DYNAMIC TIME WARPING IN HYDROACOUSTICS SIGNAL COMPARISON

ANDRZEJ ŻAK

Polish Naval Academy
Smidowicza 69, 81-103 Gdynia, Poland
a.zak@amw.gdynia.pl

The paper presents the technique of dynamic time warping used as comparison method of hydroacoustics signatures generated by moving ship. In the paper firstly the method of feature extraction from hydroacoustics signatures using calculation of Mel-Frequency Cepstral Coefficients was discussed. Next the method of dynamic time warping for comparison hydroacoustics signals was described. Dynamic Time Warping (DTW) is a well-established algorithm for comparing time series. The basic problem that DTW attempts to solve is how to align two sequences in order to generate the most representative distance measure of their overall difference. The DTW algorithm uses a dynamic programming technique to solve this problem. At the end the chosen results of research were presented on this paper.

INTRODUCTION

Dynamic time warping (DTW) is a technique that finds the optimal alignment between two time series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis. This warping between two time series can then be used to find corresponding regions between the two time series or to determine the similarity between the two time series [2, 8].

The Dynamic Time Warping algorithm (DTW) is a well-known algorithm in many areas. While first introduced in 60's and extensively explored in 70's by application to the speech recognition, it is currently used in many areas: handwriting and online signature matching, sign language recognition and gestures recognition, data mining and time series clustering (time series databases search), computer vision and computer animation, surveillance, protein sequence alignment and chemical engineering, music and signal processing [2, 8].

This work aims to benefit the software metrics analysis through the application of the Dynamic Time Warping algorithm to the hydroacoustics signal comparison.

An example of how one time series is "warped" to another is shown in Figure 1. In Figure 1, each vertical line connects a point in one time series to its correspondingly similar point in the other time series. The lines actually have similar values on the y-axis but have been separated so the vertical lines between them can be viewed more easily. If both of the time series in Figure 1 were identical, all of the lines would be straight vertical lines because no warping would be necessary to 'line up' the two time series. The warp path distance is a measure of the difference between the two time series after they have been warped together, which is measured by the sum of the distances between each pair of points connected by the vertical lines in Figure 1. Thus, two time series that are identical except for localized stretching of the time axis will have DTW distances of zero.

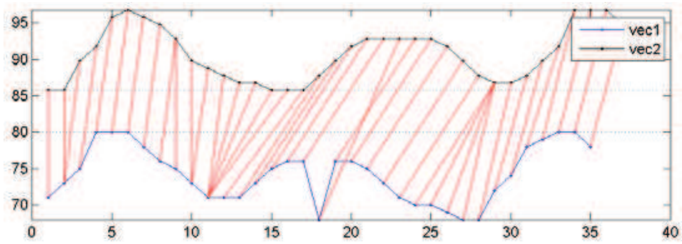

Fig.1. A warping between two time series [7]

Despite the effectiveness of the dynamic time warping algorithm, it has an $O(N^2)$ time and space complexity that limits its usefulness to small time series containing no more than a few thousand data points.

The DTW, as it was said is used to compare time series. Such comparison is one of stages in signals recognition process. The problem of acoustic signals recognition belongs to a much broader topic in scientific and engineering so called pattern recognition. The goal of pattern recognition is to classify objects of interest into one of a number of categories or classes. Hydroacoustics signals identification or classification is the process of automatically recognizing what kind of object is generating acoustics signals on the basis of individual information included in generated sounds

All signal recognition systems, at the highest level, contain two main modules (see figure 2) feature extraction and comparison. Feature extraction is the process that extracts a small amount of data from the hydroacoustics signatures that can later be used to represent each object. Comparison involves the actual procedure to identify the unknown object by comparing extracted features from input sounds with the ones from a set of known stored in some kind of database.

As a hydroacoustics signals in this paper will be understood only sound made by ships in motion and propagated in water environment. Hydroacoustics signals have the great significance because its range of propagation is the widest of all physics field of ship. Controlling and classification of acoustic signals of vessels is now a major consideration for researchers, naval architects and operators. The advent of new generations of acoustic intelligence torpedoes and depth mines has forced to a great effort, which is devoted to classify objects using signatures generated by surface ships and submarines. It has been done in order to increase the battle possibility of submarine armament. Its main objectives are to recognize the ship and only attack this one which belongs to opponent.

