

A Statistical Model for Burst Detection

Lee F. Richardson

January 9, 2019

1 Introduction

This document proposes a statistical model and estimation procedure for to detect “Bursts” in Local Field Potential (LFP) signals. By bursts, I mean local (in time) oscillations at particular frequencies. The motivation for burst detection comes from Lundqvist et al. (2016), who finds that bursts, *not* sustained oscillations, underly working memory.

2 Literature Review

3 Data

4 A Statistical Model for Bursts

A graphical example of a burst in our data-set is shown in Figure 1. Figure 1 shows that, for a short time window, the LFP signal has oscillations in the 40-60 Hz range. This short oscillation is what we refer to as a “Burst”. This observation suggests a simple deterministic model for a “burst” in our LFP signal

$$g_t = A \cos\left(\frac{2\pi Ft}{N} + \phi\right) \cdot \mathbb{1}_{S,S+L-1}(t) \quad (1)$$

Which is simply a cosine function multiplied by a box-car window. Equation 1 has 5 parameters:

- A : Amplitude. $A \in [0, \infty]$
- F : Frequency. Number of cycles in a length N signal. $F \in [0, \frac{1}{2}]$
- ϕ : Phase. $\phi \in [0, 2\pi]$
- S : Start of oscillation. $S \in [0, 1, \dots, N-2]$
- L : Length of oscillation. $L \in [1, 2, \dots, N-2]$

We turn Equation 1 into a statistical model by introducing an error term;

$$x_t = g_t + \epsilon_t \quad (2)$$

To keep things simple, assume $\epsilon_t \sim N(0, \sigma^2)$ and ϵ_t is i.i.d.

