

# Comparative Analysis of DTW based Outlier Segregation Algorithms for Wrist Pulse Analysis

Nidhi Garg and Amandeep Bisht\*

UIET, Panjab University, Chandigarh, India; nidhi\_garg@pu.ac.in, amandeepbisht@gmail.com

## Abstract

**Background/Objectives:** Quantification of Wrist Pulse Signals is helpful to take benefit of ancient approach i.e. Pulse Diagnosis. The objective of this paper is to effectively segregate outliers present within wrist pulses. **Methods/Statistical Analysis:** This work presents modification in Dynamic Time Warping (DTW) algorithm. The existing DTW algorithm searches for an optimal path using squared Euclidean distance to measure the local distance between segments. Here, we are discussing and integrating different local distance measures such as Correlation Distance, Manhattan Distance, Kendall's  $\tau$  Distance and Canberra Distance with DTW. All the discussed local distance measures were compared with existing Euclidean based DTW algorithm on the basis of Similarity Index parameter. **Findings:** Results shown that Manhattan Distance and Canberra Distance based DTW algorithm was efficient in optimal path selection and segregation of segments which lose their shape characteristics. In euclidean based DTW, outlier segregation was difficult as all values lied between 0 to 1. Correlation distance and Kendall's tau distance algorithm were inappropriate in detecting outliers as results were not matched with visual observations. It was noticed that combination of Manhattan Distance and Canberra Distance based DTW algorithm were giving better outlier finding.

**Keywords:** Canberra Distance, DTW, Euclidean Distance, Manhattan Distance, Similarity Index, Wrist Pulse outliers

## 1. Introduction

In Traditional Indian Medicine, the organ under study is zeroed down by sensing the palpation from the three fingers (index, middle and ring) placed on the radial artery (vata, pitta, kapha) to analyze body variation<sup>1,2</sup>. To add consistency and objectivity in this concept, a computerized wrist pulse signal is required which can be utilized as a substitute for practitioners job and provide reliable results. The pulses basically captures cardiac muscle activity which is influenced by the characteristics of blood and vessel that make it effective for studying both cardiac and non-cardiac diseases. A segment of wrist pulse comprise of 3 main waves: Percussion, Tidal and Dicrotic waves<sup>3</sup>. Variation in the amplitude and interval of these waves describes imbalance in doshas (vata, pitta, kapha) which is nothing but combination of panchtattva<sup>4</sup>.

The generated Pulse sequence is a semi-periodic physiological signal. One of the frequent problems faced in biomedical signal processing is noise that comes with signal during acquisition process<sup>5</sup>. It is necessary to eliminate disturbances in wrist pulse series so as to maintain the quality and time-frequency information of signal. Although even after noise removal, there exist some irregular pulses which distort key characteristics of pulses. Before feature extraction, a practical outlook for pulse segmentation was followed as done by Xia et al.<sup>6</sup> followed by ensemble averaging of segments so that indistinctness in segments could be clearly pointed<sup>7</sup>. For removal of outliers, Euclidean Distance based Dynamic Time Warping algorithm<sup>8</sup> is used for dissimilar pulse identification and subtraction as performed by Bhasker et al.<sup>9</sup>. To improve this existing DTW, various other similarity measure algorithms such as Correlation Distance, Manhattan Distance,

\*Author for correspondence

Kendall's  $\tau$  Distance and Canberra Distance<sup>10,11</sup> were used as local distance measure to outline better outlier removal algorithm using similarity index. These algorithms are applied indifferent applications such as image, speech processing, character recognition, pattern recognition and feature selection etc<sup>12,13</sup>. In this paper, we have bring about the comparative analysis of existing different outlier algorithms to find the best for wrist pulse analysis and validating the results by studying ensemble averaging of segments.

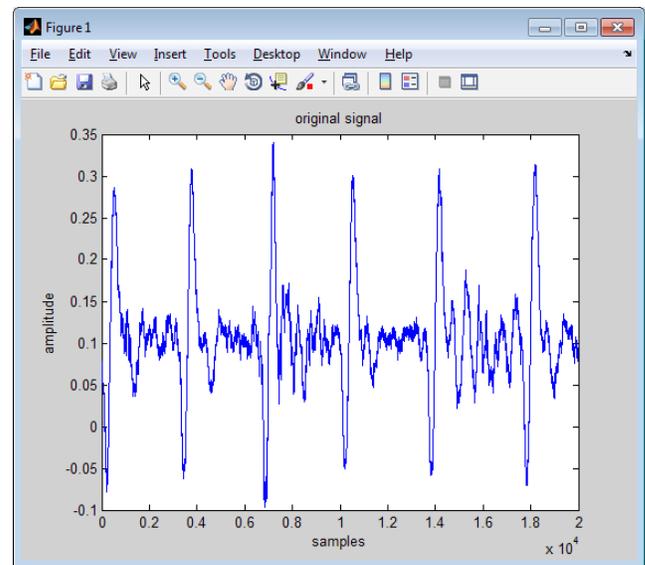
This paper is divided into five different sections as mentioned. Section II describes database collection and pre-processing. In section III, outliers removal algorithms are discussed, Section IV shows results with parameters described in a tabular form and Section V concludes the paper.

