A linear-discriminant-analysis-based approach to enhance the performance of fuzzy c-means clustering in spike sorting with low-SNR data

Chien-Wen Cho[1] gustafcho@gmail.com
[1] *Department of Electrical and Control Engineering, National Chiao Tung University, No. 1001, Ta-Hsueh Rd., Hsinchu City, Taiwan 300, R.O.C.*

Wen-Hung Chao[1, 2] wenhong@mail.ypu.edu.tw
[2] *Department of Biomedical Engineering, Yuanpei University, No.306, Yuanpei St., Hsinchu City, Taiwan 300, R.O.C.*

You-Yin Chen[2,*] irradiance@so-net.net.tw
[1] *Department of Electrical and Control Engineering, National Chiao Tung University, No. 1001, Ta-Hsueh Rd., Hsinchu City, Taiwan 300, R.O.C.*
[*] *Corresponding author: Tel: +886-3-571-2121 ext 54427; Fax: +886-3- 612-5059.*

Abstract

Spike sorting is of prime importance in neurophysiology and hence has received considerable attention. However, conventional methods suffer from the degradation of clustering results in the presence of high levels of noise contamination. This paper presents a scheme for taking advantage of automatic clustering and enhancing the feature extraction efficiency, especially for low-SNR spike data. The method employs linear discriminant analysis based on a fuzzy *c*-means (FCM) algorithm. Simulated spike data [1] were used as the test bed due to better a priori knowledge of the spike signals. Application to both high and low signal-to-noise ratio (SNR) data showed that the proposed method outperforms conventional principal-component analysis (PCA) and FCM algorithm. FCM failed to cluster spikes for low-SNR data. For two discriminative performance indices based on Fisher's discriminant criterion, the proposed approach was over 1.36 times the ratio of between- and within-class variation of PCA for spike data with SNR ranging from 1.5 to 4.5 dB. In conclusion, the proposed scheme is unsupervised and can enhance the performance of fuzzy *c*-means clustering in spike sorting with low-SNR data.

Keywords: Spike sorting; spike classification; fuzzy c-means; principal-component analysis; linear discriminant analysis; low-SNR.

1. INTRODUCTION

The recording of neural signals is of prime importance to monitoring information transmission by multiple neurons. The recorded waveform usually consists of action potentials (i.e., spikes) from several neurons that are in close proximity to the electrode site [2]. A number of reports in systems neuroscience have assumed that brain encodes information in the firing rate of neurons. Consequently, detecting the spiking activity in electrophysiological recordings of the brain is seen to be essential to the decoding of neural activity.

The analysis of neuronal recordings consists of two main general steps: (1) detecting and confirming waveform candidates that are possible action potentials, and (2) distinguishing different spikes and generating a series of spike trains according to the temporal sequence of action potentials. The unique and reproducible shape of spikes produced by each neuron allows the spiking activity of different neurons to be distinguished [2-3].

During the past 3 decades, various classification methods ranging from simple amplitude discrimination to a neural-network classifier have been applied to spike sorting [4-7]. Conventional methods such as principal-component analysis (PCA), fuzzy *c*-means (FCM), and the use of simple ad-hoc features such as the peak-to-peak amplitude and spike duration can be useful for spike recordings when the signal-to-noise ratio (SNR) is sufficiently high [4, 8]. However, these methods become inadequate for discriminating spikes in the presence of high background noise.

Many efforts – including supervised and unsupervised approaches – have been dedicated to improving spike sorting in the presence of high noise. The use of a supervised classifier produces acceptable results even under a very high background noise [9-10]. However, an automated method for spike sorting is necessary, at least in the initial analysis of experimental data that sets a basis for further analysis using some form of supervised classification algorithm for spike sorting. Kim and Kim [11-12] demonstrated an unsupervised approach comprising a spike detector, negentropy-maximization-based projection pursuit feature extractor, and an unsupervised classifier using a mixed Gaussian model. One of the key concepts is that feature extraction and dimensionality reduction can be combined together using a linear transform expressed as $\mathbf{y} = \mathbf{W}^T\mathbf{x}$, where $\mathbf{x}$ and $\mathbf{y}$ are the observed data and feature vectors, respectively, and $\mathbf{W}$ is a linear projection matrix such that $\mathbf{y}$ becomes discriminative so as to aid separation of the clusters. Many types of optimization criteria can be used to determine an appropriate $\mathbf{W}$, such as maximizing the variance, non-Gaussianity, negentropy, or the ratio of between- and within-class variations [13-14]. The ratio of between- and within-class variations (Fisher's linear discriminant criterion) appears to be an especially valid index since it allows simultaneous balancing of the maximization and minimization of the between- and within-class variations. Based on Fisher's linear discriminant criterion, linear discriminant analysis (LDA) then produces a linear projection matrix, and greatly enhances the classification ability. Inspired by this idea, we have designed an unsupervised spike sorting system that is capable of detecting and classifying spikes even under a low-SNR condition.

Our method combines action potential detectors [15], LDA-based feature extraction, and FCM clustering. This combination is unsupervised because the spike features are automatically clustered by the FCM algorithm. The proposed scheme can also resolve the low-SNR problem thanks to the high discriminative ability provided by LDA.

2. MATERIALS AND METHODS

2.1 The overall spike sorting system

The proposed system, as illustrated in Fig. 1, is an automated neural spike sorting system that does not require interactive human input, and shows high performance under a low-SNR condition. The system can be divided into three main stages. In the first stage, all possible action potentials are obtained using a detector. In the second stage, the projection matrix is trained for feature extraction. This is because although LDA can offer more effective discriminating features than PCA or when analyzing the original signal domains, it has the drawback of needing supervised knowledge about the targets of action potentials. Therefore, FCM clustering is used to analyze the obtained action potentials beforehand to form classification targets for LDA training. In the third stage, an LDA projection matrix projects all the detected action potential candidates to a new domain, the canonical space, which is named the LDA space in this paper for convenience.

