

# A Statistical Approach to 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.  $F \in [0, \frac{1}{2}]$
- $\phi$ : 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  $\epsilon_t$  is i.i.d.

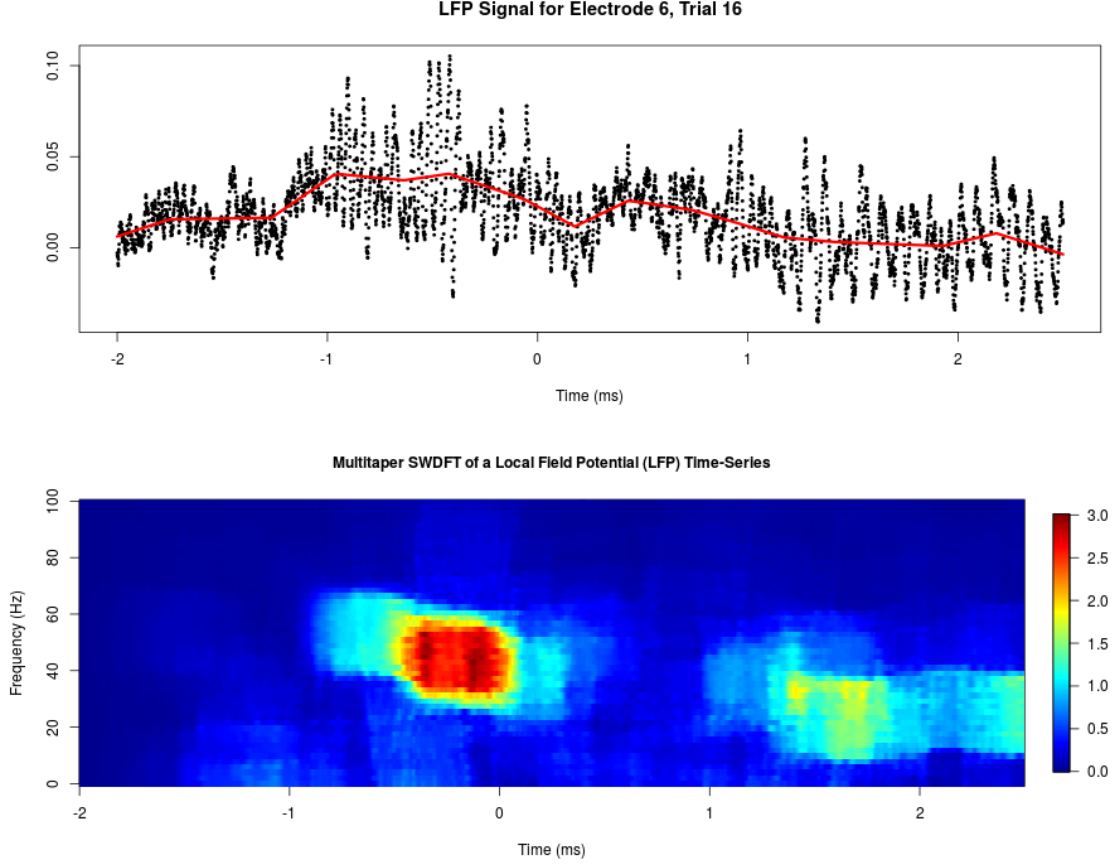


Figure 1: Top: An example Local Field Potential Signal in our data-set that displays “bursting” behavior. The red line is a trend-filter estimate of the slow trend in the signal, which we remove for spectral analysis. Bottom: Multi-taper SWDFT of the de-trended LFP signal. We see bursting behavior around 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 5.1 derives analytic solutions for three of the six parameters ( $A$ ,  $\phi$ , and  $\sigma$ ), using a standard trick from harmonic regression. Section 5.2 shows how to solve for the remaining three parameters ( $S$ ,  $L$ , and  $F$ ) numerically, using a Spectrogram.

### 5.1 Estimating $A$ , $\phi$ , and $\sigma$

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:

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

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 S, L, and F are known, and solves for A,  $\phi$ , and  $\sigma$  (the next section shows how to estimate S, L, and F). 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  $\sigma$ , 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

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

This looks exactly like least squares, but the cosine factor is non-linear. We can linearize the cosine factor 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 we define the coefficients

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

And there is a one-to-one correspondence between  $(\beta_1, \beta_2)$  and  $(A, \phi)$ :

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

With this substitution, we can re-write  $J(A, \phi)$  with  $J(\beta_1, \beta_2)$

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

And now Equation 12 is multiple linear regression. This means we can solve for  $\hat{\beta}_1, \hat{\beta}_2$  analytically. To make thing simpler, we switch to matrix-vector notation with the following definitions

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

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

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

And our estimate for  $\hat{\beta}$  is

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

Which transform into  $\hat{A}$  and  $\hat{\phi}$  using Equation 11. For  $\sigma$ , 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  $\sigma$  with normal errors.

