

## Single-Snippet Analysis for Detection of Postspike Effects

**Sagi Perel**

*sagi@cmu.edu*

*Department of Biomedical Engineering and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.*

**Andrew B. Schwartz**

*abs21@pitt.edu*

*Department of Neurobiology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, and Center for Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15231, U.S.A.*

**Valérie Ventura**

*vventura@stat.cmu.edu*

*Department of Statistics and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15231, U.S.A.*

Corticomotoneuronal cells (CMN), located predominantly in the primary motor cortex, project directly to alpha motoneuronal pools in the spinal cord. The effects of CMN spikes on motoneuronal excitability are traditionally characterized by visualizing postspike effects (PSEs) in spike-triggered averages (SpTA; Fetz, Cheney, & German, 1976; Fetz & Cheney, 1980; McKiernan, Marcario, Karrer, & Cheney, 1998) of electromyography (EMG) data. Poliakov and Schieber (1998) suggested a formal test, the multiple-fragment analysis (MFA), to automatically detect PSEs. However, MFA's performance was not statistically validated, and it is unclear under what conditions it is valid. This paper's contributions are a power study that validates the MFA; an alternative test, the single-snippet analysis (SSA), which has the same functionality as MFA but is easier to calculate and has better power in small samples; a simple bootstrap simulation to estimate SpTA baselines with simulation bands that help visualize potential PSEs; and a bootstrap adjustment to the MFA and SSA to correct for nonlinear SpTA baselines.

### 1 Introduction ---

Most of the cortical output to muscles originates in the motor cortex (Fritsch & Hitzig, 1870/1960; Rathelot & Strick, 2009). Motoneuronal pools in the spinal cord combine inputs from cortical, subcortical, and spinal areas and are responsible for muscle contraction. The traditional method detecting

the effect of cortical spikes on spinal motoneuronal excitability is the spike-triggered average (SpTA; Fetz, Cheney, & German, 1976; Fetz & Cheney, 1980; McKiernan, Marcario, Karrer, & Cheney, 1998): short EMG segments (snippets) from a muscle of interest are collected around the times of cortical spikes, and their average is computed. If a characteristic waveform appears in the SpTA, it is taken as physiological evidence of connectivity between the neuron and muscle. The multiple-fragment analysis (MFA; Poliakov & Schieber, 1998) formalizes this visual inspection; it consists of measuring the magnitudes of waveforms that appear in the SpTA around 6 ms to 16 ms postspike and testing their statistical significance. The MFA requires preprocessing the data: the EMG snippets are grouped in fragments, and SpTAs are calculated in each of them. We show that these steps can be eliminated if we use a related but simpler procedure, the single-snippet analysis (SSA).

The MFA and SSA are parametric tests that rely on distributional assumptions, but their robustness to these assumptions was never studied. We conduct a large simulation to study the statistical properties of the MFA and SSA. We develop bootstrap procedures to estimate nonlinear SpTA baselines, which are known to affect the MFA (Davidson, Odell, Chan, & Schieber, 2007), and to adjust the tests for such baselines.

## 2 Methods

---

We first describe the multiple-fragment analysis and introduce new, related tests. We then explain how we estimate their statistical powers, so that their properties can be validated and their relative performances compared. Finally, sections 2.3 and 2.4 contain bootstrap procedures to check the tests' assumptions and estimate SpTA baselines.

**2.1 The Multiple-Fragment Analysis and Related Tests.** Given  $K$  cortical spike times and a simultaneously recorded electromyography (EMG) signal, we collect EMG snippets around every spike. We use their absolute values to minimize the cancellation caused by averaging overlapping positive and negative components of motor unit action potentials; rescale time so that  $t = 0$  is the time of the spike triggers; and compute the spike-triggered average (SpTA) by averaging these "rectified" snippets across time according to

$$SpTA(t) = K^{-1} \sum_{k=1}^K rEMG_k(t). \quad (2.1)$$

Figure 1 illustrates this. A large excursion of  $SpTA(t)$  from its baseline around 6 ms to 16 ms postspike suggests that a postspike effect (PSE) is present at that latency. The multiple-fragment analysis (MFA; Poliakov &

Schieber, 1998) serves to test the significance of such excursions: the  $K$ -rectified EMG snippets are divided into  $G$  groups (fragments), and a contrast between the excursion and the SpTA baseline in each fragment is computed as

$$X_g = \overline{SpTA}_g([6, 16] \text{ ms}) - \frac{1}{2}(\overline{SpTA}_g([-4, 6] \text{ ms}) + \overline{SpTA}_g([16, 26] \text{ ms})), \quad (2.2)$$

where  $SpTA_g(t)$  is the fragment SpTA in fragment  $g$  and  $\overline{SpTA}_g([a, b] \text{ ms})$  its average over time  $t \in [a, b]$ . Values of  $X_g$  that deviate from zero provide evidence of a PSE around 6 ms to 16 ms postspike. Poliakov and Schieber (1998) formalize this by calculating the test statistic

$$T_{MFA} = \bar{X} / \widehat{se}(\bar{X}), \quad (2.3)$$

where  $\bar{X}$  is the mean of the  $X$ 's and  $\widehat{se}(\bar{X})$  its estimated standard error, and comparing the deviations of  $T_{MFA}$  from zero to the standard normal distribution. ( $T_{MFA}$  is  $t$ -distributed, which is practically gaussian for the sample sizes considered here.) The MFA effectively consists of testing if the true mean of the  $X$ 's is significantly different from zero via a  $t$ -test. Note that we can extend this test trivially to detect excitatory or suppression PSEs—they appear in the SpTA as bumps in Figures 2C, 3C, and 3F and troughs in Figure 3I, respectively—with a one-sided rather than two-sided  $t$ -test.

