Journal of Neuroscience Methods 259 (2016) 1–12

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

Accurate detection of low signal-to-noise ratio neuronal calcium transient waves using a matched filter

Agnieszka F. Szymanska[a,*], Chiaki Kobayashi[b], Hiroaki Norimoto[b], Tomoe Ishikawa[b], Yuji Ikegaya[b,c], Zoran Nenadic[a]

[a] Department of Biomedical Engineering, University of California Irvine, Irvine, CA 92697, USA
[b] Laboratory of Chemical Pharmacology, Graduate School of Pharmaceutical Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[c] Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita City, Osaka 565-0871, Japan

HIGHLIGHTS

- We develop a matched filter for multi-unit calcium event detection in neurons.
- We tested the detector on simulated and experimentally recorded calcium imaging data.
- The detector had near perfect performance on simulated data with SNR as low as 0.2.
- The detector also performed well on experimentally recorded data.
- The detector is written in MATLAB and freely available.

ARTICLE INFO

Article history:
Received 10 June 2015
Received in revised form 6 October 2015
Accepted 27 October 2015
Available online 10 November 2015

Keywords:
Calcium transients
Detection
Multineuron calcium imaging
Matched filter
Low SNR
Dendritic spines
Somatic calcium fluctuations

ABSTRACT

Background: Calcium imaging has become a fundamental modality for studying neuronal circuit dynamics both *in vitro* and *in vivo*. However, identifying calcium events (CEs) from spectral data remains laborious and difficult, especially since the signal-to-noise ratio (SNR) often falls below 2. Existing automated signal detection methods are generally applied at high SNRs, leaving a large need for an automated algorithm that can accurately extract CEs from fluorescence intensity data of SNR 2 and below.
New method: In this work we develop a Matched filter for Multi-unit Calcium Event (MMiCE) detection to extract CEs from fluorescence intensity traces of simulated and experimentally recorded neuronal calcium imaging data.
Results: MMiCE reached perfect performance on simulated data with SNR $\geq$ 2 and a true positive (TP) rate of 98.27% ($\pm$ 1.38% with a 95% confidence interval), and a false positive(FP) rate of 6.59% ($\pm$ 2.56%) on simulated data with SNR 0.2. On real data, verified by patch-clamp recording, MMiCE performed with a TP rate of 100.00% ($\pm$ 0.00) and a FP rate of 2.04% ($\pm$ 4.10).
Comparison with existing method(s): This high level of performance exceeds existing methods at SNRs as low as 0.2, which are well below those used in previous studies (SNR $\simeq$ 5–10).
Conclusion: Overall, the MMiCE detector performed exceptionally well on both simulated data, and experimentally recorded neuronal calcium imaging data. The MMiCE detector is accurate, reliable, well suited for wide-spread use, and freely available at sites.uci.edu/aggies or from the corresponding author.

1. Introduction

Functional multineuron calcium imaging (fMCI) has been particularly successful as a means of studying the interactions of multiple (often hundreds) of neurons in large populations (Takahashi et al., 2007) with specific applications both *in vitro* (Cossart et al., 2003; Ikegaya et al., 2004; Sasaki et al., 2006, 2007, 2008), and *in vivo* (Kerr et al., 2005; Takahashi et al., 2012). Regardless of the details

* Corresponding author. Tel.: +1 5756406761.
E-mail addresses: aszymans@uci.edu (A.F. Szymanska), chi5.a1.ki6@gmail.com (C. Kobayashi), hi.norimoto@gmail.com (H. Norimoto), tomoe0178@gmail.com (T. Ishikawa), ikegaya@mol.f.u-tokyo.ac.jp (Y. Ikegaya), znenadic@uci.edu (Z. Nenadic).

http://dx.doi.org/10.1016/j.jneumeth.2015.10.014
0165-0270/

in fluorescence imaging, or focus on either synaptic or somatic calcium waves, calcium imaging in neurons often suffers from low signal-to-noise ratios (SNR). This makes signal detection very difficult, prompting the development of many automated or partially automated detection approaches.

Pioneers in the field first represented calcium imaging data as fluorescence changes ($\Delta F/F$), and established a link between somatic calcium transient waves and action potentials (APs) (Smetters et al., 1999; Kerr et al., 2005). Subsequent studies linked the amplitude of calcium transient waves to AP firing patterns and rates (Yaksi and Friedrich, 2006; Moreaux and Laurent, 2007; Greenberg et al., 2008). Although it is now universally accepted that relative amplitudes of fluorescent CEs are indicative of the frequency or number of APs fired by the cell, algorithms back-calculating AP firing rates are difficult to generalize. The amplitude of a given CE depends on the amount of fluorescent calcium indicator taken up by the cell, as well as the proximity of the region of interest (ROI) to the focal imaging plane. When performing imaging population studies, such as those using fMCI (Takahashi et al., 2007), it is impractical and often impossible to calculate the exact distance from the focal plane and to quantify the amount of indicator present in each cell or ROI. Comparison of CE amplitudes across ROIs or data sets is therefore unreliable. Furthermore, low-SNR signals where CEs are generated from 5 or fewer APs often fail to be detected (Moreaux and Laurent, 2007). When they are detected, the low-SNR CEs often lack the necessary resolution for AP firing rate estimation. Although Sasaki et al. overcame many of these issues with their support vector machine approach, they still fell victim to rapid performance degradation at low SNR (Sasaki et al., 2008). It may therefore be more advantageous to sacrifice precision in AP firing rate estimation for better accuracy in CE detection at low SNR.

Other approaches to CE detection less focused on AP firing rates include thresholding the time derivative of fluorescent changes in given ROIs (Ramdya et al., 2006; Ikegaya et al., 2004), Hanning filters (Cossart et al., 2003), and matched filters (Kerr et al., 2005). These approaches work well in high-SNR settings (SNR $\geq$10) but their performance is either not reported at low SNRs or quickly degrades. Since a large fraction of CEs fall below SNR 2, there is a pressing need to develop automated methods for low-SNR CE detection. Some denoising algorithms have been developed to enhance the SNR of calcium imaging data (Joucla et al., 2013; Malik et al., 2011). However, they are not applied or tested as detection tools, but rather as post hoc analyses. Due to the various shortcomings of these methods, as well as availability and implementation hurdles, manual CE detection remains the primary means of scoring calcium imaging data even though it is both prone to bias and incredibly laborious.

Our work focuses on the development of a Matched filter for Multi-unit Calcium Event (MMiCE) detection specifically tailored for low-SNR CEs. Unlike the matched filter presented by Kerr et al. (2005), the algorithm developed here is completely data driven. It is also applied and tested in environments with SNR as low as 0.2, well below the range of existing algorithms. MMiCE was tested under 3 paradigms: (1) simulated somatic calcium imaging data (ground-truth available), (2) experimentally recorded simultaneous somatic fMCI and patch-clamp data (ground-truth available), as well as (3) experimentally recorded, from now on referred to as real, somatic fMCI and dendritic spine fMCI data (no ground-truth available). Although only tested on fMCI data here, due to its data-driven design, the algorithm is easily generalizable to other fluorescent neural imaging modalities, such as voltage-sensitive-dye imaging (Szymanska et al., 2015). In order to ease implementation and applicability hurdles, MMiCE was developed in Matlab with an intuitive graphical user interface. MMiCE, along with a tutorial, is openly available at sites.uci.edu/aggies/downloads or upon request from the corresponding author.

2. Methods

2.1. Tissue preparation

All animals used in this study were treated according to The University of Tokyo guidelines for the care and use of laboratory animals. All performed experiments were approved by the experiment ethics committee at the University of Tokyo, approval numbers: P24-5 and P24-8.

Acute Slices. Acute slices were prepared as described in Ueno et al. (2002), Norimoto et al. (2012). Briefly, 400 μm horizontal slices of the hippocampus from 3 week old C57BL/6J mice were prepared using a vibratome in ice-cold oxygenated cutting solution. The slices then rested in oxygenated artificial cerebrospinal fluid (aCSF) at room temperature for 1.5 h prior to imaging. For more details on the preparation and solutions used please see Ueno et al. (2002), Norimoto et al. (2012).

Ex vivo cultures. In order to facilitate the simultaneous visibility of many dendritic spines during functional multispine calcium imaging (fMCI), organotypic slice cultures were used in this study. *Ex vivo* rat hippocampal slice cultures were prepared as described in Takahashi et al. (2012) from 7 day old Wistar/ST rats. Briefly, 300 μm entorhinal-hippocampal organotypic slices were cut using a vibratome, placed on Omnipore membrane filters (JHWP02500; Millipore, Bedford, Massachusetts, USA), and incubated (5% CO_2, 37 °C) in culture medium (50% minimal essential medium, 25% Hanks' balanced salt solution, 25% horse serum, antibiotics) for 12–19 days prior to imaging. The medium was changed every 3.5 days. For more details please refer to Takahashi et al. (2012).

2.2. Dye loading

Simultaneous somatic fMCI and patch-clamp. Ex vivo slice cultures were transferred into a dish (35-mm diameter) containing 2 ml of the dye solution and were incubated for 1-h in a humidified incubator at 35 °C under 5% CO_2. The dye solution was aCSF containing 10 μl of 0.1% Oregon Green BAPTA1-AM (OGB1) dissolved in DMSO, 2 μl of 10% Pluronic F-127/DMSO and 2 μl of 5% Cremophor EL/DMSO. After being washed, the cultured slices were incubated at 35 °C for 40 min and were mounted in a recording chamber.