In sum, if S, L, and F are known, we can estimate  $\hat{A}$ ,  $\hat{\phi}$ , and  $\hat{\sigma}$  analytically. But in practice, we don't know S, L, and F, and the next section covers how to estimate them using the Spectrogram.

## 5.2 Estimating S, L, and F

Section 5.1 shows that, if we *knew* S, L, and F, then we could analytically determine the maximum likelihood estimates of A,  $\phi$ , and  $\sigma$ . So the next question is, how can we estimate S, L, and F? To answer this, recall the quadratic term we minimized in the previous section (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 include 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):

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

This 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 Equation 21 as

$$\arg \max_{S, L, F} \begin{bmatrix} \mathbf{c}^T\mathbf{x} \\ \mathbf{s}^T\mathbf{x} \end{bmatrix}^T \begin{bmatrix} \mathbf{c}^T\mathbf{c} & \mathbf{c}^T\mathbf{s} \\ \mathbf{s}^T\mathbf{c} & \mathbf{s}^T\mathbf{s} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T\mathbf{x} \\ \mathbf{s}^T\mathbf{x} \end{bmatrix}$$

We can approximate the diagonal terms in the  $2 \times 2$  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:

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

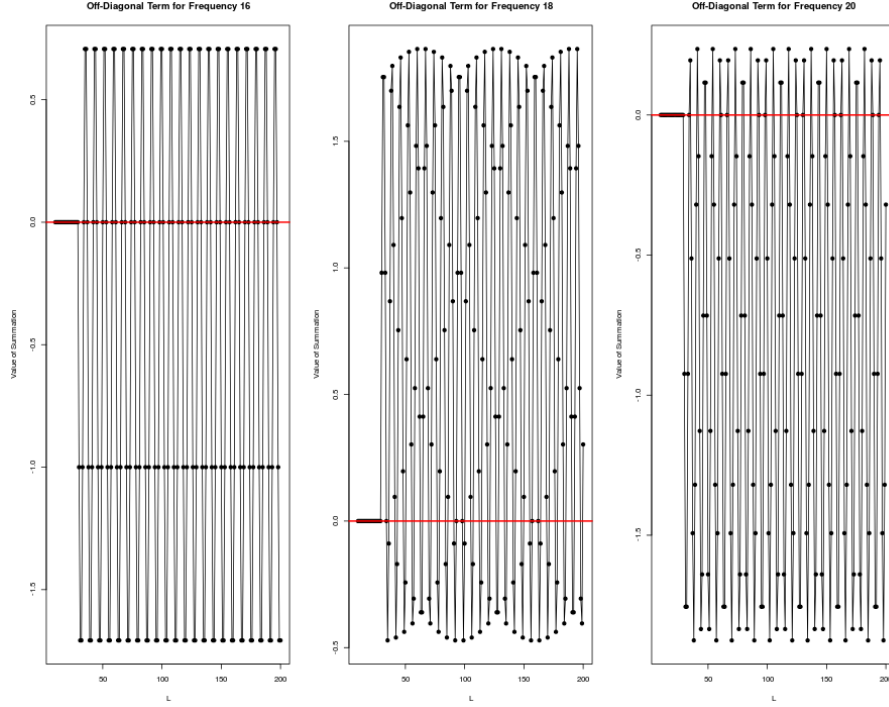


Figure 2: Off Diagonal Terms.  $N=256$ ,  $L=10-200$ ,  $S=30$ ,  $F=16,28,20$ .

This is actually the key difference between our problem and the example in Kay (1993). In Kay (1993), there is no indicator term inside the summation, and since this summation covers a sine function over a period, Equation 22 is  $\approx 0$ . In our problem, the off-diagonal terms only cancel out when  $L$  has a complete number of cycles (see Figure 2 for a demonstration), so we need to leave it in the expression.

