# A Spike Classification with a posteriori Confidence Estimation

Yi-Lin Li (李易霖)[1], Wei-Chang Shann (單維彰)[2] and Meng-Li Tsai (蔡孟利)[3]

[1]Department of Mathematics, National Taiwan University
[2]Department of Mathematics, National Center University
[3]Department of Biomechatronic Engineering, National Ilan Univeristy

**摘要**

論文描述一個分類單一電極所記錄到不同波型動作電位的方法。此法在假定已知雜訊的平均值和標準差的前提下，分類被雜訊干擾後之點集。其優點在於分類後可進一步利用已知的參數來估計可能的分類錯誤率。本文中將推導點集分類法的數學處理過程，進而實際應用至一組虛擬動作電位訊號之分類。最後並以與 PCA 點圖之對照來驗證此方法之效度。

**關鍵詞**：動作電位分類、神經訊號、電極

**Abstract**

We propose a method for spike classification based on the hypothesis that the mean and variance of a normally distributed noise is known. The proposed classification method is based on a fundamental method that classifies a set of randomly perturbed points. With some practically computable parameters, the later estimates its confidence of performance by giving an upper bound of error rates. We derive this fundamental method mathematically, apply it to the task of spike classification, and show how it works by a numerical experiment which compares with a PCA scattering plot.

**Keywords:** spikes classification, neural signals, electrodes

## Introduction

While many powerful imaging techniques have been used in neuroscience, extracellular recording remains the only choice that provides resolution at single neuron activity in the brain. However, multiple single-unit extracellular recordings are useful only if the spikes generates by different neurons can be sorted correctly. The usual assumption for spikes sorting is that all spikes generated by a specific neuron are characterized by a similar waveform and the waveform is unique and conserved for each neuron during a stationary recording. Among different methods used for spike-sorting, template matching is one of the most popular procedures. The usual practice to produce templates is to use a "supervisor," that is, an experienced and knowledgeable operator, to give a preliminary classification of the waveforms following a selection of templates corresponding to distinct neurons. A few methods have dealt with unsupervised templates creation, complicated statistical methods were largely introduced in these methods.

In this short report we propose a classification method based on mathematical properties of the statistical model for the spikes and noise. We assume that the neural action potentials are recorded properly with a digital filter at a fixed sampling rate. The recorded digital data is then processed by certain spike detection and alignment programs [1], the result is a set of short signals with same length. It is commonly assumed that these signals are neural action potentials contaminated by a normally distributed noise [2]. The task here is to classify these signals into different wave forms.

In the following, we first derive mathematically a fundamental method that classifies a set of randomly perturbed points into two groups. This method will estimate its confidence of performance by giving practically computable upper bounds of error rates. Then we apply this method to classify noisy action potential signals with confidence on error rates. One set of experiments is included to demonstrate how this method works and how does it compare to the scattering plot of first two principal components. We refer the readers to [3] for more mathematical details and experimental results.

## Classifying two groups of points

Let $a$ and $b$ be two distinct real numbers, $\boldsymbol{N_a}$ and $\boldsymbol{N_b}$ be two sets of numbers sampled from a normal distribution with mean 0 and standard deviation $\sigma$. Then $\boldsymbol{A} = a+\boldsymbol{N_a}$ and $\boldsymbol{B} = b+\boldsymbol{N_b}$ are sets of perturbed points around $a$ and $b$. Consequently $\boldsymbol{A}$ and $\boldsymbol{B}$ are also normally distributed with means $a$ and $b$, respectively, and with standard deviation $\sigma$. We mix the points of $\boldsymbol{A}$ and $\boldsymbol{B}$ into $\boldsymbol{C}$. The task here is to classify all points of $\boldsymbol{C}$ into either $\boldsymbol{A}$ or $\boldsymbol{B}$.

Theoretically, we can't classify $\boldsymbol{C}$ perfectly because the noise is not known before-

hand and hence is not removable. So practically we should evaluate "how correct" the classification is. We will present a method for classification together with an estimation of the percentage of correctness.