Poliakov and Schieber (1998) fragment the data to be able to estimate the variance of  $\bar{X}$  in equation 2.3, which they do according to  $\widehat{se}^2(\bar{X}) = S_X^2/G$ , where  $S_X^2 = \sum_g (X_g - \bar{X})^2 / (G - 1)$  is the sample variance of the  $X$ 's. This calculation assumes that the  $X$ 's are independent, which we discuss below. They use  $G = \sqrt{K}$  fragments to strike a good balance between having enough EMG snippets per fragment to calculate  $SpTA_g$  in equation 2.2 and enough  $X$ 's to estimate  $\widehat{se}^2(\bar{X})$ . They form the fragments by dividing the experimental time into  $G$  periods of equal length and assigning to fragment  $g$  all the snippets in period  $g$ . Davidson et al. (2007) note that some fragment SpTAs are noisier than others because the fragments have different sizes, so they recommend forming  $G = \sqrt{K}$  fragments with equal number of snippets by grouping every  $n = \sqrt{K}$  consecutive snippets. We refer to this method as MFAE, yielding test statistic  $T_{MFAE}$  in equation 2.3, where the "E" stands for "equal number of snippets." Calculating  $T_{MFAE}$  is easier than calculating  $T_{MFA}$ : the latter requires the times of the spike triggers to form the fragments; the former requires only their ordering.

But fragmentation is not needed. Indeed, consider a contrast similar to equation 2.2, calculated for each EMG snippet  $k$ :

$$Y_k = \overline{rEMG_k}([6, 16] \text{ ms}) - \frac{1}{2}(\overline{rEMG_k}([-4, 6] \text{ ms}) + \overline{rEMG_k}([16, 26] \text{ ms})). \quad (2.4)$$

Then  $X_g$  in equation 2.2 is the average of the  $Y$ 's in fragment  $g$ . Furthermore,  $\bar{X} = \bar{Y}$  when the fragments have equal sizes: there is no need to fragment the data to calculate the numerator of  $T_{MFAE}$ . Fragmentation is not needed either to estimate  $se(\bar{X})$  in its denominator, since  $\bar{X} = \bar{Y}$  implies  $se(\bar{X}) = se(\bar{Y})$ : when the  $Y$ 's are independent,  $\widehat{se}^2(\bar{Y}) = S_Y^2/(K-1)$  is an unbiased estimate of  $se^2(\bar{Y})$ , and therefore of  $se^2(\bar{X})$ , where  $S_Y^2$  is the sample variance of the  $Y$ 's. But we must be cautious: if the interspike interval (ISI) between two consecutive spike triggers is less than 30 ms, the corresponding EMG snippets used to calculate the  $Y$ 's in equation 2.4 overlap, so the  $Y$ 's are dependent; if three consecutive ISIs are less than 30 ms, then the three corresponding  $Y$ 's are mutually dependent; and so on. In that case, the variance estimate  $\widehat{se}^2(\bar{Y}) = S_Y^2/(K-1)$  suggested earlier may be biased, and if that estimate is used in the denominator of  $T_{MFAE}$ , then  $T_{MFAE}$  may not be standard normal under the null hypothesis of no PSE. It is therefore important to obtain an unbiased estimate of  $se(\bar{Y})$ , as discussed below.

*2.1.1 Blocking the Data to Attenuate Serial Correlations.* Blocking consists of forming  $G = K/n$  blocks of  $n$  consecutive  $Y$ 's, calculating the block means,  $Z_g$ , their sample variance,  $S_{Z_g}^2$ , and the sample variance of  $\bar{Z}$  according to  $\widehat{se}^2(\bar{Z}) = S_Z^2/G$ . Then because  $\bar{Z} = \bar{Y}$ ,  $\widehat{se}^2(\bar{Z})$  will be unbiased for  $se^2(\bar{Z}) = se^2(\bar{Y})$ , provided the  $Z$  are independent. Moreover, the  $Z$ 's should be independent when  $n$  is large because the  $Y$ 's are only serially correlated. We do not expect other types of correlations: the EMG signal is the sum of the activities of many motor units, so the noise should be random, and the underlying EMG signal was subtracted from the  $Y$ 's, since they are contrasts, so there should be no signal correlation. We refer to the blocking approach as the fixed-fragment analysis (FFA), with test statistic

$$T_{FFA} = \bar{Z}/\widehat{se}(\bar{Z}).$$

Note that MFAE is a special case of blocking with  $n = \sqrt{K}$ , although the original motivation to fragment or block was not to reduce serial correlations.

To select a block size  $n$  that yields independent  $Z$ 's, we plot  $\widehat{se}^2(\bar{Z}) = S_Z^2/G$  versus  $n$  (see Figures 4A to 4C), we identify  $n_0$  such that  $\widehat{se}^2(\bar{Z})$  is approximately constant for  $n \geq n_0$ , and we use  $n = n_0$  as the block size. Indeed,

letting  $n_0$  be such that the corresponding block means  $Z$  are independent, then the means in larger blocks will also be independent since the correlations are serial; hence,  $\widehat{se}^2(\bar{Z}) = S_Z^2/G$  should be a constant equal to the true variance of  $\bar{Y}$ , within random error, for all  $n \geq n_0$ . Note that we do not have to pick  $n_0$  very carefully. Indeed, assuming that the  $Z$ 's are independent and identically distributed (i.i.d.),  $T_{FFA}$  has a Student- $t$  distribution, which is practically normal when  $G \geq 30$ . Then the FFA test should be invariant to  $n$ , provided  $n \geq n_0$  and  $n \leq K/G$  with  $G = 30$ . In our data, we observed that FFA was not very sensitive to  $n$  in that range—in particular, the MFAE and FFA often performed comparably (see Figure 3)—although  $n$  closer to  $K/30$  often reduced its power.