## 2. Preprocessing

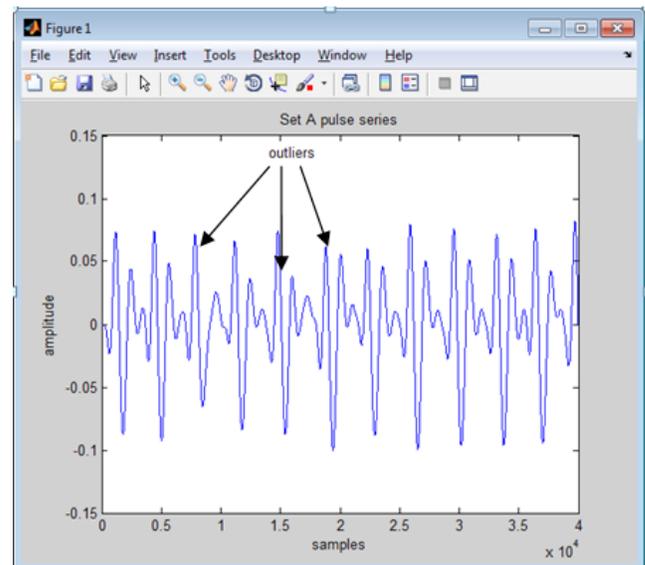
A pressure sensor is used to collect wrist pulses from subject and tools used for acquisition is °LabVIEW<sup>14,22</sup>. Notch filter has been used for removing 50 Hz powerline interference<sup>15</sup>, linear detrending for baseline removal<sup>16</sup> and Daubechies (db) wavelet transform<sup>17</sup> for high frequency noise removal followed by band pass filter. Whole pre-processing has been done in MATLAB Software. Figure 1a and 1brepresents original and denoised pulse signal for Set A that comprises of 11 segments. Ideally, data acquisition setup needs to be perfect that no variation in segments occur even with movement of subject. But practically it's unavoidable. Even slight movement by subject causes these variations which are difficult to avoid. We can observe in figure 1 that all segments are not absolutely similar to each other which is because of body movements. By using appropriate algorithms, segregation of these segments is achievable.

## 3. Outlier Removal Algorithm

Outliers are irregular segments present in pulse series which mainly occurs due to motion artifacts (such as hand movements)<sup>2</sup> as can be visually identified from figure 1b. We can determine heart rate from pulse series which is nothing just beat to beat variations. To study different features such as pulse percussion wave amplitude, tidal wave amplitude, dicrotic notch amplitude, dicrotic wave amplitude, their onsets, timing interval and beat frequency<sup>18,19</sup>, it is compulsory to compute these



(a)



(b)

**Figure 1.** (a) Acquired Wrist pulse - Set A. (b) Denoised Wrist pulse - Set A.

parameter for each pulse segment or after taking average of all the segments. It becomes a necessity to eliminate segments which loses their character as it is not possible to extract features from these irrelevant pulses. Further, we are discussing different outlier segregation algorithms which were implemented to find the best method.

### 3.1 Dynamic Time Warping (DTW)

One of the advantages of DTW is that it can be applied to segments of uneven length that means no need for

zero padding DTW calculates optimal distance for segments of uneven length. It utilizes euclidean distance to match samples in one segment to other in time series by nonlinear alignment<sup>20</sup>. Its main advantage is that it reduces distortion effect by allowing stretching (elastic) transformations of time series for detecting shape similarity for different phases. A warping path is computed in a non-linear fashion by adding local distance with global distance (minimum distance of adjacent element) as shown:

$$\begin{aligned}
 d_{ij} &= a_{ij} + \min(d_{i-1, j-1}, d_{i, j-1}, d_{i-1, j}) \\
 a_{ij} &= (p_i - q_j)^2 \\
 \text{where; } i &= 1, 2, \dots, N; \\
 j &= 1, 2, \dots, M
 \end{aligned}$$

The warping path is bounded to some constraints:

➤ Boundary conditions: If  $D = d_1, d_2, \dots, d_k$

$$d_1 = (1, 1) \text{ and } d_k = (M, N)$$

This requires the warping path to start and finish in diagonally opposite corner cells of the matrix.

