

Automatic scan test for detection of functional connectivity between cortex and muscles

Sagi Perel, Andrew B. Schwartz and Valérie Ventura

J Neurophysiol 112:490-499, 2014. First published 23 April 2014; doi:10.1152/jn.00800.2011

You might find this additional info useful...

This article cites 18 articles, 10 of which can be accessed free at:

</content/112/2/490.full.html#ref-list-1>

Updated information and services including high resolution figures, can be found at:

</content/112/2/490.full.html>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

This information is current as of October 2, 2014.

Automatic scan test for detection of functional connectivity between cortex and muscles

Sagi Perel,^{1,2} Andrew B. Schwartz,^{2,3} and Valérie Ventura^{2,4}

¹Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania; ²Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania; ³Department of Neurobiology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania; and ⁴Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania

Submitted 30 August 2011; accepted in final form 16 April 2014

Perel S, Schwartz AB, Ventura V. Automatic scan test for detection of functional connectivity between cortex and muscles. *J Neurophysiol* 112: 490–499, 2014. First published April 23, 2014; doi:10.1152/jn.00800.2011.—Postspike effects (PSEs) in averages of spike-triggered EMG snippets provide physiological evidence of connectivity between CMN cells and spinal motoneurons innervating skeletal muscles. They are typically detected by visual inspection of spike-triggered averages (SpTAs) or by multiple-fragment/single-snippet analyses [MFA (Poliakov AV, Schieber MH. *J Neurosci Methods* 79: 143–150, 1998) and SSA (Perel S, Schwartz AB, Ventura V. *Neural Comput* 26: 40–56, 2014)]; the latter are automatic tests that yield *P* values. However, MFA/SSA are only effective to detect PSEs that occur at about 6–16 ms posttrigger. Our first contribution is the scan test, an automatic test that has the same utility as SpTA, i.e., it can detect a wide range of PSEs at any latency, but it also yields a *P* value. Our second contribution is a thorough investigation of the statistical properties of PSE detection tests. We show that when the PSE is weak or the sample size is small, visual inspections of SpTAs have low power, because it is difficult to distinguish PSEs from background EMG variations. We also show that the scan test has better power and that its rate of spurious detections matches the chosen significance level α . This is especially important for investigators because, when a PSE is detected, this guarantees that the probability of a spurious PSE is less than α . Finally, we illustrate the operational characteristics of the PSE detection tests on 2,059 datasets from 5 experiments. The scan test is particularly useful to identify candidate PSEs, which can then be subject to further evaluation by SpTA inspection, and when PSEs are small and visual detection is ambiguous.

electromyography; postspike effect; scan test; single-snippet analysis; spike-triggered average

THE MOTOR CORTEX WAS ORIGINALLY defined as an area of cortex from which muscle contraction is elicited (Fritsch and Hitzig 1870). We now know that both direct and indirect projections originate from the motor cortex and terminate in the motoneuronal pools of the primate spinal cord. Projections from subcortical and spinal structures also converge in these pools and, together, determine the excitability of the motor units responsible for muscle contraction.

The effect of a neuron on motoneuron excitability is typically detected using spike-triggered averaging (SpTA; Fetz et al. 1976; Fetz and Cheney 1980; McKiernan et al. 1998): after each cortical spike, a short electromyography (EMG) snippet

from a target muscle is collected; if the cortical neuron generates a postspike effect (PSE) in the EMG, a characteristic waveform will emerge in the average of these EMG snippets. Because many factors in addition to the recorded spikes contribute to motoneurons excitability, the effect of a single neuron on the EMG is weak and may require many snippets (at least 2,000, but sometimes up to 20,000 spikes; Fetz and Cheney 1980) to yield a clear waveform in the SpTA.

The multiple-fragment analysis (MFA; Poliakov and Schieber 1998) is an alternative PSE detection test that is automatic, i.e., does not rely on visual inspection or subjective judgment. It is especially useful in small samples, when PSEs are difficult to distinguish from the EMG baseline activity, and it can be applied automatically and quickly to many datasets. However, the MFA is not as flexible as SpTA visual inspections: the former is designed to detect PSEs around the 6–16 ms postspike latency, the latter allows PSEs to be detected at any latency.

In this article, we develop the scan test, a formal statistical test that has the same utility as SpTA visual inspections, i.e., it can detect PSEs at any posttrigger latency, but that has the advantage of being fully automatic, so it can be applied quickly when many datasets must be analyzed. We perform an extensive study to show that the scan test has the expected rate of spurious detections and has more power to detect PSEs than SpTA visual inspections. Power is especially important in small samples, when PSEs are difficult to distinguish from the EMG background.

This work was primarily motivated by the need to inspect over 1,700 SpTAs in the Grasp experiment described in RESULTS, a tedious task that is also particularly difficult in these datasets, because most contain between 700 and 5,000 spike triggers, a number often too small to visually identify PSEs unequivocally. This work is also generally useful because large datasets of neuron-EMG recordings are becoming increasingly prevalent with the use of chronically implanted electrode arrays. An automatic and efficient PSE detection method helps prescreen those data to look for candidate PSEs, which can then be subject to further evaluation by SpTA or other means.

METHODS

We first review the current methods to detect PSEs: visual inspection of SpTAs and MFA/single-snippet analyses (SSA). We then extend the MFA/SSA to an automatic scan test that can detect PSEs at any latencies.

Address for reprint requests and other correspondence: S. Perel, 14400 Fifth Ave., Rm. 110D, Dept. of Biomedical Engineering, Carnegie Mellon Univ., Pittsburgh, PA (e-mail: sagi@cmu.edu).

PSEs Can Be Detected at Any Latency by Visual Inspection of a SpTA

An SpTA shows the changes in the EMG activity of a muscle following spikes from a cortical neuron. Given K cortical spikes and simultaneously recorded EMG activity, spike-triggered EMG snippets are collected around every spike and rectified to minimize the cancellations caused by averaging overlapping positive and negative components of motor unit action potentials. The SpTA is computed by averaging these rectified snippets across time according to

$$\text{SpTA}(t) = \frac{1}{K} \sum_{k=1}^K \text{rEMG}_k(t), \quad (1)$$

where typically, t ranges from -20 to 40 ms relative to the trigger at $t = 0$. A wide interval allows the baseline mean (M) and standard deviation (SD) of the SpTA in the pre and posttrigger periods to be estimated. A PSE is usually considered present at time t^* if $\text{SpTA}(t)$ exceeds $M \pm 2SD$ around $t = t^*$. The times when the SpTA exits and reenters the bands are called the PSE onset and offset; the extremum between onset and offset is called the peak of the effect; the peak width at half-maximum (PWHM) is the excursion width at half the height between the PSE peak and the baseline mean M . PSEs are divided into postspike facilitation and suppression effects, depending on whether the waveform appearing in the SpTA is a peak or a trough. PSEs are further classified as pure or synchrony effects: a PSE is often considered pure if it begins at a latency consistent with conduction times from cortical neurons to spinal motoneurons and is narrow enough to reflect monosynaptic or disynaptic connectivity, although caution must be taken when using these criteria (Smith and Fetz 2009). When the PSE onset is earlier than the shortest possible conduction time from cortex to the spinal cord, it is classified as synchrony PSE.

SpTA visual inspection. PSE detection by visual inspection of an SpTA is subjective. We followed these guidelines:

1) We de-trend the SpTA: we fit a linear regression to, and subtract it from, the SpTA (Bennett and Lemon 1994).

2) If a PSE with onset times in the $[-5, 20]$ -ms range appears clearly, we record it.

3) If the SpTA exhibits deviations that are not obvious PSEs:

i) we calculate the M and SD of the detrended SpTA in three baseline windows: $[-5, 5]$ ms around the spike-trigger (Kasser and Cheney 1985), $[-20, -10]$ ms pretrigger (McKiernan et al. 1998), and $[-30, -10]$ ms pretrigger (Schieber 2005);

ii) if the largest deviation exceeds at least one of the three $M \pm 2SD$ bands, if it is not excessively narrow, and if the SpTA displays no other oscillations with similar excursions outside the bands, we deem this deviation a PSE.