Let $\alpha = \#A/\#C$ where $\#A$ is the number of elements in $\boldsymbol{A}$, and so on. Then $1-\alpha = \#B/\#C$. Let $\overline{\mu}$ and $\overline{\sigma}$ be the mean and standard deviation of points in $\boldsymbol{C}$, respectively. Without loss of generality, we assume $a > b$. Let $K$ be a positive number such that $\boldsymbol{C} \subset [\overline{\mu} - K\overline{\sigma}, \overline{\mu} + K\overline{\sigma}]$. Let $L$ be another positive number such that $\boldsymbol{A} \subset [a - L\sigma, a + L\sigma]$ and $\boldsymbol{B} \subset [b - L\sigma, b + L\sigma]$. Since $\boldsymbol{C}$ is a finite set, such numbers $K$ and $L$ must exist. Furthermore, since the noise is assumed to be normally distributed, $K$ and $L$ can actually be chosen between 1 and 3 in practice. In this model, $a$, $b$ and $\alpha$ are unknown, we are given the set $\boldsymbol{C}$ and the standard deviation $\sigma$; note that $\overline{\mu}$ and $\overline{\sigma}$ are computable from $\boldsymbol{C}$.

Here is the method of classification. Choose two numbers $s_1$ and $s_2$—we will give guidelines for the choice later. We classify $\boldsymbol{C}$ into two groups by the following criteria. Take a point $p \in \boldsymbol{C}$, if $p > \overline{\mu} - K\overline{\sigma} + s_2$ then we assign $p \in \boldsymbol{A}$. If $p < \overline{\mu} + K\overline{\sigma} - s_1$ then we assign $p \in \boldsymbol{B}$. Otherwise $p$ is not classified.

The choice of $s_1$ and $s_2$ has to avoid the error of overlapped classification; that is, a point is assigned to both $\boldsymbol{A}$ and $\boldsymbol{B}$. It suffices to require $s_1 + s_2 \geq 2K\overline{\sigma}$; if we want a total classification (every point is classified), replace $\geq$ by $=$. When $s_1$ and $s_2$ are larger, the rate of correct classification is higher. Unfortunately, the number of unclassified points is also higher. When $s_1$ and $s_2$ are too large, we might not get any point classified. So there should be an upper bound for the choice of $s_1$ and $s_2$. We suggest

$$\begin{cases} s_1 \leq K\overline{\sigma} + L\sigma - (1-\alpha)(a-b) \\ s_2 \leq K\overline{\sigma} + L\sigma - \alpha(a-b) \end{cases} \tag{1}$$

This suggestion is justified by (2) below, and we shall estimate $\alpha$ and $a-b$ later.

The percentage of correct classification is 1 − (percentage of not classified) − (percentage of incorrect classification). Let $N_\mu(x)$ be the probability density function of normal distribution with mean $\mu$ and standard deviation $\sigma$. By the parameters defined above, the percentage of incorrect classification is by definition $\alpha \int_{a-L\sigma}^{\overline{\mu}+K\overline{\sigma}-s_1} N_a(x)\,dx + (1-\alpha)\int_{\overline{\mu}-K\overline{\sigma}+s_2}^{b+L\sigma} N_b(x)\,dx$. It can be simplified to

$$\begin{aligned} &\alpha \int_{-L\sigma}^{K\overline{\sigma}-(1-\alpha)(a-b)-s_1} N_0(x)\,dx \\ &+(1-\alpha)\int_{-K\overline{\sigma}+\alpha(a-b)+s_2}^{L\sigma} N_0(x)\,dx \end{aligned} \tag{2}$$

which is computable once $\alpha$ and $a-b$ are estimated.