For notational purposes, let  $z(N, F, S, L)$  be the off-diagonal term. To be concise, we call this  $z$  for the remainder of the section. Then we have

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

The inverse of the  $2 \times 2$  matrix is approximately

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

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). Another way to think about this is that  $\mathbf{c}^T \mathbf{x}$ , when

viewed as “the correlation between  $\mathbf{c}$  and  $\mathbf{x}$ ”, would be maximized when  $\mathbf{x} = \mathbf{c}$ , and similarly  $\mathbf{s}^T \mathbf{x}$  would be maximized when  $\mathbf{s} = \mathbf{x}$ . Then using Equation 22, we can bound the term by

$$\begin{aligned} \frac{8z}{L^2}(\mathbf{c}^T \mathbf{x})(\mathbf{s}^T \mathbf{x}) &= 8z\left(\frac{\mathbf{c}^T \mathbf{x}}{L}\right)\left(\frac{\mathbf{s}^T \mathbf{x}}{L}\right) \\ &\leq 2z \\ &\leq 4 \end{aligned} \tag{23}$$

In any case, I am arguing that the second term in Equation 23 is negligible, and we can focus on maximizing the first term. The key insight is that we can connect the first term to the Spectrogram. Notice that

$$\begin{aligned} \frac{2}{L}[(\mathbf{c}^T \mathbf{x})^2 + (\mathbf{s}^T \mathbf{x})^2] &= \frac{2}{L}\left[\left(\sum_{t=0}^{N-1} x_t \cos\left(\frac{2\pi Ft}{N}\right) \mathbb{1}_{S,S+L-1}(t)\right)^2 + \left(\sum_{t=0}^{N-1} x_t \sin\left(\frac{2\pi Ft}{N}\right) \mathbb{1}_{S,S+L-1}(t)\right)^2\right] \\ &= \frac{2}{L}\left|\sum_{t=0}^{N-1} x_t \omega_N^{-tF} \mathbb{1}_{S,S+L-1}(t)\right|^2 \\ &= \frac{2}{L}\left|\sum_{t=S}^{S+L-1} x_t \omega_N^{-tF}\right|^2 \end{aligned} \tag{24}$$

This is strikingly similar to the Periodogram (which is the MLE estimate in Kay (1993)), the only difference is that the summation indices are different. To see how this connects with the Spectrogram, recall the definition of the Spectrogram computed using a Sliding Window Discrete Fourier Transform (SWDFT) with length  $n$  windows

$$\begin{aligned} a_{k,p,n} &= \sum_{j=0}^{n-1} x_{p-n+1+j} \omega_n^{-jk} \\ k &= 0, 1, \dots, n-1 \\ p &= n-1, n, \dots, N-1 \end{aligned} \tag{25}$$

And if we let  $\frac{f}{L} = \frac{F}{N}$ , we can interpret Equation 24 as a SWDFT with length  $L$  windows

$$\arg \max_{S,L,F} \frac{2}{L} \left| \sum_{t=S}^{S+L-1} x_t \omega_L^{-tf} \right|^2 \tag{26}$$

Equation 26 tells us that our estimates of  $S$ ,  $L$ , and  $F$  will be the values that maximize the squared modulus of SWDFT coefficients across varying window sizes.

So, how would we make these estimates in practice? Notice that if we find the maximum squared modulus SWDFT coefficient across window sizes

$$(\hat{k}, \hat{p}, \hat{n}) = \arg \max_{k,p,n} |a_{k,p,n}|^2 \tag{27}$$

We can translate these into estimates of S, L, and F

$$\begin{aligned}\hat{S} &= \hat{p} - \hat{n} + 1 \\ \hat{L} &= \hat{n} \\ \hat{F} &= \frac{\hat{k}}{\hat{n}}\end{aligned}\tag{28}$$

Our overall estimation approach is summarized in Algorithm ??

---

**Algorithm 1:** MLE Estimation of Bursts Model

---

**input** :  $x$ , a length  $N$  signal  
-  $(\hat{k}, \hat{p}, \hat{n}) = \arg \max_{k,p,n} |a_{k,p,n}|^2$   
-  $(\hat{S}, \hat{L}, \hat{F}) = f(\hat{k}, \hat{p}, \hat{n})$   
- Analytic solutions for  $\hat{A}, \hat{\phi}, \hat{\sigma}$  using Section 5.1  
**output:** MLE estimates  $\hat{\theta} = [\hat{A}, \hat{\phi}, \hat{F}, \hat{S}, \hat{L}, \hat{\sigma}]$

---

The computationally intensive part of this approach is repeatedly computing the SWDFT of varying window sizes. We are currently looking into algorithms to speed this up. Finally, further numerical optimization may be performed around the  $\hat{F}$ , which allows us to search beyond the Fourier Frequencies.

## 6 Results

## 7 Discussion

## 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*.