



# HHS Public Access

Author manuscript

*Proc IEEE Int Conf Acoust Speech Signal Process.* Author manuscript; available in PMC  
2018 March 08.

Published in final edited form as:

*Proc IEEE Int Conf Acoust Speech Signal Process.* 2016 March ; 2016: 749–753. doi:10.1109/ICASSP.

2016.7471775

## CLUSTERING OF INTERICTAL SPIKES BY DYNAMIC TIME WARPING AND AFFINITY PROPAGATION

John Thomas<sup>a</sup>, Jing Jin<sup>a</sup>, Justin Dauwels<sup>a</sup>, Sydney S. Cash<sup>b</sup>, and M. Brandon Westover<sup>b</sup>

<sup>a</sup>Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore 639798

<sup>b</sup>Neurology Department, Massachusetts General Hospital and Harvard Medical School, USA

### Abstract

Epilepsy is often associated with the presence of spikes in electroencephalograms (EEGs). The spike waveforms vary vastly among epilepsy patients, and also for the same patient across time. In order to develop semi-automated and automated methods for detecting spikes, it is crucial to obtain a better understanding of the various spike shapes. In this paper, we develop several approaches to extract exemplars of spikes. We generate spike exemplars by applying clustering algorithms to a database of spikes from 12 patients. As similarity measures for clustering, we consider the Euclidean distance and Dynamic Time Warping (DTW). We assess two clustering algorithms, namely, K-means clustering and affinity propagation. The clustering methods are compared based on the mean squared error, and the similarity measures are assessed based on the number of generated spike clusters. Affinity propagation with DTW is shown to be the best combination for clustering epileptic spikes, since it generates fewer spike templates and does not require to pre-specify the number of spike templates.

### Index Terms

Epilepsy; Electroencephalogram; Affinity propagation; K-means; Dynamic time warping; Euclidean distance

## 1. INTRODUCTION

Epilepsy refers to a group of chronic brain disorders, which can be characterized by unprovoked seizures. It affects 50 million people worldwide as reported in 2004 [1]. The spikes in the electroencephalogram (EEG) are the biomarkers for epilepsy [2]. Diagnosis based on spike detection is a tedious task as experts have to monitor the EEG recordings of patients continuously for hours or even days. The spike patterns show a large variation in morphology between patients.

Template matching [3] is one of the common methods in automated spike detection. Due to time constraints and the need for high accuracy, it is crucial to select suitable spike templates. Clustering spikes [4] based on morphological characteristics and selecting templates will help in building a template library. In this paper, we explore two clustering algorithms, viz., K-means clustering [5] and affinity propagation [6], for extracting exemplars of spikes from a large database. We consider two different similarity measures:

the Euclidean distance and Dynamic Time Warping (DTW) [7]. DTW is a robust method for calculating the similarity between temporal sequences. DTW has been applied for face detection [8], template matching [9], clustering [10] and many other applications. DTW has also been shown to improve the performance of various clustering methods such as K-means clustering, hierarchical clustering, and fuzzy clustering [11-13]. Affinity propagation is an efficient clustering technique which has a wide range of applications such as traffic network analysis, audio pattern recognition, gene pattern detection, etc. [6]. In this paper, we extract spike templates by applying clustering algorithms. From our numerical experiments, we observe that the combination of affinity propagation and DTW is highly adequate for identifying spike templates. These templates can in turn be used in semi-automated and automated spike detection systems. A proper choice of spike templates may lead to faster and more accurate spike detection, and ultimately, more reliable diagnosis of epilepsy.

In Section 2, we elaborate on the EEG data analyzed in this paper, and review the K-means and affinity propagation (AP) clustering methods, in addition to the Euclidean distance (ED) measure and Dynamic Time Warping (DTW). In Section 3, we discuss our results for extracting spike exemplars, obtained by combining different clustering algorithms (K-means and AP) and different similarity measures (ED and DTW). In Section 4 we conclude the paper with ideas for future work.

## 2. METHODS

### 2.1. EEG Data

We analyzed an EEG dataset recorded at the Massachusetts General Hospital in Boston, consisting of 30-minute scalp EEG recordings of 12 epileptic patients. The data was recorded according to the international 10-20 electrode system, where the Common Average Reference (CAR) montage was applied. The EEG data was sampled at 128 Hz and an IIR notch filter centered at 60 Hz was applied to remove the powerline noise. A high-pass filter at 1 Hz was applied for removing baseline variations. Epileptic spikes were annotated by two neurologists independently. We consider here only those waveforms that were labelled as spikes by both neurologists. In total, 9240 spikes were extracted, each with a length of 500 milliseconds (equivalently, 64 samples).

### 2.2. Clustering

Two clustering methods were applied to the set of epileptic spikes: K-means clustering [5] and affinity propagation [6]. K-means clustering is one of the simplest clustering algorithms in which the data points are grouped into a specified number of clusters with an objective of reducing the squared error between the data points and cluster centers [5]. The number of clusters is specified apriori. Let  $x = (x_1, x_2, \dots, x_n)$  be a set of  $n$  observations, which is to be clustered into  $k$  (with  $k \leq n$ ) sets  $S = (S_1, S_2, \dots, S_k)$ . The algorithm aims to minimize the objective function:

$$J = \sum_{i=1}^k \sum_{j=1}^n \|x_j - C_i\|^2, \quad (1)$$

where  $C_j$  is the cluster center and  $\|x_j - C_j\|^2$  is the distance between the data point  $x_j$  and  $C_j$ . The cluster center  $C_j$  is determined as the mean of the points in set  $S_j$ . Initially a set of cluster centers is chosen randomly, and then it is updated repeatedly till convergence. Data points are grouped together by minimizing the squared error. The algorithm is terminated after the squared error is below a specific threshold. The main drawback of the algorithm is the need for pre-specifying the number of clusters. In many applications, including the clustering of spikes, the natural number of clusters is unknown apriori. Moreover, oscillations can occur, therefore, properly choosing the initial set of cluster centers is a challenge.

