A Supervised Multi-Sensor Matched Filter for the Detection of Extracellular Action Potentials

Agnieszka F. Szymanska[1], Michael Doty[1], Kathryn V. Scannell[2], Zoran Nenadic[1,3]

Abstract— **Multi-sensor extracellular recording takes advantage of several electrode channels to record from multiple neurons at the same time. However, the resulting low signal-to-noise ratio (SNR) combined with biological noise makes signal detection, the first step of any neurophysiological data analysis, difficult. A matched filter was therefore designed to better detect extracellular action potentials (EAPs) from multi-sensor extracellular recordings. The detector was tested on tetrode data from a locust antennal lobe and assessed against three trained analysts. 25 EAPs and noise samples were selected manually from the data and used for training. To reduce complexity, the filter assumed that the underlying noise in the data was spatially white. The detector performed with an average TP and FP rate of 84.62% and 16.63% respectively. This high level of performance indicates the algorithm is suitable for widespread use.**

I. INTRODUCTION

Before any signal analysis can be performed, action potentials (APs), representative of neurophysiological activity, must be identified from the data. Detection is therefore a crucial component of any neurophysiological signal processing algorithm. Biological noise in extracellular neurophysiological recordings is composed of the activity of background neurons and ion-channel noise. It is therefore both correlated with, and statistically similar to the AP signal. This compounds the detection problem and makes most standard signal detection tools used in other fields unsuitable.

Template matching, or more generally the matched filter, have proven to effectively detect spikes from various types of neurophysiological data [1]–[5]. The general basis behind template matching is to generate a template representative of neural activity, and then compare the data to the template. This approach can be used with a single template for detection purposes or many different templates for purposes of simultaneous detection and classification [6].

Matched filtering can be either supervised or unsupervised. Traditional template approaches are supervised, using both spike and noise training measurements for various parameter estimation before detection. The template is then usually generated from extracellular action potentials (EAPs) selected by an analyst from the data [3], [6]. Un-supervised algorithms using wavelets as templates, and assuming no prior knowledge about the signal, are also becoming increasingly popular [4], [5], [7], [8].

[1]Department of Biomedical Engineering, UCI, CA, 92697, USA aszymans@uci.edu
[2]Center for Complex Biological Systems, UCI, CA, 92697, USA
[3]Department of Electrical Engineering and Computer Science, UCI, CA, 92697, USA

Although widely applied to single-sensor extracellular data, there are few examples of matched filers applied to multi-sensor data. Gozani and Miller [6] developed a technique using multiple matched filters to simultaneously detect and classify multi-sensor EAPs. Their technique tries to both maximize EAP signal-to-noise ratio (SNR) and minimize the interference between EAPs. However, this method relies on the construction of many EAP templates, which requires some prior knowledge of the number of recorded neurons, and may be very time consuming, as a single tetrode in a neuron-dense region may record activity from up to 20 neurons [9]. Furthermore, minimizing the interference between EAPs may compromise their detectability.

The work presented here concentrates on the generation of a more reliable and simpler supervised matched filter for use on multi-sensor extracellular recordings. The filter takes into account the noise statistic of the data and is designed to use a single template to detect EAPs from multiple different neurons.

II. METHODS

A. Data Collection

Data used in this experiment is publicly available online [10] and the full data collection procedure is described by Pouzat et al. in [11]. Briefly, a planar silicon probe was placed below the surface ($\sim$50-100 μm) of an adult locust's antennal lobe and used for recording. Recorded data was sampled at 15 kHz and bandpass filtered from 300 - 5,000 Hz. A total of 20 seconds of data was provided from four of the probe tip sensors. All subsequent data analysis was performed in MATLAB.

B. Ground-Truth Estimation

Half of the collected data was used for training (training data), and the remaining half was used for further analysis (test data). Three trained analysts independently tagged all of the spikes in the test data.

C. Detection Algorithm Design

In the case of extracellular signals, the two models being investigated represent noise only data, and data containing both EAPs and noise. More formally, EAP 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 an EAP and noise.

Given a C-sensor signal of length N, where N is the number of samples spanned by an EAP, we can express the hypotheses mathematically as

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

where $x \in \mathbb{R}^{C \times N}$ is a multi-sensor signal, $s \in \mathbb{R}^{C \times N}$ is an EAP, and $n \in \mathbb{R}^{C \times N}$ is zero-mean sensor noise. The likelihood ratio test (LRT) for this problem can accordingly be expressed as

$$L(x) = \frac{p(\underline{\mathbf{X}}|H_1)}{p(\underline{\mathbf{X}}|H_0)} \quad , \quad \begin{array}{l} \text{accept } H_1 \text{ if } L(x) > \gamma \\ \text{accept } H_0 \text{ if } L(x) < \gamma \end{array}$$

where $\underline{\mathbf{x}} \in \mathbb{R}^{1 \times CN}$ is the row vector form of the signal matrix x, and γ is the threshold.