Somatic fMCI. Acute slices were loaded locally with OGB1 dissolved in DMSO containing 10% Pluronic F-127 to yield a concentration of 2 mM. Immediately before use, this solution was 10 diluted with aCSF and loaded into pipettes (3–5 MΩ). The tip of the pipette was inserted into the CA1 stratum pyramidale, and a pressure was applied using a 10-ml syringe pressurizer (50–60 hPa for 5 min).

Dendritic spine fMCI. CA3 pyramidal neurons were selected for spine imaging from *ex vivo* slice cultures. Selected neurons were voltage-clamped at −30 mV (MultiClamp 700B amplifier and a Digidata 1440A digitizer controlled by pCLAMP 10.4 software) and loaded with a Fluo-4 solution (97.3 $CsMeSO_4$, 42.7 CsCl, 10 HEPES, 10 phosphocreatine, 4 MgATP, 0.3 NaGTP, and 0.2 Fluo-4, all in mM). The voltage clamp facilitated channel currents mediated by NMDA receptors and calcium-permeable AMPA receptors.

2.3. Patch-clamp recording

CA3 pyramidal neurons selected for simultaneous somatic fMCI and patch-clamp recordings were voltage-clamped at 0 mV (Axopatch 700B amplifier) using a borosilicate glass pipettes (4–9 MΩ) filled with aCSF, and recorded from at a sampling frequency of 20 kHz (MultiClamp 700B amplifier and a Digidata 1440A digitizer controlled by pCLAMP 10.4 software).

2.4. Optical recording

Fluorophores were excited at 488 nm with a laser diode (HPU50101PFS, FITEL, Tokyo, Japan) and visualized using a 507-nm long-pass emission filter. Videos were taken at 10 Hz for the simultaneous somatic fMCI and patch-clamp data and at 50 Hz for the somatic fMCI and dendritic spine fMCI data using a Nipkow-disk confocal microscope (CSU-X1; Yokogawa Electric, Tokyo, Japan), and a cooled EM-CCD camera (iXon DU897, Andor, Belfast, UK). A 16× objective was used for somatic fMCI data and a 60× objective was used for the dendritic spine fMCI data (CFI75LWD16xW and CFI75LWD60xW Nikon, Tokyo, Japan).

One data trace was acquired for simultaneous somatic fMCI and patch-clamp (fMCI-PC) recordings. The recording lasted 300-s and measured spontaneous activity from a CA3 pyramidal neuron from the *ex vivo* slice cultures. Three somatic fMCI data sets were acquired measuring spontaneous activity from CA1 pyramidal neurons from the acute slices. Recordings for each data set lasted 60-s and contained from 75 to 190 ROIs. Five dendritic spine fMCI data sets were acquired measuring spontaneous synaptic inputs from CA3 pyramidal neurons from the *ex vivo* slice cultures. Recordings ranged from 54-s to 111-s and contained from 157 to 313 ROIs. Only spines located within 200 μm from the soma were monitored to avoid the space clamp problem.

2.5. Optical data pre-processing

ROIs for all somatic fMCI and dendritic spine fMCI data sets were identified manually using custom software in Microsoft Visual Basic (Ikegaya et al., 2004). A single analyst (Analyst 5 from Table 4) identified all ROIs for the somatic fMCI data sets, and another analyst (Analyst 6 from Table 4) identified all ROIs for the dendritic spine fMCI data sets. These common sets of ROIs were used in all further analysis. The diameter of each ROI was tailored to the visible size of the neuron soma or dendritic spine being identified. For fMCI-PC data, the ROI was drawn to exclude the micropipette.

The average fluorescence for each ROI was then calculated and used to determine the change in fluorescence, $\Delta F/F = (F_1 - F_0)/F_0$, where F_1 is the fluorescence intensity at any time point, and F_0 is the average baseline fluorescence intensity 1-s before and after F_1. This normalization with a 2-s epoch around the target frame is needed to compensate for photobleaching. The resulting fluorescence intensity trace, X, for each ROI was then used for further analysis.

2.6. Manual CE identification

Somatic and dendritic spine fluorescence intensity traces, X, derived as described in Section 2.5, were independently plotted in Matlab. Six trained human analysts then identified CE peak times from the traces. On average, the human analysts spent 24 working hours identifying CEs in a given data set, where each data set contained between 75 and 313 ROIs (179 ROIs on average). CEs were not manually identified for fMCI-PC data.

2.7. Matched filter design and implementation

The matched filter for multi-unit calcium event detection (MMiCE) was implemented in Matlab and is freely available along with a graphical user interface tool at sites.uci.edu/aggies/downloads or from the corresponding author.

2.7.1. General likelihood ratio test

According to the Neyman–Pearson lemma, the likelihood ratio test (LRT) is the most powerful discriminant of two underlying models from a deterministic signal. In the case of calcium signals, the two models being investigated represent noise-only data, and data containing both CEs and noise. More formally, CE detection can be interpreted as a hypothesis testing problem, where under the null hypothesis, H_0, the signal contains noise only, and under the alternative hypothesis, H_1, the signal contains both a CE and noise. Given a signal of length N, where N is defined as the number of samples spanned by a CE, we can express the hypotheses mathematically as

$$H_0: \quad x = n$$
$$H_1: \quad x = s + n$$

where $x \in \mathbb{R}^{1\times N}$ is a signal of length N, $s \in \mathbb{R}^{1\times N}$ is a CE, and $n \in \mathbb{R}^{1\times N}$ is zero-mean noise. The LRT for this problem can now be expressed as

$$L(x) = \frac{p(x \mid H_1)}{p(x \mid H_0)} \quad , \quad \begin{matrix} H_1 & \text{if} & L(x) > \gamma_o \\ H_0 & \text{if} & L(x) < \gamma_o \end{matrix}$$

where $x \in \mathbb{R}^{1\times N}$ is the fluorescent signal, and γ_o is the threshold.

2.7.2. General matched filter

Assuming Gaussian noise statistics, the LRT takes on the form of a generalized matched filter (GMF)

$$S(x) = s\Sigma^{-1}x^T \quad , \quad \begin{matrix} H_1 & \text{if} & S(x) > \gamma \\ H_0 & \text{if} & S(x) < \gamma \end{matrix} \tag{1}$$

where $\Sigma \in \mathbb{R}^{N\times N}$ is the noise covariance, and γ, which subsumes γ_o and data-independent terms, is the threshold. Note that the test statistic, $S(x)$, is linearly dependent on the signal x. Even if the noise is not Gaussian, $S(x)$ has the highest achievable SNR of all other linear combinations of the data (Kay, 1998).

2.7.3. Sliding window approach

The matched filter described in Section 2.7.2 performs a hypothesis test on $x \in \mathbb{R}^{1\times N}$ where N, as stated in Section 2.7.1, is defined as the number of samples spanned by a typical CE. To test a full fluorescence intensity trace, $X \in \mathbb{R}^{1\times T}$, where $T \gg N$, we have to apply the matched filter iteratively across the entire trace. This type of approach is referred to as a sliding window. Initially the filter tests a segment of the signal $[X(1), X(2), \ldots, X(N)]$. The filter output given this window is a scalar that indicates the likelihood of a CE in the center of the window, $S(N/2)$. Once finished, the test segment is advanced by one sample to $[X(2), X(3), \ldots, X(N+1)]$ and this new segment is filtered. The process is repeated until the full data set X has been filtered.

2.7.4. White Gaussian noise assumption

The $N \times N$ parameters of the covariance matrix Σ may be difficult to estimate reliably given a finite amount of data available for conditioning. If necessary, we can reduce the number of parameters by making the simplifying assumption that the noise statistics are temporally white. This implies that each noise sample is not dependent on any sample that came before or after it. Under the white Guassian noise (WGN) assumption all parameters, $\Sigma_{i,j}$, of the covariance matrix Σ, are zero if $i \neq j$. The remaining parameters $\Sigma_{i,j}$ with $i = j$ represent each sample point's autocorrelation, which

is equivalent to the noise variance, σ^2. The covariance matrix then takes on the form

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_{N\times N}$$

where σ is the only parameter being estimated from the noise conditioning sample. Our generalized matched filter is then reduced to a matched template

$$S(x) = s\frac{1}{\sigma^2}x^T \quad , \quad \begin{matrix} H_1 & \text{if} & S(x) > \gamma \\ H_0 & \text{if} & S(x) < \gamma \end{matrix} \tag{2}$$

2.7.5. Conditioning protocol

The MMiCE detector is completely data driven, allowing it to maintain high performance even if large variations exist between data sets. Both s and Σ from Eqs. (1) and (2) are estimated from the data. First, 20 high-SNR CEs (N = 1.2-s for real data, and N = 1.6-s for simulated data) are identified manually, aligned to their peak values, and averaged to estimate s for each data set. In order to ensure that the entire data set is well represented, the CEs are always selected from multiple different ROIs (when possible), all of which exhibit high SNR. The CEs were also selected to match the traditional sharp rise and slow decaying shape of a calcium transient wave as much as possible. The specific SNR values and CE shapes available for estimating s varied between data sets. Examples of s from both somatic fMCI and dendritic spine fMCI data are shown in Fig. 1(A). Due to the limited CEs available in the single fMCI-PC data trace acquired, only 10 CEs were selected for conditioning this data.

