

QP-DTW: Upgrading Dynamic Time Warping to Handle Quasi Periodic Time Series Alignment

Imen Boulnemour* and Bachir Boucheham*

Abstract

Dynamic time warping (DTW) is the main algorithms for time series alignment. However, it is unsuitable for quasi-periodic time series. In the current situation, except the recently published the shape exchange algorithm (SEA) method and its derivatives, no other technique is able to handle alignment of this type of very complex time series. In this work, we propose a novel algorithm that combines the advantages of the SEA and the DTW methods. Our main contribution consists in the elevation of the DTW power of alignment from the lowest level (Class A, non-periodic time series) to the highest level (Class C, multiple-periods time series containing different number of periods each), according to the recent classification of time series alignment methods proposed by Boucheham (Int J Mach Learn Cybern, vol. 4, no. 5, pp. 537-550, 2013). The new method (quasi-periodic dynamic time warping [QP-DTW]) was compared to both SEA and DTW methods on electrocardiogram (ECG) time series, selected from the Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH) public database and from the PTB Diagnostic ECG Database. Results show that the proposed algorithm is more effective than DTW and SEA in terms of alignment accuracy on both qualitative and quantitative levels. Therefore, QP-DTW would potentially be more suitable for many applications related to time series (e.g., data mining, pattern recognition, search/retrieval, motif discovery, classification, etc.).

Keywords

Alignment, Comparison, Diagnosis, DTW, Motif Discovery, Pattern Recognition, SEA, Similarity Search, Time Series

1. Introduction

A time series is a continuation of couples $\langle (v_1, t_1), (v_2, t_2), \dots, (v_i, t_i), \dots \rangle$ where v_i is a value or a vector of values taken at a moment t_i . This notation can be abbreviated in $\langle v_1, v_2, \dots, v_i, \dots \rangle$ when the reference to the time does not need to be clarified [1].

There are an increasing amount of works related to time series, given their usefulness in different areas, especially in finance, where the calculation of changes in the exchange rate is very important. For example, the modeling of time series for financial forecasting was carried out by Barbulescu and Bautu [2]. In economics, Awad et al. [3] performed prediction of future values of time series related to economic activities. In industry, profiling and fault detections in systems have been performed by [4,5].

* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received March 10, 2016; first revision July 5, 2016; accepted August 9, 2016.

Corresponding Author: Imen boulnemour (boulnemourimen@hotmail.fr)

* Dept. of informatics, 20 août 1955 University of Skikda, BP 26, Skikda 21000, Algeria (boulnemourimen@hotmail.fr, boucheham_bachir@yahoo.fr)

Classification and identification of electrocardiograms (ECGs) of patients with heart diseases is also a very active area related to time series [6-8]. The detection of abnormalities in ECG by aligning time series is much appreciated. It has particularly been performed by the Shape Exchange Algorithm (SEA) method of Boucheham [9,10]. Time series motif discovery from unbounded streams is also a growing field [11].

Studies in the field of time series data mining evolve around two key issues: the choice of distance/similarity measures to match time series (effectiveness), and the mechanisms to accelerate the efficiency of the match. Concerning the first question, the matching of time series is divided into two categories: comparison and alignment. The difference between the two is that the comparison is based on a measure of distance/similarity that renders a numerical value reflecting the degree of resemblance between the two time series. However, alignment is used to specify exactly where the time series differ and where they are similar. In other words, alignment is a kind of explanation of time series, one by the other, which, obviously, is a more difficult problem to resolve.

This study is dedicated to alignment of quasi-periodic time series and especially to heartbeats time series provided from the ECG records. Quasi-periodic time series are concatenations of quasi-similar forms called periods [9].

Fig. 1 presents the characteristics of an ECG composed of three quasi-regular periods reflecting three heart cycles. Each cycle (period) is itself composed of three consecutive basic patterns: P wave, QRS complex, and T wave.

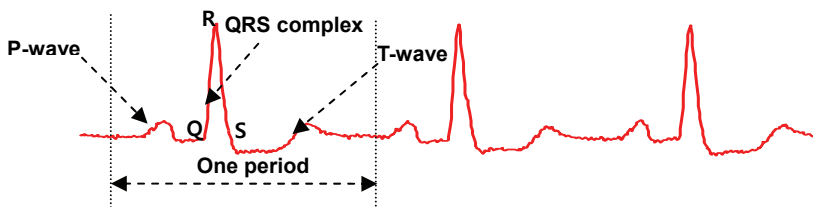


Fig. 1. A typical ECG segment with three periods.

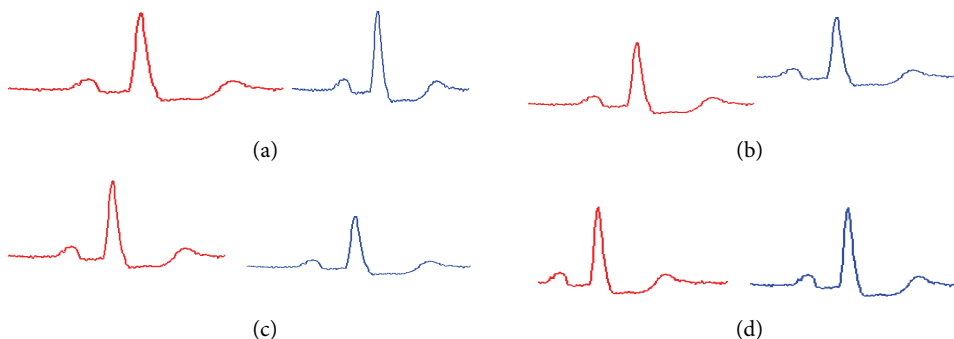


Fig. 2. Difficulties encountered in time series alignment. (a) Different time scaling, (b) different amplitude scaling, (c) shift on the amplitude axis, and (d) shift on the time axis.

Alignment of time series and search of a query of time series pose difficult problems. These problems are: difference in time axis scales (Fig. 2(a)), difference in amplitude axis scales (Fig. 2(b)), shift on the

amplitude axis (Fig. 2(c)) between the series, shift on the time axis (Fig. 2(d)) and different type of noise. Particularly, ECG traces are characterised by morphological changes due to the problems of variability of ECG measurement.

Time series are also characterized by very big sizes of collected data, big dimensionality and the need of permanent update [12]; what makes search, query, classification, which are all based on time series comparison and alignment, a very difficult task.

Dynamic time warping (DTW) [13] is generally recognized by several researchers [14–16] to be the most effective alignment technique for time series, because it handles the problems of distortion on the time and amplitude axis with great elegance.

However, the DTW has difficulties to deal with these problems when it comes to align quasi-periodic time series with different number of periods (cycles) each. In particular, the method completely fails in the alignment of quasi-periodic time series which are significantly phase shifted as illustrated in Fig. 3(a).

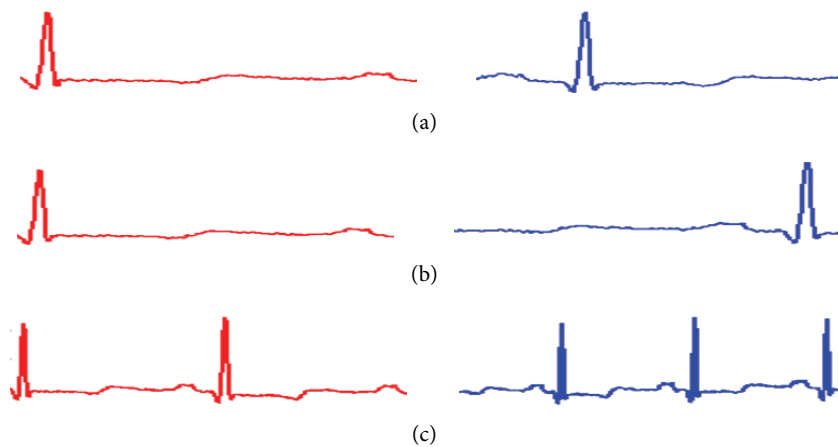


Fig. 3. Different situations of quasi-periodic time series: DTW cannot deal with the class B and C, whereas SEA and QPDTW match all cases. (a) One period no phase shift, (b) one period with phase shift, and (c) periodic-many-periods with phase shift.

SEA [9] is a method that aligns quasi-periodic time series. However, when it comes to align time series which are seriously contaminated by noise, the SEA method shows some weaknesses as illustrated in Fig. 4.

To remedy to the problems of DTW and SEA, we propose in this work a novel algorithm that combines the advantages of these methods. Our method deals effectively with the quasi-periodic time series even when they are significantly phase shifted and contaminated by noise. The new method is then called QP-DTW (quasi-periodic dynamic time warping).

The tests were performed on ECG traces, selected from the public database “Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH)” [17] and from the PTB Diagnostic ECG Database [18]. The results show that the proposed algorithm is more efficient than DTW and SEA in terms of alignment accuracy, quantitatively and qualitatively.

The remaining part of this paper is organized as follows. Section 2 presents some related works

related to time series comparison/alignment. Section 3 presents the DTW method. Section 4 presents the SEA method. In Section 5, we present our proposed method QP-DTW. Section 6 is devoted to application of our method and to experimental results. Finally, this work is terminated with a general conclusion and future plans.

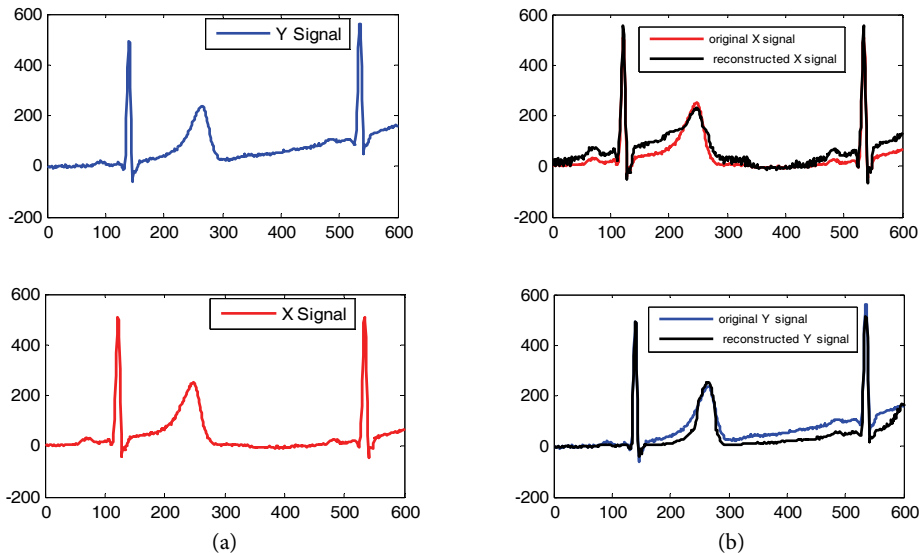


Fig. 4. Illustration of the problems of the SEA method. (a) Original signals and (b) alignment of the original signals.