➤ Continuity: If  $d_k = (x, y)$  then  $d_{k-1} = (x', y')$

Then,  $x - x' < 1$  and  $y - y' < 1$ . This restricts the allowable steps in the warping path to adjacent cells (including diagonally adjacent cells).

➤ Monotonicity: If  $d_k = (x, y)$  then  $d_{k-1} = (x', y')$ ;

Where,  $x - x' > 0$  and  $y - y' > 0$ . This forces the points in C to be monotonically spaced in time.

$$\text{Path cost } (p, q) = \sum_{i,j} d(i, j)$$

Here path cost is considered as a similarity parameter. Warping path cost for similar segments will be zero and its value will go up as dissimilarity increases. Different outlier removal algorithms has been discussed and implemented on the Set A to find a better solution for outlier segregation.

### 3.2 Euclidean distance

Euclidean distance is most extensively used similarity distance measure of all presented approach. Distance is calculated as the square root of sum of squares of the difference between corresponding samples of two segments of same length. Zero padding is done for equivalent length<sup>11,21</sup>. Lower the value of distance more similar are pulse segments while rise in value increases dissimilarity

between segments. For segment p and q of length,  $k = 1, 2, 3, \dots, N$ ;

$$d_e(p, q) = \sqrt{\sum_{k=1}^N [p(k) - q(k)]^2}$$

where;  $d_e(p, q)$  = Euclidean distance

In DTW, Squared Euclidean distance is used as local distance.

### 3.3 Manhattan Distance

Manhattan distance, sometimes acknowledged as City Block distance, is similar to Euclidean except that it cannot move with points diagonally, rather moves horizontally and vertically in grid based system<sup>11,21</sup>. It is given by:

$$d_m(p, q) = \sum_{k=1}^N |p(k) - q(k)|$$

### 3.4 Correlation Distance

It is defined as one minus the correlation coefficient of segment<sup>11,21</sup>. If p and q are two segments with mean as  $\bar{p}$  and  $\bar{q}$  resp., then correlation distance is:

$$d_{corr}(p, q) = 1 - \frac{\sum_{k=1}^N (p(k) - \bar{p}) * (q(k) - \bar{q})}{\sqrt{\sum_{k=1}^N [p(k) - \bar{p}]^2 * \sum_{k=1}^N [q(k) - \bar{q}]^2}}$$

Its range lies within [0,1]. Values closer to 1 represents greater dissimilarity and vice versa.

### 3.5 Kendall's τ Distance

The Kendall tau rank distance is also known as bubble sort distance and is defined as a metric that counts the number of pair wise disagreements between two data points<sup>11</sup>. Larger the distance, the more dissimilar are data points.

$$d_\tau(p, q) = \sum_{k=1}^N [p(k) - q(k)]$$

### 3.6 Canberra Distance

Canberra distance is defined as weighted version of Manhattan Distance and utilized as a metric for comparing ranked lists. It is very sensitive for values close to zero that's why often used for data points scattered around origin<sup>10,11,21</sup>.

$$d_c(p, q) = \sum_{k=1}^N \frac{|p(k) - q(k)|}{|p(k)| + |q(k)|}$$

### 4. Results

To check similarity between segments, Similarity Index parameter is calculated which tells about resemblance and regularity of pulse segments. Suppose we have n segments in pulse series, a distance matrix D of n × n dimensions is formed for all above algorithm which represents similarity distance between segments.

Similarity Index is defined as:

$$Sim\ Inx(v) = \frac{1}{n} \sum_{u=1}^n D(u,v)$$

Where; u = 1, 2, .....n, v = 1,2,.....n

A threshold point (Th) of Similarity Index is computed by averaging corresponding values in column as shown in Table I. Most similar segments are highlighted bold whereas outliers are highlighted red.