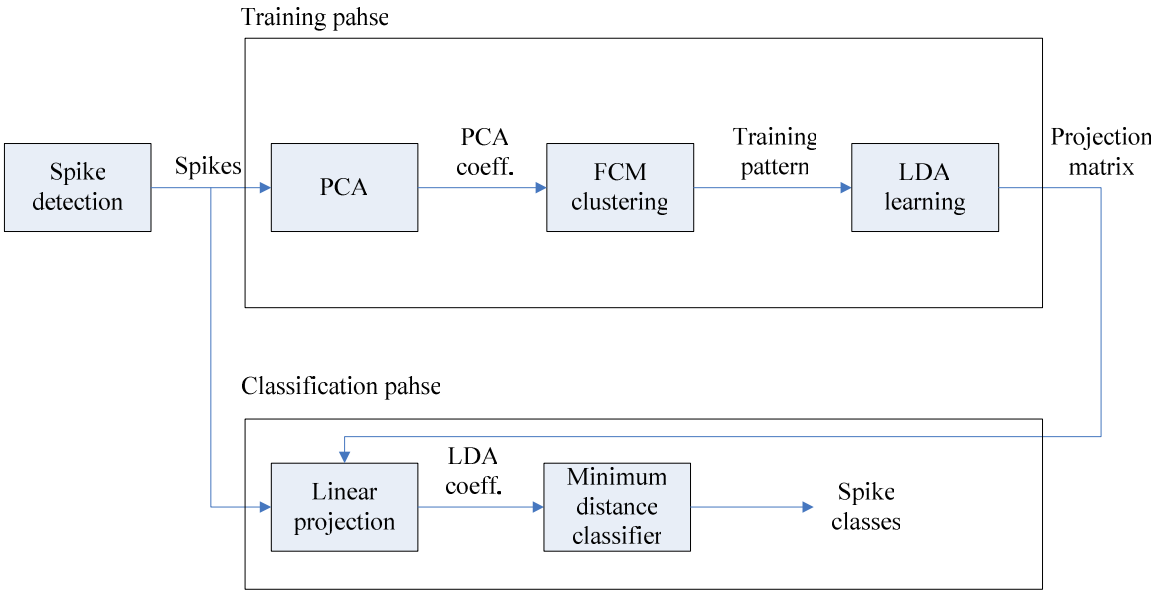

FIGURE 1: Overall structure of the proposed spike sorting system.

2.2 Construction of simulated signals

Simulated signals were constructed from the data records provided by Quiroga et al. [1], which have 594 different average spike shapes. Background noise was generated by randomly selecting spikes from the database and superimposing them at random times and amplitudes. This method was employed so as to mimic the background noise of actual recordings as generated by the activity of distant neurons. Next, a train of three distinct spike shapes from the database were superimposed on the noise signal at a random peak value of 1. In Eq. (1), we represented the noise level using the SNR as follows:

$$\text{SNR (dB)} = 20 \log_{10} \frac{\text{peak value of action potential with minimum amplitude}}{\text{root-mean-square value of pure noise segment}} \quad (1)$$

By simulation, the interspike intervals of the three distinct spikes followed a Poisson distribution with a mean firing rate of 20 Hz. Note that constructing noise from spikes also ensures that the noise and spikes exhibit similar power spectra.

2.3 Spike detection

Spike detection, as the first step of processing the obtained action potentials, can be categorized into three main groups, based on (1) the peak-to-peak threshold, (2) template matching, and (3) statistical strategies [15-17]. To avoid the artificial operations needed in the first and second method categories, we adopted the method proposed by Donoho and Johnstone [15]. The threshold (Th) is selected as

$$T_h = 4\sigma_n \quad (2)$$

where σ_n is an estimate of the standard deviation of the background signal

$$\sigma_n = \text{median}\left\{\frac{|\mathbf{x}|}{0.6745}\right\} \quad (3)$$

and $\mathbf{x}$ is the signal comprising spikes and noise. It is noted that taking the standard deviation of the signal could lead to very high threshold values, especially in cases with high firing rates and large spike

amplitudes. In contrast, by using the estimation based on the median, the interference of the spikes is diminished (under the reasonable assumption that spikes amount to a small fraction of all samples).

2.4 Feature extraction

As reported in the review of Lewicki [3], early studies on spike sorting simply detected spikes using the height of an action potential. The width and peak-to-peak amplitude were also used to characterize the shape features of spikes when the computing resources were very limited. However, choosing features based on this intuitive approach often results in poor cluster separation. This prompted the use of PCA to find an ordered set of orthogonal basis vectors that capture the directions in the data of largest variation [4, 17]. LDA is another commonly used approach for discrimination [13, 18, 19], and is reported to be more efficient than PCA experimentally except for very small training sets [18, 20-21]. LDA aims to find an optimal transformation by maximizing the between-class distance and simultaneously minimizing the within-class distance, thus achieving maximum discrimination:

$$\max_{\mathbf{W}} \text{Trace}\{(\mathbf{W}^{\text{T}}\mathbf{S}_{\mathbf{w}}\mathbf{W})^{-1}(\mathbf{W}^{\text{T}}\mathbf{S}_{\mathbf{b}}\mathbf{W})\} \tag{4}$$

where

$$\mathbf{S}_{\mathbf{w}} = \frac{1}{N_T}\sum_{i=1}^{L}\sum_{j=1}^{N_i}(\mathbf{x}_{ij} - \mathbf{m}_i)(\mathbf{x}_{ij} - \mathbf{m}_i)^{\text{T}} \tag{5}$$

$$\mathbf{S}_{\mathbf{b}} = \frac{1}{N_T}\sum_{i=1}^{L} N_i(\mathbf{m}_i - \mathbf{m}_{\mathbf{x}})(\mathbf{m}_i - \mathbf{m}_{\mathbf{x}})^{\text{T}} \tag{6}$$