Automatic SpTA inspection. The main goal of this article is to develop a PSE detection test that can be applied reliably, automati-

cally, and quickly to a large number of neuron-EMG datasets. An obvious candidate is an automated version of the SpTA visual inspection:

1) A computer program de-trends the SpTA and calculates the M and SD of the detrended SpTA in three baseline windows: $[-5, 5]$ ms around the spike trigger, $[-20, -10]$ ms pretrigger, and $[-30, -10]$ ms pretrigger, as in the visual inspection, *steps 1* and *3i*.

2) The computer program identifies all SpTA excursions outside the $M \pm 2SD$ bands, and retains the largest as a potential PSE. The three $M \pm 2SD$ bands are used in turn, which yields three tests we refer to as the automatic SpTA1, SpTA2, and SpTA3 inspections.

3) If the onset time of the largest excursion is in the $[-5, 20]$ -ms range, and its PWHM exceeds 5 ms, we deem the largest excursion a PSE. The 5-ms PWHM threshold is an important choice we discuss in RESULTS.

Automated SpTA inspections are not completely faithful to visual inspections, because they do not use expert detection skills that are hard to automatize: e.g., a small dip prior to the PSE onset is often present in PSEs considered monosynaptic; multiple excursions of the SpTA outside $M \pm 2SD$, which happens for the three SpTAs in Fig. 1, *B–D*, might be considered evidence against PSE significance.

PSEs Can Be Detected Automatically at Fixed Latencies by MFA/SSA

Visual inspections of SpTAs require human intervention. The MFA (Poliakov and Schieber 1998) is automatic. The K -rectified EMG snippets used in Eq. 1 are divided into $G = \sqrt{K}$ fragments, and their respective SpTAs calculated; let $\text{SpTA}_g(t)$ denote the SpTA in fragment $g = 1, \dots, G$. Then for all g , one calculates a difference between the peak and the baseline of $\text{SpTA}_g(t)$ as

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

where $\overline{\text{SpTA}_g}([a, b] \text{ ms})$ denotes the average of $\text{SpTA}_g(t)$ over $t \in [a, b]$ (the spike-trigger is at $t = 0$). Differential amplitude values X_g that deviate from zero provide evidence of a PSE around 6–16 ms postspike. Poliakov and Schieber (1998) calculate a standardized summary of these deviations:

$$T = \frac{\bar{X}}{\text{SE}(\bar{X})}, \quad (3)$$

where \bar{X} is the mean of the X_g s and $\text{SE}(\bar{X})$ its standard error, and they compare the deviations of T from zero to the standard normal distribution. They reject the null hypothesis of no PSE at the 5% significance level if $T > 1.96$, with corresponding P value

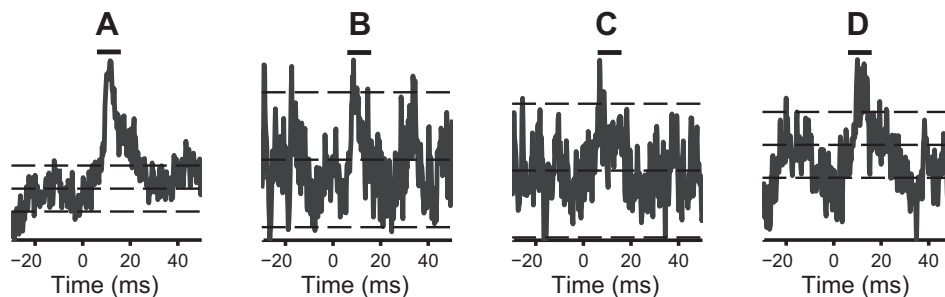


Fig. 1. Spike-triggered averages (SpTAs) of Digit Flexion datasets. *A*: SpTA of the full dataset (32,087 spikes); a PSE is clearly visible. *B*, *C*, and *D*: SpTAs of randomly subsampled datasets of sizes $K = 800$, $1,000$, and $7,000$. The vertical bars on top span the 6 – 16 ms posttrigger interval. The horizontal lines are at M and $M \pm 2SD$, where M and SD^2 are the baseline SpTA mean and variance in the $[-20, -10]$ -ms interval pretrigger. The subsampled datasets contain Postspike effects (PSEs), but they are hard to identify visually in small samples. The bootstrap scan test detects the PSEs in all cases, with P values $P < 0.0001$ (*A*), $P = 0.01$ (*B*), $P = 0.0001$ (*C*), and $P = 0.0001$ (*D*).

$$p = 2[1 - \Phi(T)], \tag{4}$$

where Φ is the standard normal cumulative distribution. The MFA is effectively a two-tailed t -test to test if the true mean of X_g is zero, so any statistical computer package will automatically output Eqs. 3 and 4. If suppression or facilitation PSEs are of specific interest, Perel et al. (2014) suggest replacing the two-sided t -test by a one-sided t -test, to increase the power of MFA to detect these specific PSEs.

Two classic MFA variants exist. Both use $G = \sqrt{K}$ fragments, but they use a different repartition of snippets per fragment: Poliakov and Schieber (1998) divide the experimental time into G periods of equal length and assign to fragment g the snippets in period g ; Davidson et al. (2007) form fragments of equal sizes by grouping every \sqrt{K} consecutive snippets. Perel et al. (2014) proposed the SSA, a related method that avoids fragmenting the data altogether. They show that SSA and the equal-sized fragment MFA have similar statistical properties and typically have better detection power than the original MFA of Poliakov and Schieber (1998).

The MFA/SSA tests rely on the assumption that T is standard normal when the data contain no PSE. Perel et al. (2014) showed that 1) the first assumption is typically met but may be questionable in small samples or when the SpTA baseline is not linear; 2) when they exist, the deviations from standard normality are small; and 3) simple bootstrap adjustments can be applied to correct the deviations.

Automatic Fixed-Latency Tests Can Be Scanned Across the SpTA To Detect PSEs at Any Latency

Let $T(l)$ denote the SpTA differential amplitude calculated on the portion of the EMG snippets centered at time l , and $p(l)$ its corresponding P value (using a one- or two-tailed tests depending on what type of PSE is of interest), so that $T(l)$ and $p(l)$ with $l = 11$ ms are T and p in Eqs. 3 and 4. When we use a fixed-latency test such as the MFA or SSA, the assumption is that the SpTA has maximal differential amplitude between 6 and 16 ms postspike, measured by $T(11)$ ms in Eq. 3. Depending on axonal and motoneuronal conduction velocities, the largest SpTA differential amplitude might occur at a different latency, which we would like to determine. A simple option consists of testing for PSEs at L postspike latencies and concluding that PSEs are present at all latencies l such that $p(l) \leq \alpha/L$; the Bonferroni corrected significance level α/L guarantees that the probability of detecting at least one spurious PSE is less than α . We reformulate this multiple test as a “scan test,” which is approximately equivalent, albeit constrained to detect at most one PSE.

Scan test. Given a neuron-EMG pair, we consider searching for PSEs at L postspike latencies spaced b ms apart, starting at $l = l_0$:

1) Perform the MFA with equal sized fragments, or the SSA, at latency l_0 , and store the P value, $p(l_0)$. Repeat at the other latencies, and store the corresponding P values, $p(l)$.

2) Evaluate the P value of the scan test:

$$p_{scan} = 1 - (1 - S)^L \tag{5}$$

(details are in APPENDIX A), where S is the smallest of the L P values:

$$S = \min_l p(l). \tag{6}$$

3) Conclusion: if $p_{scan} > \alpha$, there is no PSE in the data (or the test has insufficient power to detect it). If $p_{scan} \leq \alpha$, there is a PSE at latency peak of maximal SpTA differential amplitude $\hat{l} = \text{argmin}_l p(l)$ (it may be spurious with probability less than α).

We must choose at which postspike latencies to scan the test; for example, with $b = 2$ ms and $l_0 = 9$ ms, we scan at $l = 9, 11, 13$ ms, etc. When b is small, overlapping portions of EMG contribute to successive $T(l)$ s, so the $T(l)$ s are correlated. Then the Bonferroni correction, and hence the scan test, are conservative, as illustrated in RESULTS and APPENDIX A. This happens because the (unknown) effective number M of independent scans is smaller than the actual number

L , so we should use the less stringent significance level α/M in place of α/L . To resolve this issue, we can either increase the scan step b to avoid serial correlations in the $T(l)$ s or proceed with a small scan step, $b = 1$ ms say, and replace the parametric P value (Eq. 5) by a nonparametric bootstrap P value that does not rely on the independence assumption. We describe these two options below.