Next, 20 noise-only samples of various lengths, from multiple different ROIs, were identified manually for each data set. The noise samples were then subdivided into noise windows (N = 1.2-s for real data, and N = 1.6-s for simulated data). The number of noise windows varied between real data sets depending on the lengths of the 20 noise samples selected for conditioning. The conditioning noise samples were identical among all simulated data sets. Any data points remaining after the subdivision were discarded. Auto-covariance sequences, $r(k)$, were then calculated at lags $k_{real} \in \{-0.58\text{-s}, -0.56\text{-s}, \ldots, 0.58\text{-s}\}$ for real data sets and $k_{sim} \in \{-0.78\text{-s}, -0.76\text{-s}, \ldots, 0.78\text{-s}\}$ for simulated data sets, for each noise window. The sequences were averaged across all noise windows for a given data set, and used to generate Σ for that data set.

The size of s and each noise window was empirically selected as N = 1.2-s for real data, and N = 1.6-s for simulated data. The real data window, N = 1.2-s, ensured that most somatic and spine CEs were captured in full, although some data sets did exhibit both wider and slimmer CEs (independently of data type). The somatic CEs used to generate the simulated data set were specifically selected to have a high SNR. These CEs tended to be wider than the average CE, and therefore the window size for the simulated data was increased to N = 1.6-s. These window sizes are consistent with known typical CE durations for AP bursts in neurons (Kerr et al., 2005; Smetters et al., 1999).

2.7.6. Setting detection thresholds

The 8 fMCI data sets used in this study (excluding the fMCI-PC data trace) contained 179 ROIs on average, ranging from 75 to 313 ROIs. Because calcium indicator concentration is not uniform across all ROIs, and the independent ROIs may not all perfectly be in the focal plane of the camera, the resulting fluorescence intensity traces exhibit varying amplitudes as well as varying SNRs. Therefore a single filter threshold cannot be set for the entire data set. Similarly, it would be too laborious to manually set the filter threshold for each ROI in a data set. It is therefore necessary to devise an automated thresholding algorithm that can account for varying SNRs and amplitudes throughout a given data set.

To accomplish this goal, threshold values were calculated independently for each fluorescence intensity trace X_r, where $r \in \{1, 2, \ldots, R\}$, and R is the total number of ROIs in the data set. The thresholding parameters are calculated from the filtered signal which contains both CEs and noise. Because CEs are statistically sparse within the noise, we can effectively approximate the mean and standard deviation of the filtered noise for each X_r by respectively calculating the filter output's median, M_r, and median-based standard deviation, $\sigma_{M,r}$ (Nenadic and Burdick, 2005), with

$$\sigma_{M,r} = \frac{\mathcal{M}\left\{|\, S_r(1)) - M_r \,|, \ldots, |\, S_r(T) - M_r \,|\right\}}{\sqrt{2}Erf^{-1}(1/2)} \tag{3}$$

where S_r is the filter output, $\mathcal{M}$ is the median operator, and T is the number of time samples spanning X_r. This median-based approximation provides a more robust estimate of the filter output's noise parameters than simply taking the mean and standard deviation of the filter output. The threshold, γ_r is then set as a multiple of $\sigma_{M,r}$, above the median, M_r

$$\gamma_r = M_r + a\sigma_{M,r} \tag{4}$$

where a is the multiple and will be referred to as the thresholding sensitivity for the remainder of this article.

This thresholding approach generates thresholds that can accommodate varying signal amplitudes and SNRs for each X_r, and

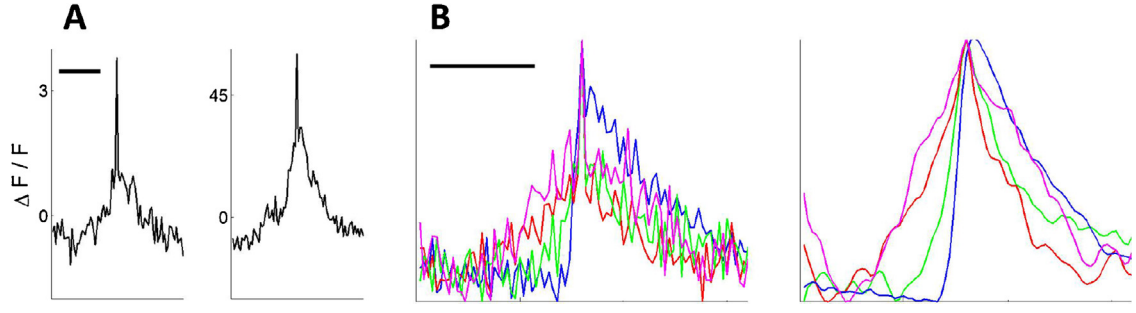

Fig. 1. (A) Examples of conditioning templates s from (*Left*) somatic fMCI data set 1 and (*Right*) dendritic spine fMCI data set 1. The scale bar corresponds to 0.5-s for both figures. Note that the dendritic spine template is wider and has a much higher amplitude than the somatic template. (B) CE templates used for generating simulated data, with s_1, s_2, s_3, and s_4 in blue, red, green, and magenta respectively. (*Left*) Normalized and unfiltered templates. (*Right*) Normalized and filtered (5 Hz low-pass) templates. The scale bar corresponds to 0.5-s for both figures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

allows the user to set a specific sensitivity value, a, for the entire data set.

2.8. Generating simulated data

In order to assess the MMiCE detector's performance at varied SNR levels and to establish a ground truth for comparison, we simulated 8 fluorescence traces such as those described in Section 2.5, each with a different average SNR. To capture as much of the real data characteristics as possible, the simulated traces, $X_{\mathrm{Sim},i}$ where $i \in \{1, 2, \ldots, 8\}$, were generated using fluorescence traces from the calcium imaged neuronal somas described in Sections 2.1–2.5.

Four different CE shapes were identified manually by human analysts from three somatic fMCI data sets (Section 2.1–2.5). 20 examples of each shape were then selected from among the three data sets, averaged, low-pass filtered at 5 Hz, and normalized to construct 4 distinct CE templates (Fig. 1(B)). The low-pass filter was necessary because averaging did not sufficiently smooth the CE templates. The CE templates, s_c, $c \in \{1, 2, 3, 4\}$, were cropped to lengths of 1-s, 1.3-s, 1.2-s, and 1.3-s respectively.

Similarly, 80 samples of varying length noise-only data were manually selected from the three somatic fMCI data sets. Each sample was normalized to be zero mean and unit variance. The noise samples were then all concatenated into a single time series of length $T_{noise} = 89{,}000 \simeq 29.7$ min at a sampling rate of 50 Hz. The four CE templates were superimposed with the noise at random time points such that no spikes overlapped, and there were 100 CEs of each spike shape for a total of 400 CEs.

The SNR level was controlled by scaling the sizes of the CE templates being added to the noise. Here we define SNR as

$$SNR = \frac{E((f \cdot s)^2)}{\sigma^2_{noise}} = f^2 E(s^2) \tag{5}$$

where E is the expectation operator, $s \in \mathbb{R}^{1\times N}$ is the CE template, f is a scaling factor, and $\sigma^2_{noise} = 1$ is the noise variance. We can then solve for the scaling factor $f_{i,c}$ for each SNR_i and each CE template s_c.

The fluorescence traces were simulated at SNR levels of 0.2, 0.5, 1, 2, 4, 6, 9, and 14, where SNR 1 indicates that the signal has the same power as the noise. The scaling factor for each template at each SNR level is listed in Table 1. Fig. 2(A) shows a 30-s excerpt containing 6 simulated CEs from each simulated data trace. Note that the only difference between each simulated data trace is the SNR of the CEs; the noise and CE times remain constant.

2.9. Performance analysis

CEs detected for each of the simulated data traces, as well as those detected from the fMCI-PC data trace, were compared against the ground-truth spike times. In the case of the fMCI-PC data, the ground-truth spike times were identified as the peak times of the first APs in an AP burst, where a single AP burst consisted of APs firing within 0.1-s of each other. The first 3.3 min of the simulated data traces were used for conditioning, and the remaining test data (26.4 min) contained 346 CEs. Similarly, 121-s of the fMCI-PC data trace was used for conditioning and the remaining test data (179-s) contained 48 AP bursts. A detected CE was considered a true positive (TP) if the detected peak time was within 0.8-s of the true peak time. This corresponds to half of the length of the simulated CEs. Otherwise, the detected CE was considered a false positive (FP). Vector tp was used to represent all of the simulated CEs, with each index assigned a value of 1 if the CE was detected (TP) and 0 otherwise. Similarly, vector fp was used to represent all of the detected CEs, with each index assigned a value of 1 if the CE was a FP and 0 otherwise. TP and FP rates for 100 incrementally increasing thresholds were then calculated as the means of tp and fp, respectively, at each of the thresholds. The false negative (FN) or omission rate can be calculated as (1 – TP Rate). The TP and FP rates at each threshold were then used to generate receiver operating characteristic (ROC) curves for each simulated data trace.

Table 1
Scaling factor f for each template at each SNR level.

	Scaling factor f			
	$c = 1$	$c = 2$	$c = 3$	$c = 4$
SNR level				
0.2	0.83	1.13	1.01	0.87
0.5	1.30	1.79	1.60	1.37
1	1.85	2.53	2.26	1.94
2	2.61	3.58	3.20	2.74
4	3.69	5.07	4.53	3.88
6	4.52	6.20	5.55	4.75
9	5.53	7.60	6.79	5.82
14	6.90	9.48	8.47	7.26

3. Results

3.1. Detection on simulated data

3.1.1. Conditioning for simulated data

The simulated data traces were all treated as independent data sets, and the MMiCE detector was separately conditioned on each set. In general, a real data set will have a wide range of SNR levels, usually with some high-SNR spikes to pick for conditioning. The presence of low-SNR spikes in the conditioning sample may result in a poorly determined template s (Section 2.7), which would adversely affect detection. However, in this case all of the CEs in each simulated data set have roughly the same SNR. The conditioning for each SNR level was kept separate in order to prove that the MMiCE detector was accurate even if the CE conditioning samples were of poor quality and very low SNR.