$\mathbf{S}_{\mathbf{w}}$ and $\mathbf{S}_{\mathbf{b}}$ are the between-class and within-class matrices, respectively. $\mathbf{x}_{ij}$ is the j-th vector corresponding to the i-th class center, $\mathbf{m}_i$ and $\mathbf{m}_{\mathbf{x}}$ are the centers of overall vectors. L, N_i, and N_T are the number of classes, the number of vectors of the i-th class, and the total number of the overall data vectors [22]. By applying eigendecomposition to the scatter matrices, the optimal transformation $\mathbf{W}$ is readily evaluated by computing the eigenvectors of $\mathbf{S}_{\mathbf{w}}^{-1}\mathbf{S}_{\mathbf{b}}$. Although LDA has an intrinsic limitation of requiring one of the scatter matrices of the objective function to be nonsingular, this problem can be overcome by using the PCA+LDA algorithm [18].

2.5 Choosing the number of classes and fuzzy *c*-means clustering

It is important to choose an appropriate number of classes. Bayesian approaches [23] can be used to estimate the probability of each model when assuming different numbers of classes given the observed data. Fuzzy approaches have also been used to estimate a suitable number of clusters, as studied by Xie and Beni [24]. In this study we determined the number of clusters for FCM by performing FCM with an increasing number of clusters, beginning with two clusters. We investigated the matrix of the Mahalanobis distances between each pair of group means; the Mahalanobis distance is a normalization technique that does not require a new threshold value to be specified for different experiments. Furthermore, according to multivariate analysis, we could also apply a sequence of tests by p values (with the threshold set to 0), which can be computed after LDA. This estimation method allows the number of clusters to be chosen without human intervention.

After selecting a reasonable number of clusters, a clustering algorithm is used to separate multidimensional data into different groups. Simple approaches such as ISODATA and k-means clustering [25] can be used. However, we adopted FCM to calculate the fuzzy centers [26], which involves finding locally optimal fuzzy clustering of the data based on locally minimizing the generalized least-square error function

$$J_m(\mathbf{U}, \mathbf{v}) = \sum_k \sum_i (u_{ik})^m \left\| y_k - v_i \right\|^2 \tag{7}$$

where u_{ik} is the membership strength of pattern y_k in cluster i, v_i is the center of the i-th cluster, y_k is the k-th training pattern, and m is a weighting exponent that sets the fuzziness. The magnitude-squared term in Eq. (7) is the Euclidean norm, and is replaced by the norm

$$d_{ik} = \|y_k - v_i\| = \left(\sum_n (y_k(n) - v_i(n))^2 \right)^{1/2}. \quad (8)$$

In our method Eqs. (7) and (8) are used to find the centers of the clusters that will partition the set of training patterns so as to minimize the least-squared error function $J_m(\mathbf{U}, \mathbf{v})$. The clustering process starts with a random set of centers. Centers and membership strengths are calculated, and the root-mean-square difference is calculated between the current partition, U^n and the previous one, U^{n-1}. The process terminates when the difference is below a specified threshold (0.001% in the experiments reported here).

2.6 Performance indices

We used two performance indices based on the scatter matrix to quantitatively compare the efficacy of the proposed method with those of other approaches based on linear transformations [12]. Each index is a function of transform matrix $\mathbf{W}$. The first index is used for Fisher's discriminant function, and is defined as follows:

$$J_1(\mathbf{W}) = \frac{\det(\tilde{\mathbf{S}}_\mathbf{b})}{\det(\tilde{\mathbf{S}}_\mathbf{w})} = \frac{\det(\mathbf{W}^\mathrm{T}\mathbf{S}_\mathbf{b}\mathbf{W})}{\det(\mathbf{W}^\mathrm{T}\mathbf{S}_\mathbf{w}\mathbf{W})} \quad (9)$$

$\mathbf{S}_\mathbf{b}$ and $\tilde{\mathbf{S}}_\mathbf{b}$ represent the between-class matrices before and after a linear projection by $\mathbf{W}$, respectively, and $\tilde{\mathbf{S}}_\mathbf{w}$ and $\mathbf{S}_\mathbf{w}$ are the associated within-class matrices. The second index, the generalized Fisher linear discriminant function, which can be interpreted in a similar context, is defined as

$$J_2(\mathbf{W}) = \mathrm{Trace}\{(\mathbf{W}^\mathrm{T}\mathbf{S}_\mathbf{w}\mathbf{W})^{-1}(\mathbf{W}^\mathrm{T}\mathbf{S}_\mathbf{b}\mathbf{W})\} \quad (10)$$

3. RESULTS

3.1 Spike sources and determining the number of clusters

The overall system was tested for three-unit clustering. The simulated signals are shown in Fig. 2A. We further extracted 300 spikes and aligned them with their highest amplitudes in the center of the waveforms, shown in Fig. 2B, by the thresholding method mentioned in Section 2.1 and preprocessed them using PCA to reduce their dimensionality. In the experiments, the total number of clusters for grouping using FCM was initially set to be 2 and was increased gradually. For demonstration purpose, we only show and compare the clustering result with the cases of the total cluster number for 2 to 4.