Determining the scan step b . The scan test P value (Eq. 5) depends on a parametric distribution for S that is valid provided the $T(l)$ s are independent. We use bootstrap diagnostics (Canty et al. 2006) to choose a scan step b large enough to satisfy this assumption. The idea is to compare the assumed parametric distribution of S (APPENDIX A, Eq. 7) to its bootstrap distribution, the latter being an approximation to the true distribution. If the two distributions match, then Eq. 5 is valid; otherwise, a nonparametric bootstrap P value must be obtained or the scan step b must be increased.

To compare the two distributions, we simulate a bootstrap sample, $s_1^*, s_2^*, \dots, s_R^*$, of size $R = 100$, as described below, and plot their ordered values against the theoretical quantiles of S , $\{1 - [1 - j/(R+1)]^{1/L}\}$, $j = 1, \dots, R$ to obtain a quantile-quantile (QQ) plot: the two distributions match only if the points lie close to the line with intercept 0 and slope 1; APPENDIX A (see Fig. A1) contains an example. For the data analyzed here, we found that $b = 8$ ms was the smallest scan step satisfying the independence assumption.

Bootstrap corrected scan test. The scan test is conservative when b is small, i.e., p_{scan} is biased upwards. This has no consequence when the test is significant ($p_{scan} \leq \alpha$) since an unbiased P value will be smaller and yield the same conclusion, or when p_{scan} is so large as to suggest that the data contain no evidence at all of a PSE. However, when $p_{scan} \in [\alpha, 5\alpha]$, we replace it with an unbiased bootstrap P value, as follows:

1) Obtain a sample of at least $R = 500$ bootstrap null values of S (Eq. 6), as follows. For $r = 1, \dots, R$:

i) Add normal jitters (SD = 30 ms) to the observed spike triggers, as justified in Perel et al. (2014) and collect the rectified EMG snippets corresponding to the jittered triggers. This is a bootstrap sample.

ii) Calculate s_r^* , the value of S (Eq. 6) in bootstrap sample r .

2) Replace p_{scan} by the nonparametric bootstrap P value:

$$p_{scan} = \#(s_r^* \leq S) / R,$$

the proportion of bootstrap values that are less than S . Details are in APPENDIX A.

3) Conclusion: if $p_{scan} > \alpha$, there is no PSE in the data. If $p_{scan} \leq \alpha$, there is a PSE at latency peak $\hat{l} = \text{argmin}_l p(l)$.

The scan test with an appropriately large step b and the bootstrap scan test are approximately equivalent, with the following differences: the former will lose power if it is not scanned exactly at the latency of maximal SpTA differential amplitude: we observe this in RESULTS, and it is not fully automatic since b is chosen by trial and error; the bootstrap scan test is fully automatic but requires a larger bootstrap simulation.

The scan tests are forms of automatic SpTA inspections, since they effectively measure and test the significance of the maximal SpTA differential amplitude, $T(\hat{l})$. Further, $T(\hat{l})$ can be used to compare PSEs across muscles and experimental conditions. Indeed, it is conceptually similar to the mean-percent increase, although it is normalized by the SD of the differential EMG values (the X values in Eq. 2) rather than by the baseline SpTA mean.

RESULTS

The methods available to detect PSEs at any posttrigger latency are SpTA visual inspections and their automated counterparts and the automatic uncorrected and bootstrap corrected scan tests (see METHODS). Here, we study the statistical properties of these tests, i.e., their spurious detections rates and

powers to detect effects, and illustrate their application to neuron-EMG data from the following five experiments.

The *grasp dataset* motivated the methods presented here. The data were collected in our laboratory, while a monkey performed reach-to-grasp movements to a variety of objects at different spatial locations and orientations. This task involved higher levels of muscle contraction than unloaded wrist/digits flexion/extension, which resulted in higher mean EMG levels. Every unique combination of object and location/orientation was presented 20 times, and we typically recorded 700 to 5,000 cortical spikes for every combination, which is often too little to identify PSEs visually. The EMG was recorded using custom made patch electrodes sutured to the epimysium (adapted from Miller et al. 1993), bandpass filtered at 100–2,000 Hz, and sampled at 4,882.8 Hz.

The two *precision-grip datasets* were collected in Dr. Roger Lemon's laboratory, while a monkey squeezed two spring-loaded levels between the thumb and index fingers. The first dataset consists of simultaneously recorded spike trains from 9 neurons and EMG activity from 7 muscles, with sizes ranging from 27,403 to 67,653 spikes (Jackson et al. 2003), and the second from 6 neurons and 12 muscles, with sizes ranging from 2,611 to 15,189 spikes (Quallo et al. 2012). The EMG was recorded using custom made patch electrodes sutured to the epimysium (adapted from Miller et al. 1993) and sampled at 5,000 Hz. Every neuron showed visible PSEs in three to five muscles.

The *wrist dataset* was collected in Dr. Peter Strick's laboratory. It consists of simultaneously recorded spike trains from 10 neurons and EMG from 12 muscles, while a monkey performed a center out, step-tracking wrist movement task (Kakei 1999). The datasets contain from 2,260 to 33,851 spikes. Every neuron showed visible PSEs in one to seven muscles. The EMG was recorded using stainless steel wire electrodes implanted subcutaneously, bandpass filtered at 300–3,000 Hz, and sampled at 4,000 Hz.

The *digit-flexion dataset* was collected in Dr. Marc Schieber's laboratory. It consists of simultaneously recorded spike trains from two neurons and EMG from nine muscles, while a monkey performed visually cued individuated flexion and extension movements of the right finger and/or wrist (Schieber 2005). The datasets contain from 19,076 to 32,087 spike triggers. One neuron showed visible PSEs in five muscles and the other in eight muscles. The EMG was recorded using stainless steel wire electrodes implanted percutaneously, bandpass filtered at 300–3,000 Hz, and sampled at 5,000 Hz.

Figure 1A shows the SpTA of a digit flexion dataset of size $K = 32,087$, and Fig. 1, B–D, shows the SpTAs of three random subsamples of sizes $K = 800, 1000$, and $7,000$ from that dataset. A PSE appears clearly in Fig. 1A. The SpTAs of the three subsamples therefore also contain PSEs, but they are harder to identify visually; it is up to the investigator to make a decision. The bootstrap corrected scan test has the statistical power to detect the PSEs in Fig. 1, B–D, with $P = 0.01, 0.0001$, and 0.0001 , respectively. Below, we investigate more fully the statistical properties of the PSE detection tests.

Properties of Tests: Rates of Spurious Detections

A statistical test of hypotheses yields either the correct decision or one of two types of error: a type 1 error is a