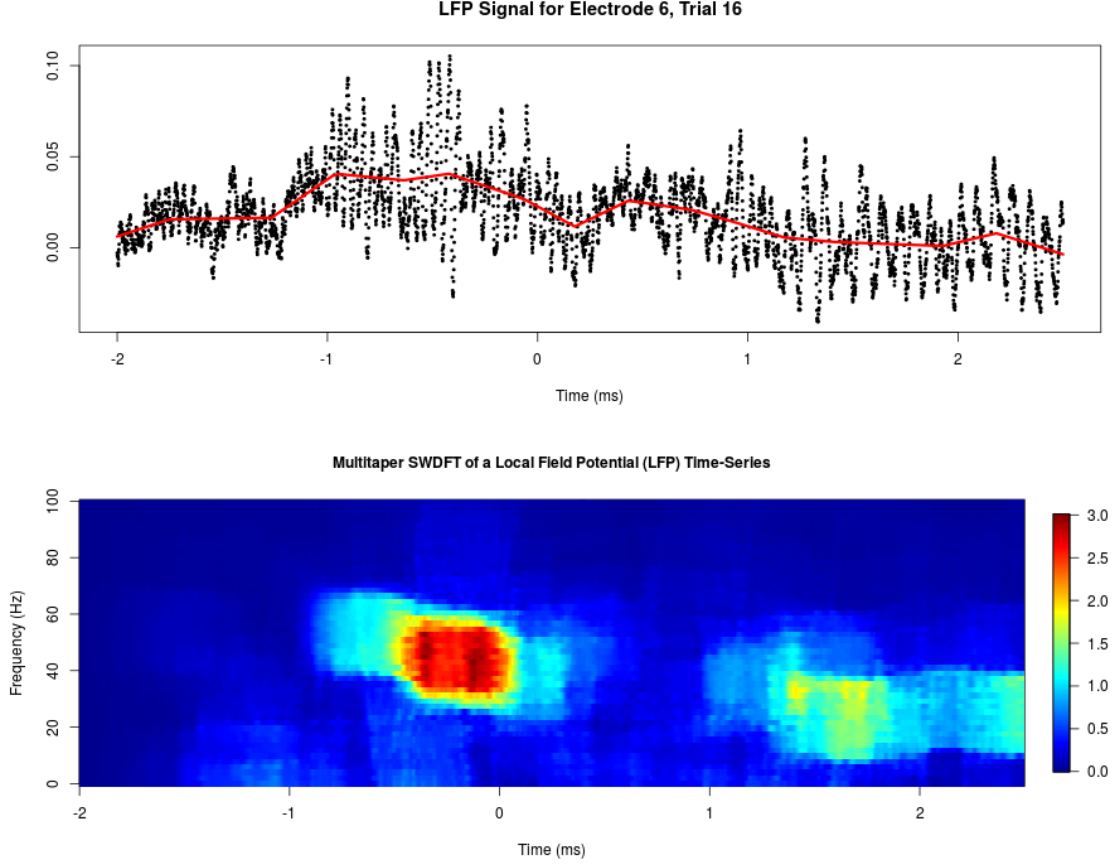


Figure 1: Top: An example Local Field Potential Signal in our data-set that displays “bursting” behavior. The red line is a trend-filter estimate of the slow trend in the signal, which we remove for spectral analysis. Bottom: Multi-taper SWDFT of the de-trended LFP signal. We see bursting behavior around 40 Hz, which corresponds to the “fixation” part of the trial.

5 Estimation

Now that we’ve proposed a statistical model (Equation 2), given an arbitrary signal, we need to estimate the parameters. Section ?? derives analytic solutions for three of the six parameters (A , ϕ , and σ), using a standard trick from harmonic regression. Section ?? shows how to solve for the remaining three parameters (S , L , and F) numerically, using a Spectrogram.

5.1 Estimating A , ϕ , and σ

Following Equation 2, we know that $x_t \sim N(g_t, \sigma^2)$. Denote $\theta = [A, F, \phi, S, L, \sigma]$ as our vector of unknown parameters, and say that we observe a length N time-series $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]$. The joint pdf of \mathbf{x} is:

$$\mathbb{P}(x_0, \dots, x_{N-1} | \theta) = \prod_{t=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_t - g_t)^2}{2\sigma^2}\right) \quad (3)$$

Therefore, the likelihood function is:

$$L(\theta | \mathbf{x}) = \prod_{t=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_t - g_t)^2}{2\sigma^2}\right) \quad (4)$$

And the log-likelihood is proportional to

$$\ell(\theta) \propto -N \log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=0}^{N-1} (x_t - g_t)^2 \quad (5)$$

The maximum likelihood estimate is:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \ell(\theta) \quad (6)$$

This section assumes that S, L, and F are known, and we will show how to solve for them numerically in the next section. Then we have

$$\ell(A, \phi, \sigma | S, L, F) \propto -N \log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=0}^{N-1} (x_t - A \cos(\frac{2\pi Ft}{N} + \phi) \cdot \mathbb{1}_{S, S+L-1}(t))^2 \quad (7)$$

Next, notice that for any value of σ , the log-likelihood is *maximized* when the quadratic term inside the sum is minimized (since there is a negative sign in front of the sum). With this, we closely follow the second example of Section 7.10 of Kay (1993), and let $J(A, \phi)$ be the term inside the sum we want to minimize:

$$J(A, \phi) = \sum_{t=0}^{N-1} (x_t - A \cos(\frac{2\pi Ft}{N} + \phi) \cdot \mathbb{1}_{S, S+L-1}(t))^2 \quad (8)$$

Which looks like least squares. The only problem is that the cosine factor is non-linear, but we can linearize it with the following trigonometric identity:

$$A \cos(\frac{2\pi Ft}{N} + \phi) = A \cos(\phi) \cos(\frac{2\pi Ft}{N}) - A \sin(\phi) \sin(\frac{2\pi Ft}{N}) \quad (9)$$

Then if we define the following coefficients:

$$\begin{aligned} \beta_1 &= A \cos(\phi) \\ \beta_2 &= -A \sin(\phi) \end{aligned} \quad (10)$$

There is a one-to-one correspondence between (β_1, β_2) and (A, ϕ) :

$$\begin{aligned} A &= \sqrt{\beta_1^2 + \beta_2^2} \\ \phi &= \arctan\left(\frac{-\beta_2}{\beta_1}\right) \end{aligned} \quad (11)$$

We can re-write $J(A, \phi)$ instead as a function of the β 's:

$$J(\beta_1, \beta_2) = \sum_{t=0}^{N-1} (x_t - \beta_1 \cos(\frac{2\pi F t}{N}) \mathbb{1}_{S, S+L-1}(t) - \beta_2 \sin(\frac{2\pi F t}{N}) \mathbb{1}_{S, S+L-1}(t))^2 \quad (12)$$