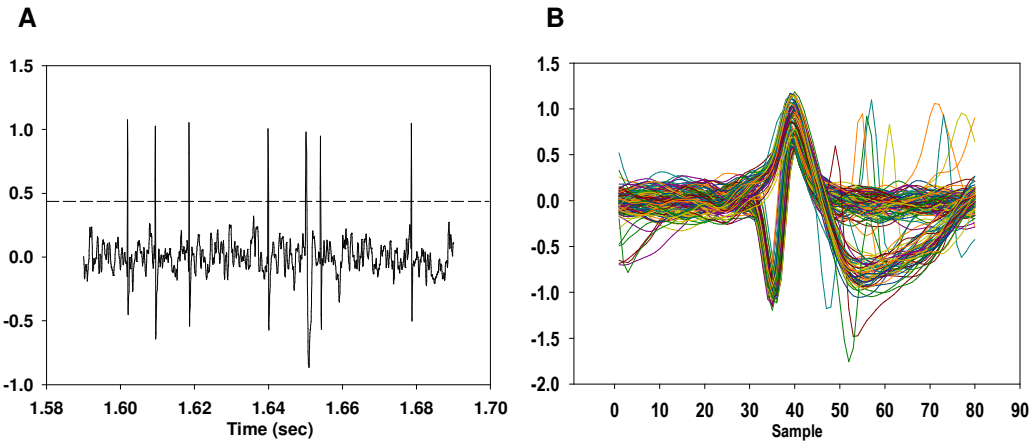

FIGURE 2: (A) Simulated signals, with the threshold shown by the horizontal line. (B) Extracted spikes.

The categorization results are illustrated, in Figs. 3-5. For the grouping with the number of clusters set to two, we observed that the first cluster had relatively consistent waveform members (red lines in Fig. 3) while the second cluster (blue lines in Fig. 3) could be further divided into more classes. The corresponding average waveforms of these two clusters of waveforms are shown in the inset of Fig. 3.

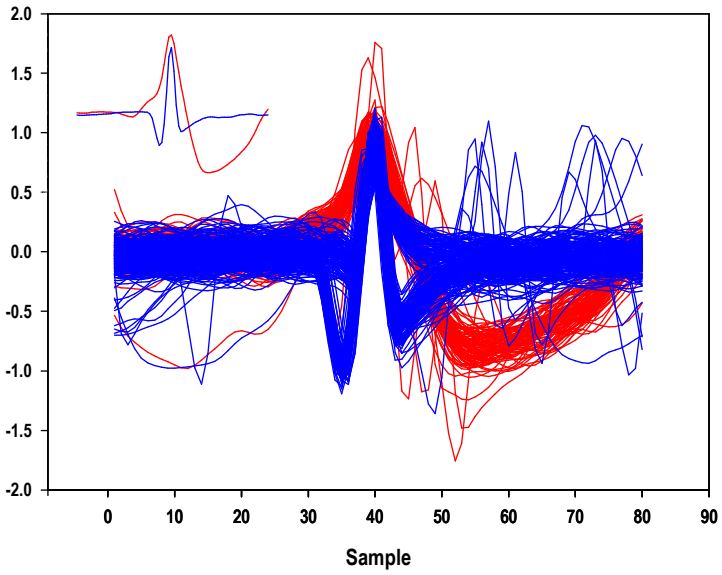

FIGURE 3: Grouping results after FCM with the desired number of clusters set to two. The obtained two clusters are indicated by the red and blue lines . The inset shows the corresponding average waveforms

Fig. 4 illustrates the results with the total number of clusters set to three, where the waveforms of the

first, second, and third clusters are indicated by the red, blue, and green lines, respectively. It is evident that this produces better clustering results than those in Fig. 3. To provide an overview of these three clusters of signals, the inset shows the associated three average waveforms.

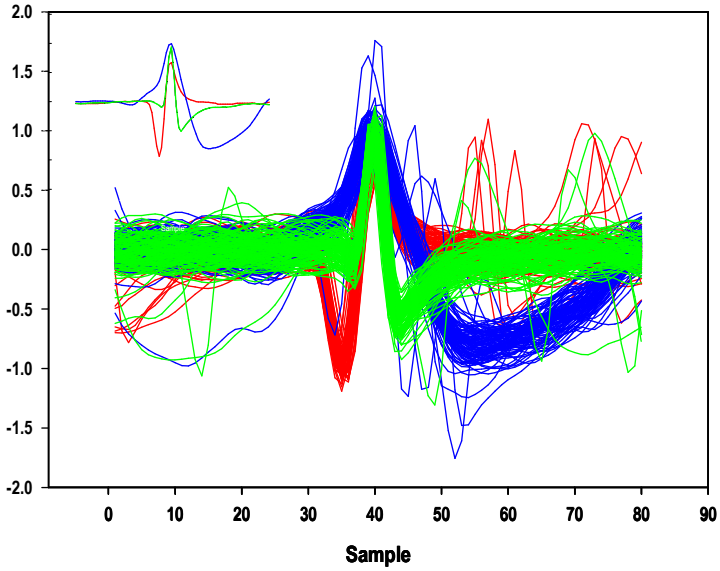

FIGURE 4: Grouping results after FCM with the desired number of clusters set to three. The obtained three clusters are indicated by the red, blue, and green lines. The inset shows the corresponding average waveforms