2. Related Works and Motivations

2.1 Related Works

According to Boucheham [14] matching methods for time series can be divided into two major classes. The first class is based on the nature of the algorithm; either they are comparison methods or alignment methods. Comparison methods render only one numerical value in the range $[0, 1]$ that reflects the similarity of the sequences. Obviously these methods are incapable to align the time series. The second class is based on the type of the used data; either they are periodic methods or non-periodic methods.

2.1.1 Comparison or alignment methods

Comparison methods

The most famous comparison method is the Euclidean distance. It is fast and simple but works only for time series of equal lengths. It has been also demonstrated by several researchers that this distance is fragile [19,20] because it is unable to handle the time offsets of a time series with respect to one other. Fig. 5 shows that even the two time series are similar, the Euclidian distance notice that they are different. This is due to the point to point alignment.

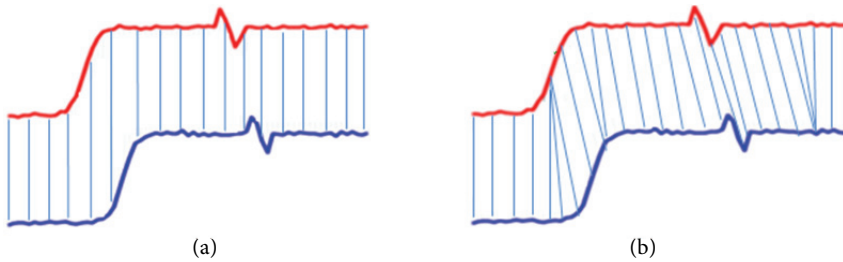


Fig. 5. (a) Inefficiency of alignment by the Euclidean distance. (b) Effectiveness of alignment by the DTW.

Another very effective method of comparison is the discrete Fourier transform (DFT) [1]. The authors represent temporal sequences in the frequency domain with the first most significant k coefficients of Fourier and they use the Euclidian distance to compare them starting from the fact that the Euclidean distance between two time series in the time domain is the same in the frequency domain. The advantage of this method is that the amplitudes of the frequencies don't change despite the time shifts. Another method of comparison is the histogram. Chen and Ozsu [21] build a histogram for each segment of both time series and compare the corresponding histograms. After that they developed multi-scale histograms. They proved that it can efficiently compare time series data even when they are of different length and contain noise or time shifting and scaling. The edit distance between two alphanumeric sequences is another comparison method. It is a way of quantifying how dissimilar are two strings (e.g., words) by counting the minimum number of operations required to transform one string into the other [22].

Alignment methods

Bozkaya et al. [23] proposed a modified version of the edit distance. They referred to the method by longest common sub-sequence (LCSS). The LCSS is an alignment method based on the number of pairs of ordered temporal sub-sequences which are similar but not overlapped. If there is enough the sequences will be regarded as similar. The method then compares each point in a series with 0 or 1 point on the other series. The LCSS is elastic and allows stretching in order to match the identical elements between them. It is mainly used for treating the noisy data because it ignores the extreme values by counting only the commons sub sequences. It is also effective for the problems of different time scaling.

The DTW is a very much appreciated alignment method. Fig. 5(b) illustrates the effectiveness of the DTW comparing to the Euclidian distance (Fig. 5(a)). Instead of comparing each point series with another point of another series, which occurs at the same time t , the measure compares each point series with one or many points of another series, these points can be shifted.

There are various applications of the DTW, we cite, ECG heart-beat clustering and analysis [24,25], ECG classification [26,27].

The SEA is a simple and parameter-free method. It was proposed by Boucheham [9] to deal with the problem of quasi-periodic time series alignment. The method is based on the sorting of time series on their amplitude axes. This sort gives the time series a kind of signature, in the sense that the resulting traces represents a stable and global characteristics of the used time series.

2.1.2 Periodic or non-periodic methods

Based on the nature of the used data, Boucheham [14] classifies alignment methods of time series in ascending order of complexity of the data.

Class A: Non-periodic or one period no phase shift

Methods in this class treat time series that are non-periodic or periodic with only one period with the condition that none of the two periodic time series is significantly phase shifted with respect to the other (e.g., Fig. 3(a)). Methods in this category are DFT, LCSS, Euclidian distance, histograms, and DTW.

Class B: One period with phase shift

Time series in this class are quasi-periodic and composed of one period. However, each time series is phase-shifted with respect to the others (e.g., Fig. 3(b)). Time series can also be composed of the same number of periods. The phase independent DTW [15] is in this class.

Class C: Multiple periods with phase shift

Time series in this class are quasi-periodic and composed of many periods each, but not necessarily the same number of periods. Indeed, time series that contain many periods but the same number of periods brings the data to class A if there is no (significant) phase shift and to class B if there is (significant) phase shift [14].

Thus, the types of data specific to this class are time series that are phase-shifted, in addition of being composed of a different number of periods each (e.g., Fig. 3(c)). Methods in this class are SEA, FANSEA, ASEAL [9,14,28] and QP-DTW.

2.2 Motivation

The DTW method suffers from several problems which are subjects of many researches. In this context, we list:

- The DTW cannot deal with the phase shift problems in quasi-periodic time series. In Fig. 6, both ECG plots are similar because they result from the same person but they do not begin at the same time. In this case, the DTW done a bad alignment; only the QRS complex has been correctly aligned, the other components of the two ECG time series (P and T waves) have just been matched with their sequentially equivalent parts. This is a case of phase shift problem.
- For the matching of objects in the images by their contours, the alignment of the shapes can be solved with the DTW, by the transformation of the 2D forms into 1D signal. However if a rotation is applied to one of the objects, the task becomes impossible to achieve by the DTW. To solve this problem Keogh et al. [15] proposed the Phase Independent DTW method. Their method is very expensive because it implements techniques of indexation to allow the verification of all the possible rotations of shape. This method deals also with the phase shifted quasi-periodic time series; but these time series cannot have a different number of periods. They can have either a single period, or the same number of periods each.
- Another problem that DTW cannot solve is the alignment of similar time series that have some

peaks of amplitudes with heights slightly difference as shown in Fig. 4, which implies difficulties in analysis and diagnosis. To solve this problem, Keogh and Pazzani [29] proposed a modification of the DTW that does not consider the values of the points of the amplitude axis, but rather the characteristic of the highest level of the shape by considering the first derivative of the sequence.

- The DTW sometimes gives pathological results because it tries to explain the variations of amplitudes by deforming the time axis. This problem is called the problem of “Singularity”. Most of the methods try to solve this problem by limiting the possible deformations [13,19,30]. However these methods cannot find the optimal warping path if the band of delimitation is too small or if it is too big [29].

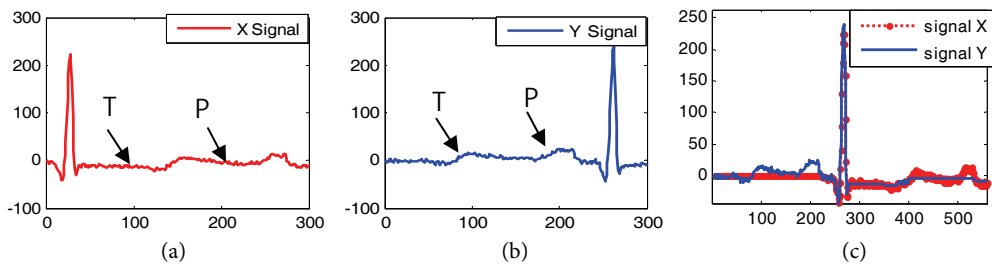


Fig. 6. Illustration of the phase shift problem of DTW.

- The SEA method has also difficulties to align similar time series that have some peaks of amplitude with heights slightly difference as shown in Fig. 4. In this figure the higher signal (blue) is significantly contaminated with low frequency noise (baseline wondering due to respiration). Improved versions of the SEA method have been proposed in [14,28,31].

2.3 Contribution

Our main contribution consists in the elevation of the DTW power of alignment from the lowest level (class A, non-periodic time series) to the highest level (class C, multiple periods with phase shift), according to the recent classification of time series alignment methods proposed by Boucheham [14]. Indeed, the DTW is inappropriate to align quasi periodic time series like ECG signals, capnogram signals (a plot reflecting the quantity of expired CO₂, extensively used in asthmatic patients surveillance, urgency medicine and also in surgery medicine) and electroencephalogram (EEG) signals (a plot measuring the electrical activity of the brain) unlike our method which is capable of aligning very complex quasi-periodic time series (according to the classification of Boucheham [14]). On the other hand and compared to the existing SEA method of Boucheham [9] the proposed QP-DTW performs significantly much better.

3. Dynamic Time Warping

The famous DTW is an alignment method. It was applied at first for the recognition of vocal sequences by [13,30] then it has been introduced in the field of data mining by Berndt and Clifford [19]

and knew a big success in this context.

The DTW method consists in establishing a non linear alignment of sequences to allow the mapping of sequences which have different lengths or suffer from the time-shift problem.

First, to align two sequences $X = (x_i), i = 1: n$ and $Y = (y_j), j = 1: m$, we construct an accumulated distance matrix $m \times n$, each cell represents the alignment between two points x_i and y_j calculated by one of the distances of Minkowski. The Euclidian distance (Eq. (1)) is the most often used metric to calculate these cells.

$$d(x_i, y_j) = (x_i - y_j)^2 \tag{1}$$

Technically, the values of the sequences to compare $X = (x_i), i = 1: n$ and $Y = (y_j), j = 1: m$ are replicated until obtaining the best matching between the sequences [32]. The obtained sequences have the same number of elements k with $\max(n, m) \leq k \leq n + m + 1$.

