational noise component and (3) trend component, each described by a separate stochastic process. For the first component, we propose to use continuous-time processes to model the underlying dynamics of the latent clean ERPs. The continuous-time models have been applied extensively in physics, biology, econometrics [6]-[7] and engineering [8], due to their advantages over their discrete-time counterparts. This also forms our motivation of applying it for modeling single-trial ERPs, with the following reasons. First, it seems more natural and appropriate to use continuous-time modeling to describe real-world dynamical phenomena [7], such as the continuous physiological process underlying the single-trial EPRs considered here. Despite that the observations are always available only at discrete-time points; it is more adequate to assume that they are generated from continuous processes. Secondly, the continuous-time models can deal with variably or irregularly sampled data problem efficiently [8]-[9], due to their flexibility in defining implicitly and consistently over observation intervals of any length, with their adjustable parameter space estimated independently of the data sampling interval. However, the discrete-time
models are specified in relation to a particular sampling interval, and therefore are fixed once estimated
and only valid for this chosen sampling rate. Thus, the variable sampled data needs to be incorporated
explicitly through model re-estimation. We illustrate these advantages of the continuous-time approach
over its discrete-time counterpart, on two real situations in ERP estimation: (1) Estimation with variably
sampled (down-sampled) single-trials, which can save computational effort without affecting the overall
estimated dynamic pattern of ERPs; and (2) Estimation with irregularly spaced single-trials with missing
observations where some trials contains very low (sometimes zeros) signals, which are currently solved
only by direct elimination [10].

We consider the commonly used continuous-time model i.e. the diffusion process driven by Brownian
noise. Diffusion processes governed by the stochastic differential equations (SDEs) can accommodate
random disturbances in the deterministic behavior, and are widely used for modeling physical and
biological dynamical systems disturbed by noise [11]. Besides, the diffusion process is a continuous-time
Markovian process to be contrasted with the discrete-time random-walk process used previously in [1]-
[5]. Various variants of diffusion models with rich modeling properties have been proposed. To
accommodate the discrete-time observation scheme of single-trial ERPs, we use a particular family of PO
diffusion processes where the continuous-time diffusions are discretely observed [11], possibly further
with noise [12]. Such formulation is suitable to describe the continuous dynamics underlying ERPs which
are observed only as discrete-time noisy single-trials. Specifically, the hidden continuous dynamic
changes in the ERP Gaussian mixture parameters are assumed to follow a multivariate diffusion process,
which is only partially observed at discrete times with additive background noise as single-trial data in the
observation model. To the authors’ knowledge, there are no studies applying this kind of models for
analyzing ERPs and its use for bio-signals in general is still very limited.

A simple special case of diffusion process is first considered, i.e. the Ornstein-Uhlenbeck (OU)
process. The use of OU model in its linear Gaussian form, assumes marginal inter-trial ERP changes that
can be captured by its normal transition density. However, this assumption is inconsistent with the
presence of both smooth and abrupt changes in the actual ERP dynamics, which implies a non-Gaussian
behavior. The OU model is unable to capture both of these changes simultaneously, where a small variance of the normal transition density fails to detect rapidly the abrupt changes, while large variance tends to produce noisy estimates. This Gaussian modeling problem has been addressed in [13] and our earlier work [14] for other types of models in discrete-time. Another limitation of the OU model is that it allows for a negative process which is unsuitable for the non-negative latency and width parameters of the Gaussian ERP components. To accommodate these non-Gaussian dynamics of ERPs, we further propose an enhancement over the linear Gaussian OU process by considering a non-linear diffusion process of Cox, Ingersoll & Ross (CIR) [15]. The non-Gaussian heavy-tailed transition density of CIR process, a noncentral chi-squared density can capture both abrupt and smooth ERP parameter changes, hence providing more accurate single-trial estimates than the OU model. We use the Gaussian OU process which can take negative values to model the amplitude parameters, and the non-negative CIR process for the latency and width parameters. Moreover, both diffusions are stationary and thus produce more stable ERP estimates than the non-stationary random-walk models of [1]-[4].

To incorporate the single-trial trend component, we proceed to use the mean-reverting variant for both diffusion processes, which approaches a stationary asymptotic distribution in long run. The advantage is that the asymptotic mean of the stationary distribution can represent the underlying trend of the trial-varying ERPs, revealing more clearly the overall morphology of the ERP waveform. To capture more accurately the changing-volatility in the noise component in order to achieve better noise-reduction, we allow the observational noise variance to be varying over trials instead of a constant one. We consider the stochastic volatility (SV) modeling [16] where the trial-varying variance is modeled directly as a latent stochastic process, which is flexible and easy to generalize to multivariate case. The idea of incorporating SV noise models for denoising of single-trial ERPs has been explored in our recent work [17]. Preliminary evaluation on normal ERPs of a single-subject showed more accurate estimates of inter-trial dynamics compared to the constant noise variance. The evolution of both the trend and noise components is assumed to follow a simple random-walk process.

The proposed PO diffusion model of ERPs is then reformulated and estimated under the state-space
framework. Estimation of the non-linear processes of latency and width parameters admits no closed-form solution and has to resort to simulation-based PF methods. Considering that the linear Gaussian process of amplitudes can be estimated analytically by KF, the partially-nonlinear ERP model is re-formulated in conditionally-linear Gaussian state-space form to enable estimation using the more efficient RBPF. Although our RBPF framework is based on [4], its implementation to the diffusion process is not as straightforward as to the discrete-time models, which requires explicit solution to the continuous-time SDEs and closed-form transition densities for the processes. Both OU and CIR diffusions considered here have transition densities in closed-form obtained via Ito integration, for which RBPF can be applied directly. We consider time-varying parameter estimation by augmenting the model parameters to state vector followed by simultaneous estimation by the generic PF [17].

We apply the proposed methods to single-trial ABR estimation, for extracting inter-trial dynamics from noise and detecting the wave V component of ABR, critical for hearing problem assessment. We demonstrate the advantages of our model over the conventional discrete-time random-walk model based on evaluation using simulated and real data, particularly in handling irregularly spaced data in ABR estimation. We also investigate the effect of the model parameter choices on the estimation performance which was not addressed much in the previous works [1]-[4].

Thus, in common with the previous studies [1]-[4], we attempt the state-space method for dynamical estimation of single-trial ERPs. In contrast to the discrete-time Gaussian random-walk modeling, we develop an enhanced framework for modeling and estimating single-trial ERPs based on continuous-time PO diffusion process. The main contributions of this paper is summarized as follows

- We introduce continuous-time diffusion processes for improved modeling of the underlying continuous dynamics of single-trial ERPs and handling of irregularly spaced data, compared to the existing discrete-time approach.
- We consider OU and CIR process which are stationary and produce more stable ERP estimates than the random-walks. The heavy-tailed density of CIR can also capture the abrupt changes better than the Gaussian models.
• We incorporate a single-trial trend model for better representation of major ERP morphology and a stochastic volatility noise model to enhance noise reduction.

• We apply the methods to estimation of single-trial ABRs and detection of wave V for hearing-loss diagnosis, with evaluation across different subjects, stimulus intensities and pathological case, instead of normal cases only as in the previous studies [1]-[5].

2. Partially Observed Diffusion Modeling of ERPs

The \(d\)-dimensional vector of single-trial ERP parameters \(o_t\) is assumed to follow a multivariate continuous-time diffusion process solution to a SDE of the form

\[
d o_t = b(o_t, \theta)dt + \sigma(o_t, \theta)dW_t
\]

where functions \(b(o_t, \theta)\) and \(\sigma(o_t, \theta)\) are drift and diffusion coefficient respectively, which depend on unknown parameter vector \(\theta\). \(W_t\) is a \(d\)-dimensional standard Brownian motion. Let \(y_n = [y_n(1), y_n(2), \ldots, y_n(k)]^T\) be measurement samples of length \(k\) of a single-trial ERP at trial \(n\), and observations \(y_1, y_2, \ldots, y_n\) occur at discrete time points \(0 \leq t_1 < t_2 < \ldots < t_n\). The continuous process (1) is only discretely observed at time \(t_n\), denoted as \(o_{t_n} \triangleq o_n\), with further noise source \(v_n\) and through mapping \(h(.)\) as noisy single-trials \(y_n\)

\[
y_n = h(o_n, v_n, \theta).
\]

Refer [12] for various observation schemes for PO diffusions.

2.1. The Observation Model

We use a special case of observation model (2), an additive noise model as in [1]-[4] with the following decomposition

\[
y_n = h(o_n) + v_n, \quad v_n \sim N(0, R)
\]

where \(v_n\) is \(k \times 1\) Gaussian white noise with mean zero and diagonal covariance matrix \(R = \sigma^2_v I_k\) where
$I_k$ denotes a $k \times k$ identity matrix. The noise component $v_n$ can represent the background EEG and artifacts.