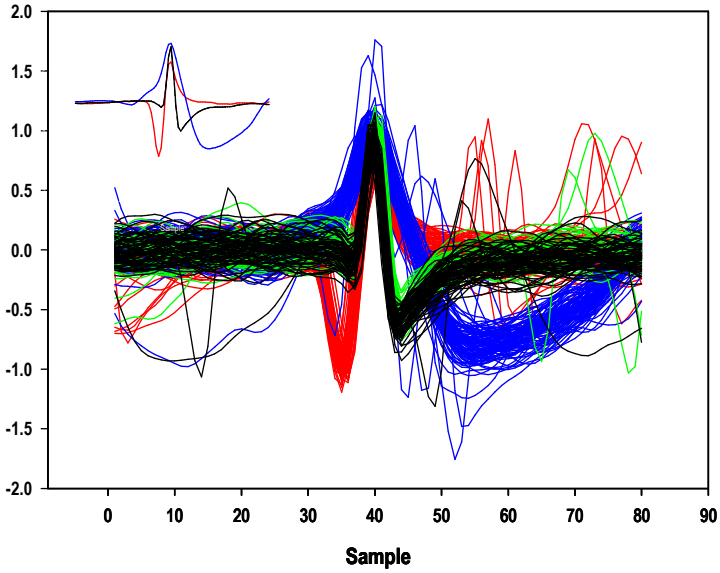

FIGURE 5: Grouping results after FCM with the desired number of clusters set to four. Only three meaningful clusters were produced (indicated by the red, blue, and green lines). The inset shows the corresponding average waveforms, which clearly indicates that the fourth cluster (black lines) was almost identical to the third cluster

Finally, Fig. 5 shows the results of having four clusters in FCM clustering. The figure shows that the third and fourth clusters had almost the same types of waveforms (green and black lines, respectively),

as also confirmed by the corresponding average waveforms plotted in the inset. This indicates that there were essentially only three units of signals. Table 1 is the matrix of Mahalanobis distances between each pair of group means with the number of clusters set to five, in order to verify the validation of the estimated number of clusters. Note that redundant clusters are indicated by a very small Mahalanobis distance (cluster 4 and cluster 5 are close to cluster 1 with Mahalanobis distance 3.24×10^{-10} and 4.12×10^{-6}, respectively).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1		1.26×10^{1}	4.06×10^{0}	3.24×10^{-10}	4.12×10^{-6}
Cluster 2	1.26×10^{1}		1.76×10^{1}	1.26×10^{1}	1.26×10^{1}
Cluster 3	4.06×10^{0}	1.76×10^{1}		4.06×10^{0}	4.06×10^{0}
Cluster 4	3.24×10^{-10}	1.26×10^{1}	4.06×10^{0}		4.20×10^{-6}
Cluster 5	4.12×10^{-6}	1.26×10^{1}	4.06×10^{0}	4.20×10^{-6}	

TABLE 1: The matrix of Mahalanobis distances between each pair of group means. Redundant clusters are indicated by a very small Mahalanobis distance.

3.2 Comparison of clustering abilities without noise contamination

Since we had determined a reasonable number of clusters by comparing the clustering results, we visualized the spike data in different planes (for best illustration, we show only two most important coefficients in each plane). Fig. 6A is a scatter plot of the spikes drawn using two of the coefficients (35th and 36th coefficients in the experiment) such that the chosen features had the largest variations to spread the spike curves. The figure shows the distribution of the spikes projected onto the PCA+LDA plane (denoted as LDA only for brevity, Fig. 6B) and the associated two components, or bases (Fig. 6D). It is evident that LDA exhibited better discriminative ability than PCA (the scatter plot in Fig. 6C and associated principal components in Fig. 6E, respectively). It is noted that, comparison of Fig. 6A and 6B indicates that PCA also had a better separating ability than the original domain.

3.3 Comparison of clustering ability under a low-SNR condition

To investigate the classification ability for very-low-SNR spike data enhanced by the LDA-based FCM algorithm, we tested spike data that were contaminated with random noise to produce SNRs of 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, and 4.5. For brevity, we only showed the scatter plot of the spikes before LDA under SNR = 1.5 in Fig. 7A. As illustrated in Fig. 7B, the spikes could only be clustered into two (and not three) groups due to the presence of the high-level noise. Fig. 7C is the scatter plot of these two groups of spikes. The clustering results of LDA and PCA are shown in Fig. 7D and 7E, respectively. It is clear that LDA exhibited the best clustering performance. To further quantitatively compare the discriminative abilities of PCA and LDA, Eqs. (9) and (10) were used as two indices to evaluate the grouping performance. Fig. 8 shows these two indices as functions of the SNR, which indicates that LDA is superior to PCA over a broad range of SNRs.