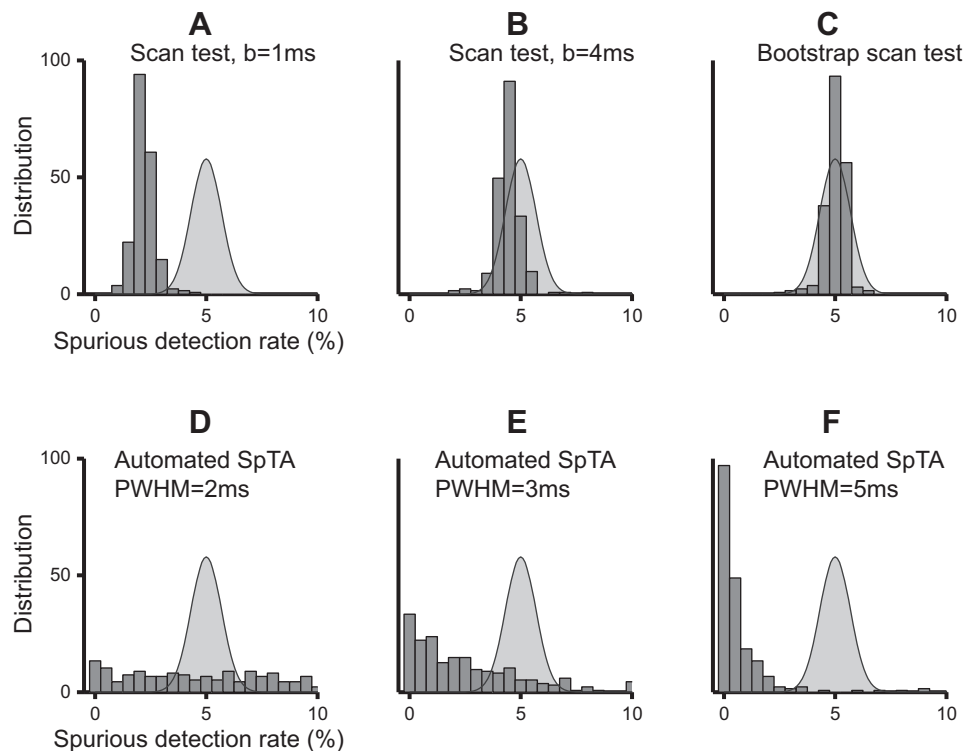
spurious detection; a type 2 error is a false negative, when the test fails to detect a real effect. The probabilities of these two error types vary in opposite directions, so errors cannot be eliminated. The accepted way to proceed is to design a test that has a prespecified maximum probability of a type 1 error, e.g., $\alpha = 5\%$, referred to as the significance level. Then, when a PSE is detected, we cannot be sure that it is real, since the truth is unknown, but we know that it is spurious with probability less than α . For a test to be reliable, it is therefore crucial that its actual spurious detections rate be equal to, or less than, α .

The spurious detections rates of the tests considered here are too difficult to derive analytically, so we estimate them by the proportions of PSEs detected in *null datasets*, i.e., datasets that contain no PSEs. The number of available experimental datasets is too small to estimate proportions accurately, and we cannot be sure whether or not they actually contain PSEs. We therefore create null datasets: given an experimental dataset, we leave its EMG trace unchanged and destroy all precisely time-locked effects by either 1) adding Gaussian white noise with variance 100^2 ms to all the spike triggers; or 2) randomizing the interspike intervals of the cortical spikes; or 3) pairing the EMG trace with a cortical spike train from another experiment.

Figure 2 shows the estimated spurious detections rates of the tests. We simulated 1,000 null datasets by jittering repeatedly one experimental dataset and recorded the proportions of PSEs detected in these datasets by the Bootstrap scan test, scan test (from $l = 8$ ms to $l = 30$ ms posttrigger, in steps of $b = 1$ and 4 ms), and automated SpTA inspections (baseline window $[-10, -20]$ ms pretrigger, PWHM threshold 2, 3, and 5 ms), using the $\alpha = 5\%$ significance level. We repeated this using 272 other experimental datasets and plotted the histograms of the 273 detection proportions obtained by each test in Fig. 2. The curve overlaid on each panel is the distribution we should observe for a test with the correct α -level; it is centered at $\alpha = 5\%$ and has some spread since the spurious detections rates are estimated. The histogram for the bootstrap corrected scan test (Fig. 2C) matches the expected distribution: this test has the correct α -level. The histogram for the scan test with $b = 1$ ms (Fig. 2A) has a similar shape but is centered $\sim 2.1\%$, less than half the nominal $\alpha = 5\%$: the test is conservative, as we anticipated, although not exceedingly so here, because we scanned the tests at a relatively small number of latencies, $L = 23$. The scan test becomes less conservative with larger scan steps: the average spurious detections rate is $\sim 4.2\%$ with $b = 4$ ms (Fig. 2B) and reaches the nominal $\alpha = 5\%$ with $b = 8$ ms (not shown, but similar to Fig. 2C). The observed spurious detections rates of automated SpTA inspections are plotted in Fig. 2, D–F. Their distributions are strikingly different than what we expect of a test with significance level $\alpha = 5\%$. The test is either exceedingly liberal, with spurious detections rates as large as 50, 38, and 22% using 2-, 3-, and 5-ms PWHM thresholds, respectively, or exceedingly conservative, with over 80% of observed spurious detections rates $< 1\%$ using a 5-ms threshold. That is, the spurious detections rates of automated SpTA inspections depend in nontrivial ways on the PWHM threshold and the shape of the PSE to be detected, which is unknown a priori.

To summarize, the bootstrap scan test and scan test with step $b = 8$ ms have the expected rate of spurious detections. The scan test becomes conservative when smaller steps are used,

Fig. 2. Observed spurious detections rates of the tests when the nominal significance level is $\alpha = 5\%$. Each histogram value is the proportion of detections in 1,000 null datasets created by jittering the cortical spikes of a particular experimental dataset. Each histogram contains values from 273 different experimental datasets. The overlaid curve is the distribution we expect of a test that has the correct α level. *A*: uncorrected scan test with scan step $b = 1$ ms has spurious detections rate lower than α : it is conservative. *B*: it becomes less conservative with larger scan steps; here $b = 4$ ms. *C*: bootstrap corrected scan test has spurious detections rate that matches the expected distribution. *D–F*: spurious detections rates of the automated SpTA visual inspections do not match their expected distribution, regardless of the chosen peak width at half-maximum (PWHM) threshold; here PWHM = 2, 3, and 5 ms, respectively.



although not exceedingly conservative if the number of scanned latencies remains small; here, the maximum number of scans is $L = 23$ when we scan every $b = 1$ ms to detect PSEs in the [8,30] ms posttrigger window. A low spurious detections rate is a very desirable property, but one that comes at the cost of low power, as we show in the next section. The automated SpTA inspection is unpredictably conservative or liberal, potentially exceedingly so, and therefore cannot be trusted. All observations about Fig. 2 extend to other commonly used significance levels, e.g., $\alpha = 1$ and 10%, to other methods of creating null datasets (interspike interval randomization and shuffling) and to other baseline windows for automated SpTA inspections (not shown).

Properties of Tests: Power

To maximize the chances of detecting PSEs, it is important to use tests with high power, i.e., high probability of detecting effects. In this section, we compare the powers of the PSE detection tests, which we estimate by the proportions of datasets in which PSEs are detected. To do that, we need many datasets with specific sample sizes and effect strengths; we construct them as follows.

Parent and test datasets. Parent datasets are experimental datasets whose SpTAs clearly display PSEs, e.g., in Figs. 1A and 3, C, F, I, and L. There are 2 Grasp, 13 Digit Flexion, 48 Precision Grip, and 24 Wrist such datasets. Their PSEs span a continuum from pure to combinations of pure-synchrony PSEs and synchrony PSEs, with PWHM values ranging from 4 to 15 ms, onsets >3 ms, and offsets >18 ms; they are classified as medium to strong, based on their mean percent increase and peak percent increase values; some PSEs are narrower and some wider than 10 ms, and some appear at latencies later than 16 ms postspike. To create a “test dataset” of size K , we choose a spike trigger at random from a parent dataset, and retain that

spike and its $(K - 1)$ successors; we wrap the spike train if fewer than $(K - 1)$ spikes follow the selected spike, and we extract the K corresponding rectified EMG snippets from the parent. This test dataset is subsampled from a parent dataset, so it contains the same PSE, although it may be difficult to detect visually when K is small, as was illustrated in Fig. 1, B–D. To create a test dataset that contains a PSE of strength $a \in [0,100]\%$, we sample a test dataset of size K , as above, jitter the spikes of a randomly selected block of $K \cdot (100 - a)$ spike triggers to remove time-locked effects in that block (normal jitter, SD 100 ms) and extract the corresponding rectified EMG snippets. This test dataset contains a PSE whose strength depends on a : there is no PSE when $a = 0\%$; when $a = 100\%$, the PSE is of strength similar to that of the parent.