Affinity propagation is a clustering method where data points are grouped by passing real valued messages between the data points [6]. The main advantage of the algorithm is that it is able to automatically determine the number of clusters and the exemplars. The latter are representative data points of each cluster. Two messages are exchanged between the data points, namely responsibility and availability. Responsibility  $r(i, k)$  is sent from the data point  $i$  to the potential exemplar point  $k$ . This message states how well suited the point  $k$  is to serve as the exemplar for the data point  $i$ , taking into consideration all the other potential points. Availability  $a(i, k)$  is sent from an exemplar  $k$  to the data point  $i$ . This message states how appropriate it is for the data point to join the cluster represented by the particular exemplar [6]. The input to the algorithm is the similarity values between the data points and the initial preference of the exemplars. At first the availabilities are initialized to zero and the responsibilities are updated as:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \text{ s.t. } k' \neq k} \{a(i, k') + s(i, k')\}. \quad (2)$$

In the first iteration, the responsibilities are initialized as the input similarities. Next, self-responsibility  $r(k, k)$  is computed. The higher the value of self-responsibility, the higher will be the likelihood to be chosen as an exemplar. Next the availabilities are updated as:

$$a(i, k) \leftarrow \min \{0, r(k, k) + \sum_{i' \text{ s.t. } i' \notin \{i, k\}} \max \{0, r(i', k)\}\}. \quad (3)$$

Self-availability  $a(k, k)$ , which reflects to what extent a point  $k$  is an exemplar, is updated as:

$$a(k, k) \leftarrow \sum_{i' \text{ s.t. } i' \neq k} \max \{0, r(i', k)\}. \quad (4)$$

The algorithm iteratively updates the values of responsibility and availability. The number of clusters is determined by the value of self-responsibility. At any iteration, the values of availabilities and responsibilities combined identify the exemplars. For a point  $i$ , the value of  $k$  which maximizes  $a(i, k) + r(i, k)$  indicates the exemplar for  $i$ . If  $k = i$ , point  $i$  is an exemplar itself. The algorithm is terminated after a fixed number of iterations or until the

cluster decisions stays unchanged for a certain number of iterations. To prevent oscillations, a damping factor is also introduced. Each new update is set to  $\lambda$  times the previous value and  $1-\lambda$  times the new value, where  $\lambda$  takes values between 0 and 1 [6].

For clustering analysis using affinity propagation, the damping factor is set to 0.5. As the initial preference for the exemplars is unknown, it was set to the median value of the similarity measures.

### 2.3. Similarity Measures

To find the similarity between two spikes, we apply the Euclidean distance (ED) and Dynamic Time Warping (DTW). The Euclidean distance  $d$  between  $s_1 = (u_1, u_2, \dots, u_n)$  and  $s_2 = (v_1, v_2, \dots, v_n)$  is defined as (see Fig. 1):

$$d(s_1, s_2) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \dots + (u_n - v_n)^2}. \quad (5)$$

DTW is a non-linear distance measure that calculates the smallest distance between two signals (see Fig. 2) [7]. First the optimal matching of points between the two signals is determined and the distance between them is calculated. Specifically, the DTW distance between  $s_1 = (u_1, u_2, \dots, u_n)$  and  $s_2 = (v_1, v_2, \dots, v_n)$  is calculated as follows. First an  $n$ -by- $n$  matrix is created with entries  $(i, j)$  equal to  $(u_i - v_j)^2$ . The Sakoe-Chiba Band [14] is applied to limit the warping window size. Next all the possible cumulative distances between the two sequences are found for every possible path, and the minimal one is chosen as the DTW distance (see Fig. 3):

$$DTW_{(s_1, s_2)} = \min_{w \in P} \sqrt{\sum_{k=1}^K d_{w_k}}, \quad (6)$$

where  $P$  is the set of all possible warping paths,  $w_k$  is the position  $(i, j)$  at the  $k^{th}$  observation of a warping path, and  $K$  is the length of the warping path [7]. The Sakoe-Chiba band is set to 0.1.

K-means clustering was carried out in two ways: First we applied K-means clustering directly to the spike waveforms; the 64 samples of the waveforms are the input features, and hence the input space is 64-dimensional. Second, we applied K-means clustering to three different features of the spike waveforms: peak-to-peak value, width of the spike, and the non-linear energy operator (NLEO).

The peak-to-peak value is illustrated in Fig. 4a. The width of the spike is obtained from the position of the three peaks [15] as illustrated in Fig. 4b. NLEO [16] is computed as:

$$NLEO(x(n)) = x(n) * x(n) - x(n-1) * x(n+1), \quad (7)$$

where the delay is chosen as 1. The feature values are normalized to the range 0 to 1 as follows:

$$\bar{e}_i = \frac{e_i - E_{min}}{E_{max} - E_{min}}, \quad (8)$$

where  $e_i$  is a feature value, and  $E_{min}$  and  $E_{max}$  are the minimum and maximum values respectively of the feature in the dataset.