A **B**

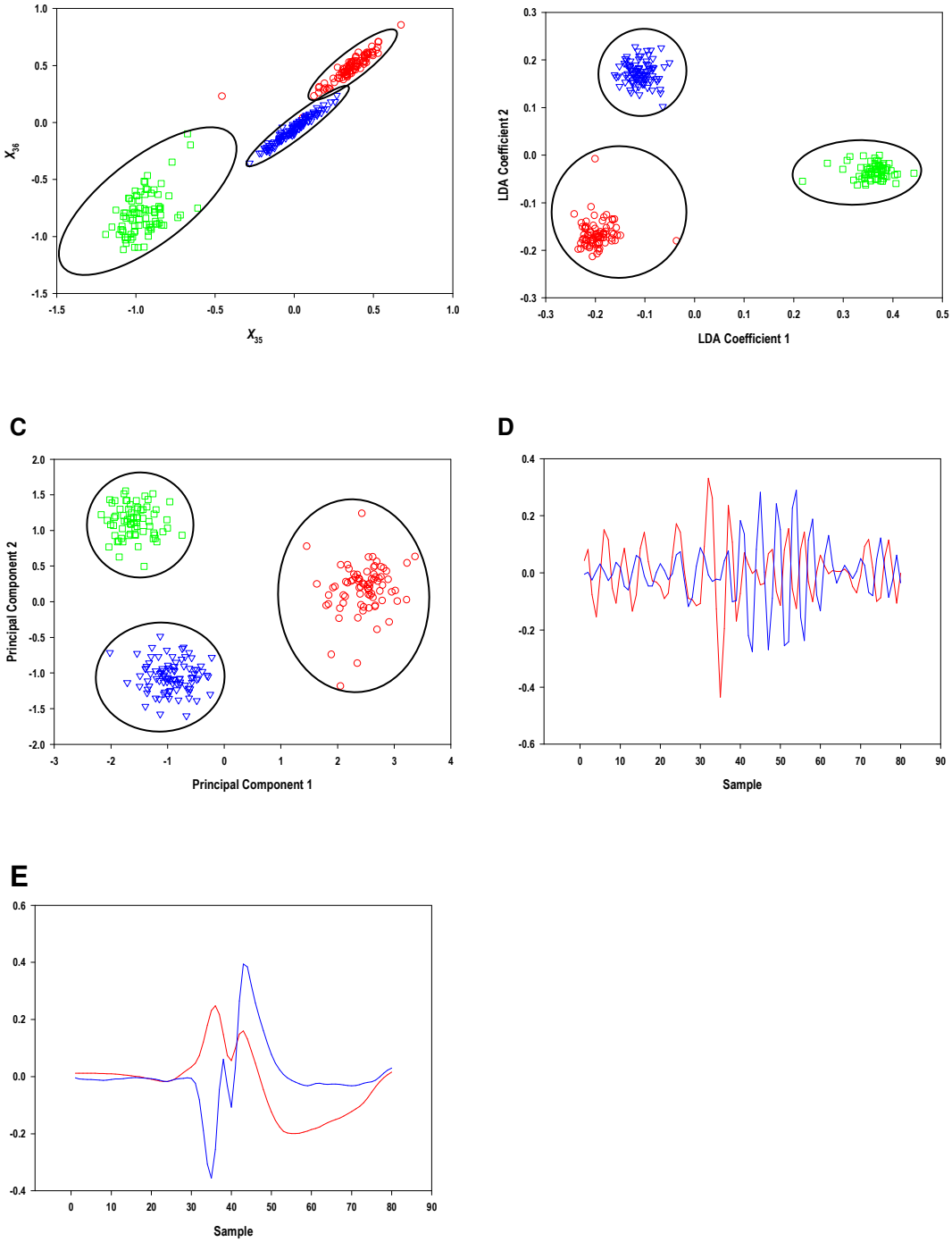

FIGURE 6: Scatter plots of the spike vectors under different spaces. (A-C) Scatter plots on the original, LDA, and PCA spaces, respectively. (D and E) Associated LDA and PCA components, respectively

A **B**

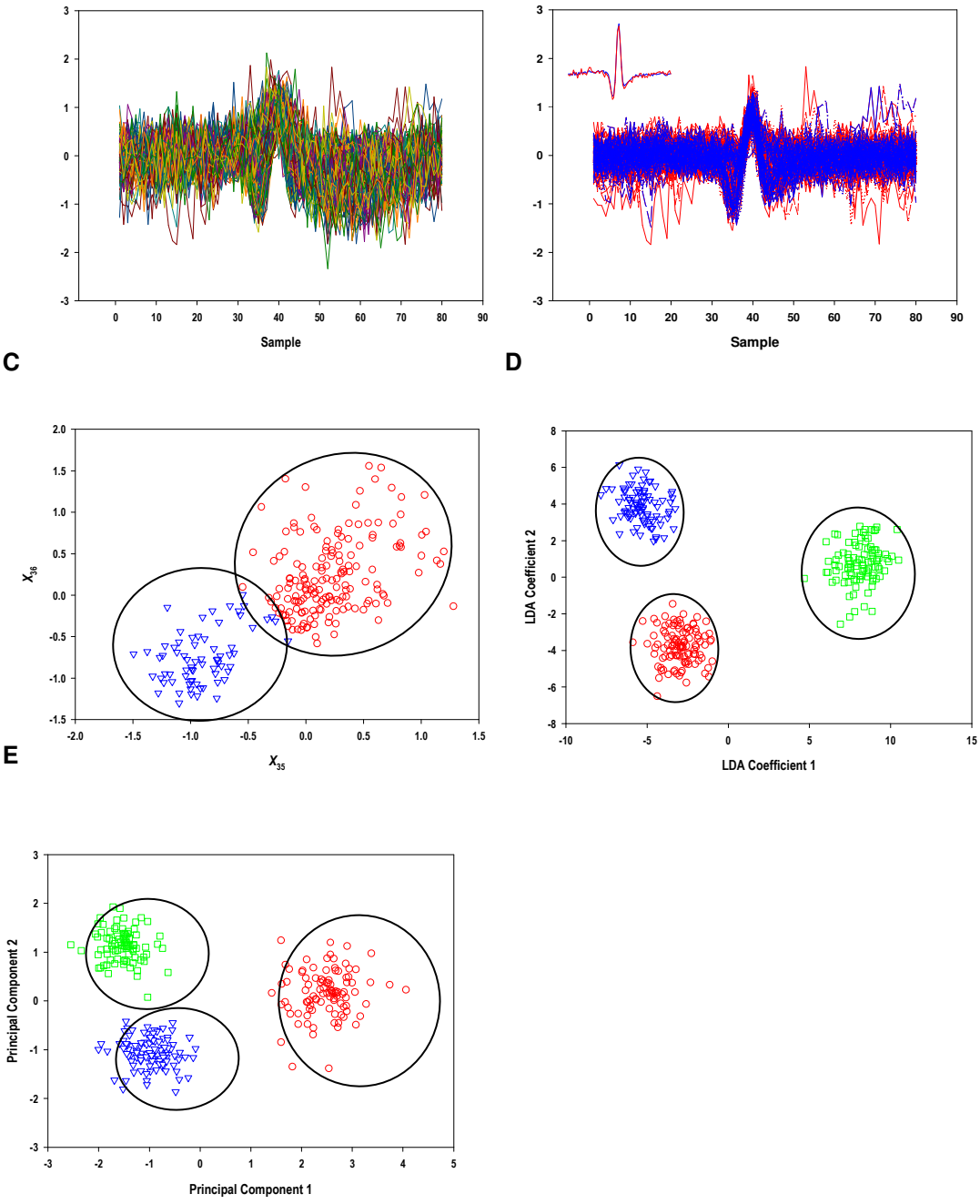

FIGURE 7: FCM alone failed to separate generate the third class when SNR = 1.5 dB. (A) Spikes before clustering. (B) First and second clusters after clustering in the time domain, with the inset showing the two corresponding average waveforms . (C) Scatter plot of clustered result of the original space. (D and E) Plots of clustering using LDA and PCA, respectively.