In the example shown in Fig. 7, the accumulated distance matrix M is presented. The sequences to compare X and Y , are presented to the underside of the matrix. After the calculation of the cumulative matrix, we calculate the minimum warping path W and we obtain the sequences X' and Y' . The values of cells are calculated as follows:

1. First cell: $M[1,1] = |x_1 - y_1|$
2. First line: $M[1, j] = |x_j - y_1| + M[1, j - 1]$
3. First column: $M[i, 1] = |x_1 - y_i| + M[i - 1, 1]$
4. All the other elements with the condition ($i, j > 1$):

$$M[i, j] = |x_j - y_i| + \min(M([i - 1, j - 1]), M([i - 1, j]), M([i, j - 1])$$

The mapping chosen for the DTW is the one that minimizes the Euclidian distance: $\sum_{i=1}^k |a_i - b_i|$.

		→ i			
↓ j	Y/X	5	8	9	7
	7	2	3	5	5
	5	2	5	7	7
	8	5	2	3	4
	7	7	3	4	3
	8	10	3	4	4

Fig. 7. The accumulated distance matrix for the sequences $X = (5, 8, 9, 7)$ and $Y = (7, 5, 8, 7, 8)$.

In this example the $DTW(X, Y) = 4$. It is represented by the value of the last cell. The cells in gray colors represent the optimal warping path “ W ”. They indicate the replicated values and the alignment of the two sequences.

The values that are close in the two sequences are replicated, such as the value 5 of the sequence X which is replicated twice in the sequence X' and put in correspondence with the values 5 and 7 of the sequence Y' .

There are several constraints on the warping path we cite:

- *Boundary Condition.* $w_1 = (1, 1)$ and $w_K = (m, n)$, to force the end points in each series to match.
- *Continuity.* Given $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a-a' \leq 1$ and $b-b' \leq 1$, to oblige the warping path to go to the adjacent cells.
- *Monotony.* Given $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a-a' \geq 0$ and $b-b' \geq 0$ to force the points of W to be spaced monotonically in time.

There are several warping path which satisfy these conditions, but we choose the one that minimizes the warping cost:

$$DTW(X, Y) = \min \left\{ \sqrt{\frac{\sum_{k=1}^K w_k}{K}} \right\} \quad (2)$$

The k^{th} element of W is defined as $w_k = (i, j)_k$ such that:

$$W = w_1, w_2, \dots, w_k, \dots, w_K \quad (3)$$

Eq. (4) illustrates the dynamic programming method used for the calculating of the optimal path which represents the DTW distance. Let D , this cumulative distance until w_k cell.

$$D(x_i, y_j) = d(i, j) + \min \{D(i-1, j); D(i, j-1); D(i-1, j-1)\}. \quad (4)$$

4. Shape Exchange Algorithm

The principle of the method is as follows:

Let $X = (x_i), i = 1: n$ and $Y = (y_j), j = 1: m$ be the two time series to match.

1. Decompose X and Y on temporal indexes and amplitude indexes.
2. Order the X and Y on their amplitudes indexes.
3. Exchange amplitude indexes between X and Y without changing their temporal indexes which are in disorder.
4. Reorder X and Y on their temporal indexes and reconstruct X and Y to X_{REC} and Y_{REC} .
5. Compare (X, X_{REC}) and (Y, Y_{REC}) .

5. The Proposed Method: Quasi-Periodic Dynamic Time Warping (QP-DTW)

Most of the methods that try to improve the DTW focus on the acceleration of its execution time [33,34]. However, the alignment quality of the DTW poses many problems [9,29], especially when the DTW comes to align quasi-periodic time series containing each a different number of quasi-similar phases. To overcome this problem Boucheham [9] propose the SEA method. However this method uses the linear mapping to equalize the series of different length. The principle of the linear

mapping is to force the longest time series to shrink by accelerating the step in the temporal axis until it can match the shorter series and vice versa, the shortest series will stretch to match the longest series by browsing slowly the temporal index. The technique of linear mapping is not appropriate for the SEA method because the longest series loses some important values when shrinking. In our method, we have done a hybrid of the SEA and DTW methods, to take advantage of both methods.

5.1 Time and Space Complexities

Time complexity: Let n and m be the lengths of the time series to align, X and Y , respectively. Given that the time complexity of DTW is of $O(nm)$ and the time complexity of SEA is of $O(\text{Max}(n^2, m^2))$, our hybrid method will have a time complexity of $O(nm + \text{Max}(n^2, m^2))$, which is of quadratic order.

Space complexity: Given that the space complexity of DTW is of $O(nm)$ and the space complexity of SEA is of $O(n+m)$, our method QP-DTW will have a space complexity of $O(nm+n+m)$, which is of quadratic order.

Step 1: Finding the optimal warping path by the DTW

The algorithm DTW is executed to equalize the length of the two time series and make the first mapping between them by the replication of certain values which are close in both time series, as illustrated in Fig. 7.

Step 2: Sorting on the amplitude

The sorting of the stretched time series, on the coordinates of their amplitude indexes is established to give them a stable signature. The result of this step is a matrix for each time series containing the sorted amplitudes in ascendant or descendent order with their equivalents temporal coordinates (not sorted). In this way, if the time series are similar but phase shifted, they will always have almost the same trace which represents their signatures. An example of signatures is illustrated in Fig. 8.

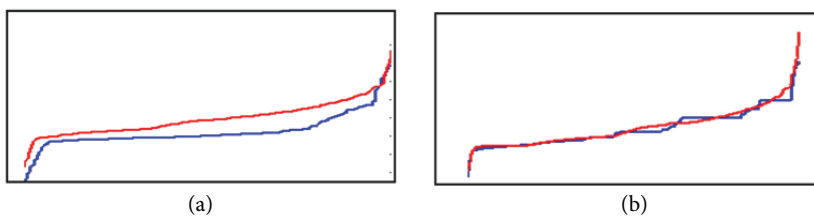


Fig. 8. Signatures of the time series X and Y established by the two methods SEA (a) and QP-DTW (b).

Step 3: Signatures exchange

This step is divided into two parts. The first one consists in making the exchange of signatures between both time series. The first time series will receive the coordinates of amplitudes (signature) of the second time series and vice versa for the second series. The second part consists in restoring the normal size of the time series after the exchange of signature which requires that the time series are of equal length. Thus, the temporal indexes replied in every new series must be deleted with their equivalent amplitudes. This part does not affect the results of alignment; it just allows having a clearer visual inspection of the traces.

Step 4: *Sorting on the time indexes and alignment*

Signatures represent well the characteristics of the time series but they do not consider the shifts and the differences of timescales. The sorting of the time series on the respective temporal indexes allows reconstructing them.

Every time series will then be aligned with its reconstructed time series by using the correlation and PRD (percent root difference) factors as objective criteria and the visual inspection as subjective criteria.

The proposed algorithm is described in Fig. 9.

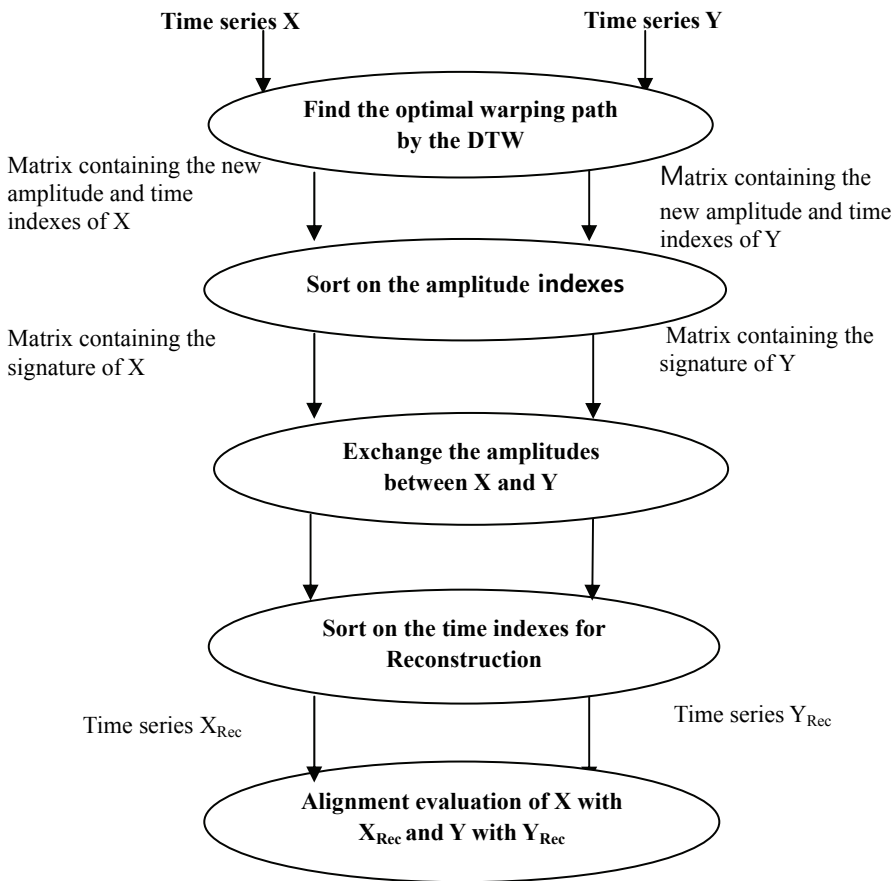


Fig. 9. Illustrative diagram of the proposed method QP-DTW.

6. Experimental Tests and Discussion

In this section, we present an illustrative example and we proceed to the numerical and the visual comparison between QP-DTW, DTW and SEA. The numerical comparison is performed by the PRD (Eq. (5)) and the correlation factor (Eq. (6)). The correlation must be between 0 and 1 and the PRD must be between 0% and 100%.