The vector function $h: \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}^k$ maps the ERP parameters $o_n$ to the observations $y_n$. The term $\bar{y}_n \triangleq h(o_n)$ is the clean or denoised single-trial of ERP. Since ERP waveform consists of subcomponents with approximately Gaussian shape, $\bar{y}_n$ can be modeled as a mixture of $M$ Gaussians using a non-linear mapping function for $h$ with each element given as [4]

$$h_k(o_n) = h_k(a_n, b_n, c_n) = \sum_{m=1}^{M} a_n(m) e^{-\frac{(k-b_n(m))^2}{2c_n^2(m)}}$$

(4)

where $a_n(m), b_n(m), c_n(m)$ are respectively the amplitude, latency and width parameters of the $m$-th Gaussian component for trial $n$. Thus, a clean single-trial ERP $\bar{y}_n$ can be represented by the set of Gaussian mixture model parameters, in a compact form $o_n \triangleq (a_n, b_n, c_n)$ where $a_n = [a_n(1), \ldots, a_n(M)]^T$, $b_n = [b_n(1), \ldots, b_n(M)]^T$, and $c_n = [c_n(1), \ldots, c_n(M)]^T$. Note that these subcomponent parameters are allowed to be trial-varying to model the inter-trial variations of ERPs.

### 2.2. The OU and CIR ERP Processes

A simple model for the trial-varying ERP parameters is the discrete-time random walk [4]

$$
\begin{bmatrix}
a_n \\
b_n \\
c_n
\end{bmatrix}
= 
\begin{bmatrix}
a_{n-1} \\
b_{n-1} \\
c_{n-1}
\end{bmatrix}
+ 
\begin{bmatrix}
\eta_{a_n} \\
\eta_{b_n} \\
\eta_{c_n}
\end{bmatrix}
\sim N
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
\sigma_a^2 I_M & 0 & 0 \\
0 & \sigma_b^2 I_M & 0 \\
0 & 0 & \sigma_c^2 I_M
\end{bmatrix}
$$

(5)

Instead of discrete-time models, it is more natural and advantageous to assume that they are generated by underlying continuous transients which can be described by continuous-time models such as (1). We consider two diffusions: a Gaussian OU process for amplitude parameters and a non-linear CIR process for the latencies and widths.

The amplitude parameters $a_n$ are assumed as discretely observed samples of an $M$-dimensional mean-
reverting OU process $a_t$ solution to the SDE

$$da_t = \beta (\mu_a - a_t) dt + \sigma_a dW_t$$

(6)

where $\beta > 0$ is speed of reversion, $\mu_a$ is long run equilibrium level of process $a_t$, $\sigma_a^2$ is the process variance. The process tends to drift in direction towards equilibrium level $\mu_a$ with speed determined by $\beta$. Refer [19]-[20] for details of OU process. By Ito formula, the explicit solution of $a_t$ can be obtained as

$$a_t = a_0 e^{-\beta t} + \mu_a (1 - e^{-\beta t}) + \alpha \epsilon_t$$

(7)

where the initial $a_0$ is normally distributed and $\epsilon_t$ is assumed as a Gaussian white noise with mean zero and covariance matrix $P = \sigma_a^2 I_M$ with $\alpha = \sqrt{(1 - e^{-2\beta t})/2\beta}$. The process has a Gaussian marginal distribution with mean and variance [19]

$$E[a_t] = e^{-\beta t} E[a_0] + \mu_a (1 - e^{-\beta t}) \rightarrow \mu_a \quad as \ t \rightarrow \infty$$

$$\text{Var}[a_t] = e^{-2\beta t} \text{Var}[a_0] + \frac{\sigma_a^2}{2\beta} (1 - e^{-2\beta t}) I \rightarrow \frac{\sigma_a^2}{2\beta} I \quad as \ t \rightarrow \infty$$

(8)

Note that $a_t$ approaches a normal stationary distribution $a_t \sim N(\mu_a, \frac{\sigma_a^2}{2\beta} I)$ as $t \rightarrow \infty$. For the dynamical estimation of ERPs, the component $\mu_a$ can be used to capture the underlying static trend of the dynamics of $a_t$, and thus provides an alignment for the ERPs which vary across trials.

We denote by $t_n$ the time point at which the $n$-th observation $y_n$ is sampled. Let $\Delta t = t_{n+1} - t_n$ be the time interval between two consecutive observations. Then, discretization of (7) gives [21]

$$a_n = a_{n-1} e^{-\beta \Delta t} + \mu_a (1 - e^{-\beta \Delta t}) + \alpha \epsilon_{a_n}, \quad \epsilon_{a_n} \sim N(0, \sigma_a^2 I_M)$$

(9)

with $\alpha = \sqrt{(1 - e^{-2\beta \Delta t})/2\beta}$. The transition density which follows directly from (9) is normal

$$a_n | a_{n-1} \sim N(a_{n-1} e^{-\beta \Delta t} + \mu_a (1 - e^{-\beta \Delta t}), \sigma_a^2 \alpha^2 I_M)$$

(10)

The dynamics of latency and width parameters can assume OU process as well with discretization.
\[
\begin{bmatrix}
 b_n \\
 c_n
\end{bmatrix} = \begin{bmatrix}
 b_{n-1} \\
 c_{n-1}
\end{bmatrix} e^{-\beta M} + \begin{bmatrix}
 \mu_{b_n} \\
 \mu_{c_n}
\end{bmatrix} (1 - e^{-\beta M}) + \alpha \begin{bmatrix}
 \varepsilon_{b_n} \\
 \varepsilon_{c_n}
\end{bmatrix}, \quad \begin{bmatrix}
 \varepsilon_{b_n} \\
 \varepsilon_{c_n}
\end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b_M}^2 I_M & 0 \\ 0 & \sigma_{c_M}^2 I_M \end{bmatrix}\right)
\]

(11)

which however is likely to generate negative values. To ensure their positivity, we use the CIR process which was originally proposed as a model for financial interest rate by [15]. Besides, the process has a non-Gaussian transition density ideally suited for both abrupt and gradual changes, which are unable to be captured by the Gaussian one in (10) whose variance can only be set either too small or too large. The latency \( b_n \), similarly for \( c_n \) is assumed to be generated from a \( M \)-dimensional mean-reverting CIR process, which has a linear drift term as the OU but with a square-root diffusion, specified by a non-linear SDE [21]

\[
db_t = \beta (\mu_b - b_t)dt + \sigma_b \sqrt{b_t} dW_t
\]

(12)

where \( \beta \) is the mean-reversion speed, \( \mu_b \) is the long-term mean and \( \sigma_b^2 \) is the volatility, similar to that of the OU process. However, the non-linear term \( \sqrt{b_t} \) suppresses the variance term to zero as \( b_t \) approaches zero, preventing the process to become negative. The process also has a Gamma marginal density. Refer [21] for the discretized model of the process.

Let assume that elements of \( b_n \) are mutually independent. The transition density for univariate case \( p(b_n \mid b_{n-1}) \) can be easily obtained from its transform \( p(s_{b_n} \mid s_{b_{n-1}}) \) where \( s_{b_n} = 2cb_n \) with

\[
c = 2\beta / (\sigma_b^2 (1 - e^{-\beta M}))
\]

The transformed transition density \( p(s_{b_n} \mid s_{b_{n-1}}) \) follows a noncentral chi-squared distribution with degree of freedom \( \nu = 4\mu_b \beta / \sigma_b^2 \) and non-centrality parameter \( \lambda = s_{b_{n-1}} e^{-\beta M} \), \( s_{b_n} \mid s_{b_{n-1}} \sim \chi^2(\nu) \) [20]. The noncentral \( \chi^2 \) conditional density can predict the rare abrupt parameter changes to high values by its heavier right-tails as well as the smooth changes by its relatively narrow central region.
2.2.1. The Motivation of Continuous-time Modeling

Use of continuous-time modeling offers several benefits over the discrete-time models both in theoretical and practical perspectives [6]-[9]. The rationale for using continuous-time models for natural data, as for the ERPs as an example, is founded on the assumption that the underlying dynamic system is continuous e.g. the physiological dynamics behind the single-trials. Analytically, differential-equation models are more easily handled than the difference models [6], owing to the stochastic calculus.

The practical advantage of continuous-time method is that it is well-suited for modeling variably or irregularly spaced data [8]-[9], due to its flexibility in defining inter-observation time $\Delta t$ of any length consistently. The solution of a differential equation such as (7) which is based on the integral of the continuous variables yields a continuous sample path that can explain the dynamics for any time-intervals. The discrete-time observations can be seen as generated by the underlying continuous process.

The parameter space of the continuous-time models is more extensive and adjustable to capture the dynamics between varying observation intervals. However, the discrete-time model parameters are fixed and inflexible to account for possible changes in $\Delta t$. In fact, the discrete-time models are approximations to their continuous counterparts, where the parameters of a discrete model are discrete samples of those of the underlying continuous model [7]. For example, the OU process of (7) is a continuous-time analogue of the discrete-time AR(1) process in a recursive form. The discretized version of OU process in (9) taken at a particular regular interval of $\Delta t$ form an AR(1) process [22]. The OU process parameters, the mean-reversion coefficient $e^{-\beta \Delta t}$ and noise variance scaling parameter $\alpha$ in (9) are time-varying according to $\Delta t$, instead of being fixed as in the ordinary discrete-time AR model [23]. As $\Delta t$ increases, $\alpha$ becomes larger to predict parameter changes for larger region of uncertainty due to a longer period of dynamics. Thus, the parameters of the continuous models can be specified and analyzed independently of the observation intervals of the samples used for estimation [6]-[8]. A continuous model estimated or pre-defined for data with a particular sampling interval can be used to analyze data of any intervals consistently. This is in contrast with the discrete models once fitted to a specific sampling rate, are rigidly fixed and invalid for
other sampling intervals. It requires re-estimation of the models for every different sampling interval. Therefore, the continuous-time models have this built-in capability to handle variable sampled data which is also present in the ERP estimation, where measurements are spaced variably at different intervals, which the discrete-time model fails to deal with.