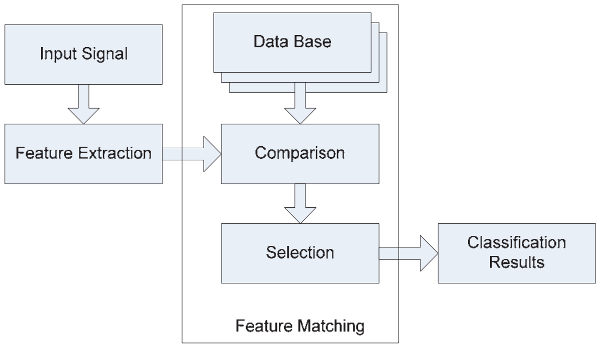

Fig.2. Signal comparison

1. SIGNAL FEATURE EXTRACTION

The purpose of signal feature extraction module is to convert the sound waveform to some type of parametric representation for further analysis and processing. This is often referred as the signal-processing front end. A wide range of possibilities exist for parametrically representing the signals for the sound recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC), and others. Mel-Frequency Cepstrum Coefficients method will be discussed in this paper.

A block diagram of the structure of an MFCC processor is given on figure 3. First step of MFCC processor is the frame blocking. In this step the continuous sound is blocked into frames of N samples, with adjacent frames being separated by M where $M < N$. The first frame consists of the first N samples. The second frame begins M samples after the first frame, and overlaps it by $N - M$ samples. Similarly, the next frames are created so this process continues until all the sound is accounted for within one or more frames [1, 3, 9].

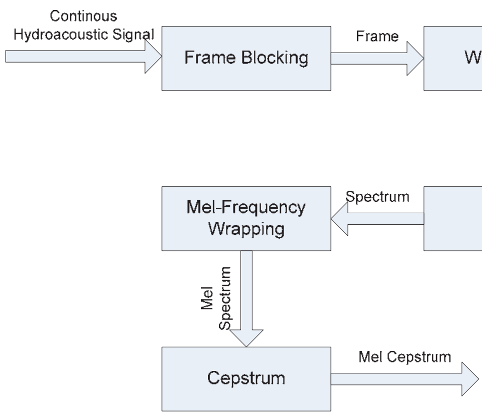

Fig.3. Block diagram of the MFCC processor

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. If we define the window as: $w(n), \quad 0 \le n \le N-1$, where N is the number of samples in each frame, and assume that used window is Hamming window, then the result of windowing is the signal [1, 9]:

$$y(n) = x(n)(0.54 - 0.46\cos(\frac{2\pi n}{N-1})), \quad 0 \le n \le N-1 \tag{1}$$

The next processing step is the Fast Fourier Transform, which converts each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT) which is defined on the set of N samples, as follow [3, 9]:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N}, \quad n = 0,1,2,\ldots,N-1 \tag{2}$$

Next step in MFCC processor is the Mel-frequency Wrapping. Each tone with an actual frequency f, measured in [Hz], a subjective pitch is measured on a scale called the “mel” scale. The mel-frequency scale is a linear frequency spacing below *1000 [Hz]* and a logarithmic spacing above *1000 [Hz]*. Therefore we can use the following approximate formula to compute the mels for a given frequency f in *[Hz]* [1, 3, 9]:

$$mel(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

In final step, we convert the logarithmic mel spectrum back to time. The result is called the mel frequency cepstrum coefficients (MFCC). The cepstral representation of the sound spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the mel spectrum coefficients, and so their logarithm, are real numbers, we can convert them to the time domain using the Discrete Cosine Transform (DCT). Therefore if we denote those mel power spectrum coefficients that are the result of the last step are $S_k, \quad k = 1,2,\ldots,K$, we can calculate the MFCC's, as [1, 3, 9]:

$$c_n = \sum_{k=1}^{K} \log(S_k)\cos(\frac{n(k-0.5)\pi}{K}), \quad n = 1,2,\ldots,K \tag{4}$$

Note that we exclude the first component, c_0 from the DCT since it represents the mean value of the input signal which carried little object specific information.