We indicate that all applications have been performed on a machine with the following specifications: Intel Pentium CPU U5400 (1.20 GHz) with 3 GB main memory, running under Windows 7. We also mention that all methods have been implemented in the MATLAB 12 programming environment.

$$PRD(X, Y) = \sqrt{\frac{\sum_{i=1}^{i=N} |x_i - y_i|^2}{\text{Max}(\sum_{i=1}^{i=N} |x_i - \bar{x}|^2, \sum_{i=1}^{i=N} |y_i - \bar{y}|^2)}}} \times 100\% \quad (5)$$

$$\text{Corr}(X, Y) = \frac{\text{cov}(X, Y)^2}{\text{var}(X) \cdot \text{var}(Y)} \quad (6)$$

where

$$\text{cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) \quad (7)$$

and

$$\text{var}(X) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^2 \quad (8)$$

We start the comparison by introducing an example of time series in the class C. In this example, we refer to the time series rec.215 of the MIT-BIH ECG database. Fig. 8(a) and (b) present the signatures of the time series X and Y for the SEA and QP-DTW methods. We see that the signatures established by QP-DTW are attached which confirm that the series come from the same recording. Unlike the signatures of SEA which are detached. Fig. 10(a) shows the original traces X and Y. Fig. 10(b) and (c) illustrate the alignment done respectively by SEA and QP-DTW methods for the time series X and its reconstructed time series (left of Fig. 10(b) and (c)) and for the time series Y and its reconstructed time series (right of Fig. 10(b) and (c)). Fig. 10(d) presents the alignment done by the DTW method for the time series X and Y. Note that the two traces are from the same time series but that they are of different lengths. They also have some pikes of amplitude of slightly different lengths and they are not taken at the same time (offset of 35 seconds for Y). The segment (X) was purposely taking at the beginning of the recording and the segment (Y) at the end of the recording to illustrate the phase shift problem in quasi periodic time series.

Formulas are applied to the two traces (X and Y) with their respective reconstructed traces (X_{rec} and Y_{rec}). As well we will report in Table 1, the $PRD(X, X_{\text{rec}})$, $\text{Corr}(X, X_{\text{rec}})$, $PRD(Y, Y_{\text{rec}})$ and $\text{Corr}(Y, Y_{\text{rec}})$ for SEA and QP-DTW methods and the $PRD(X, Y)$, $\text{Corr}(X, Y)$ for the DTW method.

Table 1. QP-DTW vs. DTW and SEA for time series rec.215 and the reconstructed time series X_{rec} (resp. Y_{rec})

	PRD(X,Y)	PRD(X, X_{rec})	PRD(Y, Y_{rec})	Corr(X,Y)	Corr(X, X_{rec})	Corr(Y, Y_{rec})
SEA	-	32.0207	50.3374	-	0.9566	0.9561
QP-DTW	-	16.8858	21.6276	-	0.9906	0.9817
DTW	68.0843	-	-	0.7582	-	-

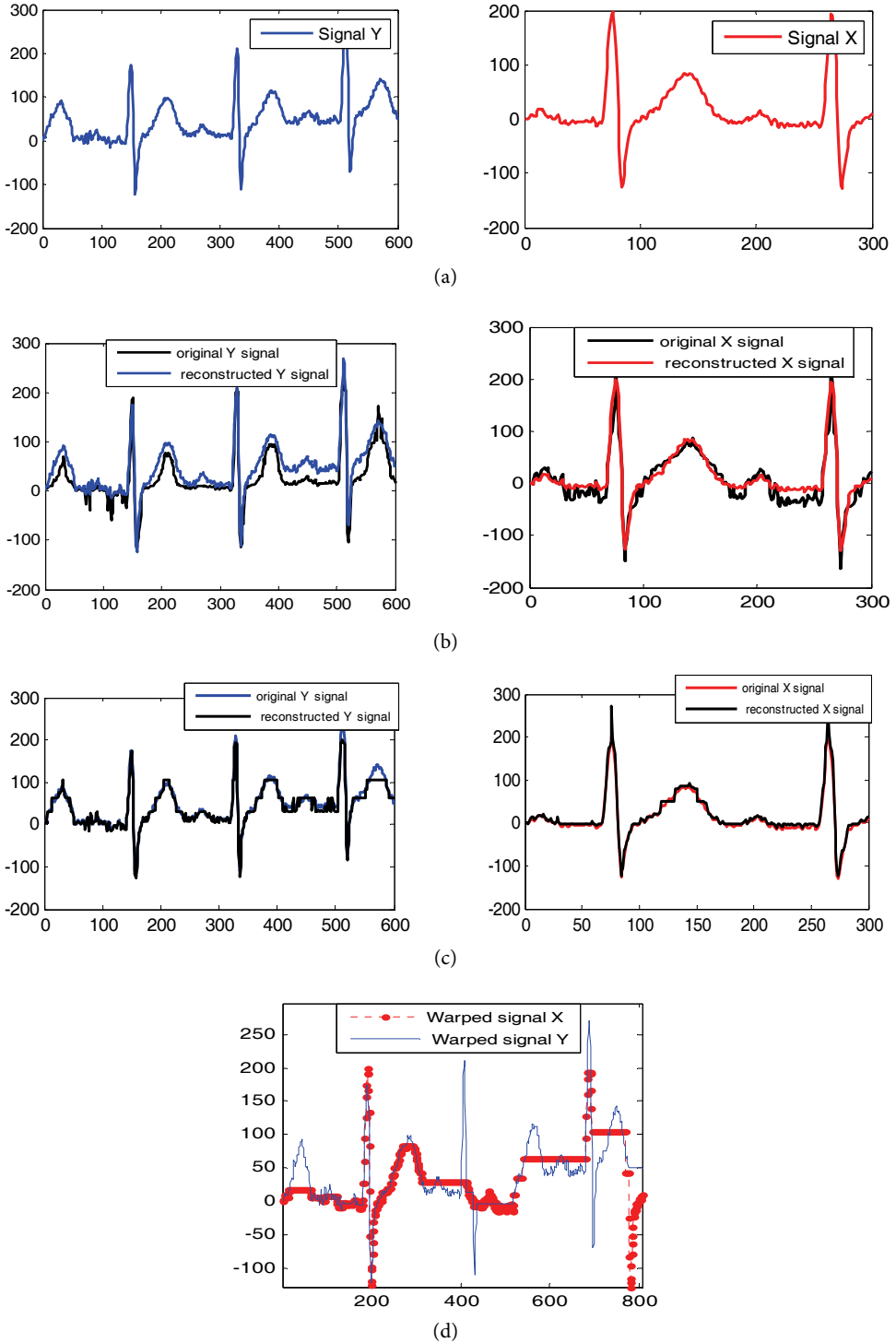


Fig. 10. Comparison of the SEA, QP-DTW, and DTW methods for the case of periodic-many-periods time series with phase shift (rec.215). (a) Original signals. (b) Original signals vs. reconstructed signals by SEA. (c) Original signals vs. reconstructed signals by QP-DTW. (d) Original signal vs. warped signal by DTW.

The visual inspection of Fig. 10 and the numerical results reported in Table 1 show that despite all the problems previously cited, QP-DTW arrives to align these particularly very complex time series. We see also that SEA method doesn't do a good alignment because time series Y and its reconstruction are almost detached, only the QRSs have been correctly aligned. The DTW done a very bad alignment because, the second QRS of the time series Y haven't been aligned with the QRSs of the time series X.

6.1. Comparison of Similar Traces from the MIT-BIH ECG Database

Fig. 11 shows the case of two plots of the same record with 1,000 samples of the time series rec.113 for X and 1,200 samples of the same time series with a shift of 30 seconds for Y. Fig. 11(a) presents the original time series. They have some peaks of amplitudes with heights slightly difference due to low frequency noise. Fig. 11(b) shows the alignment performed by SEA. We see that the original and the reconstructed time series are detached because of the problem of low frequency noise. Fig. 11(c) shows that despite all the problems mentioned above, QP-DTW done an almost perfect reconstruction, to the point that it is difficult to distinguish between the original and the reconstructed time series. Fig. 11(d) shows the misalignment of X and Y done by DTW because of the problem of quasi periodicity and low frequency noise of the time series. The numerical results presented in Table 2 confirm this state with a very low error rate (PRD) and a very high correlation in favor of QP-DTW. In contrary they indicate a high error rate (PRD) for DTW and SEA, an average correlation for SEA and a low correlation for DTW.

Table 2. Comparative table of the methods DTW, SEA, and QP-DTW concerning similar time series of Figs. 11–13 from the MIT-BIH ECG database

ECG segments	DTW		SEA		QP-DTW	
	PRD%	Corr	Mean PRD%	Mean Corr	Mean PRD%	Mean Corr
(215,215)	51.97	0.8554	35.32	0.9680	23.82	0.9887
(113,113)	40.48	0.9225	46.85	0.9689	14.55	0.9914
(104,104)	34.48	0.9389	51.83	0.8734	20.51	0.9849

Fig. 12 shows the case of two plots of the same record with 1,000 samples of the time series rec.215 for X and 2000 samples of the same time series with a shift of 35 seconds for Y. Fig. 12(a) presents the original time series. The X signal has been taken at the beginning of the recording and the Y signal has been taken at the end of the recording to illustrate the phase shift problem. Fig. 12(b) shows the alignment performed by SEA. We see that SEA method fails in face of the problem of low frequency noise which is presented in the Y signal. Fig. 12(c) shows the alignment performed by QP-DTW. Fig. 12(d) presents the bad alignment done by DTW. We see that our method resolve efficiently the problems of phase shift and noise. The factors of similarity (Corr) and of dissimilarity (PRD) clearly indicate the superiority of our method in comparison with DTW and SEA. The correlation of QP-DTW is higher than those of DTW and SEA and its PRD is the lowest. It represents the half of the value of PRD of DTW.