2.3. The Trend and Noise Volatility Process

The relatively static $3M \times 1$ trend components $(\mu_a, \mu_b, \mu_c)$ are assumed to follow a stable dynamics by adding a small noise [24], here we consider random-walk models with i.i.d. Gaussian noise

$$\mu_n = \mu_{n-1} + u_n, \quad u_n \sim N(0, \sigma^2_{u_n} I_{3M}), \quad (13a)$$

$$[\mu_n] = [\mu_{n-1}] + [u_n], \quad [u_n] \sim N([0], [\sigma^2_{u_n} I_{3M} 0 \phantom{\sigma^2_{v_n} I_{3M}}]). \quad (13b)$$

For the noise component in (3), we allow the variance (squared volatility) to vary over trials i.e. $v_n \sim N(0, R_n)$ with $R_n = \sigma^2_{v_n} I_{3M}$ to better model the changing-volatility characteristics of real noise process in ERPs. However, we still assume the variance as locally stationary within a short-segment of single-trial (i.e. identical entries in the diagonal of $R_n$). We propose to use SV modeling for $v_n$ where the trial-varying $\sigma^2_{v_n}$ is modeled directly by a separate latent stochastic process, which we assume here as a Gaussian random-walk model

$$\phi_{v_n} = \phi_{v_{n-1}} + \xi_n, \quad \xi_n \sim N(0, \sigma^2_{\phi_v}) \quad (14)$$

where $\phi_{v_n} \triangleq \log \sigma^2_{v_n}$ to ensure positivity of $\sigma^2_{v_n}$.

2.4. State-Space Formulation

The models are formulated in a discrete-time conditionally-linear Gaussian state-space form with observation and state equation respectively as

$$y_n = A_n(\theta_n) x_n + v_n, \quad v_n \sim N(0, R_n(\theta_n)) \quad (15)$$
\[
x_n = B(\theta_n)x_{n-1} + C(\theta_n)w_n, \quad w_n \sim N(0, Q(\theta_n)) \tag{16}
\]
\[
\theta_n \sim p(\theta_n | \theta_{n-1}) \tag{17}
\]

where \( \theta_n \) is a vector of unknown model parameters. It is assumed for the initial state
\[
x_0 \sim N(\hat{x}_{(0)}(\theta_0), P_{(0)}(\theta_0)) \text{.}
\]
Similar to [4], by defining an concatenated state vector \( x_n = [a_n^T, \mu_{a_n}^T]^T \) with
\[
\mu_{a_n} = [\mu_{a_n}(1), \ldots, \mu_{a_n}(M)]^T,
\]
the observation model defined by (3) can be written in a compact linear Gaussian form of (15) where \( A_n \) is a \( k \times 2M \) system matrix denoted in terms of latency and width parameters
\[
A_n = \begin{bmatrix}
g_1(b_n(1), c_n(1)) & \ldots & g_1(b_n(M), c_n(M)) & 0_{1 \times M} \\
g_2(b_n(1), c_n(1)) & \ldots & g_2(b_n(M), c_n(M)) & 0_{1 \times M} \\
\vdots & & \vdots & \vdots \\
g_k(b_n(1), c_n(1)) & \ldots & g_k(b_n(M), c_n(M)) & 0_{1 \times M}
\end{bmatrix}
\]

with nonlinear terms \( g_k(b_n(m), c_n(m)) = e^{-(k-b_n(m))^2/(2c_n^2(m))} \). The discretized linear Gaussian OU model of \( a_n \) (9) and the model for its trend (13a) can be formulated in state equation (16) with \( 2M \times 2M \) system matrices
\[
B = \begin{bmatrix}
e^{-\beta M I_M} & (1-e^{-\beta M})I_M \\
0 & I_M
\end{bmatrix}, \quad C = \begin{bmatrix}
aI_M & 0 \\
0 & I_M
\end{bmatrix}
\]

and \( 2M \times 1 \) state noise
\[
w_n = \begin{bmatrix}
e_n \\
u_{a_n}
\end{bmatrix} \sim N\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_{a_n}^2 I_M & 0 \\
0 & \sigma_{u_{a_n}}^2 I_M\end{bmatrix}\right).
\]

Note that to formulate the linear Gaussian observation equation (15) from the non-linear form of (4), \( a_n \) and \( (b_n, c_n) \) are treated separately as state and model parameters. The trial-varying parameters of latency \( b_n \) and width \( c_n \) in \( A_n \), their trends, and observation noise variance \( \sigma_{e_n}^2 \), denoted here by
\[
\theta_n = [b_n^T, c_n^T, \mu_{b_n}, \mu_{c_n}, \sigma_{e_n}^2]^T,
\]
are assumed unknown and need to be estimated. The model parameters \( (\beta, \sigma_{a_n}^2, \sigma_{u_{a_n}}^2) \) are assumed fixed and pre-specified.
Table 1 Summary of known and unknown parameters of the proposed state-space model of ERPs.

<table>
<thead>
<tr>
<th>Component Model</th>
<th>State Parameters, $z_n$ (trial-varying, unknown, estimated)</th>
<th>Model Parameters, $\Theta$ (fixed, known, pre-specified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusions for ERP parameters</td>
<td>$(a_n, b_n, c_n)$</td>
<td>$\Delta \alpha, \beta, (\sigma_a^2, \sigma_b^2, \sigma_c^2), (a_0, b_0, c_0)$</td>
</tr>
<tr>
<td>Random-walks for trends</td>
<td>$(\mu_{a_n}, \mu_{b_n}, \mu_{c_n})$</td>
<td>$(\sigma_{a_n}^2, \sigma_{b_n}^2, \sigma_{c_n}^2), (\mu_{a_0}, \mu_{b_0}, \mu_{c_0})$</td>
</tr>
<tr>
<td>Random-walk for noise log-variance</td>
<td>$\phi_{\sigma_n}$</td>
<td>$\sigma_n^2$</td>
</tr>
</tbody>
</table>

For the estimation of the time-varying model parameters, we consider a Bayesian approach where $\theta_n$ are augmented to the state vector $x_n$ as $z_n = [\theta_n^T, x_n^T]^T$ with the joint posterior density $p(z_{1:n} | y_{1:n})$ estimated using standard PF methods, as in [18], [4]. To formulate a state-space model for the augmented state $z_n$, the models for the time-evolution of $\theta_n$ (17) are pre-specified, as already in (11)-(12) and (13b)-(14). The hyper-parameters of these processes $(\sigma_b^2, \sigma_c^2, \sigma_{a_n}^2, \sigma_{b_n}^2, \sigma_{c_n}^2)$ are assumed fixed and known. The initial state $\theta_0$ is assumed to be distributed according to a density $p_0(\theta)$ which can be tailored by available prior information. Conditional on the non-linear CIR processes of $\theta_n$, (15)-(16) is a linear Gaussian state-space model.

The process of $\theta_n$ (17) together with the original state-space model (15)-(16) constitute the complete augmented state-space model for $z_n$ with all parameters summarized in Table 1. Thus, the proposed model is now fully specified. The objective is to estimate sequentially the parameters of the Gaussian ERP components for each trial $n$, and hence to uncover the de-noised ERP $\overline{y}_n$. 
3. Rao-Blackwellized PF

Let denote \( y_{1:n} = \{y_1, \ldots, y_n\} \). The inference problem here involves estimating sequentially in time the posterior density \( p(z_{1:n} | y_{1:n}) \triangleq p(x_{1:n}, \theta_{1:n} | y_{1:n}) \) where \( \{x_{1:n}, \theta_{1:n}\} \) with \( x_{1:n} = \{x_1, \ldots, x_n\} \) and \( \theta_{1:n} = \{\theta_1, \ldots, \theta_n\} \) contains the underlying clean ERPs and model parameters. We focus on estimation of its marginal, filtered density \( p(z_n | y_{1:n}) \triangleq p(x_n, \theta_n | y_{1:n}) \). Generic PF methods can be applied directly to estimate the joint density \( p(x_n, \theta_n | y_{1:n}) \) [18], [24]. However, sampling from a high-dimensional density is inefficient due to large number of samples is required for accurate approximation, which increases computational complexity. The aforementioned conditionally-linear Gaussian model structure can be utilized to facilitate the implementation of RBPF [25], [26], a more efficient variant of PF. The algorithm is based on Rao-Blackwellization method where some of the state-space variables in linear Gaussian form, which is the \( x_n \) here, can be marginalized out analytically to reduce the state-space dimension and hence the variance of the sampling estimates. This method has been used for online estimation of clean speech signals [27] and ERP subcomponents [4]. We follow the approach of these studies for the filtering of our model.