Figure 3 shows the power curves of the PSE detection tests in subsamples from the parent datasets whose SpTAs are shown in Fig. 3, C, F, I, and L; the other parent datasets produced similar results. To estimate power as a function of sample size K , we sampled 1,000 test datasets of size K from one parent dataset and recorded the proportions of PSEs detected by each test. We repeated for many values of K , capping K at half the number of spike triggers in the parent dataset to avoid simulating repeat test datasets. We proceeded similarly to estimate power as function of effect strength a , holding the sizes of all test datasets fixed at half that of the parent dataset. Figure 3 shows that power increases with sample size and effect strength, as we should expect. When the PSE is strong ($a \approx 1$) or the number K of spike triggers is large, e.g., in Fig. 3, J and K, the SpTA baseline is less variable, and all tests detect the PSEs consistently. When the effect is weaker or the sample size is smaller, the tests do not always detect the PSEs that we know exist, since the test datasets were sampled from experimental datasets that do contain PSEs. The bootstrap

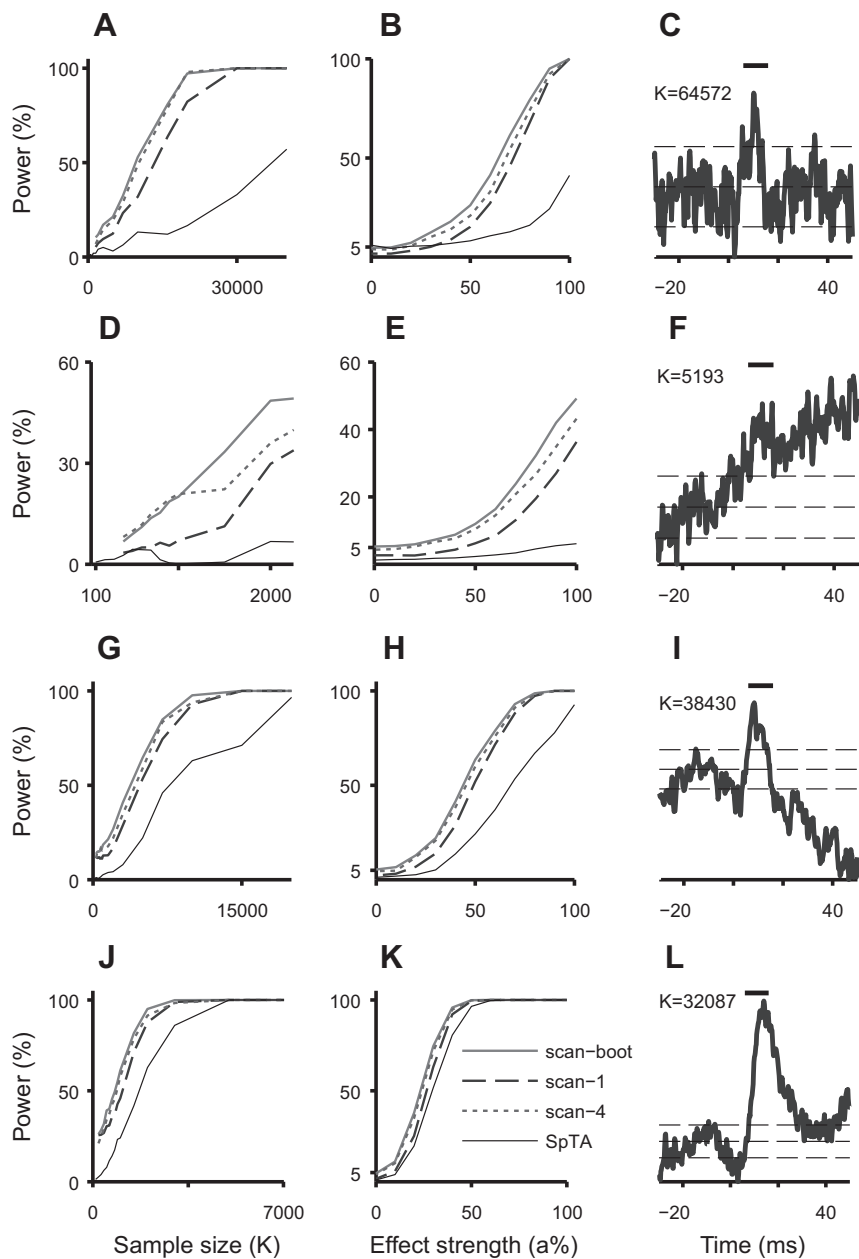


Fig. 3. Power curves of the scan tests and automatic SpTA inspections as functions of sample size K (A, D, G, and J) and effect strength a (B, E, H, and K) for the parent datasets with SpTAs in C, F, I, and L. C: narrow PSE (PWHM = 3 ms) in a Precision Grip dataset of size 64,572. F: PSE in a Grasp dataset of size 5,193. I: wide PSE (PWHM = 8.4 ms, onset 5.4 ms, offset 15 ms) in a Precision Grip dataset of size 38,430. L: wide late synchrony PSE (PWHM = 9.8 ms, onset 8.6 ms, offset 24.6 ms) in a Digit Flexion dataset of size 32,087. The vertical bars on top span the 6 – 16 ms posttrigger interval. The horizontal lines are at M and $M \pm 2SD$, where M and SD^2 are the baseline SpTA mean and variance in the $[-20, -10]$ -ms interval pretrigger. The scan tests are able to detect narrow and wide PSEs, despite using a fixed width 10-ms sliding window. They have uniformly better power than SpTA detection, most notably in small samples (K small) and when the PSE is weak (a small). The bootstrap scan test has better power than the uncorrected test. There are no PSEs in the data when $a = 0$ in B, E, H, and K, in which case the power is the rate of spurious detections: the bootstrap scan test has the correct rate, the uncorrected test is conservative, and SpTA detection can be very conservative.

corrected scan test has the best power. The scan test with $b = 1$ ms has lower power, a consequence of being conservative. Its power first increases as we increase the scan-step, and comes close to the power of the bootstrap scan test when $b \approx 4$ ms, but decreases as we increase b further (not shown), because it is then more likely to miss the location of larger SpTA amplitude. We also plotted the power of the automated SpTA inspection, even though we recommend against using a test that does not control its rate of spurious detections; we used a 5-ms PWHM threshold to prevent the test from being overly liberal. The power of this test varies as unpredictably as its spurious detection rate and has particularly low power to detect narrow or weak PSEs (Fig. 3, A, B, D, and E).

The power curves also give some information about the sample sizes needed to detect PSEs with high probability. For example, Fig. 3A shows that in datasets of size $K \approx 20,000$, the

probability that the bootstrap scan test detects the PSE is close to 100%; it is $\sim 80\%$ for the scan test with $b = 1$ ms and $< 20\%$ for automated SpTA inspections. In Fig. 3J, the bootstrap scan test has full power with 3,000 spike triggers, while the automated SpTA inspection requires twice as many. Figure 3, C, I, and L, each show a weak, medium, and strong PSE; the bootstrap scan test requires about 20,000, 10,000, and 3,000 spike triggers in each case to detect these PSEs with probability 1, which is much fewer than the samples actually collected, and fewer than needed to detect the PSEs with confidence by visual inspection of SpTAs.

Detecting PSEs in the Experimental Datasets

We have studied the statistical properties of the PSE detection tests and found that the bootstrap scan test is most powerful and controls its rate of spurious detections. Table 1

Table 1. Number of PSEs detected by the Bootstrap scan test

	Experiment				
	Digit Flexion	Precision Grip	Precision Grip 2	Wrist	Grasp
Number of datasets N	18	67	57	212	1,705
Number of detections using the Bootstrap scan test, $\alpha = 5\%$	16	51	18	60	65
Expected number of spurious detections if the data contain no PSE	[0,3]	[0,7]	[0,6]	[4,17]	[67,103]

Number of postspike effects (PSEs) detected by the Bootstrap scan test at $\alpha = 5$ and 95% confidence interval $(\alpha N \pm 2[\alpha(1 - \alpha)N]^{1/2})$ for the number of spurious detections we expect if the datasets contain no PSEs.

contains the numbers of PSEs with peaks in the [8,30] ms postspike latency range detected by that test in the experimental datasets, at the significance level $\alpha = 5\%$, and Fig. 4 includes some examples. We did not apply the uncorrected scan test or automated SpTA inspections because they have lower power or uncontrollable spurious detections rate. Some of the detections in Table 1 are undoubtedly spurious, since statistical tests are designed to allow for a controlled number of spurious detections, so they have power to detect real effects; at the extreme, a test applied to N datasets that contain no effect is expected to detect an average of αN significant effects by chance, with 95% confidence interval $[\alpha N \pm 2\sqrt{\alpha(1 - \alpha)N}]$. However, some of the detections in Table 1 are undoubtedly real PSEs because, in all but the Grasp experiment, the num-

bers of detections exceed these intervals by large margins, meaning that more effects were detected than by chance alone. In the Grasp datasets, the number of detections is close to the interval, which suggests that few or no PSEs exist or that the test lacks power to detect them, possibly because the datasets are small (700 to 5,000 spike triggers).