Let $r = \big(\sum (x-\overline{\mu})^3\big)/\#C$ where the summation takes over all $x \in \boldsymbol{C}$. Then

$$r = \int_{-\infty}^{\infty} \big(\alpha N_a(x) + (1-\alpha)N_b(x)\big)\,(x-\overline{\mu})^3\,dx$$

After a lengthy calculation we get

$$\alpha = \frac{1}{2}\Big(1 \pm \frac{r}{\sqrt{4(\overline{\sigma}^2-\sigma^2)^3 + r^2}}\Big) \tag{3}$$

and

$$a - b = \frac{\sqrt{4(\overline{\sigma}^2-\sigma^2)^3 + r^2)}}{\overline{\mu}^2 - \sigma^2} \tag{4}$$

Note that $\alpha$ has two possible values. That means the roles of $\alpha$ and $1-\alpha$ are interchangeable, which is true theoretically. In practice, however, we must distinguish between the two cases. If $\boldsymbol{C}$ has more points in the interval $[\overline{\mu}, \overline{\mu} + \overline{\sigma}]$ than that in $[\overline{\mu} - \overline{\sigma}, \overline{\mu}]$, we take the plus sign in (3); otherwise we take the negative sign.

Although it may not be necessary for a total classification in practice, we give a guideline for doing just that. The optimal choice of $s_1$ and $s_2$ in the sense of the highest rate of correct classification may not correspond to a total classification. Under the condition of total classification, parameters

$$\begin{aligned} s_1 &= K\overline{\sigma} + (1-2\alpha)\big(L\sigma - (a-b)\big) \\ s_2 &= K\overline{\sigma} - (1-2\alpha)\big(L\sigma + (a-b)\big) \end{aligned} \tag{5}$$

are nearly optimal choices.

Finally, we would like to remark that the percentage of incorrect classification (error rate) estimated by (2) is *conservative*, that is, it is an upper bound for the actual value. For instance, let us set the parameters to be $\sigma = 1$, $a = 6$, $b = 3$ and $K = L = 3$, and we take $s_1$ and $s_2$ as suggested by (5) in order to

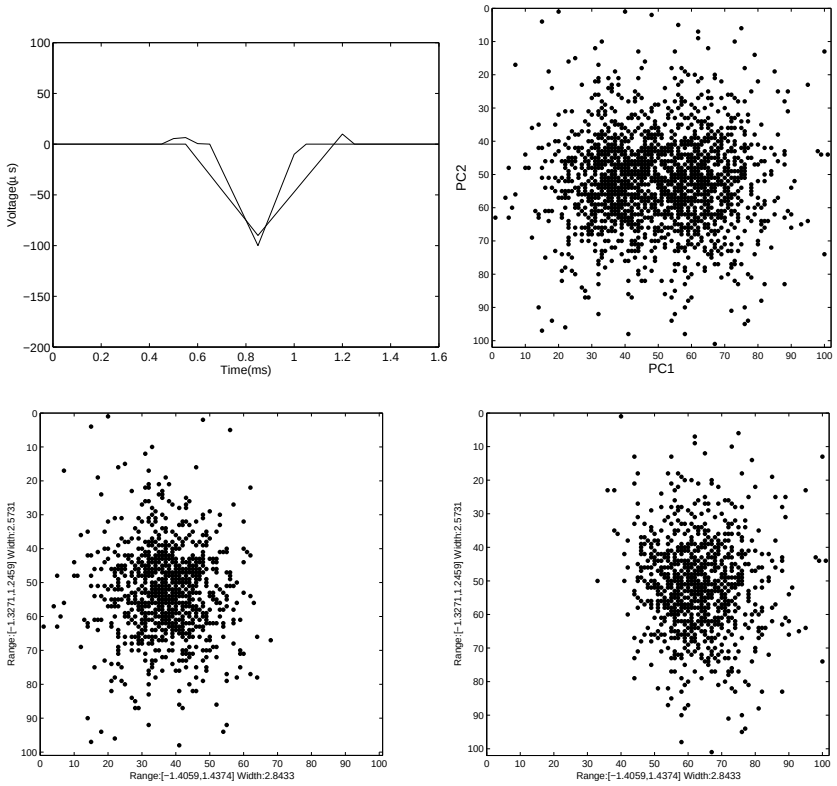


**Figure 1.** Two artificial spike waveforms (upper-left) are added by "real" noise to construct 2000 test signals. The PCA scattering plot (PC1 vs PC2) shows the difficulty of classification (upper-right). Our method separates the signals into two classes with overlapping PCA scattering (lower).