We describe here how to use the RBPF to estimate \( p(x_n, \theta_n | y_{1:n}) \) in our model; refer [25]-[27] and [4] for further details. We observe that

\[
p(x_n | y_{1:n}) = \int p(x_n | \theta_n, y_{1:n}) p(\theta_n | y_{1:n}) d\theta_n
\]  

(18)

where \( p(x_n | \theta_n, y_{1:n}) \sim N(\tilde{x}_{dn}(\theta_n), \tilde{P}_{dn}(\theta_n)) \) in which \( \{\tilde{x}_{dn}(\theta_n), \tilde{P}_{dn}(\theta_n)\} \) can be computed analytically by the KF. Then, estimating \( p(x_n, \theta_n | y_{1:n}) \) only needs to sample from a lower dimensional \( p(\theta_n | y_{1:n}) \). If we obtain by PF, an approximation of \( p(\theta_n | y_{1:n}) \)

\[
\hat{p}(\theta_n | y_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{\theta_n^{(i)}}(\theta_n), \quad W_n^{(i)} = \tilde{w}_n^{(i)} / \sum_{j=1}^{N} \tilde{w}_n^{(j)}
\]  

(19)

where \( W_n^{(i)} \) and \( \tilde{w}_n^{(i)} \) are the normalized and un-normalized weights associated with samples \( \theta_n^{(i)} \). Then,
$p(x_n | y_{1:n})$ can be approximated as a mixture of Gaussian distributions

$$
\hat{p}(x_n | y_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} N(x_n; \hat{\theta}_n^{(i)}, P_n^{(i)}) \tag{20}
$$

To obtain estimates (19) sequentially, we use sequential importance sampling (SIS) algorithm [25], [28], where $\theta_n^{(i)}$ are sampled recursively from some importance function $\pi(\theta_n | \theta_{n-1}, y_{1:n})$ given $\theta_{n-1}$, with associated weights updated recursively to [27]

$$
W_n^{(i)} = \frac{\pi \psi_n^{(i)}}{\pi \psi_{n-1}^{(i)} p(y_n | \theta_{n-1}^{(i)}, y_{1:n-1}) p(\theta_{n-1}^{(i)} | \theta_{n-1})}. \tag{21}
$$

Prior importance function $\pi(\theta_n | \theta_{n-1}, y_{1:n}) = p(\theta_n | \theta_{n-1})$ is used here with weights $\tilde{w}_n^{(i)} = \tilde{w}_{n-1}^{(i)} p(y_n | \theta_n^{(i)}, y_{1:n-1})$. To overcome weight degeneracy problem, re-sampling step [28] is applied when the effective sample size less than a specified threshold $ESS = \left( \sum_{i=1}^{N} (W_i^{(i)})^2 \right)^{-1} < N_T$.

Sampling from conditional density $p(\theta_n | \theta_{n-1})$ of our model are performed hierarchically according to

$$
\phi_{n-1} | \phi_{n-1} \sim N(\phi_{n-1}, \sigma_{\phi_n}^2),
$$

$$(\mu_{b_{n-1}}, \mu_{c_{n-1}}) | (\mu_{b_{n-1}}, \mu_{c_{n-1}}) \sim N\left( \begin{bmatrix} \mu_{b_{n-1}} \\ \mu_{c_{n-1}} \end{bmatrix}, \begin{bmatrix} \sigma_{\mu_{b_n}}^2 I_M & 0 \\ 0 & \sigma_{\mu_{c_n}}^2 I_M \end{bmatrix} \right),
$$

$$(b_n, c_n) | (b_n, c_n) \sim N\left( \begin{bmatrix} b_{n-1} \\ c_{n-1} \end{bmatrix}, \begin{bmatrix} 1 - e^{-\beta \Delta t} & \alpha^2 \sigma_{\mu_b}^2 I_M \\ \alpha^2 \sigma_{\mu_c}^2 I_M & 0 \end{bmatrix} \right),
$$

for the OU model or for the CIR model through its transform

$$
s_{b_n} | s_{b_{n-1}} \sim \chi^2_{4\mu_b \beta, \sigma_b^2},
$$

$$
s_{c_n} | s_{c_{n-1}} \sim \chi^2_{4\mu_c \beta, \sigma_c^2}, \tag{22}
$$

for each dimension based on independence assumption. Let

$$
p(x_n | \theta_n, y_{1:n}) = N(x_n; \hat{x}_{n|n-1}(\theta_n), P_{n|n-1}(\theta_n)),
$$

$$
p(x_n | \theta_n, y_{1:n-1}) = N(x_n; \hat{x}_{n|n-1}(\theta_n), P_{n|n-1}(\theta_n)) \quad \text{and}
$$

$$
p(y_n | \theta_n, y_{1:n-1}) = N(y_n; \hat{y}_{n|n-1}(\theta_n), S_{n|n-1}(\theta_n)) \tag{23}
$$
Algorithm 1: Rao-Blackwellized Particle Filter (RBPF).

1. Initialization at time $n = 0$,

   For $i = 1, \ldots, N$, sample $\theta_0^{(i)} \sim p_0(\theta_0)$ and set $\mathbf{x}_{00}(\theta_0^{(i)}) = \mathbf{0}$ and $\mathbf{P}_{00}(\theta_0^{(i)}) = \mathbf{P}_0$. The normalized initial weights are set $W_0^{(i)} = 1/N$.

2. Repeat the following steps for time $n \geq 1$:
   
   (a) For $i = 1, \ldots, N$, sample $\theta_n^{(i)} \sim p(\theta_n | \theta_{n-1}^{(i)})$.
   
   (b) For $i = 1, \ldots, N$, compute $\{\mathbf{x}_{n|n}(\theta_n^{(i)}), \mathbf{P}_{n|n}(\theta_n^{(i)})\}$ using one step of Kalman recursion, given $\{\mathbf{x}_{n-1|n-1}(\theta_{n-1}^{(i)}), \mathbf{P}_{n-1|n-1}(\theta_{n-1}^{(i)}), \mathbf{y}_n, \theta_n^{(i)}\}$.

   $\begin{align*}
   \hat{\mathbf{x}}_{n|n-1} &= \mathbf{B} \hat{\mathbf{x}}_{n-1|n-1} \\
   \mathbf{P}_{n|n-1} &= \mathbf{B} \mathbf{P}_{n-1|n-1} \mathbf{B}^T + \mathbf{C} \mathbf{Q} \mathbf{C}^T \\
   \mathbf{y}_{n|n-1} &= \mathbf{A}_n \hat{\mathbf{x}}_{n|n-1} \\
   \mathbf{S}_n &= \mathbf{A}_n \mathbf{P}_{n|n-1} \mathbf{A}_n^T + \mathbf{R}_n \\
   \hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{P}_{n|n-1} \mathbf{A}_n^T \mathbf{S}_n^{-1} (\mathbf{y}_n - \mathbf{y}_{n|n-1}) \\
   \mathbf{P}_{n|n} &= \mathbf{P}_{n|n-1} - \mathbf{P}_{n|n-1} \mathbf{A}_n^T \mathbf{S}_n^{-1} \mathbf{A}_n \mathbf{P}_{n|n-1}
   \end{align*}$

   (c) For $i = 1, \ldots, N$, evaluate the importance weights:

   $\begin{align*}
   \tilde{w}_n^{(i)} &= W_n^{-1} p(\mathbf{y}_n | \theta_n^{(i)}, \mathbf{y}_{1:n-1}) \\
   &= W_n^{-1} N(\mathbf{y}_n; \mathbf{y}_{n|n-1}(\theta_n^{(i)}), \mathbf{S}_n(\theta_n^{(i)}))
   \end{align*}$

   and normalize them:

   $W_n^{(i)} = \frac{\tilde{w}_n^{(i)}}{\sum_{i=1}^{N} \tilde{w}_n^{(i)}}$.

   (d) Resampling. If $\text{ESS} < N_T$, resample with replacement $N$ particles from the set $\{\theta_n^{(i)}, \mathbf{x}_{n|n}(\theta_n^{(i)}), \mathbf{P}_{n|n}(\theta_n^{(i)})\}_{i=1}^{N}$ according to probabilities proportional to weights $\{W_n^{(i)}\}_{i=1}^{N}$ and reset $W_n^{(i)} = 1/N$, for $i = 1, \ldots, N$.
denote respectively filtering, one-step ahead prediction, and predictive densities conditional on $\theta_n$, which are estimated by KF. For notational convenience, we write $\hat{x}_{n|n}(\theta_n) = \hat{x}_{n|n}$, $P_{n|n}(\theta_n) = P_{n|n}$, and similarly for other terms. The RBPF used in this paper to estimate $p(x_n, \theta_n | y_{1:n})$ is summarized in Algorithm 1. The algorithm requires $N$ recursions of KF at each iteration. Instead of computing $N$ ‘full’ KF recursions, we calculate only once $P_{n|n}(\theta_n) = P_{n|n}$ and $P_{n|n-1}(\theta_n) = P_{n|n-1}$ for $i = 1, \ldots, N$, as observed in [29].