Next, we inspected visually the SpTAs of all datasets, ignoring the scan test results to avoid biasing our judgment, and cross-classified the PSEs discovered by both methods in Table 2: the bootstrap scan test detected all the PSEs identified visually, with five exceptions (*row 3*) that are all very narrow, with PWHM < 3 ms; they are shown in Fig. 4. We return to the detection of narrow PSEs in the DISCUSSION. We then reexamined carefully all the SpTAs for which the bootstrap scan test

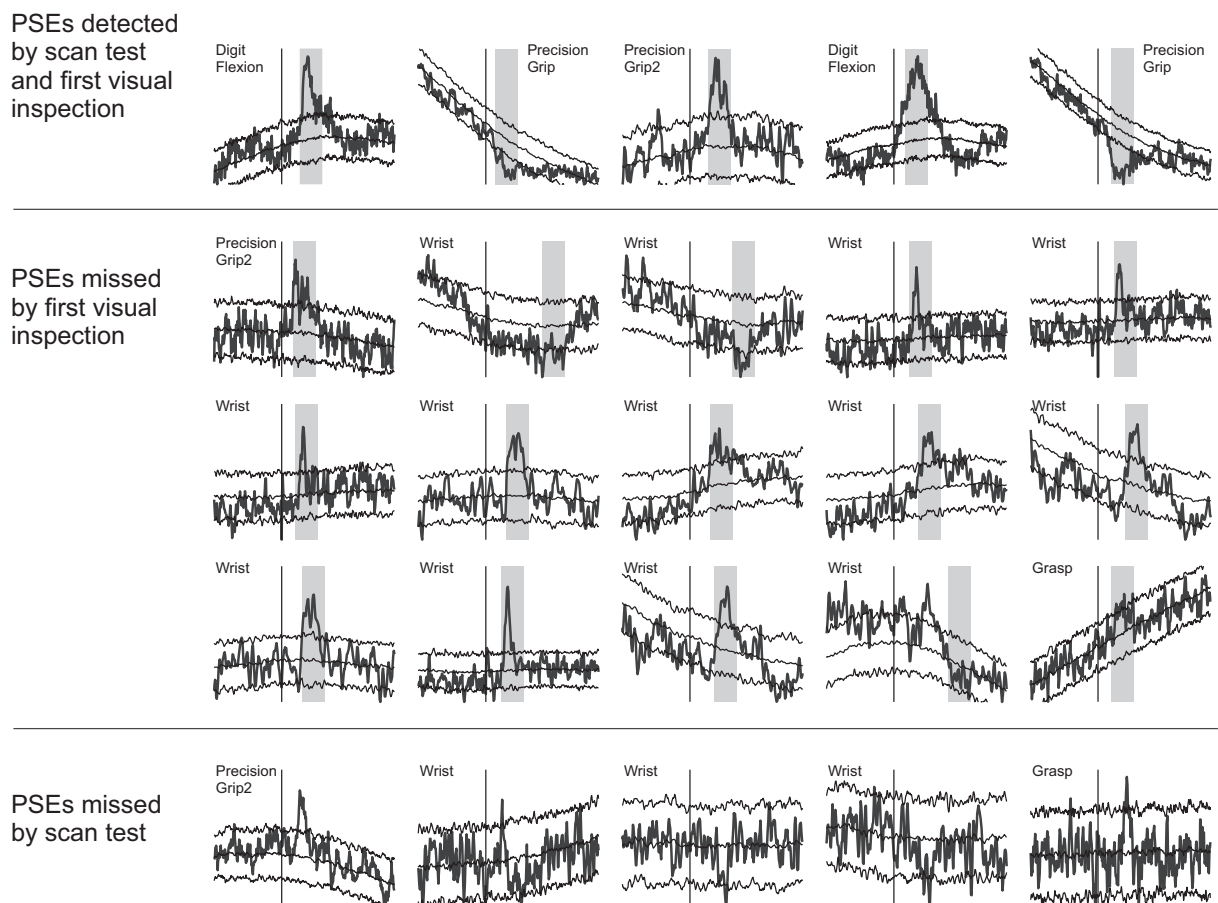


Fig. 4. Experimental datasets SpTAs. The x-axis spans 30 ms pre- to 50 ms postspike trigger; the trigger is marked by a vertical line. The SpTAs are scaled to fill the same vertical space. Bootstrap SpTA baselines and 95% bootstrap null confidence bands (Perel et al. 2014) are overlaid to aid visual PSE detections. The vertical grey bands span the 10-ms detection windows where the scan test detected PSEs. *Top*: 5 of the PSEs that were identified both visually and by the automatic tests. *Middle*: small PSEs that we did not identify visually but that were detected by the bootstrap corrected scan test. *Bottom*: 5 narrow PSEs that the scan test failed to detect.

Table 2. Number of PSEs detected visually by the authors and number of these detections that were and were not detected by the scan test

	Number of Detections/Misses				
	1	2	3	4	5
Number of SpTA visual detections	13	36	9	25	1
Number also detected by the scan test	13	36	8	22	0
Number missed by the scan test	0	0	1	3	1

Row 1 equals the sum of rows 2 and 3. The scan test missed 5 visually identified PSEs, which are all very narrow; they are shown in Fig. 4.

detected significant effects, using in addition to visual inspections the SpTA bootstrap confidence bands of Perel et al. (2014). We thus discovered 16 additional PSEs (Table 3, row 1) that we did not initially identify visually, either because their shapes were not clear or the numbers of spike triggers were small. They are shown in Fig. 4. Forty additional deviations could also be PSEs (row 2), but their shapes are more ambiguous.

To summarize, the bootstrap scan test detected all PSEs that were easily identified visually, with the exception of some very narrow PSEs, and helped detect additional PSEs that we had initially missed. Therefore, our recommended course of action when there are many SpTAs to inspect is to run the scan test first and examine only the SpTAs that are found to have effects. Many of these SpTAs might nevertheless contain no effect, since statistical tests allow $\alpha\%$ spurious detections. If one wishes to limit their number, a smaller significance level could be used, or the false discovery rate (FDR) procedure of Benjamini and Hochberg (1995) could be applied, whose objective is to cap the expected proportion of spurious detections out of all the detections. This is detailed in APPENDIX B.

DISCUSSION

Two analytical tools are available to detect PSEs: SpTA and MFA/SSA. The former consists of visually inspecting the SpTA for departures from its baseline and assessing their significance using prior experience or the more formal procedures suggested by Kasser and Cheney (1985) and Lemon et al. (1986). The MFA/SSA is fully automated and relies on a *P* value rather than subjective judgment to assess the statistical significance of PSEs. However, MFA/SSA is designed to find PSEs at a predetermined latency around 6–16 ms postspike and cannot detect late PSEs, which includes monosynaptic PSEs that appear at later latencies due to slow axonal velocities of corticospinal neurons or spinal motoneurons, synchrony effects and disynaptic suppression PSEs.

In this article, we proposed tests that have the same utility as SpTA visual inspections, i.e., they can detect a wide range of PSEs, but that are fully automatic and thus fast to apply to many datasets. They rely on *P* values to assess PSE significance; they do not depend on subjective judgment or prior experience. We considered two kinds of automatic tests: automated versions of SpTA visual inspections and MFA/SSA scan tests, which consist of performing the MFA/SSA in a sliding detection window. We performed a large study to characterize the statistical properties of these tests. We showed that automated SpTAs inspections do not control their rates of spurious detections; they tend to be very conservative but can be unexpectedly liberal. On the other hand, the scan tests maintain

their rates of spurious detections at or below the chosen significance level α . Among these tests, the bootstrap corrected scan test has the highest power. These properties are not particularly important when sample sizes are so large that the presence or absence of PSEs can be seen easily in the SpTAs. However, in small datasets, or when PSEs are weak (small excursion from baseline), the bootstrap scan test has higher probability to detect them, and when a PSE is detected, we know that it is spurious with probability is less than α .