2. DYNAMIC TIME WARPING

A distance measurement between time series is needed to determine similarity between time series and for time series classification. Euclidean distance is an efficient distance measurement that can be used. The Euclidian distance between two time series is simply the sum of the squared distances from each nth point in one time series to the nth point in the other. The main disadvantage of using Euclidean distance for time series data is that its results are very unintuitive. If two time series are identical, but one is shifted slightly along the time axis, then Euclidean distance may consider them to be very different from each other. Dynamic time warping (DTW) was introduced to overcome this limitation and give intuitive distance measurements between time series by ignoring both global and local shifts in the time dimension [2].

The dynamic time warping problem is stated as follows: given two time series $X = (x_1, x_2, \dots, x_N),\ N \in \mathbb{N}$ and $Y = (y_1, y_2, \dots, y_M),\ M \in \mathbb{N}$ represented by the sequences of values (or curves represented by the sequences of vertices) DTW yields optimal solution in the $O(MN)$ time which could be improved further through different techniques such as multiscaling. The only restriction placed on the data sequences is that they should be sampled at equidistant points in time (this problem can be resolved by re-sampling).

If sequences are taking values from some feature space Φ than in order to compare two different sequences $X, Y \in \Phi$ one needs to use the local distance measure which is defined to be a function [4]:

$$d = \Phi \times \Phi \to \mathbb{R} \geq 0 \tag{5}$$

Intuitively d has a small value when sequences are similar and large value if they are different. Since the Dynamic Programming algorithm lies in the core of DTW it is common to call this distance function the "cost function" and the task of optimal alignment of the sequences becoming the task of arranging all sequence points by minimizing the cost function (or distance).

Algorithm starts by building the distance matrix $C \in \mathbb{R}^{N \times M}$ representing all pairwise distances between X and Y. This distance matrix called the local cost matrix for the alignment of two sequences X and Y [4]:

$$C_l \in R^{N \times M} : c_{i,j} = \left\| x_i - y_j \right\|, i \in [1, \dots, N], j \in [1, \dots, M] \tag{6}$$

Once the local cost matrix built, the algorithm founds the alignment path which runs through the low-cost areas – "valleys" on the cost matrix, (see figure 4). This alignment path (or warping path, or warping function) defines the correspondence of an element $x_i \in X$ to $y_j \in Y$ following the boundary condition which assigne first and last elements of X and Y to each other, (see figure 5).

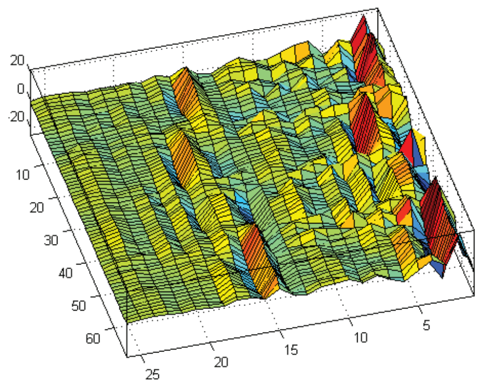

Fig.4. Time series alignment, cost matrix heatmap [7]

Formally speaking, the alignment path built by DTW is a sequence of points $p = (p_1, p_2, \ldots, p_K)$ with $p_l = (p_i, p_j) \in [1, \ldots, N] \times [1, \ldots, M]$ for $l \in [1, \ldots, K]$ which must satisfy to the following criteria [4, 5]:

- boundary condition: $p_1 = (1, 1)$ and $p_K = (N, M)$. The starting and ending points of the warping path must be the first and the last points of aligned sequences;
- monotonicity condition: $n_1 \leq n_2 \leq \ldots \leq n_K$ and $m_1 \leq m_2 \leq \ldots \leq m_K$. This condition preserves the time-ordering of points;
- step size condition: this criteria limits the warping path from long jumps (shifts in time) while aligning sequences. We will use the basic step size condition formulated as $p_{l+1} - p_l \in \{(1, 1), (1, 0), (0, 1)\}$.