*2.1.2 Estimating the Serial Correlations.* Alternatively, we can estimate  $se^2(\bar{Y})$  directly. We start with the variance formula for an average:

$$se^2(\bar{Y}) = K^{-2} \left( \sum_{k=1}^K se^2(Y_k) + 2 \sum_{k=1}^K \sum_{l=k+1}^K Cov(Y_k, Y_l) \right),$$

and estimate the terms on the right-hand side by their sample counterparts,

$$\widehat{se}^2(\bar{Y}) = K^{-1} \left( S_Y^2 + 2 \sum_{l=1}^L AC(l) \right), \quad (2.5)$$

where  $AC(l) = \sum_{k=1}^{K-l} [(Y_k - \bar{Y})(Y_{k+l} - \bar{Y})] / (K - l)$  is the autocovariance of the  $Y$ 's at lag  $l$ ,  $S_Y^2 = AC(0)$  is their sample variance, and  $L$  is the number of successive significant autocovariances that can be determined from an autocorrelation plot (see Figures 4D–4F). Note that  $AC(l)$  is easily interpretable for time series (i.e., data recorded at regular intervals). Here the  $Y$ 's are recorded at random cortical spike times, so  $AC(l)$  estimates the average autocovariance between the  $Y$ 's of spikes that are  $l$  random ISI apart.

We refer to the direct variance estimation approach as the single-snippet analysis (SSA), with test statistic

$$T_{SSA} = \bar{Y} / \widehat{se}(\bar{Y}), \quad (2.6)$$

with  $\widehat{se}^2(\bar{Y})$  in equation 2.5. If the firing rate of the cortical neuron is such that no ISI is shorter than 30 ms, then equation 2.5 reduces to  $se^2(\bar{Y}) = S_Y^2/K$ , and SSA reduces to FFA with block size  $n = 1$ . The data sets analyzed here contain many ISIs shorter than 30 ms, and we included up to  $L = 4$  significant covariance terms. Note that  $L$  does not have to be chosen very carefully because the autocovariances  $AC(l)$  in equation 2.5 are divided by

a typically large  $K$ , so they contribute little to  $\widehat{se}^2(\bar{Y})$  when they are not significant.

**2.1.3 Summary.** The MFA and MFAE are designed to detect PSEs around 6 ms to 16 ms postspike. They require fragmenting the data and calculating fragment SpTAs. The default fragment size,  $n = \sqrt{K}$ , was not chosen to achieve a particular criterion. We proposed two related tests: the FFA, which also requires fragmenting the data, but with block size chosen to achieve independence between the block means. The SSA requires no fragmentation at all. Ease of calculation aside, the only difference between the MFAE, FFA, and SSA tests is how the true variance of  $\bar{Y}$  is estimated: the former two use blocking, the latter a direct estimate. It is easy to show that the three variance estimates are equal on expectation when the  $Y$ 's are i.i.d. However, the EMG snippets are typically very heterogeneous, so the  $Y$ 's are unlikely to be i.i.d., and the tests are unlikely to be exactly equivalent. We describe how to compare their performances in the next section.

**2.2 Statistical Properties of Tests.** A desirable test has high power (i.e., high probability of detecting existing effects) and a spurious detections rate that matches the significance level  $\alpha$ , so that if a PSE is detected, we know that it is spurious with probability less than  $\alpha$ . To compare the tests, we compare their powers and spurious detection rates. Power is typically intractable, so we estimate it by the proportion of data sets in which PSEs are detected. Because the number of data sets provided by real experiments is too small to estimate proportions accurately, we construct a large pool of realistic test data sets by subsampling experimental “parent” data sets.

**2.2.1 Experimental Parent Data Sets.** Our parent data sets come from three experiments. The digit-flexion data set was collected in Marc Schieber’s lab. It consists of simultaneously recorded spike trains from two neurons and EMG activity from nine muscles while a monkey performed visually cued individuated flexion and extension movements of the right fingers or wrist (Schieber, 2005). The grasp data set, collected in our lab, consists of spike trains from 171 neurons and EMG activity from 16 muscles, as a monkey performed reach-to-grasp movements to a variety of objects at different spatial locations and orientations. The precision-grip data set was collected in Roger Lemon’s lab and consists of spike trains from nine neurons and EMG activity from seven muscles, recorded while a monkey squeezed two spring-loaded levels between the thumb and index fingers (Jackson, Gee, Baker, & Lemon, 2003).

**2.2.2 Test Data Sets.** To create a test data set of size  $K$  that contains a PSE, we select a parent data set whose SpTA clearly displays a PSE—Figure 2C

or Figures 3C, 3F, and 3I). We string together  $K/100$  blocks of 100 spikes, sampled at random from the parent data set, and extract the  $K$ -corresponding EMG snippets. Because the parent data set contains a PSE, the test data set contains a PSE of similar size, although it is harder to detect when  $K$  is small. To reduce the size of the PSE in a test data set of size  $K$ , we destroy the time-locked effects in a randomly selected subblock of size  $K \cdot (100 - a)\%$ , where  $a \in [0, 100]\%$  codes for effect size:  $a = 0\%$  means that the test data set contains no time-locked effect; when  $a = 100\%$ , the effect size is the same as the parent data set. To destroy time-locked effects, we leave the EMG trace unchanged and either add a large random jitter (SD 100 ms) to the spikes in the subblock or reflect the entire spike train in the subblock (its end becomes its start, and vice versa) to preserve the serial correlations in the  $Y$ 's we would expect to see in real data sets. Our results did not depend on which method we used.