The first 3.3 min of the simulated data traces were used for conditioning. Conditioning CEs were selected with prior knowledge of simulated CE peak times. Noise samples for conditioning were selected manually, and were the same for all $X_{\mathrm{Sim},i}$. The conditioning noise was subdivided into 56 noise windows ($N = 1.6$ – s), and auto-covariance sequences from each window were then averaged to generate the noise covariance matrix Σ_{Sim} (Section 2.7.5). The data used for conditioning was omitted when assessing MMiCE detection performance.

3.1.2. MMiCE detector performance on simulated data

The simulated data sets were filtered using the MMiCE detector and CE times were identified as the peaks of the filter output above a given threshold. Performance was assessed at 100 incrementally increasing thresholds with sensitivity values, a, ranging from 0.5 to 50.0 (Eq. (4)). The resulting TP and FP rates were used to generate ROC curves for all 8 simulated data sets (Figs. 2(B) and 3). The MMiCE detector performed exceptionally well on the simulated data, even at SNR levels as low as 0.2. Performance metrics are listed as TP or FP Rate (95% confidence interval).

WGN application. The MMiCE detector was first applied in its simplest form, assuming WGN statistics (Section 2.7.4, Eq. (2)). Under this assumption, the MMiCE detector achieved perfect performance (TP = 100 (0.00)%, FP = 0 (0.00)%) for SNR $\geq$ 2. At SNR 1 the performance was slightly affected (TP = 100 (0.00)%, FP = 0.29 (0.57)%), with a further dip in performance at SNR 0.5 (TP = 97.98 (1.49)%, FP = 0.88 (0.99)%) and SNR 0.2 (TP = 90.46 (3.11)%, FP = 12.81 (3.47)%). The MMiCE detector's ROC curves, depicting performance at all thresholds under the WGN assumption, are shown for each data set in Fig. 2(B). The best performances at each SNR level, and the corresponding threshold sensitivity values are shown in Table 2.

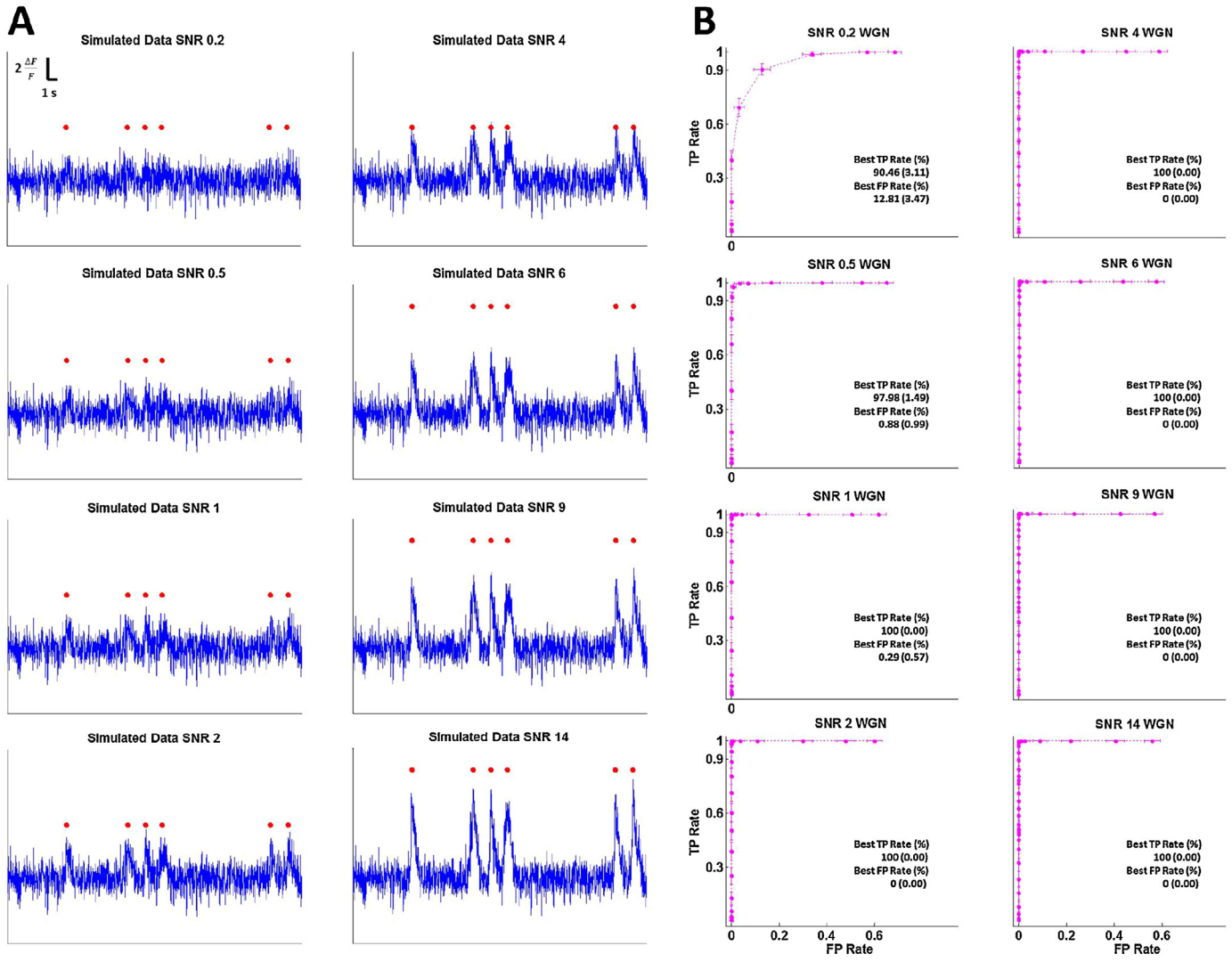

Fig. 2. (A) 30-s excerpts of simulated data for SNR 0.2, 0.5, 1, 2, 4, 6, 9, and 14. Red circles above each graph represent the peak locations of 6 inserted CE templates. Going from left to right the inserted templates are s_1, s_4, s_1, s_4, s_1, s_3. (B) Receiver operating characteristic (ROC) curves for all simulated data sets, across 100 incrementally increasing thresholds, with the MMiCE detector assuming WGN statistics. The error bars represent 95% confidence intervals. The optimal performance for each simulated data set is presented under each curve with 95% confidence intervals in parenthesis. Note that perfect performance (TP = 100 (0.00) %, FP = 0 (0.00) %) is reached for SNR 2–14. Performance only begins degrading significantly at SNR 0.2 (TP = 90.46 (3.11) %, FP = 12.81 (3.47) %). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Full covariance application. To see if even better outcomes could be achieved for SNR 0.2–2, the MMiCE detector was then applied in its full complexity, using the full noise covariance to construct the matched filter (Section 2.7.2). For simulated data sets of SNR 1 and SNR 2, performance was not affected. This is due to the fact that performance was already perfect for SNR 2 and very near perfect for SNR 1 under the WGN assumption. Performance was improved for SNR 0.5, with TP = 99.71 (0.57)% (from 97.98 (1.49)% before), and FP = 1.15 (1.12)% (from 0.88 (0.99)% before). Although the FP Rate slightly increased, overall this is closer to perfect performance than under the WGN assumption. An even greater improvement was seen for SNR 0.2, with TP = 98.27 (1.38)% (from 90.46 (3.11)% before), and FP = 6.59 (2.56)% (from 12.81 (3.47)% before). The MMiCE detector's ROC curves, depicting performance at all 100 thresholds when the detector employs the full noise covariance for conditioning, are shown in Fig. 3 for simulated data sets of SNR 0.2–2. The best performance at each SNR level, and the corresponding threshold sensitivity values for SNR 0.2–2 are shown in Table 2.