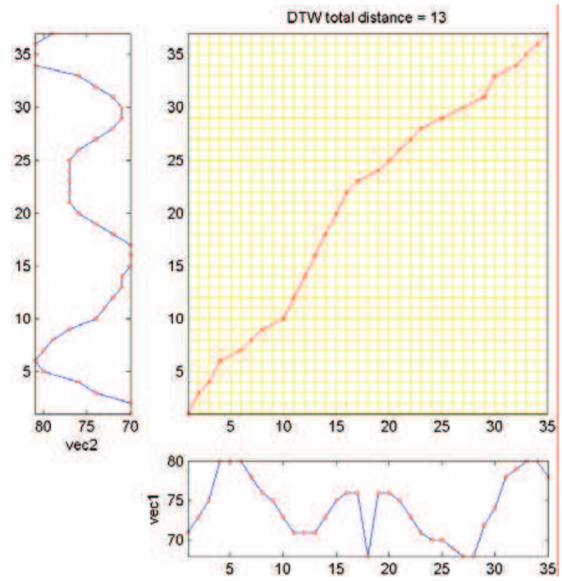

Fig.5. The optimal warping path aligning time series [7]

The cost function associated with a warping path computed with respect to the local cost matrix (which represents all pairwise distances) will be [4]:

$$c_P(X, Y) = \sum_{l=1}^{L} c(x_{n_l}, y_{m_l}) \quad (7)$$

The warping path which has a minimal cost associated with alignment called the optimal warping path. We will call this path P^*. By following the optimal warping path definition in order to find one, we need to test every possible warping path between X and Y which could be computationally challenging due to the exponential growth of the number of optimal paths as the lengths of X and Y grow linearly. To overcome this challenge, DTW employs the Dynamic Programming - based algorithm with complexity only $O(MN)$.

The Dynamic Programming part of DTW algorithm uses the DTW distance function [4]:

$$DTW(X,Y) = c_{p^*}(X,Y) = \min\{c_P(X,Y),\ p \in P^{N\times M}\} \tag{8}$$

where $P^{N\times M}$ is the set of all possible warping paths and builds the accumulated cost matrix or global cost matrix D which defined as follows [4, 5, 6]:

- First row: $D(1, j) = \sum_{k=1}^{j} c(x_1, y_k),\ j = [1,\ldots,M]$
- First column: $D(i, 1) = \sum_{k=1}^{i} c(x_k, y_1),\ i = [1,\ldots,N]$
- All other elements:
  $D(i, j) = \min\{D(i-1, j-1),\ D(i-1, j),\ D(i, j-1)\} + c(x_i, y_j),\ i \in [1,\ldots,N],\ j \in [1,\ldots,M]$.

The time cost of building this matrix is $O(NM)$. Once the accumulated cost matrix built the warping path could be found by the simple backtracking from the point $p_{end}(M,N)$ to the $p_{start}(1,1)$ following the greedy strategy. Summing the values of elements on the warping path it can be defined the unnormalized total distance between processed signals. Normalization of this parameters can be made by dividing it by the length of warping path [5, 6].

3. THE RESULTS OF RESEARCH

During the researches few ships were measured on the Polish Navy Test and Evaluation Acoustic Ranges. Next for acquired hydroacoustics signals the features' process extraction were conducted, according to the schema presented on figure 3. Possessing Mel-Frequency Cepstrum Coefficients for every hydroaocustics signal the comparison process begun. For selected pairs of signals, using algorithm of dynamic time warping the accumulated cost matrix were calculated. Basis on this matrix the warping path has be found and distance between compared signals. This distance is coefficient of similarity between compared signals. Using it we can presume about signal adhesion to the group of signals and the same we can make object classification. On the pictures bellow it was shown the chosen results (accumulated cost matrix and warping path) of calculations.

4. CONCLUSION

The main aim of paper was to check usefulness of dynamic time warping in hydroacoustics signals processing. As it is shown on results using of dynamic time warping is useful for hydroacoustics signal comparison. This method could be used in solving sailing object classification problem. Presented in paper part of researches is quite simple because in investigation there were not take into account that object sounds change with time, efficiency conditions (e.g. some elements of machinery are damaged), sound rates, etc. It doesn't consider the influence of changes of environment on acquired hydroacoustics signals. Therefore these cases should be investigated in future research.