*2.2.3 Estimating Power Curves.* The power of a test increases with effect size and sample size. We obtain power curves as functions of sample size by subsampling over 2000 test data sets of size  $K$  from a parent data set, applying all detection tests to the test data sets, and estimating their powers by the respective proportions of times they each detect PSEs. We repeat for other values of  $K$  and plot the detection proportions of each test against  $K$ . To obtain power curves as functions of effect size, we fix  $K$  around half the number of spike triggers in the parent, and, given a fixed  $a \in [0, 100]\%$ , create over 2000 test data sets of size  $K$  with effect size  $a$ . We apply all detection tests to the test data sets and estimate their powers by the respective detection proportions. We repeat this for several values of  $a \in [0, 100]\%$  and plot the detection proportions against  $a$ . (See Figure 3.)

**2.3 Assumptions of the Tests.** Let  $T$  refer to the test statistic of any detection test, and  $T_{MFA}$ ,  $T_{MFAE}$ ,  $T_{FFA}$ , and  $T_{SSA}$  refer to specific tests. All tests rely on the assumption that  $T$  is standard normal under the null hypothesis of no PSE. There is no reason to question either the normality assumption, which is justified by the central limit theorem in the large samples we have in typical experiments, or the unit variance of  $T_{FFA}$  and  $T_{SSA}$ , since we take great care to estimate the variance of  $\bar{Y}$  used in their denominators. The variances of  $T_{MFA}$  and  $T_{MFAE}$  might deviate from one in samples small enough to make the default block size  $n = \sqrt{K}$  too small to attenuate the serial correlations in the  $Y$ 's. However,  $K$  is rarely less than 1000, and  $n = \sqrt{1000} = 31$  was large enough for all the data sets investigated here. The only contentious assumption is that the null mean of  $T$  is zero; indeed, it relies on the SpTA baseline being linear in  $t$  (Davidson et al., 2007), which is often violated to some degree (e.g. Figures 3C and 3I). Questionable assumptions can bias  $p$ -values and in turn bias the spurious detections rates of the tests. Biased  $p$ -values that are much larger or much smaller than the significance level  $\alpha$

are of no concern since an adjustment will not change the outcome of the test (i.e., respectively retain or reject the null hypothesis of no PSE). But we adjust the tests when  $p$ -values are within few percentage points of  $\alpha$  using the following bootstrap simulation:

1. Jitter the observed spike triggers using a normal jitter with SD 30 ms and extract the EMG snippets corresponding to the jittered triggers; this is the  $r$ th bootstrap sample. See the appendix for a justification.
2. Calculate  $N_r^*$ , the numerator of  $T$  in bootstrap sample  $r$ , that is, calculate  $N_r^* = \bar{Y}_r^*$  for the MFAE, FFA, and SSA tests. The MFA test requires fragmenting the bootstrap sample.
3. Repeat steps 1 and 2 for  $r = 1, \dots, R$ , with  $R = 100$ .
4. Calculate  $m = R^{-1} \sum_{r=1}^R \bar{N}_r^*$ , the sample mean of the  $N^*$ 's.
5. Subtract  $m$  in the denominator of  $T$  (e.g.,  $T_{SSA} = (\bar{Y} - m) / \widehat{se}(\bar{Y})$ ).

To understand the adjustment to  $T$ , note that  $\bar{Y}$  is effectively the difference between the average SpTA 6 ms to 16 ms postspike, where a PSE might appear, and the SpTA baseline, estimated by the average SpTA in the two adjacent time windows. Hence,  $\bar{Y} = 0$ , within random error, when the SpTA is linear, and large deviations of  $\bar{Y}$  from zero provide evidence of a PSE. But if the SpTA baseline is not linear,  $\bar{Y}$  should be compared not to zero but to the expected value of  $\bar{Y}$  calculated from the baseline; this value is  $m$ . We therefore standardize the tests by subtracting  $m$  in the denominator of  $T$ .

**2.4 Bootstrap Estimates of SpTA Baselines.** The detection tests presented so far are automatic—they do not require visual inspections of SpTAs—and formal—they yield  $p$ -values. But PSE detection is often initially performed by assessing visually the deviations of an SpTA from its baseline. The simplest and most frequently used baseline estimate is a line fitted to the SpTA (Bennett & Lemon, 1994). The method of Davidson et al. (2007) applies more generally to estimate baselines that are nonlinear. We propose an alternative bootstrap SpTA baseline estimate, which has the advantage of providing simulation bands that help assess the significance of SpTA excursions from their baselines (see Figures 2C, 3C, 3F, and 3I). We proceed as follows:

1. Jitter the observed spike triggers (normal jitter with SD 30 ms, as justified in the appendix). This is the same step 1 in the bootstrap simulation of section 2.3.
2. Calculate the average of the rectified EMG snippets of the jittered triggers; this is a null bootstrap SpTA.
3. Repeat steps 1 and 2  $R = 100$  times, and calculate the sample mean of the  $R = 100$  null bootstrap SpTAs (this is a bootstrap SpTA baseline) and their sample pointwise variance.