3.2. Detection on simultaneous fMCI and patch-clamp data

3.2.1. Conditioning for fMCI-PC data

A single simultaneous somatic fMCI and patch-clamp data trace was used for this experiment. 121-s of the trace was used for conditioning. Due to the limited amount of data available, only 10 CEs with relatively high SNRs, and smooth shapes could be manually selected to generate the conditioning template, s. Similarly, the 20 manually selected noise samples were relatively short, therefore only 91 noise windows were available for conditioning (Table 3). This under-sampled conditioning set could result in a poorly determined s as well as Σ, and thereby adversely affect detection performance. The average SNR of the conditioning CEs was calculated as

$$SNR_{CE} = \frac{E(CE^2)}{\sigma^2_{Noise}} \tag{6}$$

where σ^2_{Noise} was calculated from the conditioning noise samples, and $E(CE^2)$ is the expected value of a CE^2. Note that unlike in

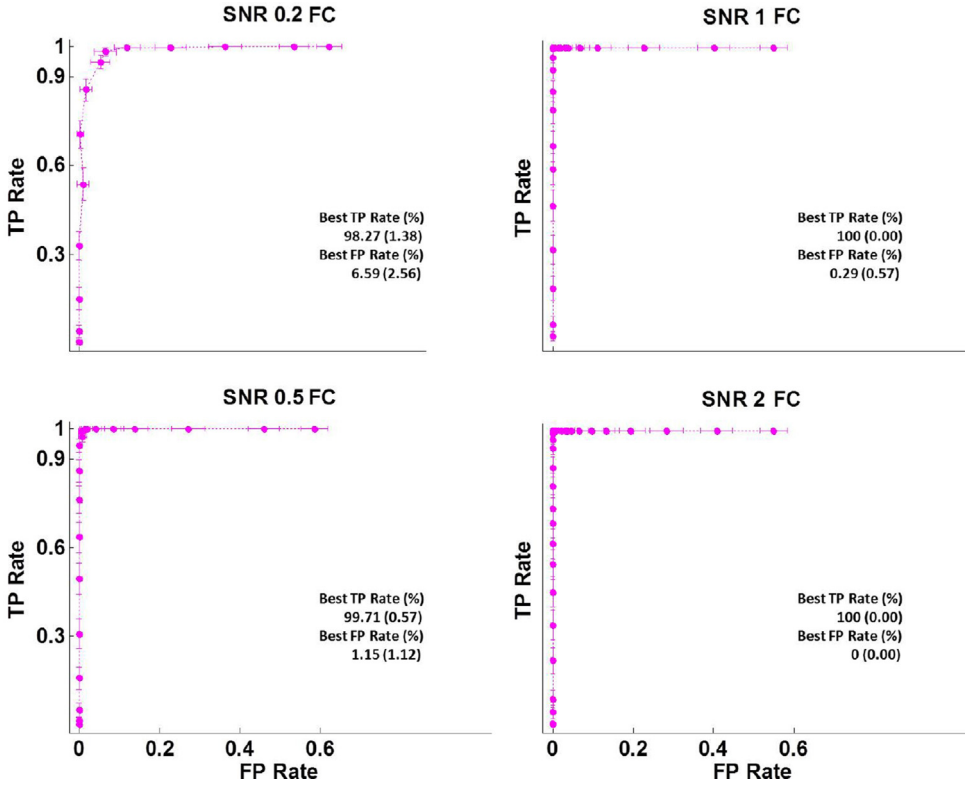

Fig. 3. ROC curves for simulated data sets of SNR 0.2–2, across 100 incrementally increasing thresholds, with the MMiCE detector employing the full noise covariance. The error bars represent 95% confidence intervals. The optimal performance for each simulated data set is presented under each curve with 95% confidence intervals in parenthesis. Note that perfect performance (TP = 100 (0.00)%, FP = 0 (0.00)%) and near perfect performance (TP = 100 (0.00)%, FP = 0.29 (0.57)%) is still reached for SNR 2, and SNR 1, just as in the WGN case. However, performance improved significantly from the WGN case for SNR 0.5 (TP = 99.71 (1.49)%, FP = 1.15 (1.12)%) and SNR 0.2 (TP = 98.27 (1.38)%, FP = 6.59 (2.56)%).

Table 2
Best MMiCE detector performances, and the associated threshold sensitivity values *a* for each simulated data set. Both performances under the WGN assumption and using the full noise covariance are shown, with 95% confidence intervals in parenthesis. Because perfect performance was achieved under the WGN assumption for SNR 4 and above, the full noise covariance detector was not applied to these data sets. This is reflected by the lack of threshold sensitivity values for SNR 4–14 under full covariance.

	Best MMiCE detector performance					
	WGN			Full covariance		
	TP %	FP%	*a*	TP%	FP%	*a*
SNR level						
0.2	90.46 (3.11)	12.81 (3.47)	2	98.27 (1.38)	6.59 (2.56)	3
0.5	97.98 (1.49)	0.88 (0.99)	3.5	99.71 (0.57)	1.15 (1.12)	4.5
1	100 (0.00)	0.29 (0.57)	4	100 (0.00)	0.29 (0.57)	7
2	100 (0.00)	0 (0.00)	4-8	100 (0.00)	0 (0.00)	10-11
4	100 (0.00)	0 (0.00)	4.5-13	100 (0.00)	0 (0.00)	–
6	100 (0.00)	0 (0.00)	5-17.5	100 (0.00)	0 (0.00)	–
9	100 (0.00)	0 (0.00)	8-22.5	100 (0.00)	0 (0.00)	–
14	100 (0.00)	0 (0.00)	4.5-29	100 (0.00)	0 (0.00)	–

Eq. (5), the CE in this case contains a noise component. Therefore the SNRs calculated from the real data will be biased. Due to this bias an SNR of 2 calculated from the real data will roughly correspond with an SNR of 1 calculated for the simulated data. The average SNR of the fMCI-PC conditioning CEs is 8.23 (Table 3). All data selected for conditioning were omitted from the performance analysis.

3.2.2. *MMiCE detector performance on fMCI-PC data*

The ground-truth CEs for the fMCI-PC data trace, identified as described in Section 2.9 had an average SNR of 10.54. As explained in Section 3.2.1, this roughly corresponds with simulated data of SNR 9. As in the simulated data case, the fMCI-PC data trace was filtered using the MMiCE detector under the WGN assumption as well as using the full noise covariance. Performance for both approaches was assessed at 100 incrementally increasing thresholds with sensitivity values, *a*, ranging from 0.5 to 50.0 (Eq. (4)). The resulting ROC curves for both the WGN case, and using the full noise covariance are shown in Fig. 4 *Left*. Best performance under the WGN assumption was TP = 95.83 (5.86)% and FP = 2.13 (4.28)%. The corresponding *a* range was 3.5–4. This constitutes a total of 3 errors, 1 false positive, and 2 false negatives or omissions. The false positive was a result of misidentifying an artifact of photobleaching as a CE resulting from cell activity. The first false negative was a failure of the MMiCE detector to identify an AP burst that resulted in fMCI activity of SNR 2.07, which is much lower than average for this data trace, and as explained in Section 3.2.1 is equivalent to SNR 1 in the simulated data. The second false negative was the MMiCE detector's failure to identify one of two overlapping CEs. The first CE was identified but the second was omitted. Best performance using the full noise covariance was TP = 100.00 (0.00)%

Table 3
Conditioning sample information for each real data set. Average conditioning CE SNR is the average SNR of the CEs selected for conditioning for a given data set. The number of windows the conditioning noise sample for each data set was split into is also listed.

Data set		Average cond. CE SNR	Number cond. noise windows
fMCI-PC	1	8.23	91
Somatic fMCI	1	1.80	328
	2	1.22	159
	3	1.51	152
Dendritic spine fMCI	1	2.51	162
	2	3.84	124
	3	2.95	183
	4	8.87	116
	5	2.34	211

and FP = 2.04 (4.10)% at $a = 2.0$. This constitutes 1 error, which is the same false positive as in the WGN case, i.e. misidentifying an artifact of photobleaching as a CE resulting from cell activity. The two false negatives, however, were resolved when using the full noise covariance. An excerpt of the spike train as determined by patch-clamp recordings, the fMCI data trace, detected CEs, and the filter output using the full noise covariance are shown in Fig. 4. Overall, the MMiCE detector did exceptionally well in CE detection from somatic fMCI data, as verified via patch-clamp recording.

3.3. Detection on somatic fMCI and dendriti spine fMCI data

The CE SNR in fMCI-PC data used for validation of the MMiCE algorithm in Section 3.2 was 10.54. Although CEs of this SNR do appear for some ROIs in many data sets, they are not common. Therefore MMiCE detector performance was also tested in more realistic SNR settings with SNR ranging from 1 to 3 in real fMCI data. Three somatic fMCI and 5 dendritic spine fMCI data sets were used for this experiment. Note that no ground-truth was available for this data.

3.3.1. Conditioning for real data

The MMiCE detector was separately conditioned for each data set. Because the fluorescent trace, X_r, for a given ROI, r, is too short to serve as an adequate conditioning sample, all of the fluorescent traces for a given data set were concatenated into a single time series for conditioning. For each data set, 20 CEs with relatively high SNRs, and smooth shapes were then manually selected from each concatenated data and used to generate conditioning templates s_l, for each data set l. An example of both a somatic and spine template is shown in Fig. 1(A). The concatenated data were also used to manually select 20 conditioning noise samples for each data set. The length of each noise sample varied depending on the availability of long uninterrupted noise segments in the data. The average SNR of each data set's conditioning CEs, as well as the number of windows the conditioning noise could be subdivided into, is listed in Table 3. The SNR was calculated as shown in Eq (6). All data selected for conditioning were omitted from the performance analysis.

3.3.2. Manual CE identification in real data

Six trained analysts, two of whom also collected the fMCI data, manually tagged CEs in the 3 somatic and 5 spine data sets. The average CE SNR for each data set and each analyst is shown in Table 4. The SNR was calculated as in Eq. (6), with σ^2_{Noise} calculated from the conditioning noise sample from each corresponding data set. Agreement between analysts was defined as the percentage of CEs that all analysts who worked with the data set agreed are CEs and not noise. The agreement % for each analyst and each data set is listed in Table 4.

On average, the manually identified CEs in the somatic fMCI data sets had an SNR of 1.53. Similarly, the CEs identified in the spine data sets had an average SNR of 2.43. As mentioned previously, this very low SNR level is fairly typical of neural imaging data. Although some ROIs exhibited SNR levels as high as 30, this was very rare.