The mean of the filtered joint density $E[x_n, \theta_n | y_{1:n}]$ is the optimal estimator of $(x_n, \theta_n)$. Given respective approximations in (19) and (20), the means of the marginal densities $p(\theta_n | y_{1:n})$ and $p(x_n | y_{1:n})$ can be approximated

$$E[\theta_n | y_{1:n}] = \sum_{i=1}^{N} W_n^{(i)} \theta_n^{(i)}$$

and

$$E[x_n | y_{1:n}] = \sum_{i=1}^{N} W_n^{(i)} E[x_n | \theta_n^{(i)}, y_{1:n}] = \sum_{i=1}^{N} W_n^{(i)} \hat{x}_{n|n}(\theta_n^{(i)}) .$$

The marginal likelihood of $y_{1:j}$ is given as [24]

$$\ell(y_{1:j}, \hat{\lambda}) = \prod_{n=1}^{j} p(y_n | y_{1:n-1})$$

where, given approximation of the one-step ahead prediction $\hat{p}(\theta_n | y_{1:n-1}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{\theta_n^{(i)}}(\theta_n)$ [25], the predictive likelihood can be estimated by

$$\hat{p}(y_n | y_{1:n-1}) = \int p(y_n | \theta_n, y_{1:n-1}) p(\theta_n | y_{1:n-1}) d\theta_n$$

$$= \sum_{i=1}^{N} W_n^{(i)} p(y_n | \theta_n^{(i)}, y_{1:n-1}) = \sum_{i=1}^{N} W_n^{(i)} .$$

4. Experimental Results

In this section, we evaluate the performance of the proposed method in tracking the dynamic changes of ERP parameters and in estimating the single-trials on a simulated and a real data set. We focus on a
particular kind of ERPs i.e. the ABRs which comprise the early portion of auditory evoked potentials elicited by acoustic stimulus to the ear. ABR is composed of several waves or peaks labeled with roman numerals I-VII. This paper focuses on the study of waves III and V which are major peaks generated from auditory regions in caudal and rostral brainstem respectively. The ABR waveform characteristics such as the peak latencies and amplitudes provide useful information for objective assessment of hearing loss as well as pathologies affecting auditory brainstem pathways. Refer to [30] for further details. The ABR vary across trials, different subjects and stimulus intensities [31]. We focus on estimating the single-trial variability of ABRs and identifying the absence or presence of the wave V for objective diagnosis of hearing loss. We study the chirp-evoked ABRs which can induce stronger wave V [32].

4.1. Simulation Results

We compare the performance of PO diffusion ERP models with OU and CIR latency process in estimating single-trial ABRs on a simulated data. The simulation analysis adopted here is based on [1] and [4]. Major waves III and V of ABR are simulated, sufficiently by using three Gaussian mixture components $M = 3$ in model (4). The amplitude, latency and width parameters of the Gaussian components are set to $a = [0.65, -0.4, 1]^T$, $b = [50, 120, 200]^T$ and $c = [30, 30, 50]^T$ samples respectively, which serve as constant baselines for the dynamic variations simulated as the following. The trial-to-trial variability of ABRs is simulated by artificially varying over trials their subcomponent parameters. To simulate abrupt parameter changes, the dynamic variations of amplitudes and latencies for the first two components are set to follow i.i.d. noises distributed respectively as $U[-0.2, 0.2]$ and $U[-5, 5]$. $U[a,b]$ denotes uniform distribution over interval $[a,b]$. To further illustrate the benefit of CIR models, we add a burst signal of magnitude 20 to the latency parameters, which exhibits more prominent abrupt transitions at both ends. To simulate the smooth changes, the amplitude and latency variations for the third component are represented as sinusoidal functions with respective magnitudes of 1 and 5 samples, plus uniformly distributed i.i.d. noise with smaller interval. In this simulation, the width parameters are
assumed fixed with time and not estimated. The simulated clean ERPs constructed from the trial-varying parameters are artificially added with Gaussian white noise \( v_n \sim N(0, \sigma_v^2 I) \) with the log-variance modeled as in (14) with parameters \( \phi_{0v} = -2.5 \) and \( \sigma_{\phi v}^2 = 5 \times 10^{-4} \), to generate noisy observations. Note that we use the nonlinear term \( \exp(-(t-h_{n}(m))^2/\sigma_{n}^2(m)) \) in (4) as in [4] for this simulation.

The estimators of the ‘true’ ERP parameters, trend components and observation noise variance are obtained simultaneously from the mean of the filtering distribution of \( z_n \) estimated by the RBPF. The estimated ERP parameters are then used to reconstruct the clean ERPs. The choice of model parameters \( \Theta \) determines the performance of the estimators, which can be measured in term of estimation error from the true values. The behaviors of the ERP parameter estimators are determined by the noise variances of the diffusion processes \( (\sigma_a^2, \sigma_b^2, \sigma_c^2) \). Generally, larger values allow larger prediction to capture abrupt changes while small values enable to follow the smooth one. In this simulation, these parameters are set empirically to give sufficiently good estimates that are able to detect satisfactorily abrupt changes while with less noisy estimates for the smooth changes, which is measured by the mean square error (MSE) of the parameter estimates and guided by visual inspection. The average MSE of the parameter estimates for each ERP subcomponent is defined by

\[
\text{MSE}(m) = \frac{1}{RJ} \sum_{r=1}^{R} \sum_{n=1}^{J} (\hat{a}_{n}^{r}(m) - a_{n}(m))^2
\]

(28)

where \( \hat{a}_{n}^{r}(m) \) is estimator for a parameter of subcomponent \( m \) at trial \( n \) \( E[a_{n}(m) \mid y_{1:n}] \) of replication \( r \) and \( a_{n}(m) \) is the true value.

Table 2 shows the effect of the choice of the process variance \( \sigma_b^2 \) on MSEs for the OU and CIR modeling of latency parameters. The results for the latency parameters of the second and third Gaussian components (abrupt and smooth changes) are presented. Generally, for the second component, there is a steady improvement in MSE as \( \sigma_b^2 \) increases to capture faster the abrupt changes, until a point where
Table 2  Selection of process variance for OU and CIR model of latency parameters (averaged over 5 independent runs with \(N=250\))

<table>
<thead>
<tr>
<th>(\sigma^2,\text{OU})</th>
<th>MSE (m=2)</th>
<th>MSE (m=3)</th>
<th>(\sigma^2,\text{CIR})</th>
<th>MSE (m=2)</th>
<th>MSE (m=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>78.50</td>
<td>7.78</td>
<td>0.85</td>
<td>224.78</td>
<td>11.57</td>
</tr>
<tr>
<td>30</td>
<td>78.23</td>
<td>8.27</td>
<td>0.95</td>
<td>65.48</td>
<td>9.68</td>
</tr>
<tr>
<td>50</td>
<td>52.14</td>
<td>7.84</td>
<td>1.05</td>
<td>31.66</td>
<td>8.67</td>
</tr>
<tr>
<td>70</td>
<td>40.89</td>
<td>7.00</td>
<td>1.15</td>
<td>24.89</td>
<td>8.01</td>
</tr>
<tr>
<td>90</td>
<td>30.24</td>
<td>8.07</td>
<td>1.25</td>
<td>28.81</td>
<td>10.70</td>
</tr>
<tr>
<td>110</td>
<td>36.70</td>
<td>8.41</td>
<td>1.35</td>
<td>41.32</td>
<td>12.94</td>
</tr>
</tbody>
</table>

Further increase results in performance drops for the third component, probably due to spurious estimates on the smooth changes. We suggest \(\sigma^2_b = 70\) for the OU model and \(\sigma_b = 1.15\) for the CIR model as a compromise to satisfactorily detect both type of changes.

The initial distributions are assumed as \(x_0 \sim N(0, 2I)\) and \(b_0 = \mu_b \sim N([50, 120, 200]^T, 5I_M)\). For the amplitude parameters, the variance of the OU model is set \(\sigma^2_a = 10\). For the latency parameters, the process variances of the OU and CIR model are set respectively as \(\sigma^2_b = 70\) and \(\sigma_b = 1.15\). Tuning of \(\sigma^2_u\) and \(\sigma^2_{\phi_u}\), which are only hyper-parameters for Bayesian estimation is less critical. For the random-walk models for the trend components, small noise variances \((\sigma^2_{u_s}, \sigma^2_{u_b}) = (1 \times 10^{-3}, 1 \times 10^{-3})\) are used to maintain the stability of \(\mu\). We set \(\Delta t = 1\) and \(\beta = 1\) for all diffusion processes. We assume \(\phi_{\sigma_0} \sim U[-2.5, 0]\) and \(\sigma^2_{\phi_s} = 5 \times 10^{-4}\) for the model of \(\sigma^2_{v_0}\). Number of particles \(N = 1000\) and re-sampling threshold \(N_T = N/4\) are used throughout this paper.

Fig. 1 shows the ABR estimation results for one realization of simulation. The results are obtained by averaging over 10 independent replications of RBPF. The tracking of ABR subcomponent parameters and their trend components using the OU and CIR model of latency is shown in Fig. 1(a). Both models are able to track reasonably well the dynamic variability including abrupt and smooth changes in both amplitude and latency parameters, with higher accuracy observed for the amplitudes. Larger deviation from the true values can be seen for the latency estimates. The reason is that the estimator of latency \(E[b_n | y_{1:n}]\) is approximated from simulation by PF and thus suffers estimation variance, in contrast to the
amplitude estimates which are computed exactly by KF separately in the RBPF.

For the latency parameters with more abrupt changes, the use of non-linear CIR model exhibits faster tracking over the Gaussian OU model which suffers tracking lag. The Gaussian transition density of OU model performs better in tracking the smooth changes, however, cannot capture the abrupt ones. It can be seen that the CIR model estimates more accurately resemble the true values especially for the second component, as evident by the better tracking of the abrupt random fluctuations and the faster capturing of the sudden jumps at step transitions. This is due to its heavy-tailed transition density which can allow larger deviation to predict rare abrupt changes. It induces slightly noisier estimates on the smoothly changing sinusoid, which is however still acceptable. Fig. 1(a) also shows that the stationary trends for both parameters are estimated comparably well by both models. The estimates of the static trends are not affected by the abrupt step transitions. We can see that the estimated trends take the initial 15 trials of “burn in” period to converge from initial estimates to the constant actual values. Faster convergence is seen for the latency trends since their true values are assumed known a priori and are incorporated in the mean of initial distribution of $p_{b_0}$.
Fig. 2. Estimated trial-varying $\sigma_{v_n}^2$ by the CIR-latency model: $\rightarrow$, true parameters; $\rightarrow$, RBPF estimators.