Pointwise 95% simulation bands for null SpTAs are drawn around the bootstrap baseline estimate at  $\pm 2$  pointwise standard deviations. The

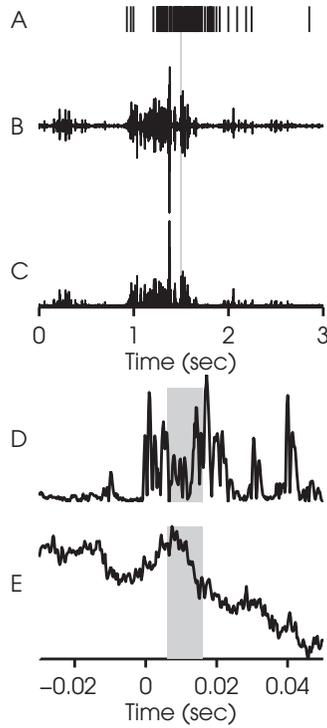


Figure 1: A typical data set. (A, B) Portion of simultaneously recorded cortical spike train and EMG trace from a precision-grip data set. (C) Rectified EMG trace. (D) Rectified EMG snippet of the spike marked by the vertical gray line in panels A–C. (E) SpTA for this neuron-EMG pair. The gray bands in panels D and E mark the 6 ms to 16 ms postspike trigger period.

observed SpTA exiting the bands suggests that there may be a PSE, although this test is informal because the bands have 95% pointwise rather than joint coverage. The bootstrap could also be used to simulate approximate joint simulation bands (Ventura, 2010), but would require a simulation size considerably larger than  $R = 100$ .

### 3 Results

---

Figure 1 illustrates the type of data analyzed in this paper. Panels A to C show a portion of simultaneously recorded cortical spike train and EMG trace and the corresponding rectified EMG trace. Figure 1D displays one rectified EMG snippet; time was rescaled so that  $t = 0$  corresponds to the time of the spike trigger. Figure 1E shows the SpTA for this neuron-EMG

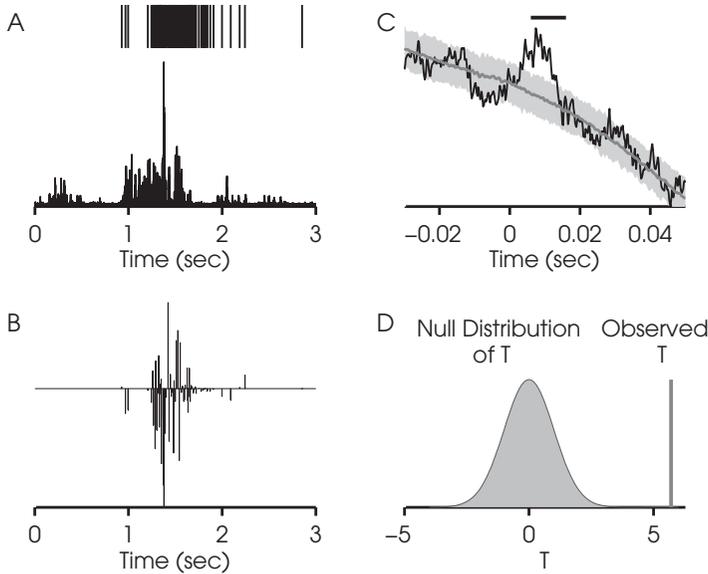


Figure 2: PSE detection methods. (A) Simultaneously recorded spike train and rectified EMG trace (see Figures 1A and 1C). (B) The  $Y$ 's, equation 2.4, corresponding to the EMG snippets extracted from panel A. (C) Same SpTA as in Figure 1E, with bootstrap baseline estimate and 95% pointwise simulation bands (gray) (see section 2.4). (D) Observed value of  $T_{SSA}$  in this data set (right) and standard normal null distribution of  $T_{SSA}$  (left). The corresponding  $p$ -value is  $<10^{-4}$ , which confirms the PSE we clearly see in panel C.

pair, that is, the average of all rectified EMG snippets (see equation 2.1). Averaging reduced the variability, and we can detect a small bump around the gray band at 6 ms to 16 ms postspike.

Figure 2 illustrates the PSE detection methods described in this paper. Figure 2A is the portion of spike train and rectified EMG trace shown in Figure 1AC. Figure 2C shows the same SpTA as in Figure 1E, on which we overlaid a bootstrap baseline estimate and 95% pointwise bootstrap simulation bands in gray (see section 2.4). The SpTA clearly exceeds the bands, which provides evidence of a PSE around 6 ms to 16 ms. The MFA, MFAE, FFA, and SSA tests measure this evidence with a  $p$ -value. Figure 2B shows the contrasts  $Y$  (see equation 2.4) for each of the EMG snippets extracted from Figure 2A, which are used to calculate the test statistics  $T$ . The observed value of  $T_{SSA}$  for that data set is plotted in Figure 2D, along with its null distribution: the observed  $T_{SSA}$  is extreme, with  $p$ -value less than  $10^{-4}$ . We reject the null hypothesis of no PSE at the 5% significance level.