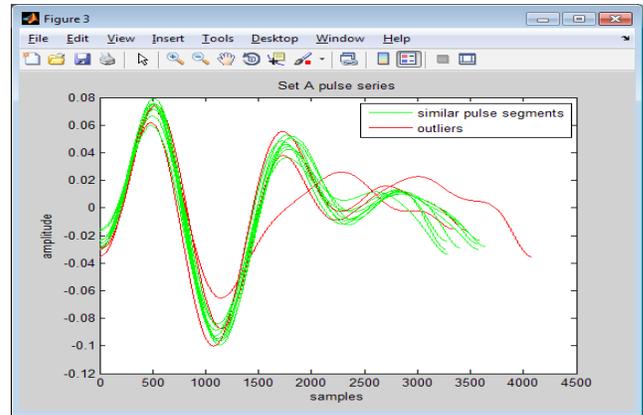
Table 1 describes Similarity Index of different algorithms for 11 segments of Set A. Valley to valley pulse segmentation is done so that similar pulses and outliers are noticeable as shown in Figure 2.

After thresholding, we have removed S3, S5, S6 and S11 from Set A. All the segments with outliers of Set A are

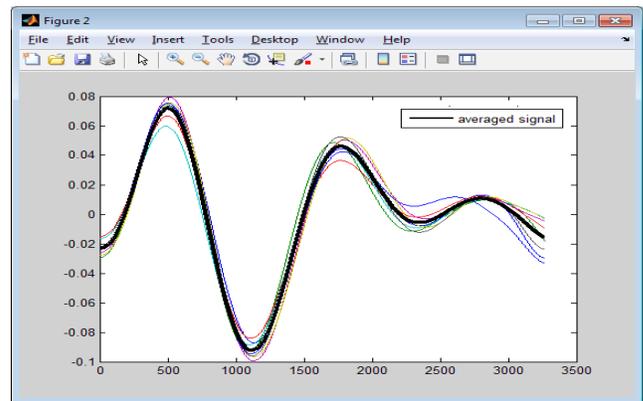
**Table 1.** Similarity Index of different algorithm for Set A

Seg./ Signal	A1	A2	A3	A4	A5	Visual inspection
S1	<b>0.0455</b>	<b>5.4001</b>	0.1801	<b>-356.1</b>	<b>186.70</b>	-
S2	<b>0.0601</b>	6.2988	0.1769	-374.3	241.41	-
S3	0.2886	18.245	<b>0.1285</b>	<b>-317</b>	548.25	Outlier
S4	0.0716	7.9071	0.1762	-358.6	269.79	-
S5	0.0996	10.297	0.1643	-408.3	315.58	Outlier
S6	0.1136	9.729	<b>0.1532</b>	-370.1	269.43	Outlier
S7	0.072	7.4709	0.1643	-356.7	233.12	-
S8	0.0765	6.8917	0.1769	-391.5	214.87	-
S9	0.0692	<b>6.2788</b>	0.1795	-384.8	<b>196.51</b>	-
S10	0.0729	6.7343	0.1872	-372.09	227.89	-
S11	0.0808	8.061	0.1732	-359.7	400.57	-

A1 = Euclidean Distance  
 A2 = Manhattan Distance  
 A3 = Correlation Distance  
 A4 = Kendall's τDistance  
 A5 = Canberra Distance



(a)



(b)

**Figure 2.** (a) Averaging of segments along with outliers for Set A. (b) Averaging of segments without outliers for Set A.

superimposed on each other to easily witness outliers in ensemble average as shown in figure 2a while figure 2b represents ensemble average of pulses without outliers for Set A.

### 5. Conclusion

DTW is a reliable algorithm for similarity measure as it considers elastic transformation of time series and can use for dissimilar length segments. DTW while searching for optimal path considers surrounding points of time series resulting in non-linear alignment which is an advantage over others metric based algorithm. From the collective analysis of results, Euclidean, Manhattan and Canberra based DTW computes S3 as most dissimilar segment and S1 as most similar one. It is difficult to distinguish all prominent outliers with existing DTW as its similarity index lies between 0 to 1 making analysis unreliable for our database. Better Similarity Index has been found for A2 and A5 algorithm and it was validated from ensemble averaging where S3, S5, S6 and S11 were dissimilar

segments. Comparing with visual analysis, segments S3, S5 and S6 were easily differentiable but S11 was difficult to analyze with naked eyes. Correlation distance based Algorithm showed S1 and S10 as outlier, S3 as most similar segment. From tabular data, it is observed that Correlation distance and Kendall's tau distance based DTW algorithm appears to be less effective for detecting shape irregularities in which Manhattan and Canberra based DTW algorithm proved to be better one followed by existing Euclidean based DTW. Overall, it can be concluded that combination of Manhattan and Canberra distance measure with DTW algorithm is robust in terms of detecting small contraction or expansions whereas Correlation distance and Kendall's tau distance algorithm appears to be inefficient in our database for wrist pulse. For future work, efficiency of outlier removal could be improved by using DDTW with above local distance measure to improve shape resemblance and spatial variations.