Moreover, we applied affinity propagation with two different similarity measures, i.e., negative of ED and negative of DTW distance. The performance in each case is evaluated based on mean squared error, which is the mean distance between data points and the cluster centers. The results for K-means clustering depend on the initial selection of cluster centers. Consequently, we run the K-means clustering 100 times and compute the average error.

### 3. RESULTS AND DISCUSSION

The mean squared error for affinity propagation with ED and K-means clustering is shown in Fig. 5.

To obtain these results, we first applied affinity propagation. Next, we applied K-means clustering with the same number  $K$  of clusters as generated by affinity propagation. In this manner, we can make a fair comparison between the different approaches. The error values for K-means clustering applied to spike features and to the spike waveforms are similar. The values are significantly lower for affinity propagation in comparison with K-means clustering, for all the 12 patients. From this observation, we conclude that affinity propagation outperforms the K-means algorithm for clustering spikes. Moreover, as the number of epileptic spike templates is unknown, affinity propagation has an added advantage over the K-means algorithm as it does not require the number of clusters to be specified a priori.

Affinity propagation with DTW generates a significantly smaller number of clusters than with ED (see Fig. 7). The spikes of the same patient tend to be morphologically more similar in shape. It is well known from the literature that “within patients, interictal discharges tend to be fairly stereotyped” [17]. As DTW yields fewer clusters, it suggests that this distance measure captures the morphological similarity between spikes more effectively. Information about the 6 largest clusters is displayed in Fig. 8. The four largest clusters are depicted in Fig. 6. The combination of affinity propagation with DTW groups epileptic spikes most efficiently. The resulting templates can be used in semi-automated (e.g. spikeGUI [17]) and automated spike detection systems.

### 4. CONCLUSIONS

In this paper, we have shown that affinity propagation in conjunction with DTW is an effective approach for clustering epileptic spikes and identifying the exemplars. In contrast to K-means clustering, the number of clusters does not need to be pre-specified in affinity

propagation. We plan to apply the same approach to a database of EEG recordings from hundreds of epilepsy patients. The resulting exemplars will then provide an operational definition of interictal spikes, summarizing the morphology of hundreds of thousands of spikes. These exemplars can also be utilized in semi-automated and automated detection algorithms.

In future work, we will develop algorithms to select the proper number of exemplars. As Fig. 7 shows, the number of exemplars grows with the number of patients, and seems to gradually saturate in the case of DTW. We will develop an objective measure to select a suitable number of exemplars for the purpose of semi-automated and automated spike detection.

## References

1. Ngugi AK, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Estimation of the burden of active and life-time epilepsy: A meta-analytic approach. *Epilepsia*. 2010; 51(5):883–890. [PubMed: 20067507]
2. Celesia GG, Chen R-C. Parameters of spikes in human epilepsy. *Diseases of the nervous system*. 1976
3. Kim S, McNames J. Automatic spike detection based on adaptive template matching for extracellular neural recordings. *Journal of neuroscience methods*. 2007; 165(2):165–174. [PubMed: 17669507]
4. Nonclercq A, Foulon M, Verheulpen D, De Cock C, Buzatu M, Mathys P, Van Bogaert P. Cluster-based spike detection algorithm adapts to interpatient and inpatient variation in spike morphology. *Journal of neuroscience methods*. 2012; 210(2):259–265. [PubMed: 22850558]
5. MacQueen, J., et al. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*; Oakland, CA, USA. 1967. p. 281-297.
6. Frey BJ, Dueck D. Clustering by passing messages between data points. *science*. 2007; 315(5814): 972–976. [PubMed: 17218491]
7. Niennattrakul V, Ratanamahatana CA. Learning dtw global constraint for time series classification. arXiv preprint arXiv:0903.0041. 2009
8. Sahbi, H., Boujemaa, N. *Biometric Authentication*. Springer; 2002. Robust face recognition using dynamic space warping; p. 121-132.
9. Chang W-D, Im C-H. Enhanced template matching using dynamic positional warping for identification of specific patterns in electroencephalogram. *Journal of Applied Mathematics*. 2014; 2014
10. Huang H-C, Jansen B. Eeg waveform analysis by means of dynamic time-warping. *International journal of bio-medical computing*. 1985; 17(2):135–144. [PubMed: 4055119]
11. Niennattrakul, V., Ratanamahatana, CA. On clustering multimedia time series data using k-means and dynamic time warping. *Multimedia and Ubiquitous Engineering, 2007 MUE'07 International Conference on*; IEEE; 2007. p. 733-738.
12. Hu, J., Ray, B., Han, L. An interweaved HMM/DTW approach to robust time series clustering. *Proceedings of the 18th International Conference on Pattern Recognition*; Washington, DC. IEEE; 2006. p. 145-148.
13. Chen, G., Wei, Q., Zhang, H. Discovering similar time-series patterns with fuzzy clustering and dtw methods. *IFSA World Congress and 20th NAFIPS International Conference, 2001 Joint 9th*; IEEE; 2001. p. 2160-2164.
14. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*. 1978; 26(1):43–49.
15. Quiroga RQ, Nadasdy Z, Ben-Shaul Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural computation*. 2004; 16(8):1661–1687. [PubMed: 15228749]