Equation 12 is just multiple linear regression, which means we can solve for β analytically. We now write $J(\beta_1, \beta_2)$ in matrix-vector notation. Define the following

$$\begin{aligned} \mathbf{c} &= [\cos(\frac{2\pi F \cdot 0}{N}) \mathbb{1}_{S, S+L-1}(0), \cos(\frac{2\pi F \cdot 1}{N}) \mathbb{1}_{S, S+L-1}(1), \dots, \cos(\frac{2\pi F \cdot (N-1)}{N}) \mathbb{1}_{S, S+L-1}(N-1)]^T \\ \mathbf{s} &= [\sin(\frac{2\pi F \cdot 0}{N}) \mathbb{1}_{S, S+L-1}(0), \sin(\frac{2\pi F \cdot 1}{N}) \mathbb{1}_{S, S+L-1}(1), \dots, \sin(\frac{2\pi F \cdot (N-1)}{N}) \mathbb{1}_{S, S+L-1}(N-1)]^T \\ \mathbf{U} &= [\mathbf{c}, \mathbf{s}] \\ \mathbf{x} &= [x_0, x_1, \dots, x_{N-1}] \\ \beta &= [\beta_1, \beta_2] \end{aligned} \quad (13)$$

$J(\beta_1, \beta_2)$ can now be written as:

$$J(\beta_1, \beta_2) = (\mathbf{x} - \mathbf{U}\beta)^T (\mathbf{x} - \mathbf{U}\beta) \quad (14)$$

And the estimate for $\hat{\beta}$ is:

$$\hat{\beta} = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{x} \quad (15)$$

Which we transform into \hat{A} and $\hat{\phi}$ using Equation 11. For σ , we have:

$$\frac{\partial \ell}{\partial \sigma} = \frac{-N}{\sigma} + \frac{1}{\sigma^3} \sum_{t=S}^{S+L-1} (x_t - g_t)^2 \quad (16)$$

Setting this to 0 and plugging in \hat{A} and $\hat{\phi}$ gives:

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{t=0}^{N-1} (x_t - g_t)^2} \quad (17)$$

Similar to standard estimates of σ with normal errors.

So, if we assume S , L , and F are known, we can estimate \hat{A} , $\hat{\phi}$, and $\hat{\sigma}$ analytically. But in practice, we don't know the values of S , L , and F , and we need to estimate them as well. This is where we use the Spectrogram.

5.2 Estimating S, L, and F

Section 5.1 shows that, if we *know* S, L, and F, we have analytic solutions to the maximum likelihood estimates of A, ϕ , and σ . So the next question is, how can we estimate S, L, and F? Recall the sum of squares term we minimized in Equation 14

$$J(\hat{\beta}_1, \hat{\beta}_2) = (\mathbf{x} - \mathbf{U}\hat{\beta})^T(\mathbf{x} - \mathbf{U}\hat{\beta}) \quad (18)$$

Since \mathbf{U} includes S, L, and F, we can extend this minimization to these three parameters

$$J(\hat{\beta}_1, \hat{\beta}_2, S, L, F) = (\mathbf{x} - \mathbf{U}\hat{\beta})^T(\mathbf{x} - \mathbf{U}\hat{\beta}) \quad (19)$$

Plugging in Equation 15 gives (Kay (1993) Equation 7.65):

$$J(\hat{\beta}_1, \hat{\beta}_2, S, L, F) = \mathbf{x}^T(\mathbf{I} - \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T)\mathbf{x} \quad (20)$$

Which is equivalent to finding the values of S, L, and F that maximize

$$\arg \max_{S, L, F} \mathbf{x}^T\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{x} \quad (21)$$

Since $\mathbf{U} = [\mathbf{c}, \mathbf{s}]$, we can re-write this maximization as

$$\begin{bmatrix} \mathbf{c}^T\mathbf{x} \\ \mathbf{s}^T\mathbf{x} \end{bmatrix}^T \begin{bmatrix} \mathbf{c}^T\mathbf{c} & \mathbf{c}^T\mathbf{s} \\ \mathbf{s}^T\mathbf{c} & \mathbf{s}^T\mathbf{s} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T\mathbf{x} \\ \mathbf{s}^T\mathbf{x} \end{bmatrix}$$