We illustrated the application of these tests to 2,059 datasets from 5 different experiments. We inspected all SpTAs visually and detected 86 PSEs. The bootstrap scan test detected all but five of them, all of which were very narrow with PWHM <3 ms. The scan test uses a fixed 10-ms wide sliding detection window, and while it still has power to detect PSEs that are narrower and wider than 10 ms, as in Fig. 3, C, F, and L, very narrow PSEs are problematic. In future work, we will allow the width of the scan test detection window to adapt automatically to the dataset under investigation. The bootstrap scan test also detected effects that we did not identify visually. We re-inspected carefully the corresponding SpTAs and, relying in addition on the SpTA bootstrap confidence bands of Perel et al. (2014), we discovered 16 additional smaller PSEs that we had previously missed or dismissed. Some Wrist datasets also included stimulation triggers instead of spike triggers, which the scan test detected and which we dismissed as PSEs after inspecting their SpTAs.

Without the scan test, it was particularly difficult to inspect the Grasp datasets, because they are many and most contain <2,000 spike triggers, which is often too small to visually identify PSEs unequivocally. The SpTAs failed to produce clear shapes, and without an objective measure of significance, we could only be confident that one dataset contained a PSE. The bootstrap scan test failed to detect it because it was too narrow. After examining again the SpTAs deemed significant by the scan test, we identified another small PSE that we had previously dismissed as background noise. It is very likely that other PSEs exist. Indeed, we applied intracortical microstimulations to 113 cortical neurons during the course of the experiment and recorded stimulus triggered effects for 37 of them. The scan test could not detect more effects either because the datasets are too small, the effects too weak or too narrow, or the EMG signal too noisy due to the type of EMG recording electrodes we used: patches instead of intramuscular wires, which record many more overlapping motor units. It would be useful to determine, preexperiment, the number of samples to collect to detect PSEs with high confidence.

Finally, the scan test does not replace visual inspections of SpTAs. Rather, it is an automatic and efficient PSE detection method that helps prescreen those data to look for candidate PSEs, which can then be subject to further evaluation by SpTA

Table 3. Second inspection of the SpTAs found significant by the scan test but not by the initial visual inspection

	Number of Detections				
	0	1	2	3	4
Number of clear PSEs	0	1	1	13	1
Number of possible PSEs	1	9	6	16	8
Number of likely spurious detections	2	5	3	9	56

The sum of the three rows equals Table 1, row 1 minus Table 2, row 2.

or other means. The scan test is useful when PSEs are small and visual detection ambiguous; combined with the bootstrap SpTA baseline and 95% confidence bands estimates of Perel et al. (2014), it helped detect small PSEs that we had dismissed as background variation. The scan test is also useful because large datasets of neuron-EMG recordings are becoming increasingly prevalent with the use of chronically implanted electrode arrays.

APPENDIX A: SCAN TEST DETAILS

Parametric Distribution for the Scan Test

The scan test statistic, S in Eq.6, is the smallest of the L P values $p(l)$ of the MFA/SSA test in a sliding detection window. Small values of S provide evidence of a PSE, so the scan test P value is the area in the left tail of the null distribution of S : $p_{scan} = \int_0^S f_S(s)ds$. Under the null hypothesis, the P values $p(l)$ are uniform on $[0,1]$, and when they are independent, a standard probability calculation yields

$$f_S(s) = L(1 - s)^{L-1}, s \in [0, 1], \tag{7}$$

so that $\int_0^S f_S(s)ds = 1 - (1 - S)^L$ (Eq. 5). Equations 5 and 7 are valid under the assumption that the $T(l)$ s are standard normal and independent. Perel et al. (2014) showed that the first assumption is typically met and that the occasional deviations from standard normality are small. The independence assumption is violated when the test is scanned at close latencies. Questionable assumptions can bias P values and in turn the rate of spurious detections of the test. It is therefore important to adjust the test when needed.

Bootstrap Test and Diagnostics

We diagnose deviations from the assumptions by comparing the assumed parametric null distribution of S (Eq. 7) to its bootstrap null distribution using the QQ-plot described in METHODS. Figure A1 shows this QQ-plot for the scan test applied to a dataset at $L = 104$ postspike latencies spaced $b = 1$ ms apart: the points clearly deviate from the line with intercept 0 and slope 1, which means that the bootstrap and theoretical null distributions of S do not match. Moreover, the points lie above the line: the parametric distribution is shifted to the left of the bootstrap distribution and implies that p_{scan} (Eq. 5) is larger than it should be, i.e., that the scan test is conservative. We need to either use a larger step b to decrease the serial dependence between the scanned values, or calculate a bootstrap P value. Figure A1B is the diagnostic QQ-plot when we run the scan test with step $b = 4$ ms. The theoretical and bootstrap quantiles of S now match better; the match becomes perfect when $b = 8$

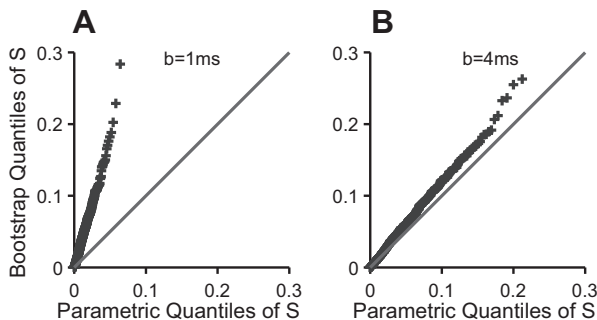


Fig. A1. Bootstrap diagnostic quantile-quantile (QQ) plot for the scan test. The plot is pathological in A, when the test is scanned every $b = 1$ ms, which means that the parametric P value (Eq. 5) is biased; a bootstrap P value should be calculated, or the scan step b increased. The plot looks better in B, when the test is scanned every $b = 4$ ms. The points overlay the line when $b \geq 8$ ms (not shown), in which case the parametric P value can be trusted.

Table A1. Number of PSEs detected by the Bootstrap scan test with a 20% false discovery rate and expected number of these detections that are spurious

	Experiment				
	Digit Flexion	Precision Grip	Precision Grip 2	Wrist	Grasp
Number of datasets N	18	67	57	212	1,705
Bootstrap scan test, FDR = 20%	16	54	22	64	1
Expected number of spurious detections	[3]	[11]	[4]	[13]	[0]

FDR, false discovery rate.

ms (not shown): in this dataset, the parametric P value (Eq. 7) can be trusted when $b = 8$ ms.

Instead of increasing b , we prefer to retain $b = 1$ ms and replace the parametric P value by a nonparametric bootstrap P value:

$$p_{scan}^* = \frac{\#s_r^* \leq S}{R},$$

where s_r^* , $r = 1, \dots, R$, are bootstrap null values of S ; p_{scan}^* is the area in the left tail of the bootstrap null distribution of S , since small values of S provide evidence of a PSE. When the parametric assumptions of the scan test are met, the parametric and bootstrap P values match within simulation error. Calculating nonparametric bootstrap P values requires a large bootstrap simulation. To avoid unnecessary simulations, we recommend replacing p_{scan} by p_{scan}^* only when $p_{scan} \in [\alpha, 5\alpha]$, since p_{scan} is biased upwards. We can also make due with a relatively small bootstrap simulation, $R = 500$ say, because the bootstrap null distribution of S is concentrated near zero, so its small quantiles are relatively accurate. We must increase R if more accurate P values are needed, for example if a small significance level is used.