In the case of Gaussian noise the LRT takes on the form of a generalized matched filter (GMF) [12]

$$S(\underline{\mathbf{x}}) = \underline{\mathbf{s}} \Sigma^{-1} \underline{\mathbf{x}}^T \quad , \quad \begin{array}{l} \text{accept } H_1 \text{ if } S(\underline{\mathbf{x}}) > \gamma\prime \\ \text{accept } H_0 \text{ if } S(\underline{\mathbf{x}}) < \gamma\prime \end{array} \quad (1)$$

where $\underline{\mathbf{s}} \in \mathbb{R}^{1 \times CN}$ is the row vector form of the EAP, $\Sigma \in \mathbb{R}^{CN \times CN}$ is the spatio-temporal noise covariance matrix calculated from the row vector form of the sensor noise n, and $\gamma\prime$ is the updated threshold subsuming γ and other data-independent constants. It is important to note that the test statistic, $S(\underline{\mathbf{x}})$, is linearly dependent on the signal $\underline{\mathbf{x}}$. Even if the noise is not Gaussian, $S(\underline{\mathbf{x}})$ still has the highest achievable SNR of all other linear combinations of the data. In this case, however, the GMF could underperform the more general LRT.

D. Parameter Estimation

Twenty five EAPs, 2 ms each ($N = 30$, $C = 4$), were manually selected from the training data, aligned to their peak values, and averaged to generate the matched template, s. Similarly, 25 noise samples, roughly 30 ms each, were manually selected from the training data, concatenated into a single four-sensor time series, n_T, and used to generate the noise covariance matrix, Σ.

To simplify the noise covariance calculation we can represent Σ as

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,C} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,C} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{C,1} & \Sigma_{C,2} & \cdots & \Sigma_{C,C} \end{bmatrix} \quad (2)$$

where the submatrix $\Sigma_{i,i} \in \mathbb{R}^{N \times N}$ is the temporal covariance matrix of noise at sensor i, and $\Sigma_{i,j} \in \mathbb{R}^{N \times N}$ ($j \neq i$) is the temporal cross-covariance matrix of noise at sensors i and j. Estimating the full covariance matrices may require a prohibitively large noise sample, however, the calculation becomes more tractable if we assume the noise is uncorrelated in space. This forces all $\Sigma_{i,j}$ ($j \neq i$) to 0. Although neurophysiological noise exhibits spatial correlations [13], our experience shows that the spatially

white noise assumption outperforms fully colored noise in detection.

The training noise n_T was subdivided into 375 noise windows ($N = 30$). Auto-covariance sequences, $r_i(k)$, were then calculated at lags $k \in [-29, 29]$ for each window and each sensor $i \in [1, 4]$. The sequences were averaged across all 375 noise windows, and used to generate each $\Sigma_{i,i}$.

E. Threshold Setting

Threshold values were determined using the median standard deviation, σ_M, of the filtered noise training data, $S(\underline{\mathbf{n}}_T)$,

$$\sigma_M = \frac{\mathcal{M} \{| S(1)) - M |, ..., | S(T) - M |\}}{\sqrt{2} Erf^{-1}(\frac{1}{2})} \quad (3)$$

where $S(\underline{\mathbf{n}}_T)$ was calculated using (1), $\mathcal{M}$ is the median operator, M is the median of the filtered noise training data $S(\underline{\mathbf{n}}_T)$, and T is the number of samples spanning n_T [8]. The threshold is then set as a multiple, a, of median standard deviations above M

$$\gamma\prime = M + a\sigma_M \quad (4)$$

a will be referred to as the threshold multiple for the remainder of this article.

The median was used for this calculation because it is less affected by spikes that may be accidentally present in the training noise than a mean. M and σ_M are therefore robust estimates of the true noise mean and standard deviation.