Fig. 1(b) and 1(c) show the simulated noiseless ABRs and their noisy observations respectively. Fig. 1(d) shows the clean ABRs reconstructed from the parameters estimated by the CIR-latency model in Fig. 1(a). The simulated noise is clearly removed. Moreover, it can be seen that the underlying original ABR dynamics in Fig. 1(b) are accurately estimated, including the abrupt structural right-shift of the second component between the trial 40 and 60. The estimated trend is shown in Fig. 1(e). The underlying ABR trend is well estimated with the inter-trial variability obviously removed. A global waveform morphology is clearly revealed where each subcomponent is more easy to localize without being obscured by the inter-trial variability. The OU-latency model gives similar results which are hard to differentiate with the CIR by naked eye, except showing a more gradual transient during the shift. Fig. 2 shows the estimated $\sigma_{v_n}^2$ by the CIR-latency model, which closely resemble the true volatilities of the simulated noise.

While MSE is used for the parameter tracking performance, the signal-to-noise ratio (SNR) is used as objective evaluation for the denoising of ERPs. The SNR for single-trial ERPs is defined as follows [33]

$$\text{SNR} = 10 \log_{10} \frac{\sigma_s^2}{\sigma_N^2}$$

(29)

where $\sigma_N^2$ and $\sigma_s^2$ are the noise and signal power estimated using averages over trials respectively as

$$\hat{\sigma}_N^2 = \frac{1}{K(J-1)} \sum_{n=1}^{J} \sum_{k=1}^{K} (y_n(k) - \bar{y}(k))^2$$

$$\hat{\sigma}_s^2 = \frac{1}{K} \sum_{k=1}^{K} \bar{y}(k)^2 - (\hat{\sigma}_N^2 / J)$$

(30)
Table 3  Comparison of estimation performance in mean-squared error and signal-to-noise ratio by the
discrete-time random-walk and diffusion ERP models with OU and CIR latency process on
simulated ABRs. The results are averages over 10 replications.

<table>
<thead>
<tr>
<th>EPR Model</th>
<th>MSE Amplitude</th>
<th>MSE Latency</th>
<th>SNR (in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m=1</td>
<td>m=2</td>
<td>m=3</td>
</tr>
<tr>
<td>Random-walk</td>
<td>0.002678</td>
<td>0.003926</td>
<td>0.002116</td>
</tr>
<tr>
<td>Diff-OU-Latency</td>
<td>0.002609</td>
<td>0.004579</td>
<td>0.002053</td>
</tr>
<tr>
<td>Diff-CIR-Latency</td>
<td>0.002619</td>
<td>0.003409</td>
<td>0.001615</td>
</tr>
</tbody>
</table>

where \( \bar{y}(k) = (1/J) \sum_{n=1}^{J} y_n(k) \) is the ensemble average. \( J \) and \( K \) are number of single-trial observations
and measurement samples respectively.

Table 3 presents the evaluation results for the continuous-time diffusion models with OU and CIR-latency processes with the discrete-time random-walk model (5) of [4] as baseline for comparison. For the random-walk model, we set the Gaussian state noise variances in (5) as \( (\sigma_a^2, \sigma_b^2) = (5, 30) \) corresponding to that of the diffusion models and assume a constant observational noise variance fixed to \( \log \sigma_v^2 = -2.5 \).

It can be seen that both diffusion models, especially the CIR model, clearly outperform the discrete-time random-walk, as evident from significant reduction in MSE for both parameters of all components and improvement in SNR of the denoised single-trials. Lower MSEs are observed for the amplitude parameters compared to the latencies, suggesting more accurate estimation by the analytical KF. For the amplitudes, both models perform comparably with improvement by the CIR model for the last two Gaussian components. For the latencies, use of CIR model significantly outperforms the OU model in estimating abrupt changes, giving relative MSE reduction of 16.1% and 36.2% for the first and second components respectively, with only slightly poorer performance for the smooth changes in the third. This is consistent with the results shown in Fig. 1(a). The removal of simulated noise by both models yields substantial SNR improvements over the noisy ABR baseline with further gain for the trend estimates.

Note that the SNR measure used here are ideal for evaluating the removal of observation noise but not the estimation of inter-trial variability which are treated in-discriminatively as noise, hence giving artificially low SNR. However, it is still applicable to the trend where this variability needs to be removed.
4.2. Estimation of Single-trial ABRs’ Dynamics

This section presents the evaluation of the proposed continuous-time diffusive ERP model on estimating real chirp-evoked single-trial ABRs. The data are obtained following the procedure in [32]. The chirp waveforms were computed for different intensity levels of sound pressure level (SPL) unit. Each stimulus was then calibrated by first calculating its peak-equivalent (pe) SPL, which was then measured by a sound level meter, via a microphone connected to an artificial ear. The artificial ear was simultaneously coupled to the headphones while producing the reference sinusoidal wave used to calculate the peSPL [31]. The ABRs were elicited by the chirps, presented at rate of 20 Hz through a headphone to the subjects. Single-trial ABRs from different subjects, under normal and pathological conditions with different stimulus intensities are studied, i.e. two subjects with normal hearing (60 dB and 50 dB SPL respectively) and a pathological subject with hearing impairment (60 dB SPL). The potentials were recorded using passive electrodes placed at positions A1, A2, Cz and the ground Fpz, at sampling frequency of 19.2 kHz. 2000 single-trials of ABR were obtained. The data was band-pass filtered with cutoffs 0.1 and 1.5 kHz and down-sampled to 14.4 kHz. We use only the first 500 trials segmented to fixed time frame where the wave III and V are located, which are further smoothed by moving-averaging every 10 trials.

As in the simulation, we use $M=3$ which is sufficient to model the major waves III and V. Model selection based on fitting Gaussian mixture curves with $M$ varied from 1 to 6 to the segment of mean-ABR of subject 1 that contains waves III and V, also suggests $M=3$ with minimum root-MSE of 0.3154. For the parameter setting for the diffusion processes of ERP parameters, we assume initial distributions $x_0 \sim N(0,2I)$, $b_0 = \mu_{b_0} \sim N([95,135, 200]^T, 5I_M)$ and $c_0 = \mu_{c_0} \sim N([20, 20, 20]^T, 5I_M)$, based on minimum prior information from the mean-ABR obtained by ensemble averaging and let the estimates adjust on their own towards the true values. Note that the units used for the latency and width parameters are in sample index. We set $\Delta t = 0.4$ s, duration of 10 trials for time interval between two average-trials and $\beta = 1$ for all diffusion processes. The noise variances of the CIR process for the widths are selected as
The choice for noise variances of the OU amplitude process and CIR latency process will be further discussed in the next section. The noise variances of the trend models are fixed \((\sigma^2_{\Omega,a}, \sigma^2_{\Omega,b}, \sigma^2_{\Omega,c}) = (1 \times 10^{-3}, 2.5, 0.025)\). For the observational noise volatility process, we use \(\phi_{0_i} \sim U[0,1.5]\) and \(\sigma^2_{\phi} = 2.5 \times 10^{-4}\).

4.2.1. Choice of Diffusion Process Variances

In this section, we shall investigate how the different choices of user-defined model parameters may affect the ERP estimation performance on real data, to provide a heuristic guideline for determining them in practice. We focus on the selection of the noise variances of the diffusive processes due to the sensitivity of the ERP parameter estimators to their variations, particularly those of the amplitude and latency process i.e. \(\sigma^2_a\) and \(\sigma^2_b\), which provide clinically more significant information on the hearing condition, compared to the widths. The selection for the noise variances of the static trend component models is also not as critical, as long as they are set sufficiently small to yield stable estimates, as suggested in [24]. The estimation performance can be assessed indirectly by the denoising performance in SNR or the model likelihood given the parameters which, however, still unable to precisely quantify the correctness of the estimators themselves on the real data where the ground-truth values are unknown. The parameter set that maximize the likelihood and SNR are ideally considered as optimal and to be chosen.

We use a parameter selection approach based on discrete grid search to determine the optimal parameters, based on these two criterions.

Table 4 shows the log-likelihoods and SNR performance for different combinations of \(\sigma^2_a\) and \(\sigma^2_b\) on the ABR data of the subject 1 with normal hearing. When \(\sigma^2_a\) and \(\sigma^2_b\) increase, the log-likelihoods increase implying a better fit to the data, however, the SNR of the denoised ABRs decreases indicating the presence of high level of noise. Thus, obtaining a parameter set that optimizes both criterions at the same time seems not possible, since it is mutually exclusive.
Table 4 Effect of different amplitude and latency process noise variances $\sigma_a^2$ and $\sigma_b^2$ of the OU-CIR ERP model to the fitted log-likelihood and denoising performance in SNR (averaged over 10 independent runs) on a single-trial ABR dataset from a subject with normal hearing. The raw data has been pre-smoothed by moving-averaging of every 10 trials, giving a SNR of -7.7947.