APPENDIX B: FALSE DISCOVERY RATE TESTING

A test applied to a large number N of datasets will detect many spurious effects: $N\alpha$ on average when the N datasets contain no effects, where α is the significance level. If one wishes to limit the number of spurious detections, a smaller significance level could be used, e.g., the Bonferroni corrected significance level α/N , which caps at α the probability of detecting one or more spurious effects in N SpTAs. This is a stringent criterion that renders tests exceedingly conservative when N is large. Benjamini and Hochberg (1995) proposed a conceptually different procedure, whose objective is to cap the FDR, defined as the expected proportion of spurious detections out of all the detections; for example, if the procedure applied with FDR = 10% detects 90 PSEs, then 9 or fewer of these are spurious, on average. Traditional testing controls the number of spurious detections out of the total number of tests, which is a different criterion entirely. Just as with traditional tests, choosing which FDR to use

Table A2. Number of PSEs detected visually by the authors and number of these detections that were and were not detected by the scan test

	Number of Detections/Misses				
	Digit Flexion	Precision Grip	Precision Grip 2	Wrist	Grasp
Number of SpTA visual detections	13	36	9	25	1
Number also detected by the scan test	13	36	9	23	0
Number missed by the scan test	0	0	0	2	1

Row 1 equals the sum of rows 2 and 3.

Table A3. *Second inspection of the SpTAs found significant by the scan test but not by the initial visual inspection*

	Number of Detections				
Number of clear PSEs	0	1	1	11	0
Number of possible PSEs	1	9	8	18	0
Number of likely spurious detections	2	8	4	12	1

The sum of the three rows equals Table 1, row 1 minus Table 2, row 2.

involves a trade-off: lowering it to avoid spurious detections also lowers the power to detect real effects, and vice versa.

To apply the FDR procedure, let $p_{(n)}$, $n = 1, \dots, N$ denote the ordered P values of the scan test applied to N datasets, with $p_{(1)}$ being the smallest and $p_{(N)}$ the largest. We then apply the usual test sequentially with a dynamic significance level: we start with the largest P value, $p_{(N)}$, and compare it to the chosen FDR (we used $FDR = 20\%$); if $p_{(N)} \leq FDR$, we stop the procedure and deem all N tests significant. Otherwise, we compare the next largest P value, $p_{(N-1)}$, to the significance level $FDR \cdot (N-1)/N$: if $p_{(N-1)} \leq FDR \cdot (N-1)/N$, we stop the procedure and deem this test, and all the $N-2$ tests with smaller P values, significant. Otherwise, we continue until the test that has the largest P value $p_{(k)}$ such that $p_{(k)} \leq FDR \cdot k/N$; we stop the procedure and reject the null hypothesis of no PSE for the test performed last, and for all the tests that have smaller P values. Ventura et al. (2004) has a useful pictorial representation of this process. Note that when $N = 1$, the FDR procedure reduces to the usual test with the significance level $\alpha = FDR$.

Tables A1, A2, and A3 contain information similar to Tables 1–3. Table A1 reports the numbers of detections using the bootstrap scan test with $FDR = 20\%$ and the expected number of these that are spurious, i.e., $16 \cdot 20\% \approx 3$ in the Digit Flexion experiment, $52 \cdot 20\% \approx 10$ in the Precision Grip experiment, and so on. The detections are mostly the same as in Table 1 (bootstrap scan test, traditional spurious detections rate control), except in the Grasp datasets, where FDR testing detected just one PSE, vs. 65 using traditional testing, to satisfy the guarantee of 20% spurious effects on average. The number of expected spurious detections (Table A1, row 3) are close to their suspected observed number (Table A3, row 3). Here and in general, traditional and FDR testing will detect mostly the same effects, but one should expect differences since the two procedures satisfy two different criteria; experimentalists should choose which criterion they care most about.

ACKNOWLEDGMENTS

We thank Dr. Marc Schieber, Roger Lemon, Dr. Andrew Jackson, Dr. Darcy Griffin, and Prof. Peter Strick for generous contribution of datasets.

GRANTS

Our work was supported by National Institute of Neurological Disorders and Stroke Grant NS-050256 (to S. Perel and A. B. Schwartz) and National Institute of Mental Health Grant 2R01-MH-064537 (to V. Ventura). Dr. M. Schieber’s work cited in this article was supported by National Institute of Neurological Disorders and Stroke Grant NS-27686. This project also acknowledges the use of the Cornell Center for Advanced Computing’s “MATLAB on the TeraGrid” experimental computing resource funded by National Science Foundation Grant 0844032 in partnership with Purdue University, Dell, The MathWorks, and Microsoft.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: S.P., A.B.S., and V.V. conception and design of research; S.P. performed experiments; S.P. and V.V. analyzed data; S.P. and V.V. interpreted results of experiments; S.P. prepared figures; S.P. and V.V. drafted manuscript; S.P. and V.V. edited and revised manuscript; S.P., A.B.S., and V.V. approved final version of manuscript.

REFERENCES

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc* 57: 289–300, 1995.

Bennett KM, Lemon RN. The influence of single monkey cortico-motoneuronal cells at different levels of activity in target muscles. *J Physiol* 477: 291–307, 1994.

Canty AJ, Davison AC, Hinkley DV, Ventura V. Bootstrap diagnostics and remedies. *Can J Stat* 34: 5–27, 2006.

Davidson A, Odell R, Chan V, Schieber M. Comparing effects in spike-triggered averages of rectified EMG across different behaviors. *J Neurosci Methods* 163: 283–294, 2007.

Fetz EE, Cheney PD. Postspike facilitation of forelimb muscle activity by primate corticomotoneuronal cells. *J Neurophysiol* 44: 751–72, 1980.

Fetz EE, Cheney PD, German DC. Corticomotoneuronal connections of precentral cells detected by postspike averages of emg activity in behaving monkeys. *Brain Res* 114: 505–10, 1976.

Fritsch G, Hitzig E. On the electrical excitability of the cerebrum. In: *Some Papers on the Cerebral Cortex*, translated and edited by von Bomin G. Springfield, IL: Thomas, 1960, p. 73–96 [originally in *Arch f Anat, Physiol und Wissenschaftl Mediz, Leipzig* 37: 300–332, 1870].

Jackson A, Gee VJ, Baker SN, Lemon RN. Synchrony between neurons with similar muscle fields in monkey motor cortex. *Neuron* 38: 115–225, 2003.

Kakei S. Muscle and movement representations in the primary motor cortex. *Science* 285: 2136–2139, 1999.

Kasser RJ, Cheney PD. Characteristics of corticomotoneuronal postspike facilitation and reciprocal suppression of EMG activity in the monkey. *J Neurophysiol* 53: 959–978, 1985.

Lemon RN, Mantel GW, Muir RB. Corticospinal facilitation of hand muscles during voluntary movement in the conscious monkey. *J Physiol* 381: 497–527, 1986.

McKiernan BJ, Marcario JK, Karrer JH, Cheney PD. Corticomotoneuronal postspike effects in shoulder, elbow, wrist, digit, and intrinsic hand muscles during a reach and prehension task. *J Neurophysiol* 80: 1961–1980, 1998.

Miller LE, van Kan PL, Sinkjaer T, Andersen T, Harris GD, Houk JC. Correlation of primate red nucleus discharge with muscle activity during free-form arm movements. *J Physiol* 469: 213–243, 1993.

Perel S, Schwartz AB, Ventura V. Single-snippet analysis for detection of postspike effects. *Neural Comput* 26: 40–56, 2014.

Poliakov AV, Schieber MH. Multiple fragment statistical analysis of postspike effects in spike-triggered averages of rectified emg. *J Neurosci Methods* 79: 143–150, 1998.

Quallo MM, Kraskov A, Lemon RN. The activity of primary motor cortex corticospinal neurons during tool use by macaque monkeys. *J Neurosci* 32: 17351–17364, 2012.

Schieber MH. A spectrum from pure post-spike effects to synchrony effects in spike-triggered averages of electromyographic activity during skilled finger movements. *J Neurophysiol* 94: 3325–3341, 2005.

Smith WS, Fetz EE. Synaptic linkages between corticomotoneuronal cells affecting forelimb muscles in behaving primates. *J Neurophysiol* 102: 1040–1048, 2009.

Ventura V, Paciorek CJ, Risbey JS. Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. *J Climate* 17: 4343–4356, 2004.

Downloaded from on October 2, 2014