We can approximate the diagonal terms in the 2×2 matrix as

$$\begin{aligned} \mathbf{c}^T\mathbf{c} &= \sum_{t=0}^{N-1} \cos\left(\frac{2\pi Ft}{N}\right)^2 \mathbb{1}_{S, S+L-1}(t) \approx \frac{L}{2} \\ \mathbf{s}^T\mathbf{s} &= \sum_{t=0}^{N-1} \sin\left(\frac{2\pi Ft}{N}\right)^2 \mathbb{1}_{S, S+L-1}(t) \approx \frac{L}{2} \end{aligned}$$

For the off diagonal terms, we have:

$$\begin{aligned} \mathbf{c}^T\mathbf{s} &= \mathbf{s}^T\mathbf{c} \\ &= \sum_{t=0}^{N-1} \cos\left(\frac{2\pi Ft}{N}\right) \sin\left(\frac{2\pi Ft}{N}\right) \mathbb{1}_{S, S+L-1}(t) \\ &= 2 \sum_{t=0}^{N-1} \sin\left(\frac{4\pi Ft}{N}\right) \mathbb{1}_{S, S+L-1}(t) \end{aligned} \quad (22)$$

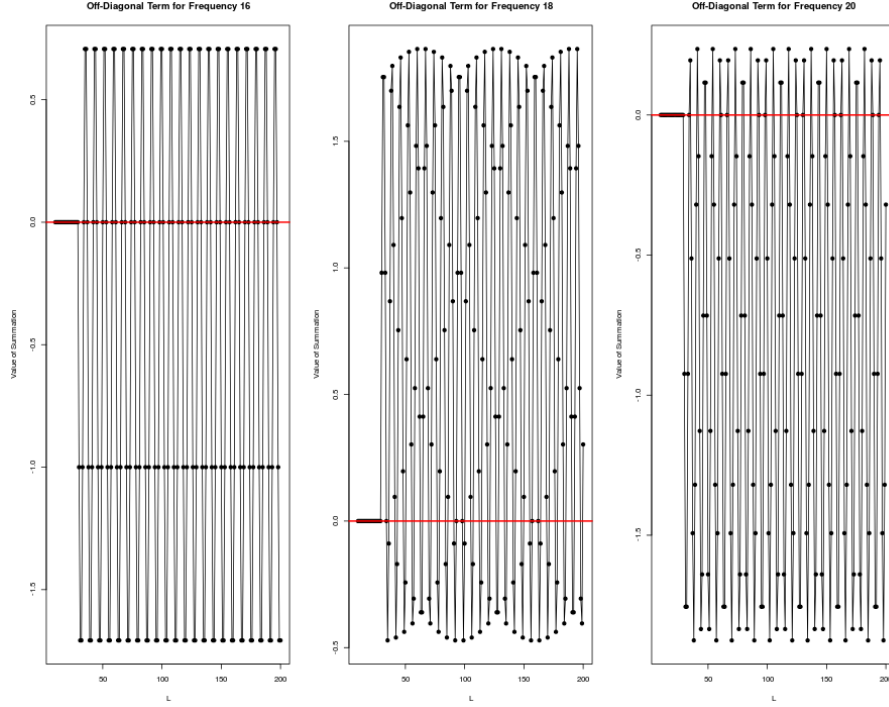


Figure 2: Off Diagonal Terms. $N=256$, $L=10-200$, $S=30$, $F=16,28,20$.