<table>
<thead>
<tr>
<th>$\sigma_a^2$</th>
<th>$\sigma_b^2$</th>
<th>Log-likelihood</th>
<th>SNR (in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Denoised ABRs</td>
</tr>
<tr>
<td>0.2</td>
<td>0.9</td>
<td>-12286.6</td>
<td>1.9124</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>-12149.1</td>
<td>1.1096</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>-12084.8</td>
<td>1.3799</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
<td>-11956.1</td>
<td>-1.6057</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>-11919.1</td>
<td>-2.0194</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>-11803.5</td>
<td>-2.4904</td>
</tr>
<tr>
<td>20</td>
<td>0.9</td>
<td>-11789.4</td>
<td>-4.3140</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
<td>-11739.7</td>
<td>-4.5003</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>-11586.4</td>
<td>-5.1146</td>
</tr>
</tbody>
</table>

Fig. 3 and Fig. 4 show respectively the effect of varying $\sigma_a^2$ with fixed $\sigma_b^2$ on the amplitude estimators and of $\sigma_b^2$ with fixed $\sigma_a^2$ on the latency estimators, corresponding to the results in Table 4. The estimators for both parameters exhibit, within each ABR sub-component, similar dynamic patterns but with larger variations as the variances increase. It can be seen that the increase of noise variances can follow more closely the detailed variations, especially the abrupt changes of the underlying parameters that generate the observation data. However, too large a value tends to over-fitting, until to the non-neuronal, random components of noise and artifacts, and hence produces spurious jumps as shown in Fig. 3(c) and Fig. 4(c).

Fig. 5 shows the influence of increasing the variances on the reconstructed single-trial estimates. Use of small variance gives smooth estimates as seen in Fig. 5(a), however, might risk the loss of trial-to-trial details. In contrast, larger variances are capable of capturing the finer structure of the inter-trial dynamics (Fig. 5(b)), but if set overly large, will induce noisy estimates as in Fig. 5(c), which is correctly translated as the SNR performance drops in Table 4. Based on these results, we select $\sigma_a^2=2$ and $\sigma_b^2=1.1$ for this particular case to achieve an optimal balance between the goodness of fit and low level of noise.
Fig. 3. Effect of different $\sigma^2_a$ on the estimation of amplitudes of the ABR subcomponents: first component (top row), second component (middle row) and third component (bottom row). (a) $\sigma^2_a = 0.2$. (b) $\sigma^2_a = 2$. (c) $\sigma^2_a = 20$. $\sigma_b^2 = 1.1$ and $\sigma_c^2 = 1.7$ are fixed. The results are obtained by averaging over 10 independent runs.
Fig. 4. Effect of different $\sigma_b^2$ on the estimation of latencies of the ABR subcomponents: first component (top row), second component (middle row) and third component (bottom row). (a) $\sigma_b^2 = 0.9$, (b) $\sigma_b^2 = 1.1$, (c) $\sigma_b^2 = 1.3$. $\sigma_a^2 = 0.2$ and $\sigma_c^2 = 1.7$ are fixed. The results are obtained by averaging over 10 independent runs.
Fig. 5. Effect of increasing $\sigma_a^2$ and $\sigma_b^2$ on the single-trial estimates. (a) $\sigma_a^2 = 0.2, \sigma_b^2 = 0.9$. (b) $\sigma_a^2 = 2, \sigma_b^2 = 1.1$. (c) $\sigma_a^2 = 20, \sigma_b^2 = 1.3$. $\sigma_c^2 = 1.7$ is fixed. The results are obtained by averaging over 10 independent runs.

4.2.2. Comparison with Discrete-time Random-walk Model

We compare the estimation performance of the continuous-time OU-CIR diffusion ERP model with the discrete-time random-walk model (5). For the diffusion model, we use the parameters pre-selected above. For the random-walk model, the time-invariant observational noise variance is fixed to $\log \sigma_c^2 = 1.5$ and the Gaussian state noise variances are set $(\sigma_a^2, \sigma_b^2, \sigma_c^2) = (0.2, 25, 0.3)$. The comparison results for one realization of RBPF on the same subject used in previous section are shown in Fig. 6. Fig. 6(a) shows the estimators of ABR subcomponent parameters and the corresponding trend components. The stationary OU and CIR model produce more stable parameter estimates oscillating within stationary distributions centered at the trends. The estimates by the non-stationary random-walk, however, drift away from the trends, as clearly seen for the amplitudes of first and second components and the width of second component, despite setting small variances to maintain its stability at the cost of tracking lag on inter-trial dynamics. This instability of the random-walk specification could also explain their poor MSE performance in Table 3. Besides, the non-linear CIR models can detect the abrupt changes. The
underlying trend estimates are well located at the central path of the single-trial dynamics, and able to converge rapidly from the initials towards stable values representing the means of the ABR Gaussian subcomponents, without requirement of very careful setting of the prior distribution to the exact values.

Fig. 6(b) shows the noisy single-trial ABR measurement data of the subject with SNR = -7.79 dB. The ensemble averaging of the noisy trials can reveal the existence of wave V expected for normal hearing subjects, as shown by epoch mean. However, the trace of wave V is barely seen from the single-trial epoch plot (bottom) and still vaguely observed in background noise from the image plot (top), despite the moving-average smoothing. The clean single-trial estimates reconstructed from the estimated parameters by the random-walk and diffusion model are shown in Fig. 6(c) and Fig. 6(d) respectively. The estimates for both models show significant reduction in noise and exhibit clearly the underlying inter-trial dynamics with more variability details captured by the diffusion model. Moreover, the trace of wave V is more pronounced. Fig. 6(e) presents the ABR stationary trend estimates, where the inter-trial variability is
greatly reduced. The wave V trace, located around 28 ms, become almost static and much clearer. Besides, larger and smoother epoch mean is obtained for the denoised ABRs and even better for its trend, compared to that of noisy one. This result suggests that the proposed method can provide better visualization and detection of wave V for the assessment of hearing condition.

Table 5 presents the comparison of log-likelihoods fitting the data and SNR performance between the two models. The log-likelihoods of the diffusion model are higher than the random-walk, implying a better fit to the data. The denoising by both models gives significant SNR improvements of 71.2% and 74.1% respectively over the noisy ABRs, with better performance by the diffusion model. More substantial SNR improvements can be achieved by decreasing the diffusion variances, as already shown in Table 4. The lower SNR of the random-walk model is due to that the resulting single-trial subcomponents are found, in some realizations, diverted substantially from their means due the non-stationary parameter estimates, as shown in Fig. 6(a). Other reason may be that the use of fixed variance fails to eliminate the highly non-stationary background noise.

<table>
<thead>
<tr>
<th>ERP Model</th>
<th>Log-likelihood</th>
<th>SNR (in dB)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Noisy ABRs</td>
<td>Denoised ABRs</td>
</tr>
<tr>
<td>Random-walk</td>
<td>-12804.4</td>
<td>-7.79468</td>
<td>-2.2472</td>
</tr>
<tr>
<td>OU-CIR</td>
<td>-11919.1</td>
<td>-7.79468</td>
<td>-2.0195</td>
</tr>
</tbody>
</table>

4.2.3. Results for Different Stimulus Intensity and Pathological Subject

We perform further evaluation on different subjects and stimulus intensities, and a pathological case to investigate the robustness of the proposed method for ERP estimation across varied subjects and conditions, and the portability of the pre-defined model parameters. The single-trial estimates for the second normal-hearing subject with 50 dB SPL using the parameters pre-selected above for the subject 1, is given in Fig. 7. It is showed that the ABR morphology and the inter-trial dynamics, although different from that of the subject 1 e.g. smaller waveform amplitudes due to lower stimulus intensity, can still be
well extracted. Fig. 8 shows the estimation results using the same parameter set for the subject with hearing impairment, which is correctly indicated by the substantial attenuation of the trace of wave V, clearly noticeable in the single-trial estimates (Fig. 8(b)) and even so in the trend estimates (Fig. 8(c)). This further suggests the usefulness of our method for hearing problem detection. These results also imply some degree of robustness of the user pre-defined model parameters for cross-subject or condition ERP estimation, which is possibly due to two reasons. First is the self-adjustability towards the true values, of the minimum mean-square time-varying estimators generated by the Bayesian RBPF. Second is the sharing of some common waveform characteristics and inter-trial dynamical behaviors within certain range, despite the variability across conditions.

We shall give further remarks on the issue of parameter determination. The primitive grid search considered here is only sub-optimal and computationally inefficient especially when the number of parameters to search through is large, which renders a re-estimation for different subjects and conditions expensive. Despite the somehow portability of the user pre-defined parameters, an automated, more accurate and efficient iterative algorithm for maximum likelihood (ML) estimation of the proposed model, should be considered. ML estimation of a linear-Gaussian dynamic model of ERPs using the iterative expectation-maximization (EM) algorithm has been studied in our earlier work [5], where the likelihood function can be evaluated analytically using the KF. However, the likelihood of the non-linear PO diffusion process considered here is not available in closed-form. Possible solution is to use simulation-based sequential Monte Carlo approximation for the ML estimation [34], which however is computationally expensive and mainly in the stage of simulation study. Improving these algorithms for practical usage in ERP estimation can be pursued in the future work.