have a total classification. In the first experiment, $\boldsymbol{C}$ has 5000 points with $\alpha = 0.8$. The estimated error rate is 12%, but it is in fact 7%. In this case, the optimal choice of $s_1$ and $s_2$ for total classification has an error rate of 5%. In another experiment, $\boldsymbol{C}$ has 2000 points with $\alpha = 0.5$. The estimated error rate is 18% while it is actually 7%, and it is almost the best rate. These experiments coincide with our claims that (i) the estimated error rate is conservative, and hence it serves as an indicator of reliability for the classification; and (ii) that the choice made by (5) is close to the optimal (lowest) error rate.

## Application to signals of neural action potentials

We include one set of numerical experiments in this short article and refer the interested readers to [3]. All numerical tests in this article are done with Matlab 6.5 on an Intel PC by programs written by the investigators. There are two artificial spike waveforms as shown in the upper-left corner of Figure 1. We construct 2000 test signals by adding "real" noise cut from a recording of biological experiment to these two waveforms. Each signal is considered a 32-dimensional vector, In order to apply the classification method from previous section, we have to select one particular position and collect numbers of that position from all signals to form the data set $\boldsymbol{C}$.

The best position to collect should be that at which waveforms have the largest difference. But this position is unknown to us since original waveforms are supposed to be contaminated by noise. One may consider applying (4) to estimate $a - b$ (the difference between the waveforms) at every position. But this approach is not desirable because the formula is unstable when $a - b$ is smaller than $\sigma$. In fact, $a - b$ is larger when $\overline{\sigma}$ is larger, it leads us to consider the position at which the signals have the largest standard deviation.

In this numerical experiment, $\alpha = 0.5$ and the standard deviation of noise is 11 ($\mu$V). The largest difference of these two waveforms ($a-b$) is about 37 ($\mu$V), but we did not use this fact during the experiment. The PCA scattering plot (PC1 vs PC2) of the 2000 test signals is shown in the upper-right corner of Figure 1. It is clear that an experienced operator may

think there should be two clusters in the plot, but it is not clear where to make a decisive cut.

With parameters chosen (we always let $K = L = 3$) or evaluated as in the previous section (for a total classification), our method classifies the test signals at the 21st position (at which position the signals have the biggest standard deviation about 22) into two classes. The PCA plots for the two classes are shown in two graphs at the bottom of Figure 1. It is clear that these two classes are slightly overlapping in the PCA plot, and it is usually hard to achieve by human operators. The estimated error rate is 14%, and it is actually only 5%.

Let's make some final remarks here. It is clear that the smaller the difference between waveforms, the worse the classification. If the largest difference between waveforms is even smaller than the standard deviation of noise, we cannot expect any useful results from this method. For instance, in the previous experiment, when the true difference is 0 at the 3rd position, the estimated value is an absurd $-62$; but when the true difference is 30 at the 14th position, the estimated value is 31. Although the estimated upper bound of error rates is a guarantee for the correctness, but it seems a little bit too conservative. It leaves some room of improvements. Finally, it is an interesting and important task to extend this method to the classification of more than two groups.

## Reference

[1] Lewicki M (1998) Network: Comput. Neural Syst. 9:R53–R78.

[2] Diedrich A, Charoensuk W, Brychta RJ, Ertl AC and Shiavi R (2003) IEEE Trans Biomed Eng. 50:41–50.

[3] Li Y-L, *On a Classification Method for Neural Signals*, Master Thesis, Dept of Math, National Central Univ, June 2005.