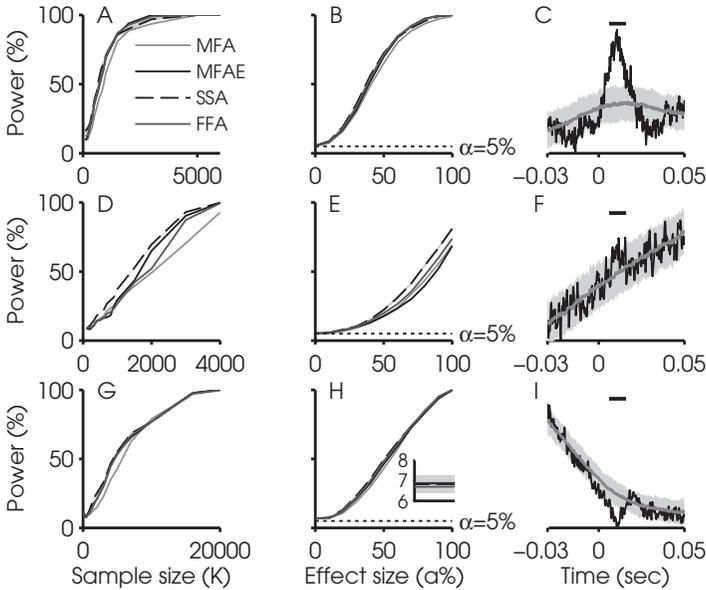


Figure 3: Power curves based on three typical experimental parent data sets (see section 2.2), whose SpTAs are in panels C, F, and I, with overlaid bootstrap baselines and 95% pointwise simulation bands (see section 2.4). (C) Strong facilitation PSE in a digit-flexion data set. (F) Weak facilitation PSE in a grasp data set. (I) Moderate suppression PSE in a precision-grip data set. (A, D, G) Power as a function of sample size. (B, E, H) Power as a function of effect size; at  $a = 0\%$ , there is no PSE; at  $a = 100\%$ , the effect size is the same as in the parent data set. All tests are approximately comparable, but MFA often has lower power, and SSA tends to have better power in smaller samples. (B, E) The spurious detection rates at  $a = 0\%$  match the significance level  $\alpha$ . (H) The spurious detections rate exceed  $\alpha$  (see inset) because the SpTA baseline is not linear; applying the bootstrap adjustment in section 2.3 systematically rectifies all such discrepancies.

Figure 3 shows the power curves of the tests applied to subsamples from the three parent data sets whose SpTAs are in Figures 3C, 3F, and 3I as functions of sample size  $K$  and effect size  $a$  (see section 2.2). The powers of all tests increase with  $K$  and  $a$ , as expected. The tests have comparable powers, with no test being uniformly more powerful. However, the MFA test is often the least powerful and the SSA test more powerful in smaller samples, as is seen most clearly in Figure 3D. These observations are also true in all other data sets we analyzed (not shown). We therefore recommend against the MFA test and favor the SSA test; the latter is also easiest to calculate. Next, Figures 3B, 3E, and 3H at  $a = 0\%$  show the spurious

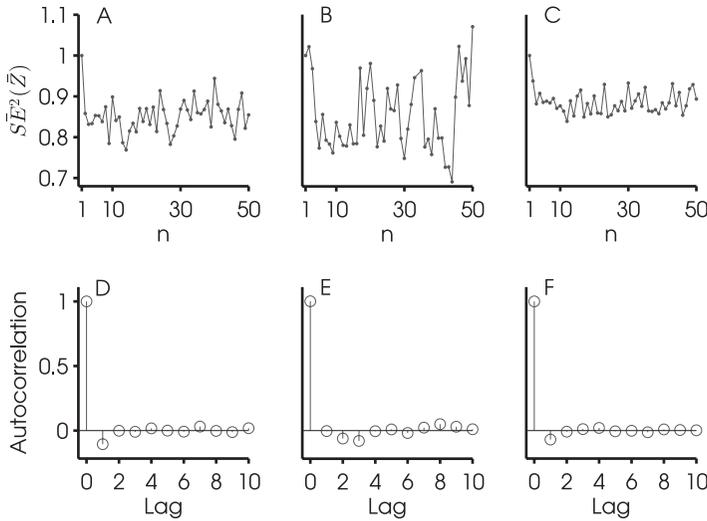


Figure 4: Variance and autocorrelation functions used to determine the parameters of the FFA and SSA tests. (A, B, C) Estimated variance of  $\bar{Z}$ ,  $\widehat{se}^2(\bar{Z})$ , as a function of block size  $n$ , scaled by  $\widehat{se}^2(\bar{Z})$  when  $n = 1$  to standardize the plot size. The curves first decrease, then stabilize, within random error, as  $n$  increases. This indicates that the EMG snippet contrasts  $Y$  are serially correlated. We pick the block size  $n$  after which the curves are stable. (D, E, F) Autocorrelations of the  $Y$ 's,  $AC(l)/AC(0)$ , for lags  $l = 0$  to 10, which measures the serial correlations in the  $Y$ 's. We include the significant terms in equation 2.5 to obtain an unbiased estimate of the variance of  $\bar{Y}$ .

detections rates of the tests—their powers when there is no effect in the data. They match the significance level  $\alpha = 5\%$  in Figures 3B and 3E, which suggests that the assumptions on which the tests rely are met. However, the spurious detection rate is significantly larger than  $\alpha$  in Figure 3H, although at about 7%, the discrepancy is not alarming. This discrepancy is due to the markedly nonlinear SpTA baseline in Figure 3I; when we reran the power simulation including the bootstrap baseline adjustment described in section 2.3, the spurious detections rates lowered to  $\alpha$  (not shown). We observed the same across all data sets (not shown). Our simulation study shows that the parametric assumptions of the tests are met except when the SpTA baseline is not linear and that performing the tests systematically with the bootstrap adjustment described in section 2.3 rectifies the problem.

Figure 4 shows the variance and autocorrelation functions used to determine the parameters of the FFA and SSA (see section 2.1). The MFA and MFAE use the default block size  $n = \sqrt{K}$ , where  $K$  is the number of spike triggers. We choose the block size for the FFA by determining when the

variance plot becomes constant up to random error: it is very approximately  $n = 20$  in Figures 4A and 4C and  $n = 10$  in Figure 4B. Recall from section 2.1 that  $n$  need not be determined very accurately. This is fortunate because the variance plots are hard to read; they are highly variable, especially in small samples (see Figure 4B), and have strong serial correlations (the errors are clearly not random). A block size  $n = 20$  seemed adequate for all data sets analyzed in this paper (not shown) and could be used as a default. For the SSA test, we determine the number  $L$  of autocovariance terms to include in equation 2.5 using the autocorrelation plots in Figures 4D, 4E, and 4F.  $L = 1$  seems adequate in Figures 4D and 4F and  $L = 3$  in Figure 4E. Recall from section 2.1 that the SSA test is not sensitive to including few more nonsignificant autocovariance terms, so we could also use  $L = 4$  or 5 as a default.