F. Performance Analysis

Detected EAPs, given 40 incrementally increasing thresholds, $a = [1, 1.5, ..., 20.5]$, were compared against those tagged by each trained analyst, with the analyst acting as the ground-truth. True positive (TP), and false positive (FP) rates for each threshold and analyst were then calculated as

$$\text{TP Rate} = \frac{TP}{Spikes_A} \quad (5)$$

$$\text{FP Rate} = \frac{FP}{Spikes_D} \quad (6)$$

where $Spikes_A$ is the total number of spikes tagged by the analyst, and $Spikes_D$ is the total number of spikes detected by the detector, and TP and FP are the total numbers of true positives, and false positives, respectively. The false negative (FN) rate can be calculated as $(1 - FP)$, however as it gives no new information about the detector performance, it is not reported here. The TP and FP rates at each threshold were then used to generate ROC curves for each analyst.

III. RESULTS

The detection algorithm performed well compared with all three analysts. The optimal threshold for each analyst (Table I) was determined by minimizing the distance between their respective ROC curve (Fig. 1) and theoretically perfect performance (100% TP, 0% FP). At the optimal threshold, the detector performed with a TP rate of 90.79% and a FP rate of 20.66% compared with Analyst 1, a TP rate of 80.00% and a FP rate of 10.31% compared with Analyst 2, and a

TP rate of 83.06% and a FP rate of 18.91% compared with Analyst 3 (Table I). The detector's average TP and FP rates were 84.62% and 16.63%, respectively.

Note that the optimal thresholds associated with the best performance for each analyst differed. This is a result of the discrepancy in spike tagging between analysts, which can be quantified in terms of signal-to-noise ratio (SNR). The SNR for each spike, u, tagged by the analysts was calculated as

$$\text{SNR}(u) = \max_i \left\{ \frac{\sigma_i^2(u)}{\sigma_i^2(n)} \right\} \qquad (7)$$

where $\sigma_i^2(u)$ is the variance of spike u at sensor i, and $\sigma_i^2(n)$ is the variance of the training noise n at sensor i. The maximum SNR across sensors was taken as u's SNR.

Analyst 1 was the most selective with a median SNR of 28, followed by Analyst 2 with a median SNR of 20, and Analyst 3 was the most liberal, with a median SNR of 7. An example of the types of spikes tagged by each analyst, as well as the detected spikes at $a = 5.5$ is shown in Fig. 2.

TABLE I

DETECTOR PERFORMANCE FOR THE THREE ANALYSTS. THE CORRESPONDING DETECTOR THRESHOLD MULTIPLE a (4) IS LISTED IN COLUMN 3 AND EACH ANALYST'S MEDIAN SPIKE SNR IS PROVIDED FOR REFERENCE IN COLUMN 4.

Analyst	Optimal Performance		Threshold Multiple 'a'	Median Analyst SNR
	TP Rate (%)	FP Rate (%)		
1	90.79	20.66	7	28
2	80.00	10.31	5.5	20
3	83.06	18.91	3.5	7

IV. DISCUSSION

A. Analyst Reliability

The data used in this study was collected *in vivo*, therefore the ground-truth in terms of spike times was not available and was instead estimated as the spikes selected by each analyst. However, agreement between analysts was as low as 39%. The variability in the analysts' tagged spikes therefore implies that human analysts may not be a reliable source of ground-truth estimation.

Besides not being consistent with each other, the analysts are also not internally consistent in spike selection. Analysts are likely to arbitrarily increase or decrease their internal thresholding criteria over time, even while tagging spikes in a single data set. Similarly, spikes with a relatively low amplitude but surrounded by noise are very likely to be selected by the analyst. However, similar spikes surrounded by other APs are less likely to be selected.