4. CONCLUSION & FUTURE WORK

We implemented an automatic neural spike sorting system that does not require interactive human input. The first step of the proposed scheme involves extracting spikes using a detector. The desired number of clusters is then iteratively changed with the obtained spikes clustered using FCM, and the distances between cluster centers are quantified using the Mahalanobis distance. Assigning too many clusters results in FCM producing clusters with almost identical centers, as shown by Fig. 5 and Table 1. As reported by Zouridakis and Tam [8], FCM is suitable for clustering spike data with a low noise level, and this is verified in Fig. 4. However, they did not investigate the grouping performance in the presence of very high noise levels. The experiments performed in this study revealed that FCM might fail to separate spikes into a sufficient number of clusters due to noise contamination (see Fig. 7B and 7C).

Besides FCM, some researchers have also used several types of linear transformation for feature extraction in the sorting of data [27-29]. Among them, PCA is arguably the best known and most widely used. PCA has previously been utilized for feature extraction by Richmond and Optican [30] in spikes generated by the primate inferior temporal cortex. However, PCA has limited utility when the signals of interest are sparsely distributed, such as when the difference between firing patterns is based on one or a small number of spikes occurring within a narrow time window. This issue arises due to PCA emphasizing global features in signals [31] - PCA focuses on computing eigenvectors accounting for the largest variance of the data are selected, but these directions do not necessarily provide the best separation of the spike classes.

In this study, we utilized FCM as the clustering strategy for implementing an unsupervised spike sorting system. We further improved the discriminating ability by incorporating LDA [13, 18-19] for feature extraction, which is a frequently used method for classification and dimension reduction. LDA and its variations thereof have been used widely in many applications, particularly in face recognition. LDA aims to find an optimal transformation by minimizing the within-class distance and maximizing the between-class distance simultaneously, thus maximizing the discrimination ability. However, such an approach cannot be implemented in an unsupervised way. In practice, it is very difficult to perform spike sorting by directly applying supervised approaches during the course of an experiment. Thus, we incorporated FCM to avoid the problem resulting from the lack of a priori knowledge of spike targets.

The clustering performances of FCM alone, PCA, and LDA were compared by visualizing scatter plots of the spikes. As shown in Fig. 6 and 7, the proposed scheme outperformed FCM and PCA. This is because FCM exhibits low noise tolerance and PCA only focuses on maximizing the variation rather than the grouping performance. A further comparison of PCA and LDA under a low-SNR condition is provided in Fig. 8. LDA exhibited better (greater) performance values for both indices J_1 and J_2. For instance, at SNR = 1.5 dB, J_1 for LDA and PCA was 50.63 and 23.86, respectively, and J_2 for LDA and PCA was 28.41 and 20.83, respectively. That is, at SNR = 1.5 dB, LDA was 2.12 times the ratio of between- and within-class variation of PCA for J_1, and 1.36 times for J_2. When SNR was 4.5, J_1 for LDA and PCA was 381.62 and 165.67, respectively, and J_2 for LDA and PCA was 42.27 and 28.25, respectively. That is, at SNR = 4.5 dB, LDA was 1.50 times the ratio of between- and within-class variation of PCA for J_1, and 1.48 times for J_2. The results revealed that LDA improves the grouping performance for low-SNR spike data.

The performance of the proposed unsupervised spike sorting system could be further improved by combining more powerful feature extraction approaches, such as wavelet and Gabor transformations [1, 32-33]. Wavelet transformation has been proven to be excellent for reducing noise and signal reconstruction, and Gabor transformation makes use of both time- and frequency-domain information. The use of more feature coefficients to represent spikes increases the importance of the efficiency of the dimension reduction technique. Incorporating such feature extraction methods into the proposed scheme should be investigated in future work.

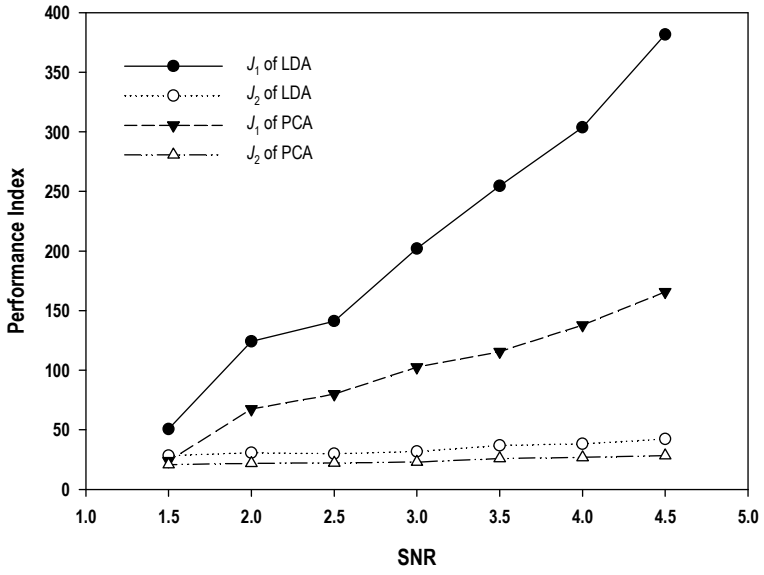

FIGURE 8: Performance comparison of LDA and PCA using two indices: J_1 and J_2.