Presently there is the DTW successors, Cluster Generative Statistical Dynamic Time Warping (CSDTW), based on dynamic time warping (DTW) and hidden Markov modeling (HMM) which demonstrates superior performance over the plain DTW or HMM and combines cluster analysis and generative statistical sequence modeling.

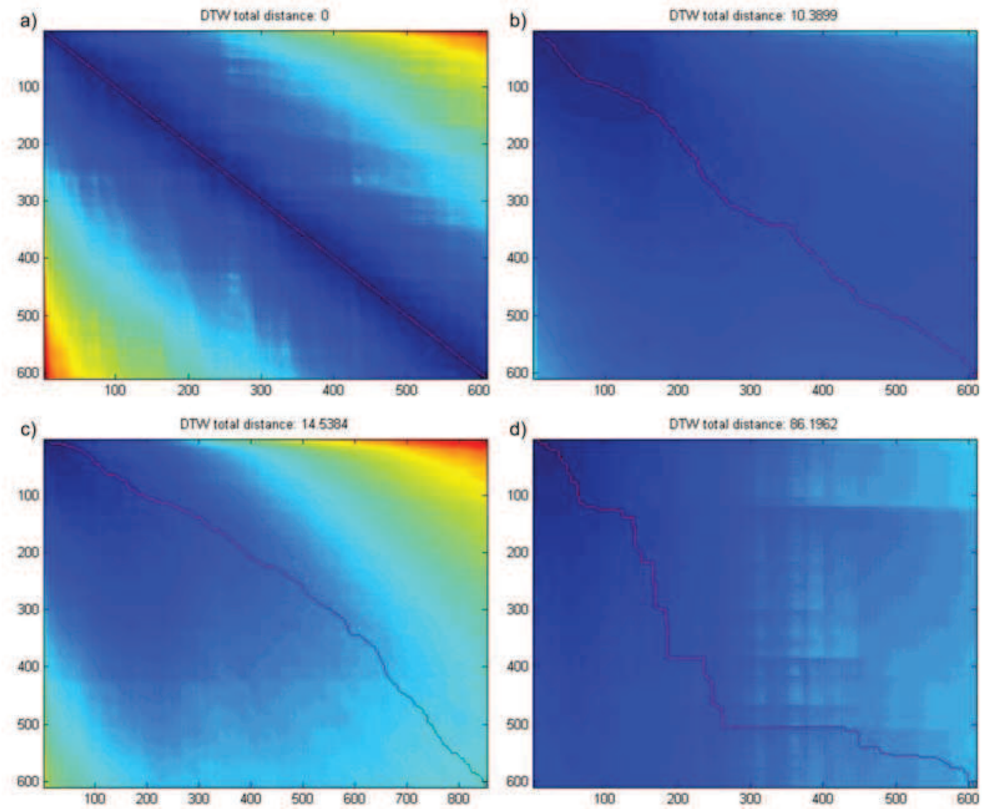

Fig.6. Accumulated cost matrix and warping path a) for the object itself; b) for the same object which move on course NS and SN, c) for the same object which moves with different speed on course NS and WE, d) for the different objects

REFERENCES

[1] C.H. Lee, F.K. Soong, K.K. Paliwal, Automatic Speech and Speaker Recognition: Advanced Topics, Kluwer, Boston, 1996
[2] S. Salvador, P.Chan, FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space, Florida Institute of Technology, 2006
[3] T. Schalk, P. J. Foster, Speech Recognition: The Complete Practical Reference Guide, Telecom Library Inc, New York, 1993
[4] P. Senin, Dynamic Time Warping Algorithm Review, University of Hawaii at Manoa, 2008
[5] Strona internetowa: http://labrosa.ee.columbia.edu/matlab/dtw/
[6] Strona internetowa: http://www.learnartificialneuralnetworks.com/speechrecognition.html #timewarping
[7] Strona internetowa: http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/ dpDtw.asp?title=14-4%20Dynamic%20Time%20Warping
[8] T. P. Zieliński, Cyfrowe przetwarzanie sygnałów, od teorii do zastosowań, WKŁ, 2009
[9] A. Żak: Creating patterns for hydroacoustics signals, Hydroacoustics Vol. 8, Gdynia 2005, pp. 265- 270