More importantly, the level of agreement between analysts was very low. For the somatic data sets, on average, analysts only agreed on 50.48% of the identified CEs. This means that on average 49.52% of CEs identified by any analyst in a given data set, were considered noise by at least one other analyst working with that data set. Similarly the average agreement rate between analysts for the spine data sets was only 31.52%. These results are even worse, meaning that 68.48%, or over two-thirds, of CEs identified by any analyst were considered noise by at least one other analyst. Examples of the kinds of CEs tagged by the analysts are shown in Fig. 5.

Overall, the analysts were not confident in their ability to accurately identify CEs at SNR levels of 3 and below, which constitute most of the events present in these data sets. The lack of agreement among analysts also shows that manual CE identification is unreliable and error-prone. As has also been concluded by Sasaki

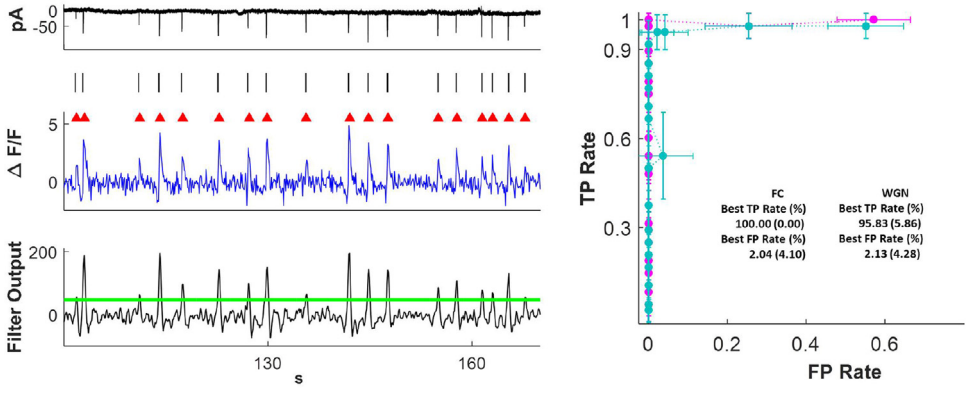

Fig. 4. MMiCE detector performance and real somatic fMCI data as verified by patch-clamp recordings. (*Left*) The top shows the ground-truth patch-clamp recordings. In the middle is the corresponding ΔF/F data with MMiCE CE peaks detected at the optimal threshold ($a = 3.5$) marked by red triangles. The spike-train derived from the patch-clamp recordings is shown above the detected CEs for easy comparison. The bottom trace shows the filter output using the full noise covariance and the optimal detection threshold as determined by the ROC curves presented on the *Right*. (*Right*) ROC curves for the fMCI-PC data shown on the *Left*, across 100 incrementally increasing thresholds. The two curves depict the MMiCE detector's performance under the WGN assumption (cyan) and when using the full noise covariance (FC, magenta). The error bars represent 95% confidence intervals. Optimal performances with 95% intervals in parenthesis are presented under the curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 4
Average SNRs of manually identified CEs for each analyst and each fMCI data set. Not all analysts identified CEs in all available data sets. The data sets not analyzed by a given analyst are shown as blanks. The agreement % among analysts for each data set is also shown. The agreement % is the fraction of CEs that all analysts who worked with the data set agree are CEs and not noise.

Data set		SNR Analyst							Agreement % Analyst						
		1	2	3	4	5	6	Avg.	1	2	3	4	5	6	Avg.
Somatic fMCI	1	1.59	1.35	1.34	1.28	2.17	–	1.55	58.20	53.38	65.03	48.77	73.26	–	59.93
	2	2.20	1.45	1.53	2.01	1.93	–	1.83	71.79	50.79	67.11	74.93	58.81	–	64.69
	3	1.37	1.14	1.11	1.19	1.30	–	1.22	47.83	18.97	28.28	6.96	31.43	–	26.83
	Average	1.72	1.31	1.33	1.50	1.80	–	**1.53**	59.27	41.05	53.70	43.89	54.50	–	
Dendritic spine fMCI	1	1.88	1.79	1.78	–	–	2.47	1.98	19.05	13.46	25.47	–	–	38.07	24.01
	2	2.84	2.11	2.17	1.54	–	2.91	2.31	41.33	29.48	38.26	9.86	–	42.06	32.20
	3	2.23	1.77	2.11	1.35	–	2.20	1.93	34.01	22.83	29.45	6.95	–	42.43	27.15
	4	3.23	4.47	5.23	3.19	–	5.07	4.24	31.84	53.22	67.70	29.21	–	60.01	48.40
	5	1.42	1.49	1.48	–	–	2.07	1.62	22.73	33.21	17.63	–	–	48.84	30.60
	Average	2.32	2.33	2.55	2.03	–	2.94	**2.43**	29.79	30.44	35.70	15.34	–	46.28	**31.51**

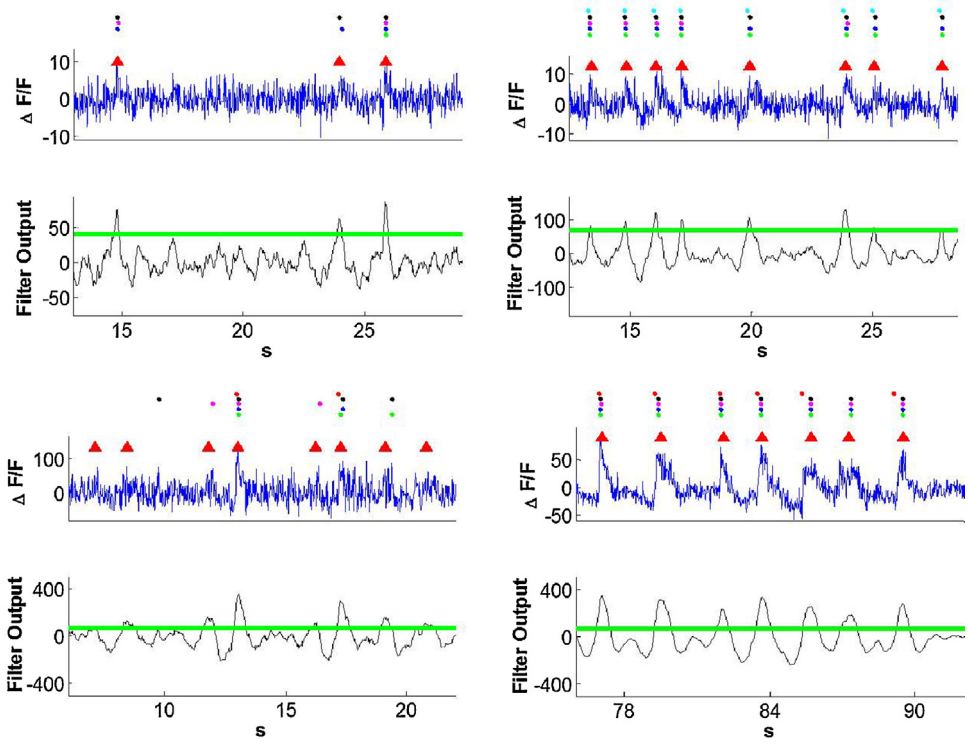

Fig. 5. Four examples of CEs detected from real data. Each example consists of a *top* trace showing the raw Δ F / F data with MMiCE detected CE peaks marked by red triangles, and a *bottom* trace showing the filter output and the detection threshold used. The circles, going from bottom to top, represent spikes identified by analysts 1–6 respectively. The top two examples are from somatic fMCI data set 1 in both a low-SNR case (*Left*) and a high-SNR case (*Right*). The bottom two examples are from spine fMCI data set 4 in both a low-SNR case (*Left*) and a high-SNR case (*Right*). All four filter outputs shown here were thresholded with $a = 3$. Note that in the spine low-SNR case (*Bottom Right*) the MMiCE detector was able to identify many different CE shapes even though only a single conditioning template was used. This is also the case where analysts did not agree on most of the spikes. In the high-SNR cases (*Right*), the MMiCE detector agrees more closely with a majority of the analysts, although even here the analysts are not unanimous about all of the spikes.

et al. (2008), manually identified CEs are not a reliable estimate of the ground truth, especially at low SNRs. Furthermore, the process proved to be extremely laborious. On average, each analyst spent roughly 24 working hours identifying CEs in a single data set.

3.3.3. MMiCE detector performance on real data

The MMiCE detector was applied using the full noise covariance, and performed very well on real data. The detector could consistently identify CEs with a specified sensitivity level. This was not true for manually identified CEs, which were not consistent between traces or across data sets. Upon reviewing detection results with varying threshold sensitivity values, each analyst identified a threshold sensitivity value, a, that they considered to be optimal for each data set. The selected values ranged from 2 to 4 for the somatic fMCI data, and 1.7–3 for the spine fMCI data. This roughly corresponds to the thresholds for the best detection results at SNR 0.2–0.5 for the simulated data sets. The average SNR of the real data is 1.53–2.43 (Table 4), which, given the bias discussed in Section 3.2.1, is about equivalent to SNR 0.5–2 in the simulated data case. These thresholds should therefore achieve near optimal performance.

Fig. 5 *Top* shows examples of detected CEs in both a low-SNR case (*Left*) and a high-SNR case (*Right*) for somatic fMCI data. Both data traces are from somatic fMCI data set 1, and both filter outputs are thresholded with $a = 3$. Similarly, Fig. 5 *Bottom* shows the same for the spine fMCI data, with both traces from spine fMCI data set 4,

and both filter outputs thresholded with $a = 3$. Note that the MMiCE detector was able to identify many CEs of various shapes (Fig. 5 *Bottom Right*), even though only a single template was used for each data set. Analysts indicated that they trust the filter output more than their own judgment in identifying low-SNR CEs.