*3.1.1 Thresholding EMG Snippets.* It is common practice to discard low-valued EMG snippets before computing the SpTA. To determine if this has an effect on the tests, we repeated the power simulation using only EMG snippets above a threshold. We used two thresholds: a constant equal to the mean of an EMG segment where the muscle was inactive (McKiernan et al., 1998) and a threshold based on the root-mean-square EMG value in an EMG segment where the muscle was inactive (Davidson et al., 2007). We found that neither method had any effect on the powers of the MFAE, FFA, and SSA tests in any data set investigated here. We did not rerun the simulation for the MFA test, because it is unclear how fragments should be formed after thresholding the EMG snippets.

## 4 Discussion

---

The procedures currently available to detect functional connections between cortical neurons and spinal motoneurons are visual inspections of SpTAs, subtracting a linear baseline estimate from the SpTA and identifying excursions greater than two standard deviations from the pretrigger baseline (Bennett & Lemon, 1994), and performing the multiple-fragment analysis (MFA, MFAE), which are automatic tests that yield  $p$ -values.

Our first contribution is the single-snippet analysis (SSA), an automatic test that has the same functionality as MFA/MFAE. The MFAE and SSA are effectively the same  $t$ -test applied to serially correlated observations. They differ in their treatment of these correlations: MFAE attenuates them by blocking the data using a default block size (although this was not the original intention), and SSA estimates them directly. The MFAE is a special case of the fixed-fragment analysis (FFA), which is also based on blocking, but for which the block size is chosen specifically to attenuate the serial correlations. The MFAE, FFA, and SSA tests are conceptually equivalent, but SSA is easier to calculate because it requires no fragmentation of the data.

Our second contribution is a power study, which shows that the MFA, MFAE, FFA, and SSA tests can be applied with confidence. Indeed, their spurious detection rates match the significance level, at least when the SpTA baseline is linear. When the baseline is not linear, we observed discrepancies smaller than 2% between a nominal 5% significance level and the observed spurious detection rates. Although these discrepancies are not very concerning, we proposed a simple bootstrap adjustment that eliminates them. The power study also revealed that the tests have comparable powers, although MFA is often the least powerful and SSA the most powerful in small samples. Finally, we investigated the effect of thresholding the EMG snippets on the powers of the tests and found none. This common preprocessing step could therefore be eliminated.

Our last contribution is a simple bootstrap simulation that provides not only an estimate of the SpTA baseline, but also simulation bands that can be used to assess informally the significance of SpTA deviations from their baselines. In particular, PSEs that might appear to be small because many EMG snippets have low values (e.g., if we do not threshold the snippets) would still exceed the simulation bands.

We consider two directions for further development. The tests are designed to detect PSEs around 6 ms to 16 ms postspike, and they use a fixed 10 ms detection window that may not be optimal to detect narrow or wide PSEs. We plan to extend them so they can detect PSEs at any postspike latency and to design a test that adapts automatically to PSE width.

## Appendix: Justifying the Bootstrap

---

The bootstrap methods in sections 2.3 and 2.4 require bootstrap samples that satisfy  $H_0$ , the null hypothesis of no time-locked PSE. There are many ways to create data that satisfy  $H_0$ , and the bootstrap baseline estimate and adjustment  $m$  will depend on which way we use (Ventura, 2010).

One option to destroy time-locked effects is to keep the EMG signal untouched and randomize the interspike intervals (ISIs) of the cortical neuron spike train. But the resulting EMG snippets might not resemble the snippets we would observe if  $H_0$  was true. In our data, the EMG signal tends to be larger when the neuron's firing rate is high, so that many of the EMG snippets have comparatively large values. Randomizing the ISIs yields data that contain a smaller proportion of large-valued EMG snippets. Figure 5 shows the observed SpTA for a digit-flexion data set and the SpTA obtained after randomizing the ISIs: the PSE has disappeared, but the pre- and post-randomization SpTAs are dramatically different.

An alternative is to jitter the spikes, which preserves approximately the firing rate of the cortical neuron (Harrison & Geman, 2009) while spreading any potential time-locked effect. Figure 6A shows the same observed SpTA as Figure 5, and SpTAs obtained after adding normal jitters with mean 0 and

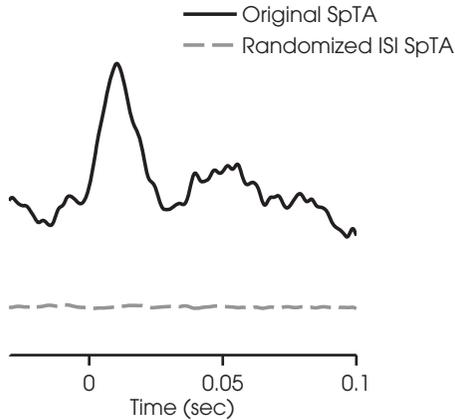


Figure 5: SpTA from a digit-flexion data set overlaid with the SpTA obtained after randomizing the cortical spikes interspike intervals (ISIs). Randomizing ISIs removes the PSE but also changes the SpTA outside the 6 ms to 16 ms window.