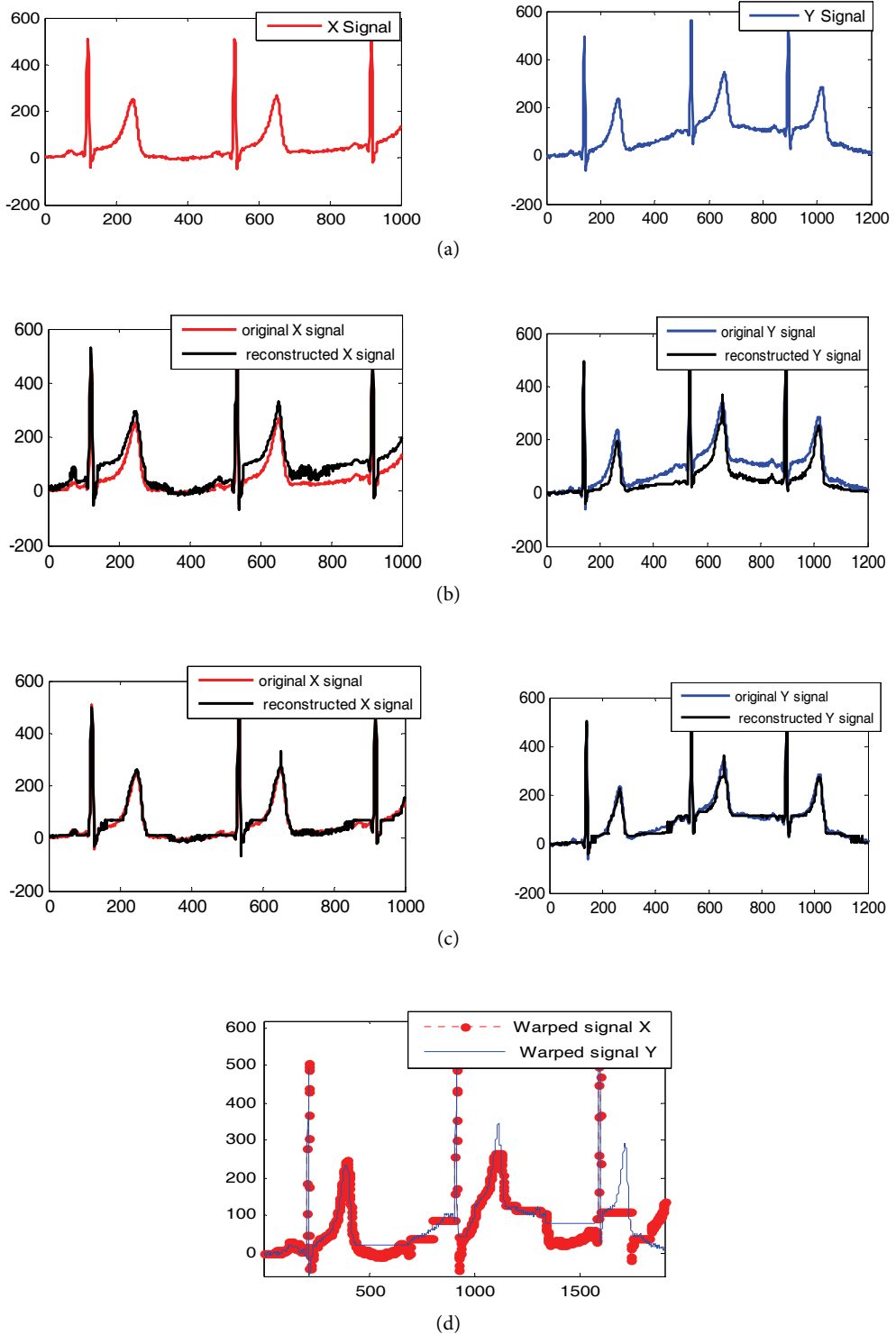


Fig. 11. Illustration of alignment by SEA, QP-DTW, and DTW of two similar traces (113,113) with shift of 30 seconds for Y. (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

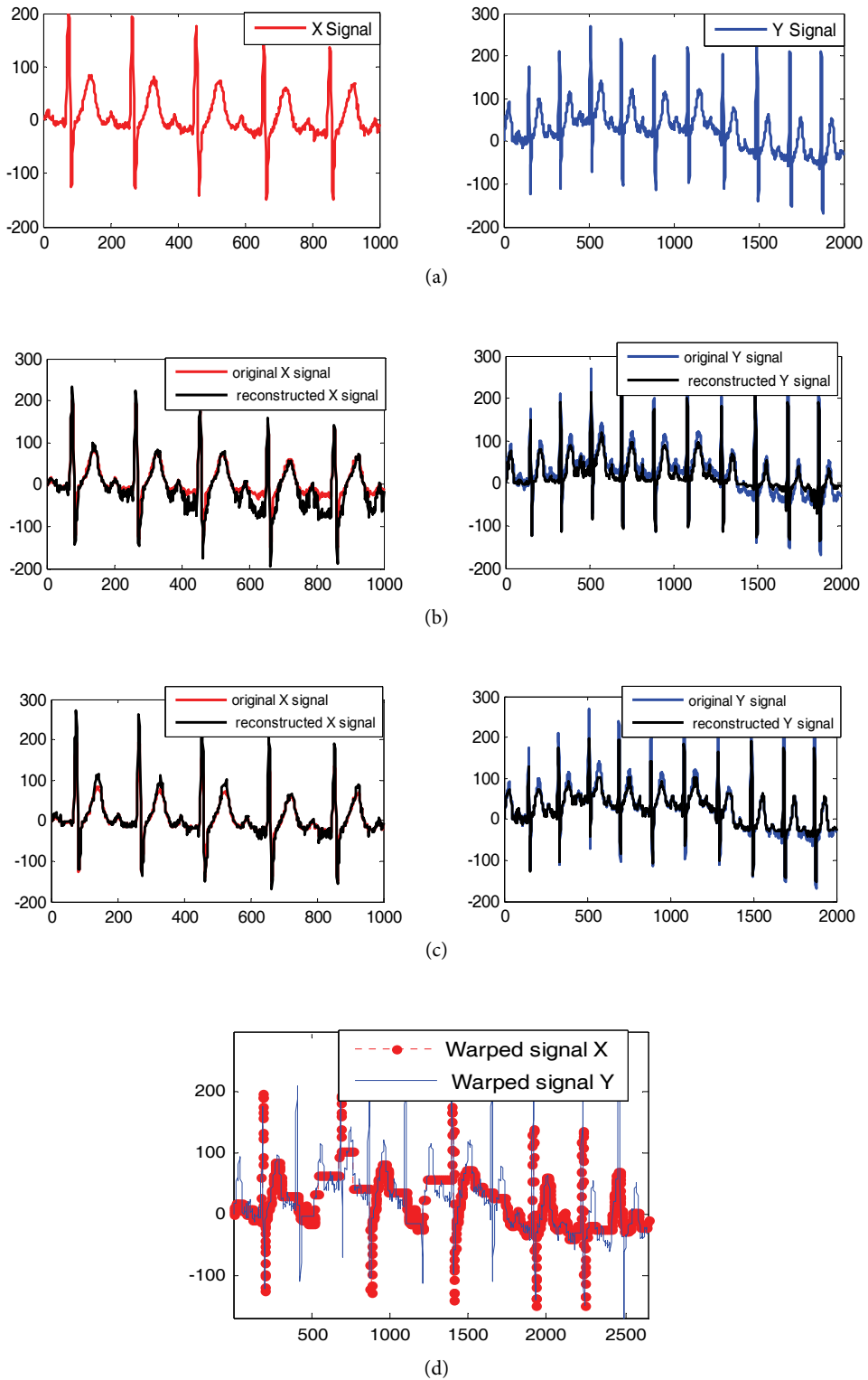


Fig. 12. Illustration of alignment by SEA, QP-DTW, and DTW of two similar traces (215,215) with shift of 35 seconds for Y. (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

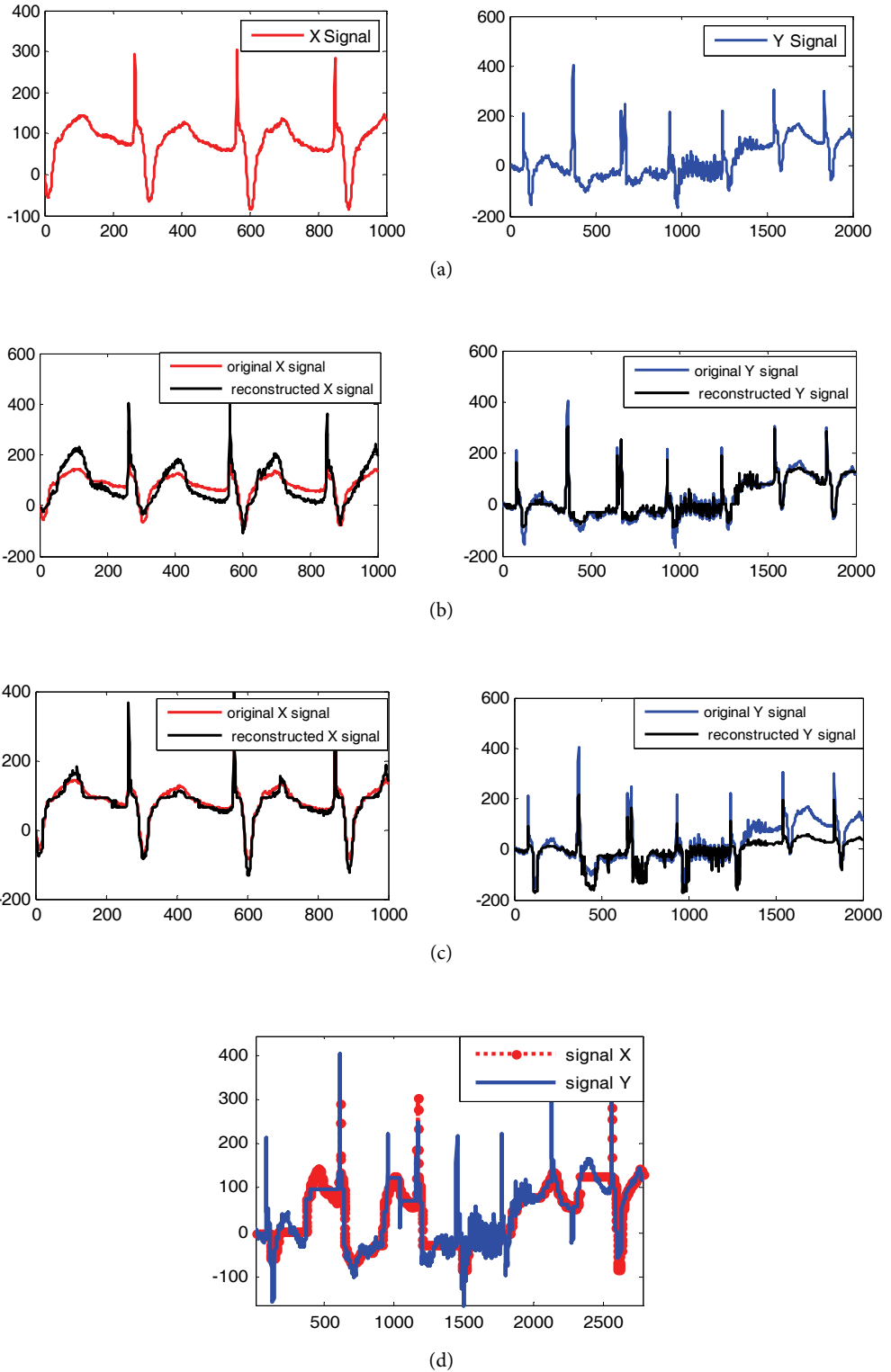


Fig. 13. Illustration of alignment by SEA, QP-DTW, and DTW of two similar traces (104,104) with shift of 20 seconds for Y. (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