Comparison with manually identified CEs. MMiCE detected CEs were compared to the CEs identified manually. In this case there is no ground-truth to assess the detector's performance, therefore the analysts' judgment was used as the ground-truth. Detected CEs did not closely match the analysts' CEs. Although the detector could always accurately identify over 90% of any analyst's CEs for any given data set at a low enough threshold, this often resulted in an extremely high FP Rate, where again here a FP is defined relative to the CEs identified by the analyst. This discrepancy is largely due to the analysts' lack of consistency in CE selection. Even taking the CEs that all analysts agreed on did not result in a close match with the MMiCE detected CEs (TP $\simeq$ 70%, FP $\simeq$ 20%, data not shown). This is largely due to the bias, and within-analyst inconsistency (i.e. the analyst is not consistent with herself) of the analysts. This is further discussed in Section 4.2. Furthermore, when analysts were shown the MMiCE detected CEs compared with their own identified CEs, they were more likely to fault themselves with an omission or false positive than to attribute an error to the detector. Lastly, the MMiCE detector greatly reduced detection time. The computation time for CE detection was tested using Matalb and is 0.0047-s per every second of recording per ROI (0.0047-s/ROI-s). This amounts to 75.0 ± 47.6-s on average per data set, depending linearly on the length of the recordings, and the number of ROIs in the data set. The same data sets took about 24 hours to tag manually.

4. Discussion

4.1. WGN assumption

The MMiCE detector was tested both under a WGN approximation, and using the full noise covariance. In general, the noise in neurological data is not white, as a large portion of it is biological noise coming from neighboring neurons (Buzsáki, 2004). Our results in simulated data, however, indicated that using a WGN approximation is sufficient for precise and accurate CE detection, especially at SNR $\geq$ 1. Once the SNR falls below 1, a WGN approximation is no longer sufficient and the general matched filter (Eq. (1)) outperforms the matched template (Eq. (2)).

4.2. Reliability of manual CE identification

As shown in Section 3.3.2, the level of agreement between analysts was very low (50.48% for the somatic data, and 31.52% for the dendritic spine data). As all analysts were thoroughly trained in identifying CEs, this discrepancy is not due to any single analyst simply mis-identifying a majority of events. For the most part, the disagreement occurred in low-SNR traces, where the signal was difficult to identify accurately. There was also disagreement, especially in the spine data sets, due to overlapping CEs and variations in CE shape. While some analysts would identify wide and oddly shaped CEs as more than one event, others considered them a single CE. This discrepancy is the reason why the spine data has a much lower agreement rate, while maintaining a higher average SNR than the somatic data. As SNR increased and overlapping CEs exhibited two or more clearly identifiable peaks, this discrepancy disappeared.

The analysts also proved to be inconsistent in CE identification. Although in one trace a given analyst may identify a CE of SNR 1.5, in another trace they may fail to identify any CEs below SNR 3. This is especially dependent on the number of higher SNR CEs in the same trace. If there are CEs of SNR 5 and above in a trace, then spikes of SNR 1.5 will be omitted by analysts. Conversely if the trace is mostly noise, or only has some CEs of SNR $\leq$ 3, then CEs of SNR 1.5 are very likely to be identified. This bias was true for all analysts. Overall, manual CE identification is unreliable and error-prone.

4.3. Effects of differing conditioning samples

Although only 20 samples of both CEs and noise segments needed to be identified for conditioning on a given data set, some may still find this process too laborious. It may therefore be advantageous to re-use conditioning data samples on several data sets. The MMiCE detector's efficacy was therefore tested with the conditioning samples and test data coming from different data sets, and the results were qualitatively compared.

When conditioning samples for different somatic fMCI data sets were applied to the somatic fMCI data, the detection results were virtually unchanged. Although the actual amplitude of the filter output changed, the relative peaks in the filter output stayed consistent. A slight dip in performance was observed when using conditioning samples from somatic fMCI data set 2. This conditioning sample contained the lowest SNR CE samples (Avg. SNR = 1.22, Table 3) of all of the data sets. The slight dip in performance, and low SNR of conditioning CEs may indicate that the conditioning template s from this data set was poorly determined.

Similarly, if spine fMCI conditioning data was varied between the spine fMCI data, the detection results remained roughly the same. No dip in performance was observed for any conditioning sample. The filter output peaks were more pronounced when using the conditioning CEs from spine fMCI data set 4 (Avg. SNR = 8.87, Table 3), but this did not result in markedly improved detection.

Detection performance did worsen when spine fMCI conditioning data was used with somatic fMCI data, but not vice versa. This is largely attributed to the fact that the spine conditioning CEs were wider than the somatic CEs. This caused the filtered CEs to be wider and lower in amplitude in the case of somatic fMCI test data, which in turn may have negatively affected detection. When somatic fMCI conditioning spikes are used on spine fMCI data, the skinnier CE template caused the filtered CEs the be thinner and higher in amplitude, which did not significantly affect spike detection.

In general, re-using conditioning samples from one data set to detect spikes in other data sets did not negatively affect detection performance. Performance only degraded if the conditioning CEs are of low quality and very low SNR. Although it is possible to use conditioning samples from one data type (such as somatic fMCI) on other data types (such as dendritic spine fMCI), it is not recommended for best performance.

Note, however, that the imaging set-up for the data collected here was consistent. Furthermore, one experimentalist collected all somatic fMCI data (Analyst 5 from Table 4) and another experimentalist collected all dendritic spine fMCI data (Analyst 6 from Table 4). Therefore, performance may degrade if conditioning samples from data collected by a given experimentalist are used on data collected by a different experimentalist whose procedure may differ. Similarly, performance may degrade if conditioning samples are used on data with a different imaging set-up or a different sampling frequency. For best performance it is recommended that conditioning samples are only re-reused on data of the same type, collected by the same experimentalist/under the same protocol, and using the same imaging set-up. Otherwise, performance may be affected. Conditioning samples from the fMCI-PC data were not re-used because this data was taken at a sampling frequency of 10 Hz, where as the other real fMCI data was taken at a sampling frequency of 50 Hz. The conditioning samples were therefore not compatible.

4.4. Automating sensitivity setting

Given the large variability in CEs identified by analysts, it may be advantageous to remove as much human involvement from the detection process as possible. Since different analysts often disagree on an appropriate threshold sensitivity setting, one of the best ways to minimize human involvement would be to completely automate threshold setting. Assuming CEs are outliers within a Gaussian noise distribution, we can define an outlier threshold as

$$\gamma_r = M_r + \sigma_{M,r}\sqrt{2\ln T} \tag{7}$$

where T is the number of time samples spanning a data trace X_r, and $\sigma_{M,r}$ is derived in Eq. (3) (Nenadic and Burdick, 2005). This threshold is essentially an upper bound on the noise distribution, therefore anything above it should belong to a different distribution, i.e. the signal. For the fMCI-PC data trace this gives $\gamma = 76.71$. This is equivalent to a threshold sensitivity setting of $a = 3.9$, which falls within the range of optimal threshold sensitivity settings ($a = 3$–4) when MMiCE is applied under the WGN assumption (Section 3.2, Fig. 4). It is higher than the optimal threshold when the full noise covariance is used, which was $a = 2$. For the 5 dendritic spine fMCI data sets used here this gives an average threshold of $\gamma = 51.48 \pm 27.49$, which is equivalent to a threshold sensitivity of $a = 4.16 \pm 0.08$. Similarly, the 3 somatic fMCI data sets had an average optimal threshold of $\gamma = 69.96 \pm 28.71$, corresponding to a threshold sensitivity of $a = 4.03 \pm 0.06$. The variability in γ is expected as it is intended to differ significantly based on signal amplitude. The threshold sensitivity, a, on the other hand should remain relatively constant throughout the data sets, which is also reflected here. This threshold sensitivity is slightly above that set by analysts (1.7–3 for dendritic spine, and 2–4 for somatic data), but corresponds well with the optimal threshold determined for the SNR 0.5 simulated data set ($a = 4.00 \pm 0.50$).

4.5. Limitations

Like many pre-existing methods (Cossart et al., 2003; Ikegaya et al., 2004; Kerr et al., 2005; Moreaux and Laurent, 2007; Ramdya et al., 2006; Sasaki et al., 2008), the MMiCE detector does not focus on image processing, and requires ROIs to be pre-determined and pre-processed into fluorescence intensity traces before being fed into the algorithm. As the number of neurons that can be simultaneously imaged continues to increase, this may prove to be a major limitation of this approach. Not only is manual ROI selection time consuming, but it may also be subject to bias. Our previous work addresses this issue by developing a stand alone algorithm for automatically extracting ROIs from somatic fMCI data (Nakae et al., 2014). Although not applied here, the MMiCE detector can be easily paired with either this method, or other algorithms for ROI extraction such as the ones presented in Mukamel et al. (2009) or Cossart et al. (2003). Note, however, that the automated ROI extraction tools listed here are developed for somatic fMCI data. To the best of our knowledge, manual ROI selection is therefore still necessary for dendritic spine fMCI data.