And this is actually the key difference between our problem and the example in Kay (1993). In Kay (1993), since this summation is over an entire cycle, Equation 22 is ≈ 0 . But for our problem, we need to include this term, since it only cancels out when L is a complete cycle (see Figure 2 for a demonstration). So for now, let $\mathbf{c}^T \mathbf{s} = z(S, L, F, N)$, or z for short. Note that we can bound z in the range $[-2, 2]$. Then our expression is

$$\approx \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \frac{L}{2} & z \\ z & \frac{L}{2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}$$

The inverse of the inner matrix is

$$\frac{1}{\frac{L^2}{4} - z^2} \begin{bmatrix} \frac{L}{2} & -z \\ -z & \frac{L}{2} \end{bmatrix} \approx \frac{4}{L^2} \begin{bmatrix} \frac{L}{2} & -z \\ -z & \frac{L}{2} \end{bmatrix}$$

Plugging this back in gives

$$\frac{4}{L^2} \left[\frac{L}{2} (\mathbf{c}^T \mathbf{x})^2 + \frac{L}{2} (\mathbf{s}^T \mathbf{x})^2 - 2z (\mathbf{c}^T \mathbf{x})(\mathbf{s}^T \mathbf{x}) \right] = \frac{2}{L} [(\mathbf{c}^T \mathbf{x})^2 + (\mathbf{s}^T \mathbf{x})^2] - \frac{8z}{L^2} (\mathbf{c}^T \mathbf{x})(\mathbf{s}^T \mathbf{x})$$

Since the first term on the RHS is normalized by $1/L$, and the second term is normalized by $\frac{1}{L^2}$, the first term dominates this maximization (and if $z = 0$, or close to it, we know that the second term on the RHS goes away). Therefore, we focus on maximizing the first term in this expression. Notice that

$$\begin{aligned}
\frac{2}{L}[(\mathbf{c}^T \mathbf{x})^2 + (\mathbf{s}^T \mathbf{x})^2] &= \frac{2}{L}[(\sum_{t=0}^{N-1} x_t \cos(\frac{2\pi F t}{N}) \mathbb{1}_{S,S+L-1}(t))^2 + (\sum_{t=0}^{N-1} x_t \sin(\frac{2\pi F t}{N}) \mathbb{1}_{S,S+L-1}(t))^2] \\
&= \frac{2}{L} |\sum_{t=0}^{N-1} x_t \omega_N^{-tF} \mathbb{1}_{S,S+L-1}(t)|^2 \\
&= \frac{2}{L} |\sum_{t=S}^{S+L-1} x_t \omega_N^{-tF}|^2
\end{aligned} \tag{23}$$

Which looks strikingly similar to the Periodogram (the MLE estimate in Kay (1993)). Recall the definition of the Sliding Window Discrete Fourier Transform (SWDFT), with length n window size

$$\begin{aligned}
a_{k,p,n} &= \sum_{j=0}^{n-1} x_{p-n+1+j} \omega_n^{-jk} \\
k &= 0, 1, \dots, n-1 \\
p &= n-1, n, \dots, N-1
\end{aligned} \tag{24}$$

And if we let $\frac{f}{L} = \frac{F}{N}$, we can interpret Equation 23 as a SWDFT with length L windows

$$\arg \max_{S,L,F} \quad \frac{2}{L} |\sum_{t=S}^{S+L-1} x_t \omega_L^{-tf}|^2$$

Equation 25 tells us that our estimates of S , L , and F will be the values that maximize the Sliding Window DFT across varying window sizes.

So in practice, how would we make these estimates? The first step would be to find the maximum squared modulus SWDFT across window sizes

$$(\hat{k}, \hat{p}, \hat{n}) = \arg \max_{k,p,n} |a_{k,p,n}|^2 \tag{25}$$

Which gives the following estimates of S , L , and F

$$\begin{aligned}
\hat{L} &= \hat{n} \\
\hat{S} &= \hat{p} - \hat{n} + 1 \\
\hat{F} &= \frac{\hat{k}}{\hat{n}}
\end{aligned} \tag{26}$$

The heavy computational component of this approach is repeatedly computing Spectrograms of varying window sizes. We are currently looking into methods to speed this up. Then with S , L , and F , you can easily obtain analytic solutions from Section 5.1 for A , ϕ , and σ . If desired, further numerical optimization may be performed around the estimate of \hat{F} , which ensures we aren't restricting F to the set of Fourier Frequencies.

References

- Kay, S. M. (1993). Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1). *PTR Prentice-Hall, Englewood Cliffs*.
- Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron*, 90(1):152–164.
- Richardson, L. F. and Eddy, W. F. (Submitted. 2018). The sliding window discrete fourier transform. *arXiv preprint arXiv:1807.07797*.