## 6. References

1. Thakkar S, Thakker B. Wrist Pulse Acquisition and Recording System. Communications on Applied Electronics (CAE). Foundation of Computer Science FCS, New York, USA. 2015 Apr; 1(6). ISSN: 2394-4714.
2. Kyung-Won K, Woo-Gwun N. Comparison of Pulse Diagnosis in Oriental and Western Medicine. Indian Journal of Science and Technology. 2015 Aug; 8(18).
3. Abhinav, Sareen M, Kumar M, Jayashree. Nadi yantra: a robust system design to capture the signal from the radial artery assessment of the autonomic nervous system. Journal of Biomedical Science and Engineering. 2009; 471-9.
4. Bisht A, Garg N. Quantification of wrist pulse signal : An Overview. Proceedings of 4th international conference on Advancements in Engineering and Technology ICAET. 2016.
5. Wang D, Zhang D, Lu G. A robust signal pre-processing framework for wrist pulse analysis. Biomedical Signal Processing and Control. 2016; 23:62-75.
6. Xia C, Li Y, Yan J, Wang Y, Yan H, Guo R. A practical approach to wrist pulse segmentation and single-period average waveform estimation. Int Conf BioMed Eng Inf 2008. BMEI 2008. 2008. p. 334-8.
7. Xia C, Li Y, Yan J, Wang Y, Yan H, Guo R, Li F. A Practical Approach to Wrist Pulse Segmentation and Single-period Average Waveform Estimation. International Conference on BioMedical Engineering and Informatics. 2008.
8. Magdy AK, Khadragei A, Saeb M, Baith Mohamed A. Analysis of DNA Signal Representation Applying Dynamic Time Warping (DTW) and Derivative Dynamic Time Warping (DDTW). The International Journal of Computer Science and Communications Security (IJCS). 2013 Jan; 3.
9. Thakkar B, Vyas AL. Outlier pulse Detection and Feature extraction for wrist pulse analysis. International Conference on Biological Science and Technologies (ICBST). 2009 Jul; 3.
10. Available from: <https://lemonzi.files.wordpress.com/2013/01/dtw.pdf>
11. Akila A, Chandra E. Slope Finder – A Distance Measure for DTW based Isolated Word Speech Recognition. International Journal of Engineering and Computer Science. 2013 Dec; 2(12).
12. Available from: <http://mmc.tudelft.nl/sites/default/files/DTW-vASCI.pdf>.
13. Wang L, Wang K-Q, Xu L-S. Recognizing wrist pulse waveforms with improved dynamic time warping algorithm. Proceedings of 2004 International Conference on Machine Learning and Cybernetics. 2004; 6.
14. Islam MK, Haque ANMM, Tangim G, Ahammad T, Khondokar MRH. Study and Analysis of ECG Signal Using MATLAB and LABVIEW as Effective Tools. International Journal of Computer and Electrical Engineering. 2012 Jun; 4(3).
15. Wang D, Zhang D. Analysis of pulse waveforms pre-processing. 2012 International Conference on Computerized Healthcare (ICCH). 2012 Dec 17-18. p. 175-80.
16. Jayant A, Singh T, Kaur M. Different techniques to remove baseline wander from ECG signal. International Journal of Emerging Research in Management and Technology. 2013 Jun; 2.
17. Mishra U, Verma L. Noise Removal from ECG Signal by Thresholding with Comparing Different Types of Wavelet. International Journal of Application or Innovation in Engineering and Management. 2014 Mar; 3(3).
18. Wang D, Zhang D, Chan JCN. Feature Extraction of Radial Arterial Pulse. International Conference on Medical Biometrics. 2014 May-Jun. p. 41-6.
19. Thakker B, Vyas AL, Tripathi DM. Time and Frequency Domain Analysis of Wrist Pulse Signals. International Journal of Biomedical Engineering and Technology (IJBT), Inderscience Publishers. 2014; 15(3):273-87.
20. Vasimalla K, Challa N, Naik S. Efficient Dynamic Time Warping for Time Series Classification. Indian Journal of Science and Technology. 2016 Jun; 9(21).
21. Cha S-H. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. International Journal of Mathematical Models and Methods in Applied Sciences. 2007; 1(4).
22. Lascu M, Lascu D. LabVIEW Based Biomedical Signal Acquisition and Processing. Proceedings of the 7th WSEAS Int Conf on Signal Processing, Computational Geometry and Artificial Vision, Athens, Greece. 2007 Aug 24-26.