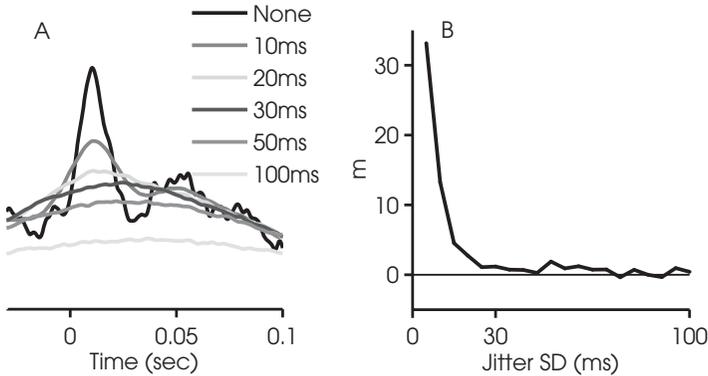


Figure 6: (A) SpTA from a digit-flexion dataset overlaid with SpTAs obtained after jittering the cortical spikes times with values drawn from a normal distribution with mean zero and SD  $\sigma$ ;  $\sigma \geq 30$  ms destroys the PSE. (B) Bootstrap mean adjustment  $m = \tilde{Y}^*$  (see section 2.3) as a function of  $\sigma$ ;  $m$  decreases from  $\tilde{Y}$  at  $\sigma = 0$  and stabilizes when  $\sigma \geq 30$  ms.

SD  $\sigma$  to the spike triggers:  $\sigma \geq 30$  ms appears to destroy the time-locked effect completely. Figure 6B shows the bootstrap adjustment  $m$  (see section 2.3) as a function of  $\sigma$ . At  $\sigma = 0$  (no jitter),  $m$  equals  $\tilde{Y}$  in the numerators of  $T_{MFAE}$ ,  $T_{FFA}$ , and  $T_{SSA}$ , which measures the strength of the evidence for a PSE. As  $\sigma$  increases,  $m$  gradually decreases to the expected value of  $\tilde{Y}$  under  $H_0$ . In Figure 6B,  $m$  stabilizes somewhat above zero when  $\sigma \geq 30$  ms; if the

SpTA baseline was linear,  $m$  would stabilize at zero. Figure 6 suggests that any  $\sigma \geq 30$  ms yields bootstrap samples that satisfy  $H_0$ . However, because the baseline flattens further as the jitter increases, we use  $\sigma = 30$  ms, the smallest  $\sigma$  that destroys the effect, so that the bootstrap samples are as close as possible to the observed data. We suggest reproducing Figure 6 for new data sets, since they might have different characteristics from ours. Note that jittering preserves approximately the mean firing rate of the cortical neuron but adds noise to its ISIs, so the bootstrap  $Y^*$ 's might not have the same serial correlations as the observed  $Y$ 's (see equation 2.4). Therefore the  $Y^*$ 's can be used to correct the mean of the null distribution of  $\bar{Y}$  using  $m$ , as described in section 2.4, but they cannot be used to correct its variance.

### Acknowledgments

---

We thank Marc Schieber, Roger Lemon, and Andrew Jackson for their contribution of data sets. Schieber's work was supported by grant NS27686. Our work was supported by grant NS050256 (S. P. and A. S.), and NIH grant 2R01MH064537 (V. V.). This project also acknowledges the use of the Cornell Center for Advanced Computing's Matlab on the TeraGrid experimental computing resource funded by NSF grant 0844032 in partnership with Purdue University, Dell, MathWorks, and Microsoft.

### References

---

- Bennett, K. M., & Lemon, R. N. (1994). The influence of single monkey corticomotoneuronal cells at different levels of activity in target muscles. *J. Physiol.*, 477 (Pt 2), 291–307.
- Davidson, A., Odell, R., Chan, V., & Schieber, M. (2007). Comparing effects in spike-triggered averages of rectified EMG across different behaviors. *Journal of Neuroscience Methods*, 163, 283–294.
- Fetz, E. E., & Cheney, P. D. (1980). Postspike facilitation of forelimb muscle activity by primate corticomotoneuronal cells. *J. Neurophysiol.*, 44(4), 751–772.
- Fetz, E. E., Cheney, P. D., & German, D. C. (1976). Corticomotoneuronal connections of precentral cells detected by postspike averages of EMG activity in behaving monkeys. *Brain Res.*, 114(3), 505–510.
- Fritsch, G., & Hitzig, E. (1870/1960). On the electrical excitability of the cerebrum. In G. van Bomin (Ed. and Trans.), *Some papers on the cerebral cortex* (pp. 73–96). Springfield, IL: Charles C. Thomas.
- Harrison, M. T., & Geman, S. (2009). A rate and history-preserving resampling algorithm for neural spike trains. *Neural Comput.*, 21(5), 1244–1258.
- Jackson, A., Gee, V. J., Baker, S. N., & Lemon, R. N. (2003). Synchrony between neurons with similar muscle fields in monkey motor cortex. *Neuron*, 38(1), 115–125.
- McKiernan, B. J., Marcario, J. K., Karrer, J. H., & Cheney, P. D. (1998). Corticomotoneuronal postspike effects in shoulder, elbow, wrist, digit, and intrinsic hand muscles during a reach and prehension task. *J. Neurophysiol.*, 80(4), 1961–1980.

- Poliakov, A. V., & Schieber, M. H. (1998). Multiple fragment statistical analysis of postspike effects in spike-triggered averages of rectified EMG. *J. Neurosci. Methods*, *79*(2), 143–150.
- Rathelot, J. A., & Strick, P. L. (2009). Subdivisions of primary motor cortex based on cortico-motoneuronal cells. *Proceedings of the National Academy of Sciences*, *106*(3), 918–923.
- Schieber, M. H. (2005). A spectrum from pure post-spike effects to synchrony effects in spike-triggered averages of electromyographic activity during skilled finger movements. *Journal of Neurophysiology*, *94*(5), 3325–3341.
- Ventura, V. (2010). *Bootstrap tests of hypotheses*. New York: Springer.

---

Received January 22, 2013; accepted June 29, 2013.