This bias artificially decreases the detector's TP rate. In general human analysts are not a reliable or consistent source of ground-truth for the detection of neurophysiological activity. This conclusion was also drawn by Harris et al. [14] where 9 analysts sorting extracellular tetrode data had FN and FP rates as high as 30%.
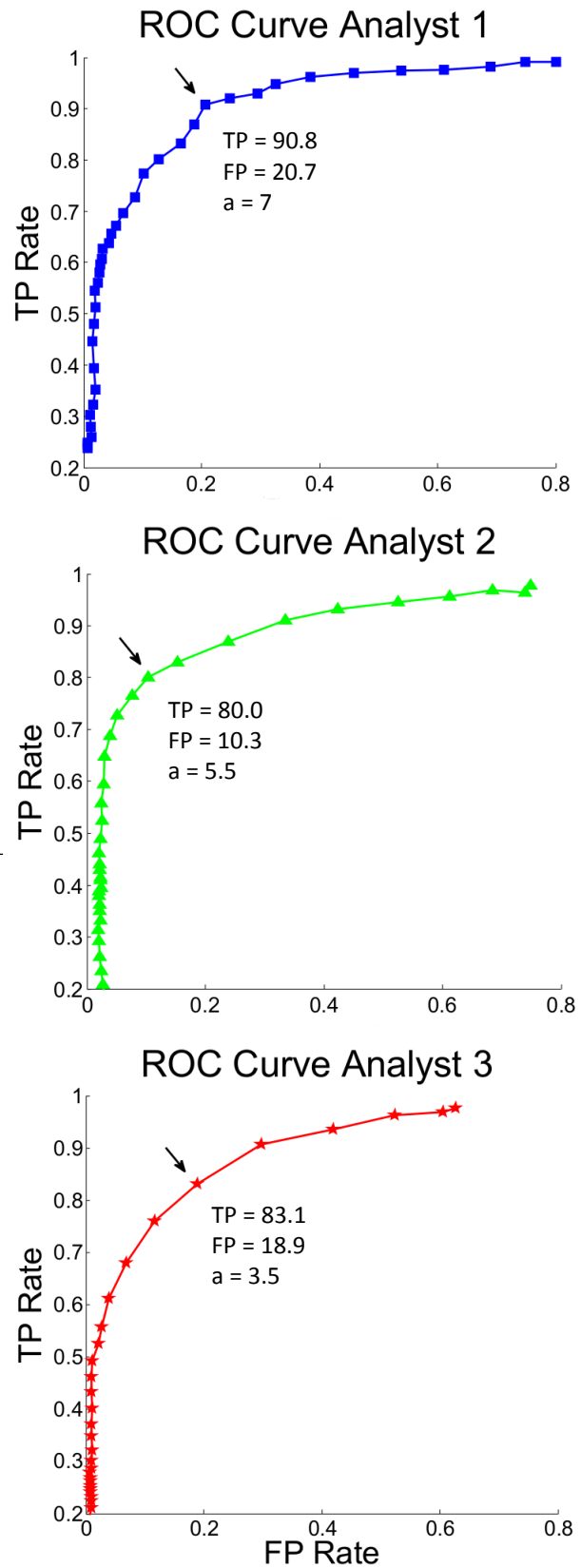

Fig. 1. ROC Curves for all three analysts. TP and FP pairs were collected for 40 incrementally increasing detection thresholds with $a = [1.0, 1.5, ..., 20.5]$, for each analyst. The black arrows point to the detectors optimal performance for each analyst, with the appropriate parameters listed below.
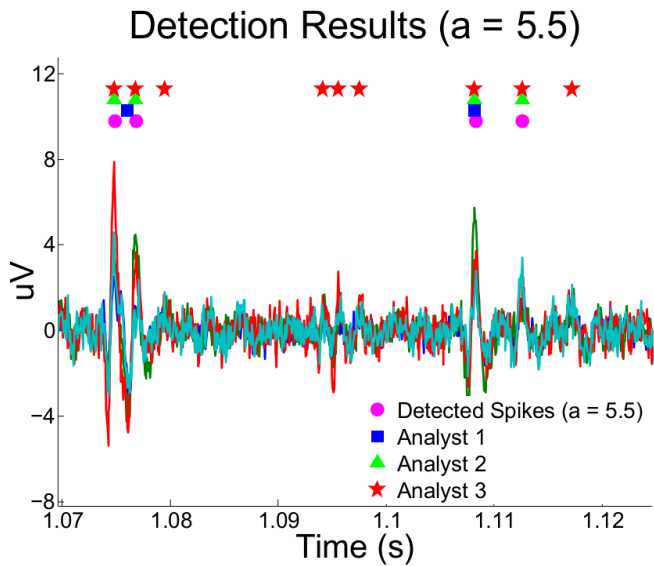
Fig. 2. Raw tetrode data with the detection results for $a = 5.5$ marked as purple circles. The spikes tagged by Analysts 1, 2, and 3 are shown as blue squares, green triangles, and red stars, respectively. Note that this threshold is optimal for detecting the spikes tagged by Analyst 2.

B. Detector Performance

The presented algorithm's performance is under estimated in this study. As discussed previously, this is partially due to the analysts' unreliability. To try to better assess the detector's performance, the FPs in the optimal cases for Analyst 1 were further inspected. The detector performed optimally when compared against Analyst 1 at $a = 7$. In this case the detector reported a 20.66% FP rate (Table I). Because Analyst 1 is the most selective, we can compare these FPs to the spikes tagged by both Analysts 2 and 3. The comparison results show that 99% of these FPs were tagged by both Analysts 2 and 3. In other words, a majority of the analysts thought that 99% of the spikes identified as FPs in this case were actually TPs. This implies that the FP rate should be closer to 0%. Similar results were derived when analyzing the FPs identified against Analyst 2.