5. REFERENCES

1. R. Q. Quiroga, Z. Nadasdy and Y. Ben-Shaul. *"Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering"*. Neural Comp, 16(8):1661-1687, 2004
2. F. Rieke et al. *"Spikes: exploring the neural code"*, Cambridge, MA: The MIT Press (1996)
3. M. S. Lewicki. *"A review of methods for spike sorting: The detection and classification of neural action potentials"*. Network: Computation Neural Syst, 9(4):R53-R78, 1998
4. B. C. Wheeler and W. J. Heetderks. *"A comparison of techniques for classification of multiple neural signals"*. IEEE Trans Biomed Eng, 29(12):752-759, 1982
5. K. Mirfakhraei and K. Horch. *"Classification of action potentials in multi-unit intrafascicular recordings using neural network patern-recognition techniques"*. IEEE Trans Biomed Eng, 41(1):89-91, 1994
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 Biomed Eng, 41(4):358-372, 1994
7. X. Yang and S. A. Shamma. *"A totally automated system for the detection and classification of neural spikes"*. IEEE Trans Biomed Eng, 35(10):806-816, 1988
8. G. Zouridakis and D. C. Tam. *"Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering"*. Comp Meth Prog Biomed, 61(2):91-98, 2000
9. R. Chandra and L. M. Optican. *"Detection, classification, and superposition resolution of action potentials in multiunit single channel recordings by an on-line real-time neural network"*. IEEE Trans Biomed Eng, 44(5):403-412, 1997
10. K. H. Kim and S. J. Kim, "*Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier*". IEEE Trans Biomed Eng, 47(10):1406-1411, 2000
11. K. H. Kim and S. J. Kim. *"Method for unsupervised classification of multiunit neural signal recording under low signal-to-noise ratio"*. IEEE Trans Biomed Eng, 50(4):421-431, 2003
12. K. H. Kim, *"Improved algorithm for fully-automated neural spike sorting based on projection pursuit and Gaussian mixture model"*. Inter J Control Autom Syst, 4(6):705-713, 2006
13. C. Liu and H. Wechsler. Enhanced Fisher linear discriminant model for face recognition. In Proceedings of Int Conf Pattern Recognition.1368-1372, 1998

14. A. Hyvarinen et al. *"Independent component analysis"*, New York: Wiley (2001)
15. D. Donoho and I. M. Johnstone. *"Ideal spatial adaptation by wavelet shrinkage"*. Biometrika, 81(3):425-455, 1994
16. R. Tasker, Z. Israel and K. Burchiel. *"Microelectrode recording in movement disorder surgery"*. New York: Thieme Medical Publishers (2004)
17. G. L. Gerstein and W. A. Clark. *"Simultaneous studies of firing patterns in several neurons"*. Science,143(3612):1325-1327, 1964
18. A. Martinez and M. Kak. *"PCA versus LDA"*. IEEE Trans Pattern Analysis Machine Intell, 23(2): 233-288, 2001
19. J. Ye, Q. Li. *"A two-stage linear discriminant analysis via QR-decomposition"*. IEEE Trans Pattern Analysis Machine Intell, 27(6):929-941, 2005
20. D. L. Swets and J. Weng. *"Using discriminant eigenfeatures for image retrieval"*. IEEE Trans Pattern Anal Machine Intell 18(8):831-836, 1996
21. P. N. Belhumeour, J. P. Hespanha and D. J. Kriegman. *"Eigenfaces vs. fisherfaces: recognition using class specific linear projection"*. IEEE Trans Pattern Anal Machine Intell, 19(7):711-720, 1997
22. H. R. Barker. *"Multivariate Analysis of Variance (MANOVA): A Practical Guide to Its Use in Scientific Decision-Making"*, AL: University of Alabama Press (1984)
23. M. S. Lewicki. *"Bayesian modeling and classification of neural signals"*. Neural Comput, 6(5):1005-1030, 1994
24. X. L. Xie and G. A. Beni. *"A validity measure for fuzzy clustering"*. IEEE Trans Pattern Anal Machine Intell 13(8):841-847, 1991
25. R. O. Duda and P. E. Hart. *"Pattern Classification and Scene Analysis"*, New York: John Wiley & Sons (1973)
26. J. C. Bezdek, R. Ehrlich and W. Full. "*FCM: The fuzzy c-means clustering algorithm*". Comput Geosci, 10(2):191-203, 1984
27. J. H. Friedman. *"Exploratory projection pursuit"*. J Ameri Statist Associat, 82(1):249-266, 1987
28. R. A. Fisher. "*The use of multiple measurements in taxonomic problems*". Ann eugenics, 7(2):179-188, 1936
29. Y. Meyer. *"Wavelet analysis book report"*. Bull Am Math Soc (New Series), 28(2):350-360, 1993
30. B. J. Richmond and L. M. Optican. *"Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform"*. J Neurophysiol, 57(1):147-161, 1987
31. P. S. Penev and J. J. Atick. *"Local feature analysis: a general statistical theory for object representation"*. Network: Comput Neural Syst, 7(3):477-500, 1996
32. S. Qian and D. Chen. *"Discrete Gabor transform"*. IEEE Trans Signal Process, 41(7): 2429-2438, 1993
33. J. C. Letelier and P. P. Weber. *"Spike sorting based on discrete wavelet transform coefficients"*. J Neurosci Methods, 101(2):93-106, 2000