Another limitation of this approach is the need for manual inspection of the data to obtain conditioning samples. Using a fully automated algorithm that relies on general templates from a library of pre-constructed templates or on canonical filters such as wavelets, would remove the need for human involvement. However, this automation step would come at the expense of lowering detection performance. While CEs across different datasets tend to exhibit stereotypical features, there are sufficient dataset-to-dataset differences that may not be optimally accommodated with general templates. Furthermore, the general templates may themselves be biased to the dataset that they were constructed from. This bias may be mitigated with the use of wavelets. However, since they have to be zero-mean functions, wavelets cannot be monophasic and are therefore suboptimally matched to the shape of CEs. Therefore, performance is expected to degrade significantly if wavelets are used to replace the data-driven CE templates presented in this article.

4.6. Graphical user interface

In order to make the MMiCE detector more accessible, it was designed with a Matlab graphical user interface. The interface is part of the package available at sites.uci.edu/aggies/downloads or from the corresponding author and includes a tutorial. It also comes with preset options such as using the full noise covariance or applying a WGN assumption to decrease computation time. These settings, along with others such as the size of conditioning samples, optimal thresholding settings, the duration of a typical CE, and graphing options are variable, and detailed in the tutorial.

5. Conclusion

In this study we presented a Matched filter for Multi-unit Calcium Event (MMiCE) detection that is designed to identify CEs in low-SNR environments. The algorithm is completely data driven, making it applicable to a wide range of data types and testing environments. The MMiCE detector was tested on simulated somatic fMCI data, and performed exceptionally well. The simulated data was constructed from noise segments and CEs captured from real neuronal somatic fMCI recordings and varied from SNR 0.2 to SNR 14. Even at SNR 0.2 the MMiCE detector reached a TP Rate of 90.46 (3.11)% and a FP Rate of 12.81 (3.47)% when applied in its simplest form, assuming WGN statistics. The performance was even better when the full noise covariance was used in conditioning, with a TP Rate of 98.27 (1.38)% and a FP Rate of only 6.59 (2.56)% at SNR 0.2. The detector reached perfect performance (TP Rate = 100 (0.00)%, FP Rate = 0 (0.00)%) at SNR 2 and above, even under WGN assumptions. The MMiCE detector's performance was also tested on simultaneously recorded somatic fMCI and patch-clamp data. The patch-clamp recording provided a ground-truth to quantitatively assess the MMiCE detector's performance on real somatic fMCI data. Under the WGN assumption the MMiCE detector reached a TP Rate of 95.83 (5.86)% and a FP Rate of 2.13 (4.28)%, which constitutes 3 errors. When the full noise covariance was used the results improved to TP = 100.00 (0.00)% and FP = 2.04 (4.10)%, which constitutes only 1 error. This high performance level was on par with and in some cases exceeded that shown by existing methods (Kerr et al., 2005; Sasaki et al., 2008; Moreaux and Laurent, 2007; Yaksi and Friedrich, 2006; Greenberg et al., 2008). Although the MMiCE detector did not attempt to extract AP firing rates from the detected CEs, it is worth noting that it achieved similar performance while being tested at SNR levels as low as 0.2, which are well below those used in previous studies (SNR $\simeq$ 5–10).

Lastly, the MMiCE detector's performance was qualitatively presented for experimentally recorded low-SNR somatic fMCI and dendritic spine fMCI data. The detector performed very well, able to consistently detect CEs with a specific sensitivity level. Furthermore, the detector was able to detect CEs of varying shapes and widths, even though only a single conditioning template was used for detection. Analysts who also identified CEs in the data sets proved to be unreliable. The level of agreement between analysts in identifying CEs was only 50.48% for the somatic fMCI data sets, and only 31.51% for the spine fMCI data sets. Analysts indicated that they trust the MMiCE detector's output over their own judgment, especially for CEs with SNR $\leq$ 3. Furthermore the MMiCE detector reduced detection time from 24 h per data set manually to 75-s per data set on average (0.0047-s/ROI-s).

In order to accommodate the varying signal amplitudes and SNR levels found in different ROIs of a given imaging data set, the MMiCE detector applied an innovative thresholding approach. The threshold was determined independently for each ROI by calculating the detection filter output's median and median standard deviation. These quantities effectively estimate the mean and standard deviation of the filtered noise, respectively. The filter threshold is then set as a multiple, a, of median standard deviations above the median, where a is a thresholding sensitivity parameter set at the onset of detection. Optimal thresholding sensitivity levels, a, set by the analysts ranged from 1.7 to 3 for spine and 2 to 4 for somatic data. Optimal thresholds for the MMiCE detector using the full noise covariance on simulated data with a similar SNR range (SNR 0.2 – SNR 1) were $a = 5.00 \pm 1.37$. The optimal thresholding sensitivity level for the fMCI-PC data ranged from 3 to 4. This thresholding approach allows the MMiCE detector to accurately and consistently extract CEs with a specified sensitivity level even across wide variations in signal and noise amplitudes.

The MMiCE detector was also shown to be very robust. Detection retains fidelity even when conditioning samples come from a data set that is not the one being tested. It is even possible to share conditioning samples across different data types, such as between somatic fMCI and dendritic spine fMCI data. However, performance does begin degrading at this point, therefore it is not recommended.

These results indicate the MMiCE detector is applicable for widespread use in detecting low-SNR neuronal CEs from imaging data. For ease of access the MMiCE detector was implemented in Matlab with a graphical user interface and is freely available with a tutorial at sites.uci.edu/aggies/downloads or from the corresponding author.

Acknowledgments

The authors would like to extend a special thanks to N. Koe, M. Doty, and K. Scannell for their help in manual CE identification. This work was supported by NSF (1056105), NSF-EAPSI (OISE-13 7307954), and JSPS (SP13054).

References

Buzsáki G. Large-scale recording of neuronal ensembles. Nat Neurosci 2004;7(5):446–51.

Cossart R, Aronov D, Yuste R. Attractor dynamics of network UP states in the neocortex. Lett Nat 2003;423:283–8.

Greenberg DS, Houweling AR, Kerr JND. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. Nat Neurosci 2008;11(7):749–51.

Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, et al. Synfire chains and cortical songs: temporal modules of cortical activity. Science 2004:304.

Joucla S, Franconville R, Pippow A, Kloppenburg P, Pouzat C. Estimating background-subtracted fluorescence transients in calcium imaging experiments: a quantitative approach. Cell Calcium 2013;54(2):71–85.

Kay SM. Fundamentals of statistical signal processing: detection theory. Englewood Cliffs, NJ: Prentice-Hall; 1998.

Kerr JND, Greenberg D, Helmchen F. Imaging input and output of neocortical networks in vivo. Proc Natl Acad Sci U S A 2005;102(39):14063–8.

Malik WQ, Schummers J, Sur M, Brown EN. Denoising two-photon calcium imaging data. PLoS ONE 2011;6(6).

Moreaux L, Laurent G. Estimating firing rates from calcium signals in locust projection neurons in vivo. Front Neural Circuits 2007;1(2):1–13.

Mukamel EA, Nimmerjahn A, Schnitzer MJ. Automated analysis of cellular signals from large-scale calcium imaging data. Neuron 2009;63(6):747–60.

Nakae K, Ikegaya Y, Ishikawa T, Oba S, Urakubo H, Koyama M, et al. A statistical method of identifying interactions in neuron-glia systems based on functional multicell Ca2+ imaging. PLoS Comput Biol 2014;10(11):e1003949.

Nenadic Z, Burdick JW. Spike detection using the continuous wavelet transform. IEEE Trans Biomed Eng 2005;52(1):74–87.

Norimoto H, Mizunuma M, Ishikawa D, Matsuki N, Ikegaya Y. Muscarinic receptor activation disrupts hippocampal sharp wave-ripples. Brain Res 2012;1461:1–9.

Ramdya P, Reiter B, Engert F. Reverse correlation of rapid calcium signals in the zebrafish optic tectum in vivo. J Neurosci Methods 2006;157(2):230–7.

Sasaki T, Kimura R, Tsukamoto M, Matsuki N, Ikegaya Y. Integrative spike dynamics of rat CA1 neurons: a multineuronal imaging study. J Physiol 2006;574(1):195–208.

Sasaki T, Matsuki N, Ikegaya Y. Metastability of active CA3 networks. J Neurosci 2007;27(3):517–28.

Sasaki T, Takahashi N, Matsuki N, Ikegaya Y. Fast and accurate detection of action potentials from somatic calcium fluctuations. J Neurophysiol 2008;100(3):1668–76.

Smetters D, Majewska A, Yuste R. Detecting action potentials in neuronal populations with calcium imaging. Methods: Companion Methods Enzymol 1999;18(2):215–21.

Szymanska AF, Heylman C, Datta R, Gratton E, Nenadic Z. Automated detection and analysis of depolarization events in human cardiomyocytes using madec. Med Biol Eng Comput 2015 (Under review).

Takahashi N, Kitamura K, Matsuo N, Mayford M, Kano M, Matsuki N, et al. Locally synchronized synaptic inputs. Science 2012;335(6066):353–6.

Takahashi N, Sasaki T, Usami A, Matsuki N, Ikegaya Y. Watching neuronal circuit dynamics through functional multineuron calcium imaging (fMCI). Neurosci Res 2007;58(3):219–25.

Ueno S, Tsukamoto M, Hirano T, Kikuchi K, Yamada MK, Nishiyama N, et al. Mossy fiber Zn2+ spillover modulates heterosynaptic N-methyl-D-aspartate receptor activity in hippocampal CA3 circuits. J Cell Biol 2002;158(2):215–20.

Yaksi E, Friedrich RW. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging. Nat Methods 2006;3(5):377–83.