Given these results, we believe that the detector would perform much better if compared against the actual ground-truth as opposed to trained analysts. This can be accomplished by patch-clamping several neurons and simultaneously recording extracellularly with a tetrode as described in [14]. The patch-clamp recordings would be temporally matched against tetrode recordings to determine precise AP times. These could then be used as the ground-truth for detection results. Any spikes recorded from non-patched neurons would not be included in such an analysis.

C. Threshold Setting

When applying this matched filter, the user can adjust the threshold multiple a to achieve their desired sensitivity. However, this threshold may vary drastically from user to user and an automated threshold may therefore be more appropriate. Assuming APs are outliers within a Gaussian

noise distribution, we can define an outlier threshold as

$$\gamma' = \sigma_M \sqrt{2 \ln l} \tag{8}$$

where l is the number of time samples spanning the full data series, and σ_M is derived in (3) [8]. 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 data used here this gives $\gamma' = 4.88$ and is roughly equivalent to $a = 6.0$, falling somewhere between the sensitivity levels of Analysts 1 and 2.

V. CONCLUSIONS

The multi-sensor matched filter developed in this study was assessed against three trained analysts and performed with average TP and FP rates of 84.62% and 16.63%, respectively. The detector's performance presents it as a great candidate for wide-spread use as one of the only multi-sensor detectors of its kind. Furthermore, the algorithm's performance is likely under estimated as a vast majority of the spikes identified as FPs for Analysts 1 and 2 were actually identified as TPs by Analyst 3. Because the analysts were proven to be inconsistent and unreliable, we believe that the detector would perform even better if compared against the ground-truth instead of trained analysts.

REFERENCES

[1] P H Thakur, H Lu, S S Hsiao, and K O Johnson. Automated optimal detection and classification of neural action potentials in extra-cellular recordings. *J Neurosci Meth*, 162(1-2):364–376, 2007.
[2] S Kim and J McNames. Automatic spike detection based on adaptive template matching for extracellular neural recordings. *J Neur Methods*, 165(2):165–174, 2007.
[3] A Szymanska, A Hajirasooliha, and Z Nenadic. Source location as a feature for the classification of multi-sensor extracellular action potentials. In *Proc. of the 6th International IEEE EMBS Conference on Neural Engineering*, pages 810–814. IEEE, 2013.
[4] Y Yuan, C Yang, and J Si. The M-Sorter: an automatic and robust spike detection and classification system. *J Neurosci Meth*, 210(2):281–290, 2012.
[5] R J Brychta, S Tuntrakool, M Appalsamy, N R Keller, D Robertson, R G Shiavi, and A Diedrich. Wavelet methods for spike detection in mouse renal sympathetic nerve activity. *IEEE Trans Bio-Med Eng*, 54(1):82–93, 2007.
[6] S N Gozani and J P Miller. Optimal Discrimination and Classification of Neuronal Action Potential Waveforms from Multiunit, Multichannel Recordings Using Software-Based Linear Filters. *IEEE Trans Bio-Med Eng*, 41(4):358–372, 1994.
[7] V Shalchyan, W Jensen, and D Farina. Spike detection and clustering with unsupervised wavelet optimization in extracellular neural recordings. *IEEE Trans Bio-Med Eng*, 59(9):2576–2585, 2012.
[8] Z Nenadic and J W Burdick. Spike detection using the continuous wavelet transform. *IEEE Trans Bio-Med Eng*, 52(1):74–87, 2005.
[9] M A Wilson and B L McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058, 1993.
[10] C. Pouzat and G. Laurent. Locust data available online at http://www.biomedicale.univ-paris5.fr/SpikeOMatic/Data.html. 2005.
[11] C Pouzat, O Mazor, and G Laurent. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *J Neurosci Meth*, 122(1):43–57, 2002.
[12] S M Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice-Hall, Englewood Cliffs NJ, 1998.
[13] A Szymanska and Z Nenadic. Wavelet-approximated generalized matched filter for the detection of multisensor extracellular action potentials. In *Proc. of the 6th International IEEE EMBS Conference on Neural Engineering*, pages 235–238. IEEE, 2013.
[14] K D Harris, D A Henze, J Csicsvari, and H Hirase. Accuracy of Tetrode Spike Separation as Determined by Simultaneous Intracellular and Extracellular Measurements. *J Neurophys*, 84:401–414, 2000.