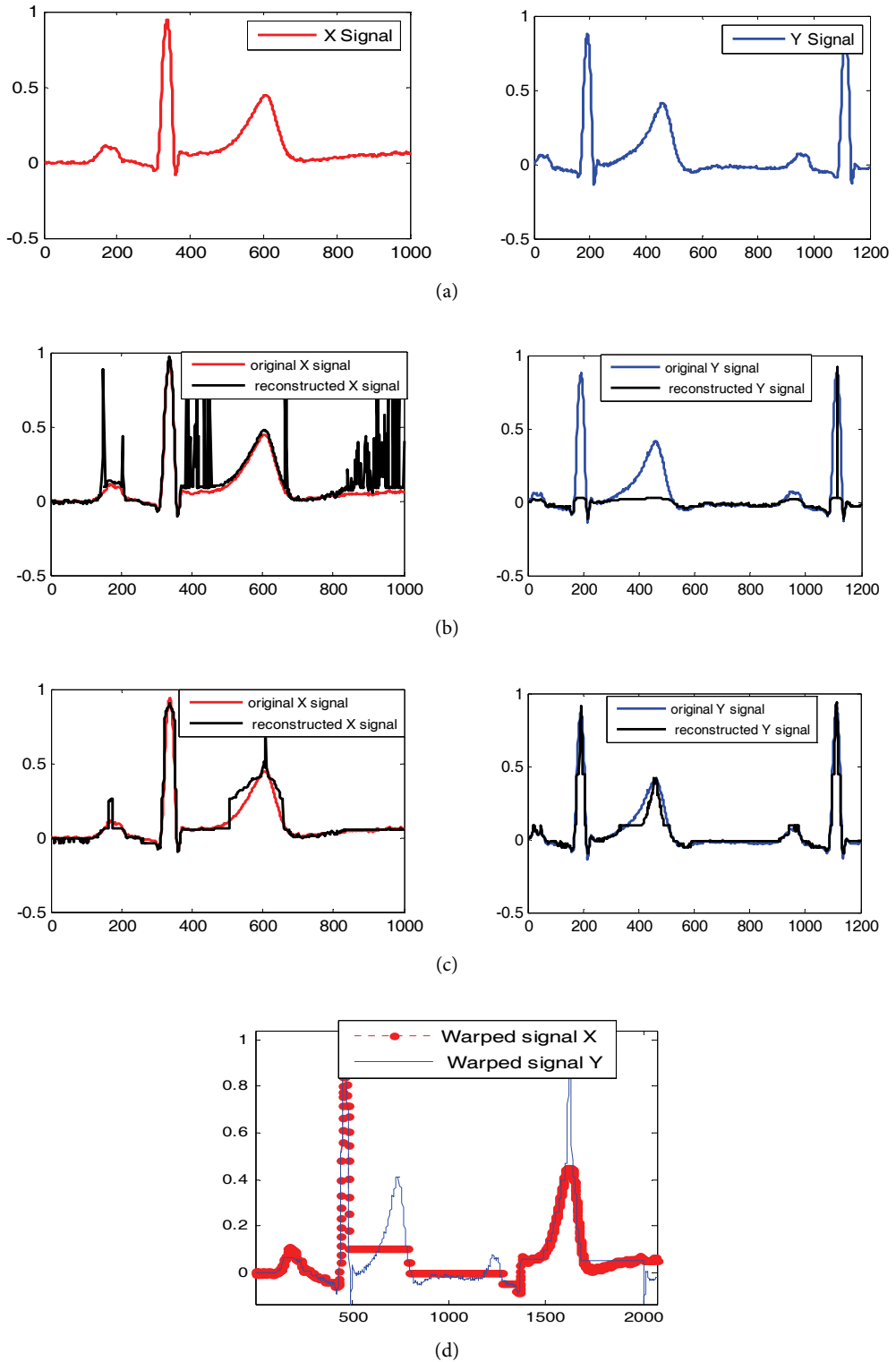


Fig. 14. Illustration of alignment by SEA, QP-DTW, and DTW of two similar traces (s0287lrem, s0287lrem). (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

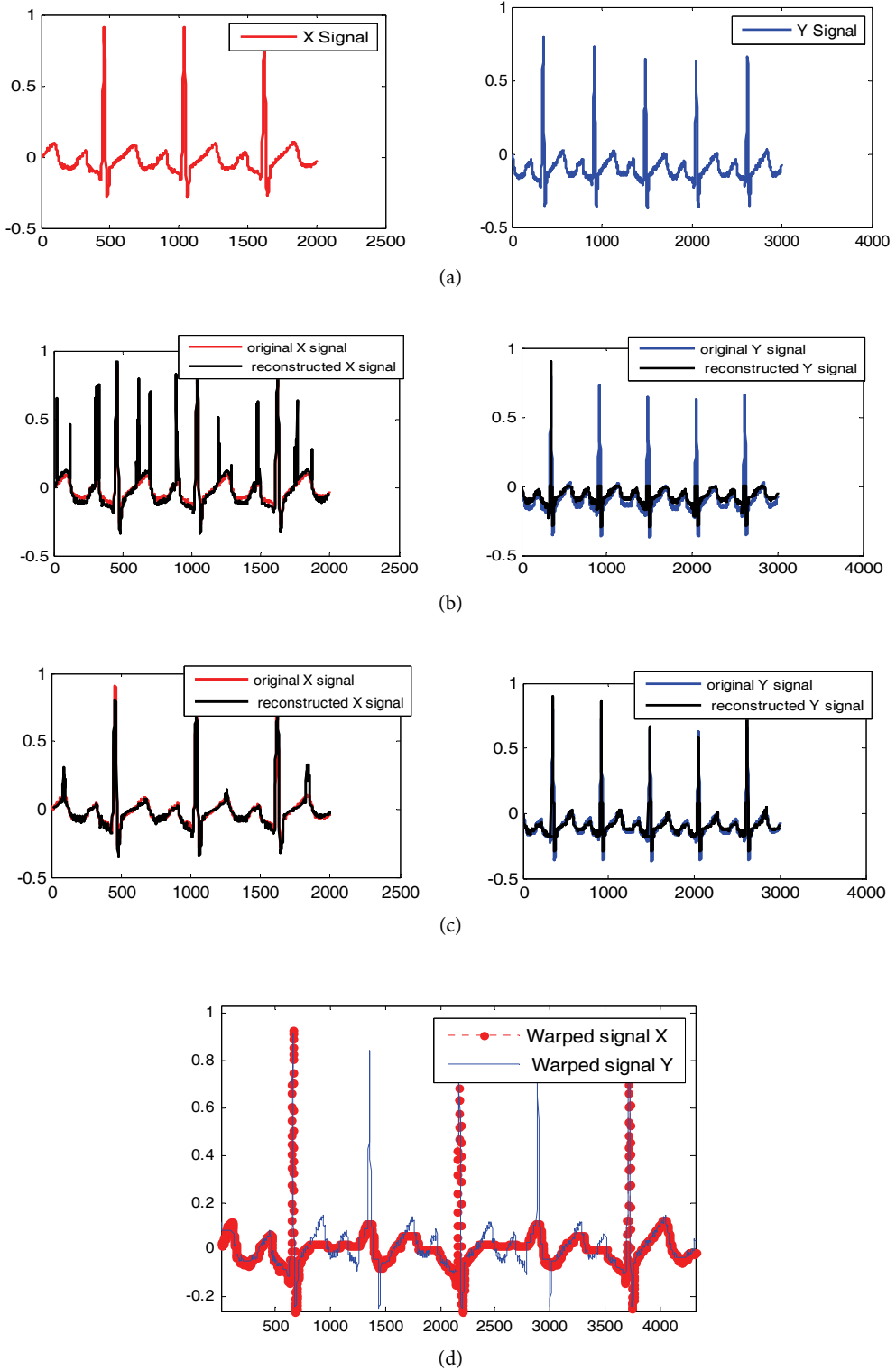


Fig. 15. Illustration of alignment by SEA, QP-DTW and DTW of two similar traces (s0273lrem, s0273lrem). (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

Fig. 13 shows the case of two plots of the same record with 1,000 samples of the time series rec.104 and 2,000 samples of the same time series with a shift of 20 seconds for Y. Fig. 13(a) presents the original time series. Each series is composed of different number of periods which are phase shifted. The Y signal presents also the problem of noise. The alignment performed by the SEA method is presented in Fig. 13(b). We see that SEA fails facing the problems of different amplitude scales and phase shift. Fig. 13(c) presents the alignment done by QP-DTW. It shows that our method resolve all the problems previously cited. The alignment done by DTW is presented in Fig. 13(d). We see that DTW fails again facing the problems of phase shift and noise. The numerical results indicate a high correlation and a low PRD for QP-DTW, a very low correlation and a high PRD for SEA and an average PRD and correlation for DTW.

6.2 Comparison of Similar Traces of the PTB Diagnostic ECG Database

Fig. 14 shows the case of two plots of the same record with 1,000 samples of the time series “s0287lrem” from the PTB Diagnostic ECG Database [18] for X and 1,200 samples of the same time series. The time series are phase shifted. The visual inspection and the numerical results indicate that the alignment done by QP-DTW is much better than the alignment done by SEA and DTW. Specifically, it can be noticed in this particular case that the SEA method deals weakly when we have one cycle in series X and two cycles in series Y, whereas the proposed QP-DTW deals much better with this complex case.