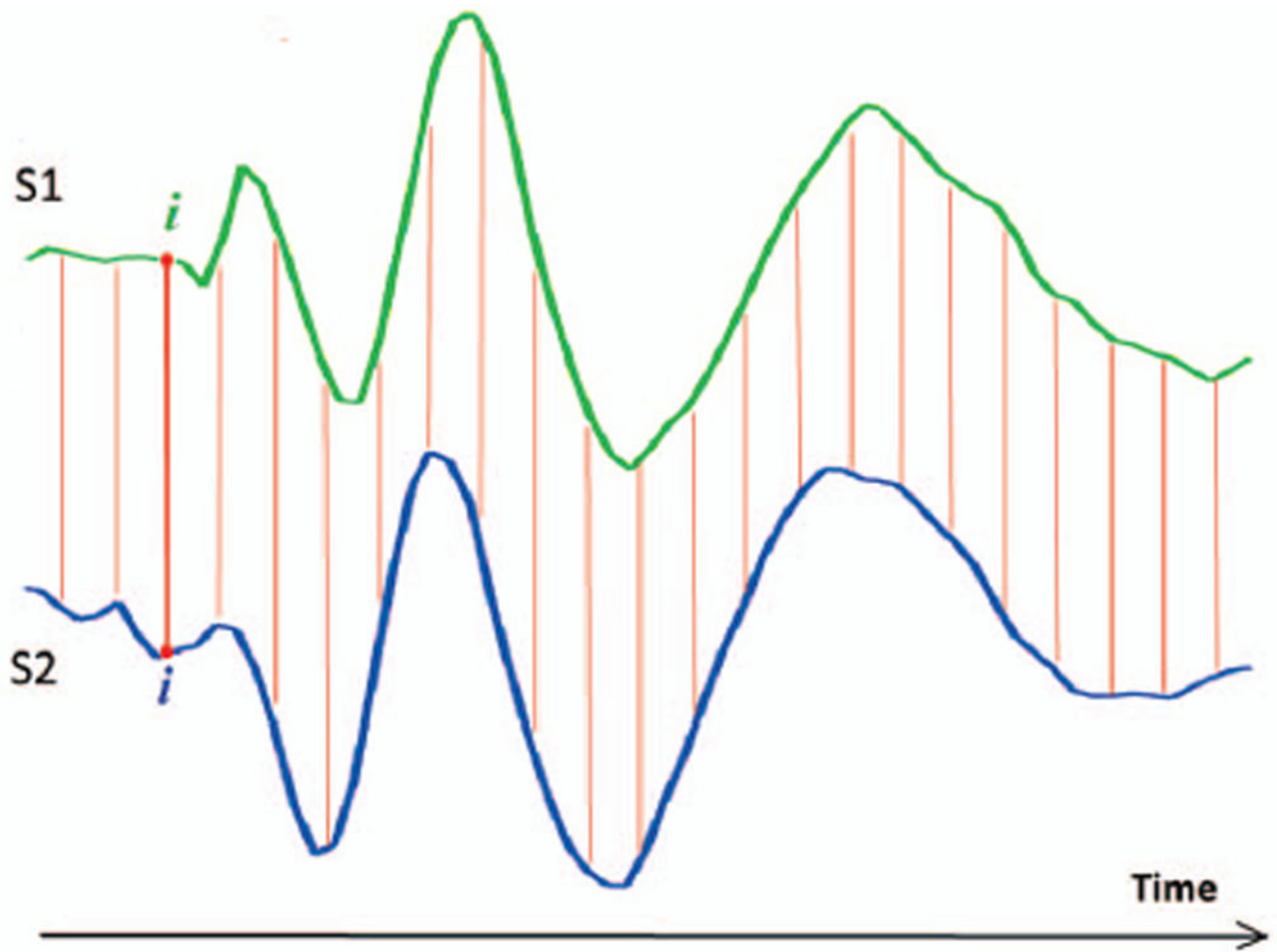
16. Mukhopadhyay S, Ray G. A new interpretation of nonlinear energy operator and its efficacy in spike detection. *Biomedical Engineering, IEEE Transactions on*. 1998; 45(2):180–187.
17. Jin, J., Dauwels, J., Cash, S., Westover, MB. Spikegui: Software for rapid interictal discharge annotation via template matching and online machine learning. *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE; IEEE; 2014*. p. 4435-4438.