We compare in further details the performance of our method with the traditional ensemble averaging technique in detecting the wave V in ABRs for diagnosis of hearing loss. Due to poor SNR condition, the conventional averaging requires large number of sweeps to cancel out random noise which, however, is time-consuming and causes much discomfort to the subject being tested especially the infants. Fast wave V detection using a small number of single-trials, yet obtaining a reliable result, is a challenging task. Fig.
Fig. 7. Single-trial ABR estimates for normal-hearing subject 2 with stimulus intensity of 50 dB SPL. (a) Noisy single-trials. (b) Denoised single-trials. (c) Estimated trends.

Fig. 8. Single-trial ABR estimates for a subject with hearing impairment, with stimulus intensity of 60 dB SPL. (a) Noisy single-trials. (b) Denoised single-trials. (c) Estimated trends.
Fig. 9. Mean-ABRs by conventional average of raw single-trials (—) and average of denoised single-trials (—) and estimated trends (—) using the proposed diffusion ERP model for a normal-hearing subject (a) and a hearing-impaired subject at a reduced number of sweeps (averaging only the first 10 trials). The results for the diffusion model are obtained by averaging over 10 independent runs.

9 shows the mean-ABRs averaged in a traditional way over the raw single-trials and after the denoising using the proposed diffusion model for the normal-hearing subject-1 and the hearing-impaired subject using a reduced number of only 10 trials, instead of 50 in the previous experiments. Even with a reduced sweeps, our method can better detect the presence of the wave V in ABRs of normal-hearing, as evident from a more prominent and smoother peak (as shown in Fig. 9(a)). Furthermore, it reflects accurately the suppression of the component expected in the ABRs of a hearing-impaired subject, compared to the conventional averaging which still produces spurious peaks around the wave V location (as shown in Fig. 9(b)), which are probably random noise unable to be eliminated with a small number of sweeps.

4.3. Handling Irregularly Spaced ABRs

In this section, we shall compare the proposed continuous-time diffusion and the discrete-time random-walk ERP model in solving two cases of irregularly or variably spaced observation problem in single-trial ABR estimation: (1) variable sampling frequencies and (2) irregularly spaced single-trial with missing data. Fig. 10 shows the estimated parameters of the third subcomponent by the discrete- and continuous-time models on the single-trial ABR data from the normal-hearing subject 1, sampled at
different inter-trial intervals of 1-trial ($\Delta t = 0.4$), 2-trial ($\Delta t = 0.8$) and 3-trial ($\Delta t = 1.2$), as shown in Fig. 11 (top). The parameter setting is the same as the previous section except that the values of $\Delta t$ in the diffusion processes can be varied corresponding to the different sampling rates of the data (not necessarily related to the original ($\Delta t = 0.4$)). We compute the mean-reversion and noise variance terms of the discretized OU process (10) corresponding to the different values of $\Delta t$ in Table 6. Note that the OU parameters, similarly for that of the CIR process, are adjustable according to $\Delta t$ for modeling data with different sampling frequencies, instead of being fixed in the discrete-time random-walk model. The continuous model is in fact discretized into discrete models of different sampling rates.

From Fig. 11 (top), it can be seen that when the single-trial data are more sparsely sampled as $\Delta t$ increases, the inter-trial changes become larger and more abrupt. The estimators of the continuous model can follow those marginal microstructure inter-trial movements of the parameters in the high-frequency data (as in Fig. 10(a)), and still able to detect rapidly the increasingly abrupt changes in the lower-frequencies data (as in Fig. 10(a) and 10(b)). However, the discrete model with parameters pre-defined for the original sampling rate, can somehow follow the small dynamics thereby, but not the more abrupt, larger changes as $\Delta t$ increases, which sees an increasingly large tracking lag compared to the continuous model. The superiority of the continuous model in capturing both small and abrupt changes is due to the adjustability of its parameter space according to the variable sampling rate of the data. Whereas, the noise variances of the discrete model pre-defined for the smoother changes in data with small $\Delta t$, unless being re-estimated, are fixed and too small for the abrupt changes when $\Delta t$ is large. These results further substantiate the argument in [8] that the discrete-time models, at a specified sampling interval, are difficult in dealing with both quick and slow dynamics which, however, can be better captured by a continuous model estimated from rapidly sampled data, where the parameters are independent of the sampling rate. Besides, the estimators trajectories of the continuous model are more stable around the trends, due to the larger mean-reversion effect, as the exponentials $e^{-\beta \Delta t}$ in (9) decreases.
Table 6 The parameters of the discretized OU process for different values of $\Delta t$ with $\beta = 1$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\Delta t = 0.4$</th>
<th>$\Delta t = 0.8$</th>
<th>$\Delta t = 1.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\exp(-\beta \Delta t)$</td>
<td>0.6703</td>
<td>0.4493</td>
<td>0.3012</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.5247</td>
<td>0.6317</td>
<td>0.6743</td>
</tr>
</tbody>
</table>

Fig. 10. Estimation of amplitudes (top), latencies (middle) and widths (bottom) parameters of the third ABR Gaussian subcomponent by the discrete-time (---) and continuous-time models (----) on single-trial data sampled at different frequencies: (a) $\Delta t = 0.4$ (1-trial interval), (b) $\Delta t = 0.8$ (2-trial interval) and (c) $\Delta t = 1.2$ (3-trial interval). The results are averages over 20 independent runs.
Fig. 11. Single-trial estimates by discrete-time (middle row) and continuous-time (bottom row) ERP model on noisy single-trial ABR data (top row) sampled at different frequencies: (a) $\Delta t = 0.4$ (1-trial interval), (b) $\Delta t = 0.8$ (2-trial interval) and (b) $\Delta t = 1.2$ (3-trial interval). The results are obtained by averaging over 20 independent runs.

The reconstructed single-trial estimates by the both models are shown in Fig. 11. It can be seen that the continuous model, while preserving the overall dynamic pattern despite the down-sampling, can retrieve well the abrupt single-trial dynamics with the increase of $\Delta t$ (as shown in Fig. 11(bottom)), which fail to be captured by the persistently smooth estimates of the discrete model (as shown in Fig. 11(bottom)). Moreover, the mean-reversion effect of the diffusions plays a role to maintain the stationarity of the growingly fluctuated single-trials around the trends, in contrast with the unsteady, constantly shifting estimates of the random-walk.

Fig. 12 compares the estimation performance between the models on the same ABR data with simulated missing observations. Three consecutive null trials are added at every 10 trials, creating an
Fig. 12. Single-trial estimates on ABR data with simulated missing observations. Segment of consecutive three trials at every 10 trials is set zero values. (a) Noisy single-trials. (b) Denoised ABRs by the discrete-time model. (c) Denoised ABRs by the continuous-time model. The results are obtained by averaging over 20 independent runs.

unevenly spaced data (Fig. 12(a)) which cannot be represented consistently by a discrete-time model with inherently constant sampling period. The RBPF is applied similarly, but with $\Delta t = 1.2$ when null-segment is encountered for the continuous model, while at a trial-to-trial basis for the discrete model. Meanwhile, $\mu_{n-\Delta t}$ is used for one-step-ahead prediction of $\mu_n$ for the trend component. From Fig. 12(b), the third subcomponents, after the zero-interval at trial 30 are incorrectly estimated using the discrete-time model, with the subsequent components diverted substantially from the right trend. Whereas, as shown in Fig. 12(c), the continuous-time model captures well the ERP dynamics including those at each interval, without being interrupted by the missing observations, and reasonably matching to that estimated on the full data in Fig. 11(a) (bottom).

5. Conclusion

We have presented a new modeling framework based on the continuous-time PO diffusion process for estimating single-trial ERP dynamics with application to analysis of ABRs. The underlying inter-trial
changes of ERP parameters are assumed to follow continuous-time diffusion processes and discretely observed in background noise. We consider the Gaussian OU process and further the non-linear CIR process with non-Gaussian transition density to capture both abrupt and gradual parameter changes. Further, we use mean-reverting variant for the diffusions to capture the trend of the dynamics and a stochastic volatility process to describe the changing-volatilities in observational noise. The proposed model formulated in a conditionally-linear Gaussian state-space form is estimated using RBPF. Results on simulated ABR data show that the CIR model gives better tracking of the non-Gaussian parameter changes. Evaluation on real data shows that the diffusive ERP model performs better in extracting the underlying inter-trial dynamics with SNR improvement, and produces more stable estimates than the random-walk models used in previous studies. We also demonstrate the robustness of the proposed method for estimation across different subjects, stimulus intensities and hearing conditions. Faster and more reliable detection of the wave V component of ABR by our method, compared to the conventional averaging technique, also suggests its use to improve the diagnosis of hearing loss. Use of continuous-time modeling for ERPs, instead of discrete-time models as in the previous works, is justified by its flexibility and efficiency in handling variably or irregularly sampled data. This has been illustrated through two situations in single-trial ERP estimation, where the continuous-time diffusive model, with adjustable parameter space independent of observation intervals, can capture more accurately both the small and abrupt changes due to variable sampling rates, as well as the interrupted dynamics by irregular intervals due to missing single-trials. Future work may consider extending the diffusion specification to the trend and noise components and incorporate other non-Gaussian non-linear diffusion models within the proposed framework. Besides, the proposed model can be applied to analyze other types of non-stationary bio-signals with hidden dynamics. Development of iterative algorithms for ML estimation in the proposed model can be also addressed in the future.
Acknowledgements

The authors would like to thank the reviewers for valuable comments. This work was supported by the Universiti Teknologi Malaysia (UTM) and the Ministry of Higher Education (MOHE) Malaysia under Research University Grant (GUP) Q.J130000.2545.04H21.

References


42