Fig. 15 shows the case of two plots of the same record with 2,000 samples of the time series “s0273lrem” from the PTB Diagnostic ECG Database for X and 3,000 samples of the same time series for Y. The visual inspection of the Fig. 15 shows that QP-DTW has done a good alignment, contrary to DTW which fails completely and to SEA which shows some weakness. In Table 3, the numerical results confirm this fact.

Table 3. Comparative table of the methods DTW, SEA, and QP-DTW concerning similar time series of Figs. 14–15 from the PTB Diagnostic ECG Database

ECG segments	DTW		SEA		QP-DTW	
	PRD%	Corr	Mean PRD%	Mean Corr	Mean PRD%	Mean Corr
(s0287, s0287)	43.18	0.9052	83.20	0.6433	27.97	0.9716
(s0273, s0273)	47.27	0.8825	60.10	0.8117	26.91	0.9673

6.3 Comparison of Dissimilar Time Series

Fig. 16 shows the case of two plots from two different records with 2,000 samples of the time series rec.103 for X and 2,000 samples of the time series rec.123 for Y and a shift of 40 seconds for Y. The visual inspection of the figure shows that the original time series and the reconstructed time series by QP-DTW are detached. The numerical results in Table 4 indicate a very low correlation and a very high PRD in favor of QP-DTW. This proves that our method is perfectly capable of differentiating dissimilar time series.

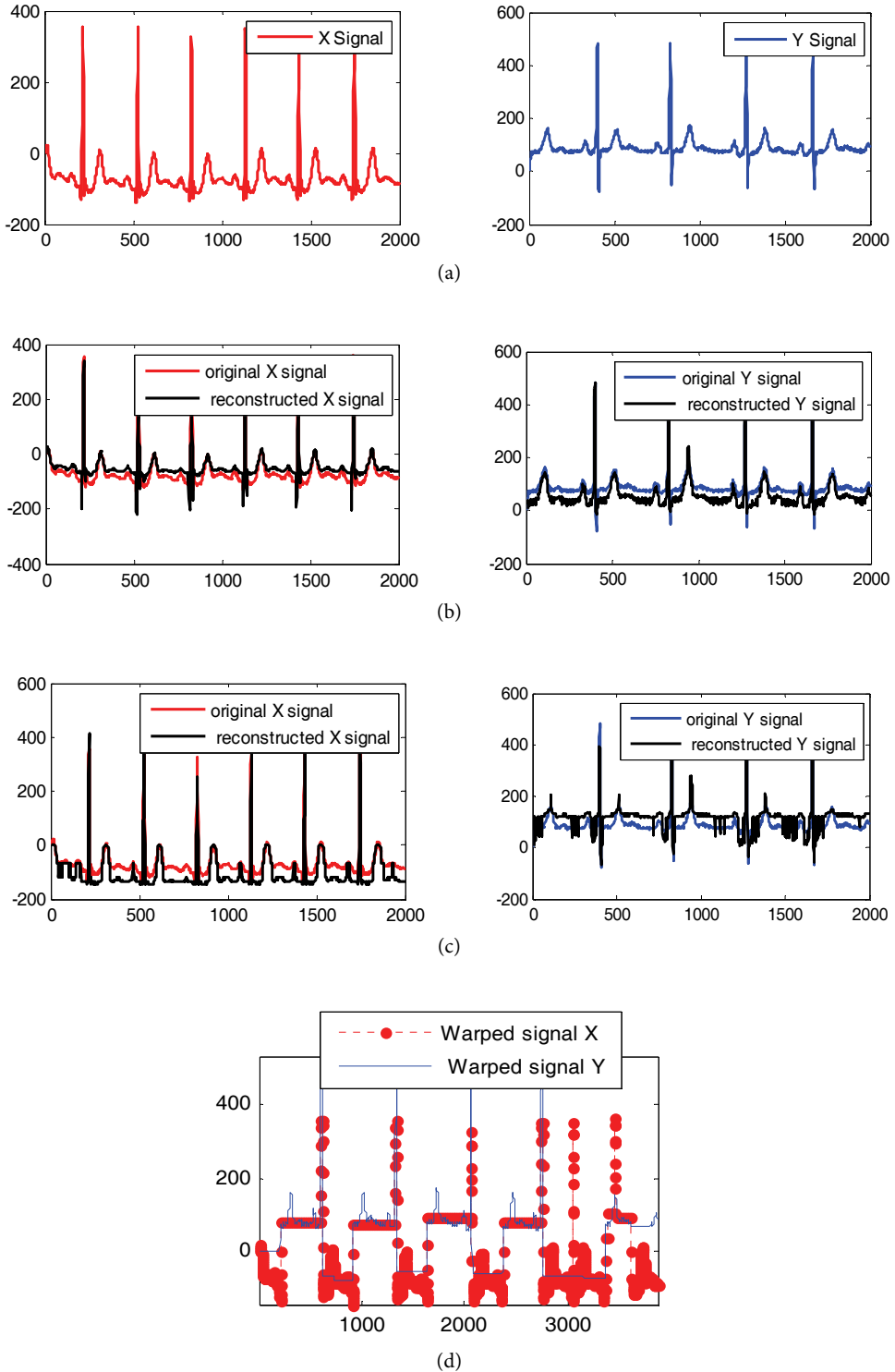


Fig. 16. Illustration of alignment by SEA, QP-DTW and DTW of two different traces (103,123) with shift of 40 seconds for Y. (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

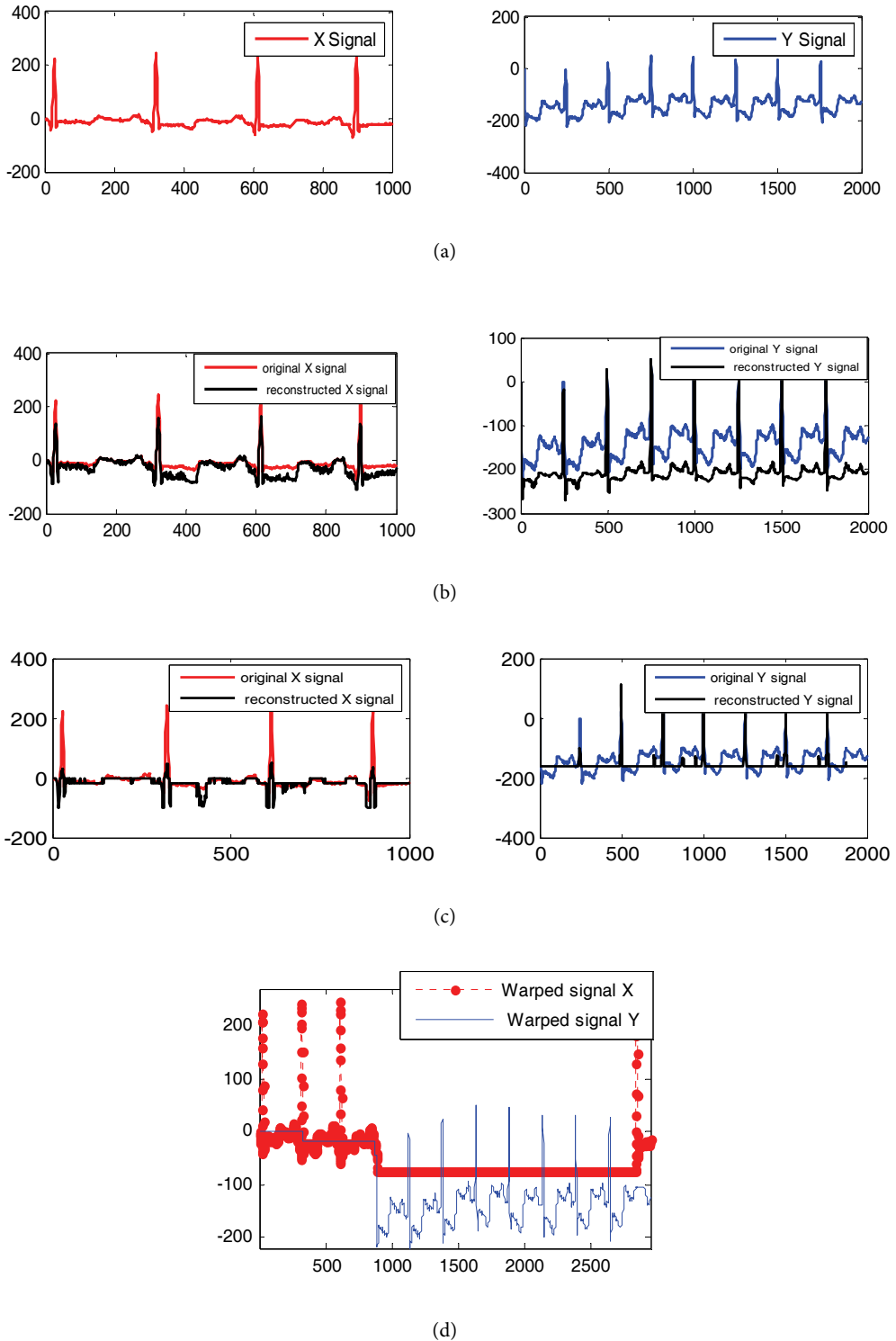


Fig. 17. Illustration of alignment by SEA, QP-DTW and DTW of two different traces (100,112) with shift of 50 seconds for Y. (a) Original traces. (b) SEA alignment. (c) QP-DTW alignment. (d) DTW alignment.

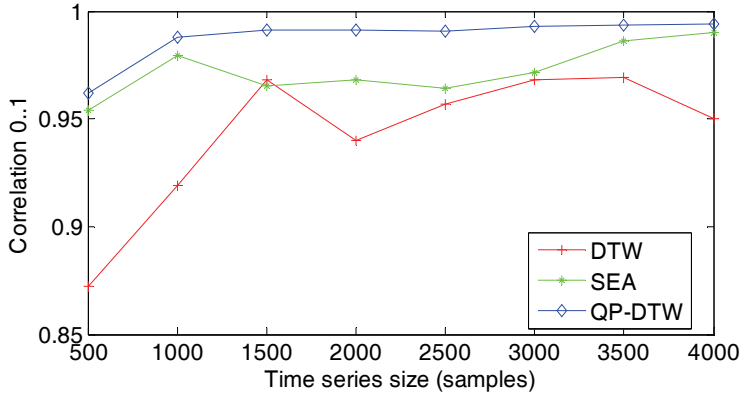


Fig. 18. Correlation factor as a function of time series length for DTW, SEA, and QP-DTW from rec.118 (MITBIH) with shift of 35 seconds for Y.