Author Manuscript

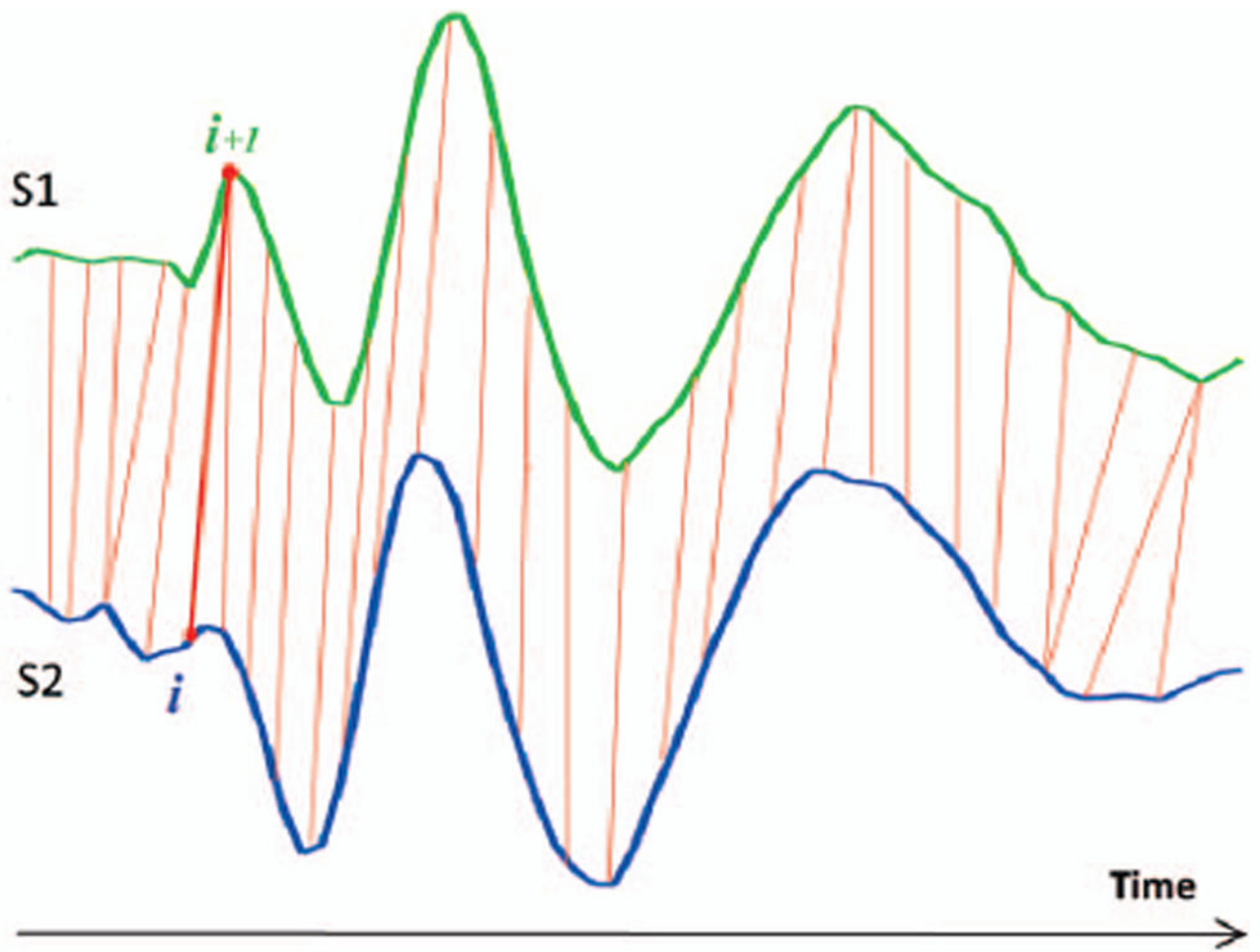
Author Manuscript

Author Manuscript

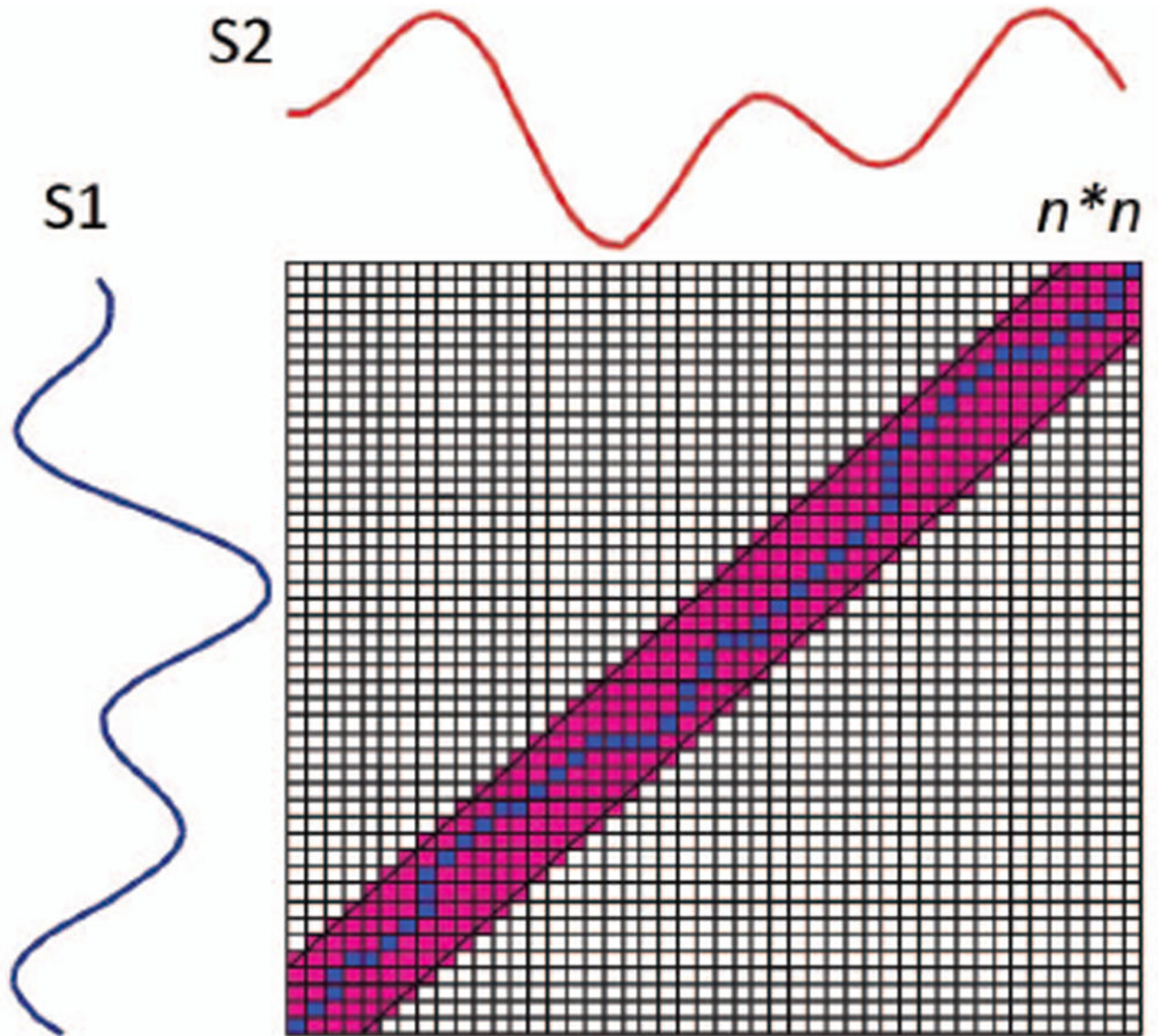
Author Manuscript



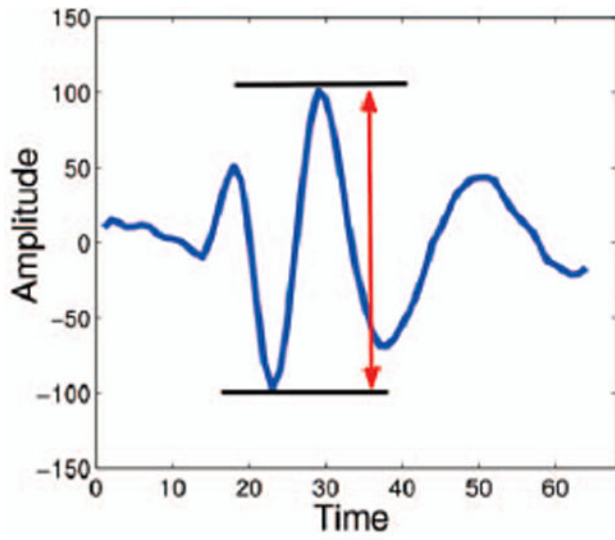
**Fig. 1.**  
The linear mapping between points of the two sequences for computing ED.



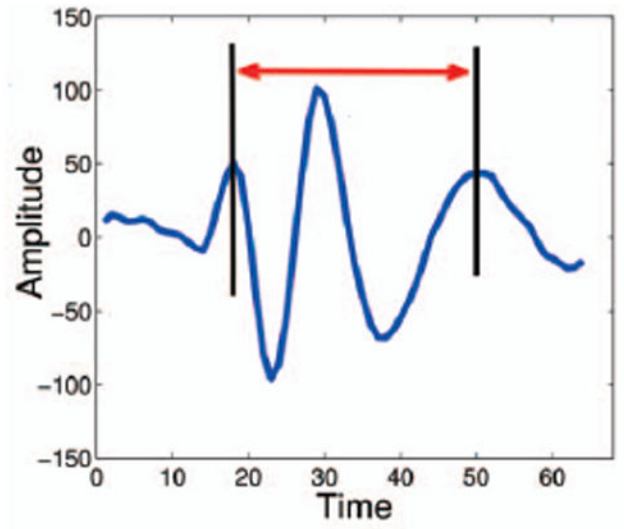
**Fig. 2.**  
The non-linear mapping in DTW between points of the two sequences.



**Fig. 3.** The  $n$ -by- $n$  matrix for computing the optimal path (blue) in DTW. The Sakoe-Chiba band is marked in purple [7].