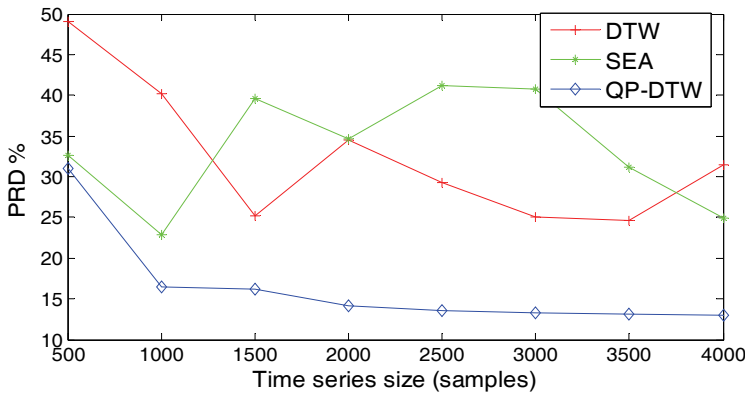


Fig. 19. PRD factor as a function of time series length (samples) for DTW, SEA, and QP-DTW from rec.118 (MITBIH) with shift of 35 seconds for Y.

Table 4. Comparative table of the methods DTW, SEA, and QP-DTW concerning dissimilar time series of Figs. 16–17

ECG segments	DTW		SEA		QP-DTW	
	PRD%	Corr	Mean PRD%	Mean Corr	Mean PRD%	Mean Corr
(103, 123)	57.24	0.8472	51.67	0.9673	69.01	0.8953
(100, 112)	80.25	0.7034	62.58	0.8709	88.38	0.5866

Fig. 17 shows two plots from two different records with 1,000 samples of the time series rec.100 for X and 2,000 samples of the time series rec.112 for Y and a shift of 50 seconds for Y. The visual inspection and the numerical results indicate that QP-DTW done the best misalignment.

Figs. 18 and 19 show respectively the changes in the correlation and in the PRD of the three methods in relation to the size of time series. Fig. 18 indicates that the correlation of QP-DTW is the best and that it is always close to 1. Fig. 19 indicates that the PRD of QP-DTW is always the lowest despite increasing data size.

7. Conclusion

In this work, we have proposed a novel alignment method for quasi-periodic time series. The proposed QP-DTW can be seen as an upgrade of the classical DTW that handles alignment of very complex time series, more effectively than the existing SEA method, quantitatively and qualitatively. Therefore, the proposed method would be a good alternative for time series applications, such as data mining, pattern recognition, search/retrieval, motif discovery and classification.

References

- [1] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," in *Foundations of Data Organization and Algorithms*. Heidelberg: Springer, 1993, pp. 69-84.
- [2] A. Barbulescu and E. Baidu, "A hybrid approach for modeling financial time series," *The International Arab Journal of Information Technology*, vol. 9, no. 4, pp. 327-335, 2012.
- [3] M. Awad, H. Pomares, I. R. Ruiz, O. Salameh, and M. Hamdon, "Prediction of time series using RBF neural networks: a new approach of clustering," *The International Arab Journal of Information Technology*, vol. 6, no. 2, pp. 138-143, 2009.
- [4] M. Z. Arshad, J. M. Nawaz, and S. J. Hong, "Fault detection in the semiconductor etch process using the seasonal autoregressive integrated moving average modeling," *Journal of Information Processing System*, vol. 10, no. 3, pp. 429-442, 2014.
- [5] L. Ulanova, T. Yan, H. Chen, G. Jiang, E. Keogh, and K. Zhang, "Efficient long-term degradation profiling in time series for complex physical systems," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, 2015, pp. 2167-2176.
- [6] A. Vishwa and A. Sharma, "Arrhythmic ECG signal classification using machine learning techniques," *International Journal of Computer Science, Information Technology, & Security*, vol. 1, no. 2, pp. 163-167, 2011.
- [7] K. Vimala and D. V. Kalavani, "Classification of cardiac vascular disease from ECG signals for enhancing modern health care scenario," *Health Informatics-An International Journal (HIJ)*, vol. 2, no. 4, pp. 63-72, 2013.
- [8] E. J. D. S. Luz, W. R. Schwartz, G. Camara-Chavez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: a survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144-164, 2016.
- [9] B. Boucheham, "Matching of quasi-periodic time series patterns by exchange of block-sorting signatures," *Pattern Recognition Letters*, vol. 29, no. 4, pp. 501-514, 2008.
- [10] B. Boucheham, "Abnormality detection in electrocardiograms by time series alignment," *Communications in Information Science and Management Engineering*, vol. 1, no. 3, pp. 6-10, 2011.
- [11] N. Begum and E. Keogh, "Rare time series motif discovery from unbounded streams," *Proceedings of the VLDB Endowment*, vol. 8, no. 2, pp. 149-160, 2014.
- [12] T. C. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164-181, 2011.
- [13] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43-49, 1978.
- [14] B. Boucheham, "Efficient matching of very complex time series," *International Journal of Machine Learning and Cybernetics*, vol. 4, no. 5, pp. 537-550, 2013.

- [15] E. Keogh, L. Wei, X. Xi, S. H. Lee, and M. Vlachos, "LB_Keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures," in *Proceedings of the 32nd International Conference on Very Large Data Bases*, Seoul, Korea, 2006, pp. 882-893.
- [16] C. A. Ratanahatano and E. Keogh, "Three myths about dynamic time warping data mining," in *Proceedings of the 2005 SIAM International Conference on Data Mining*, Newport Beach, CA, 2005, pp. 506-510.
- [17] MIT-BIH Arrhythmia Database [Online]. Available: <http://www.physionet.org/physiobank/database/mitdb/>.
- [18] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.
- [19] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of AAAI-94 Workshop on Knowledge Discovery in Database (KDD)*, 1994, pp. 359-370.
- [20] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, 2000, pp. 285-289.
- [21] L. Chen and M. T. Ozsu, "Similarity-based retrieval of time-series data using multi-scale histograms," University of Waterloo, Waterloo, Canada, *Technical Report No. CS-2003-31*, 2003.
- [22] H. Shatkay and S. B. Zdonik, "Approximate queries and representations for large data sequences," in *Proceedings of the 12th International Conference on Data Engineering*, New Orleans, LA, 1996, pp. 536-545.
- [23] T. Bozkaya, N. Yazdani, and M. Ozsoyoglu, "Matching and indexing sequences of different lengths," in *Proceedings of the 6th International Conference on Information and Knowledge Management*, Las Vegas, NV, 1997, pp. 128-135.
- [24] J. R. Annam, S. S. Mittapalli, and R. S. Bapi, "Time series clustering and analysis of ECG heart-beats using dynamic time warping," in *Proceedings of 2011 Annual IEEE India Conference (INDICON)*, Hyderabad, India, 2011, pp. 1-3.
- [25] N. Begum, L. Ulanova, J. Wang, and E. Keogh, "Accelerating dynamic time warping clustering with a novel admissible pruning strategy," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, 2015, pp. 49-58.
- [26] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, "Dynamic time warping averaging of time series allows faster and more accurate classification," in *Proceedings of 2014 IEEE International Conference on Data Mining (ICDM)*, Shenzhen, China, 2014, pp. 470-479.
- [27] G. Zhang, W. Kinsner, and B. Huang, "Electrocardiogram data mining based on frame classification by dynamic time warping matching," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 12, no. 6, pp. 701-707, 2009.
- [28] B. Boucheham, "Reduced data similarity-based matching for time series patterns alignment," *Pattern Recognition Letters*, vol. 31, no. 7, pp. 629-638, 2010.
- [29] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proceedings of the 2001 SIAM International Conference on Data Mining*, Chicago, IL, 2001, pp. 1-11.
- [30] J. B. Kruskal and M. Liberman, "The symmetric time warping algorithm: from continuous to discrete," in *Time Warps, String Edits and Macromolecules*. Reading, MA: Addison-Wesley, 1983.
- [31] I. Boulnemour and B. Boucheham, "I-SEA: improved shape exchange algorithm for quasi-periodic time series alignment," in *Proceedings of 2015 International Conference on Computer Vision and Image Analysis Applications (ICCVIA)*, Sousse, Tunisia, 2015, pp. 1-6.
- [32] A. Bahri, Y., Naija, G. Jomier, and M. Manouvrier, "Recherche par similarité de séquences temporelles dans les bases de données: un état de l'art," LAMSADE, Université Paris Dauphine, France, 2014.

- [33] L. Junkui and W. Yuanzhen, "Early abandon to accelerate exact dynamic time warping," *International Arab Journal of Information Technology*, vol. 6, no. 2, pp. 144-152, 2009.
- [34] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561-580, 2007.



Imen Boulnefour <https://orcid.org/0000-0003-3478-423X>

She received her B.S. and M.S. degrees in Computer Science from the University of Skikda in 2009 and 2011, respectively. She is now a PhD student in the Department of Computer Science of the same university. Her fields of interest are Data mining, pattern recognition, diagnosis, Time series, Optimization.



Bachir Boucheham <https://orcid.org/0000-0001-7286-659X>

He received the Doctor of Science (Ph.D.) and the HDR (post-doctoral degree for research supervision), respectively in 2005 and 2009, both in Computer Science from Mentouri University of Constantine, Algeria. He is a full professor of Computer Science at the Department of Informatics, University of Skikda, Algeria and a member of the LRES research laboratory within the same university. His main areas of interest and expertise include pattern recognition, computer vision, image processing and retrieval, time series and signals processing and retrieval, data reduction and compression. Prof. Boucheham authored/co-authored several papers in various high impact international journals and conferences and serves as reviewer for many international journals (<https://publons.com/author/489776/bachir-boucheham#profile>).