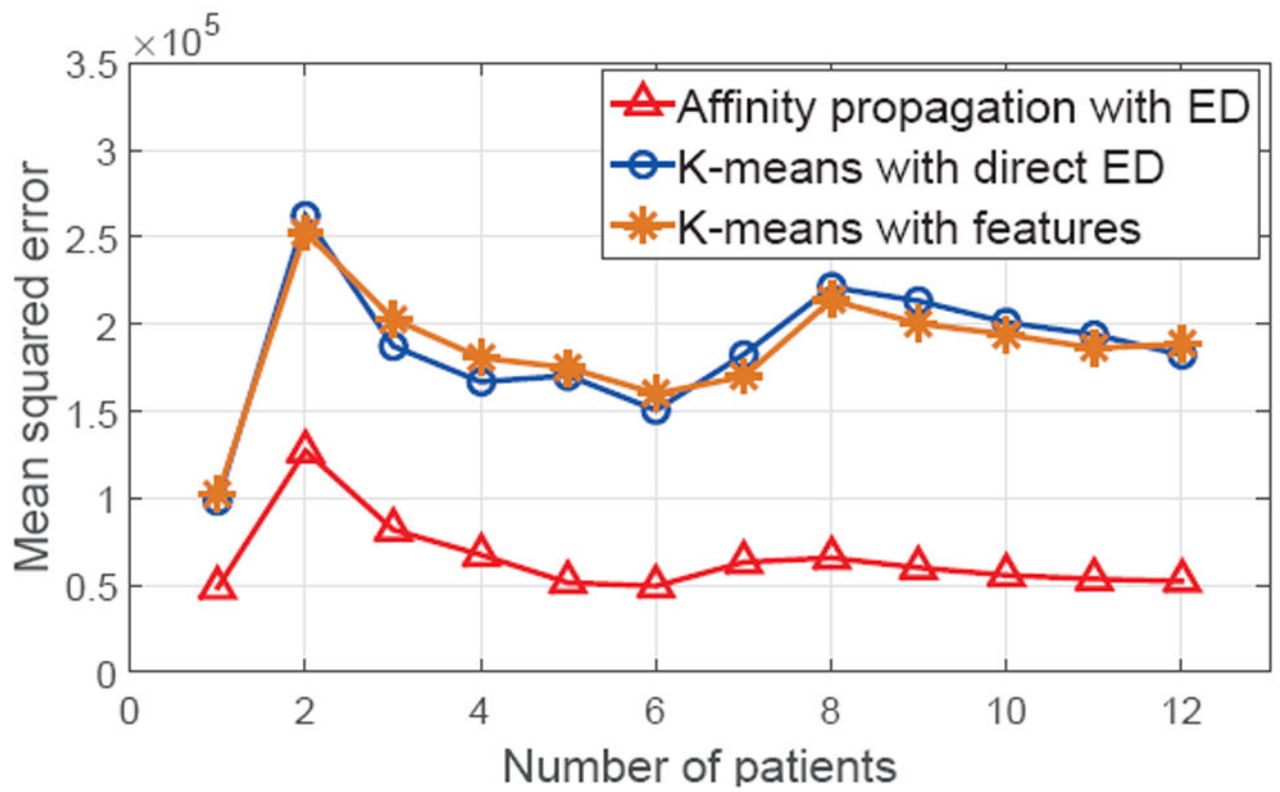


(a) Peak-to-peak value.

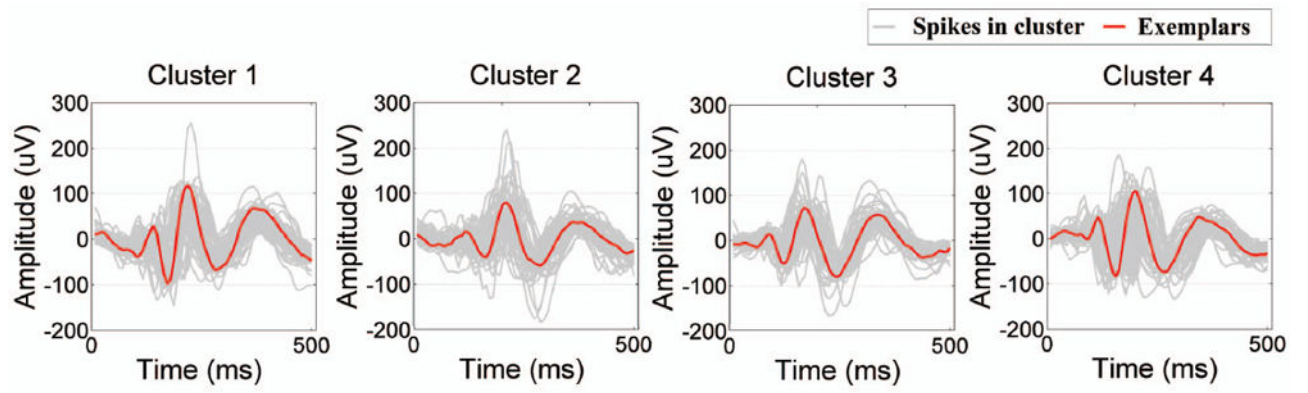


(b) Width of a spike.

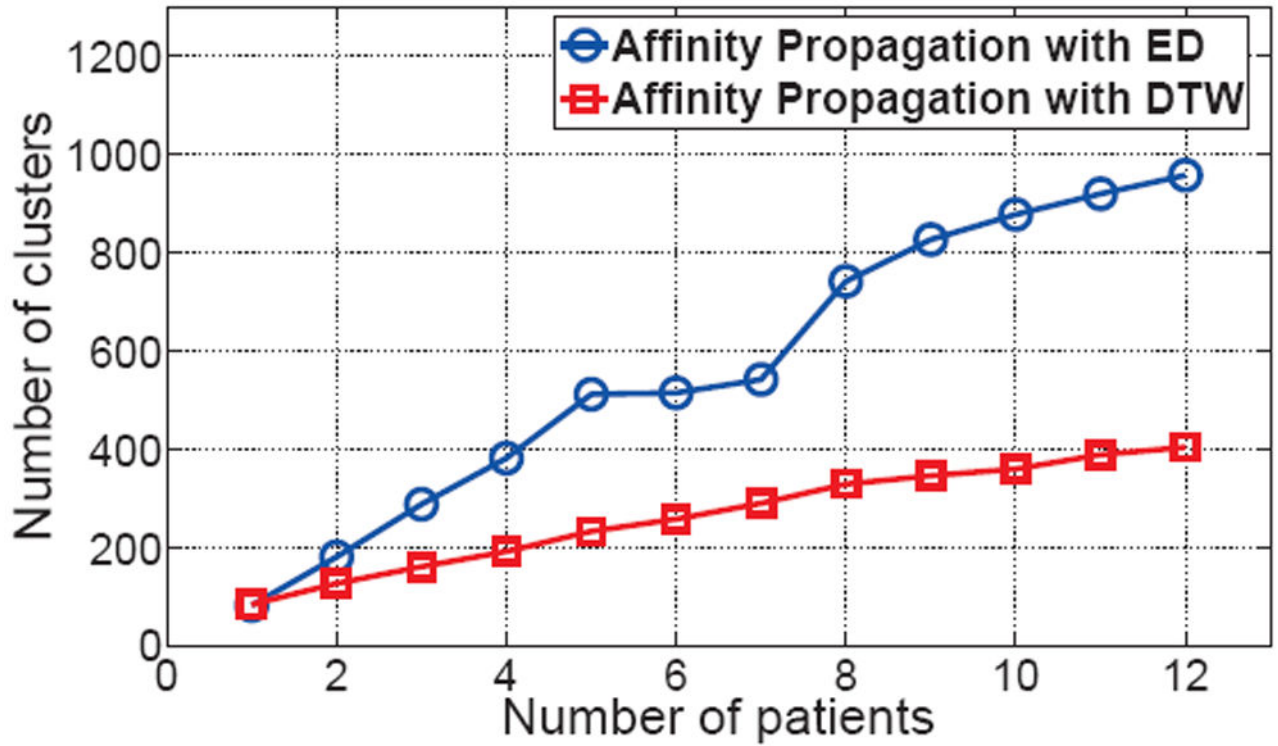
Fig. 4.  
The spike features.



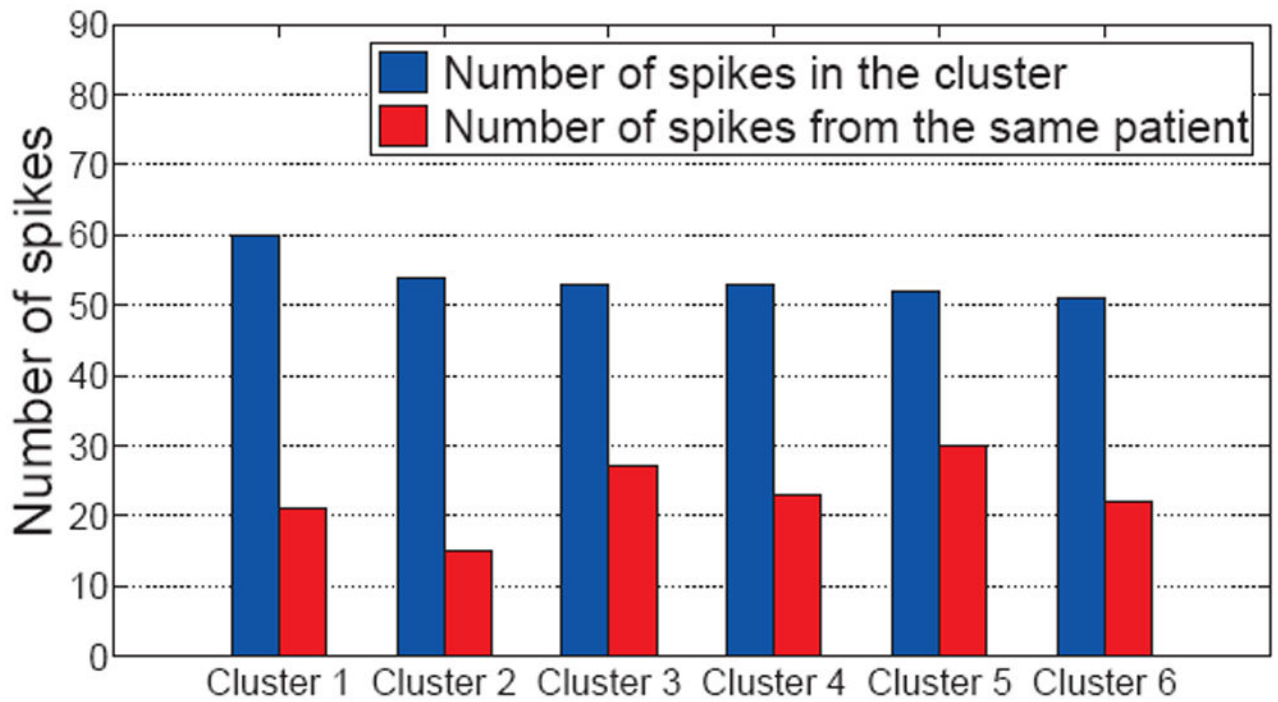
**Fig. 5.**  
Mean squared error plot for affinity propagation and K-means clustering.



**Fig. 6.** Spikes in the largest four clusters generated by affinity propagation with DTW. The exemplars are highlighted in red.



**Fig. 7.**  
The number of clusters produced by affinity propagation with DTW and ED for all the 12 patients.



**Fig. 8.** (Blue) Number of spikes in the six largest clusters generated by affinity propagation in combination with DTW. (Red) Number of spikes belonging to the same patient as the exemplar.