Automatic Template Matching for Classification of Dolphin Vocalizations

Gao Rui[*1], Mandar Chitre[‡], S.H. Ong[*2] and Elizabeth Taylor[†]

[*] Department of Electrical and Computer Engineering,
National University of Singapore,
E4-05-48, 4 Engineering Drive 3, Singapore 117576
[1]elegr@nus.edu.sg
[2]eleongsh@nus.edu.sg

[‡]Acoustic Research Laboratory, Tropical Marine Science Institute,
National University of Singapore, 12a Kent Ridge Road, Singapore 119223
mandar@arl.nus.edu.sg

[†]Marine Mammal Research Laboratory, Tropical Marine Science Institute,
National University of Singapore, 14 Kent Ridge Road, Singapore 119223
mdcohe@leonis.nus.edu.sg

***Abstract*—Whistle classification is a key step in many studies of dolphin vocalizations. Automatic whistles tracing algorithms have been developed but tracing errors such as breaks and outliers are usually unavoidable. Local variations of whistle contours occur even in whistles of the same type. In this paper, we describe a modified dynamic time warping (DTW) algorithm for dynamic non-linear matching. It exhibits a good performance in matching against a template whistle. The modifications to the basic DTW algorithm provide improved tolerance to noise and breaks in tracing. Together with automatic de-noising, this template matching is used to classify vocalizations of Indo-Pacific dolphins (*Sousa chinensis*). We believe this method can be applied for large scale analysis of whistles for species recognition, dolphin training and other dolphin studies.**

***Index Terms*—Dynamic Time Warping, Pattern matching**

I. Introduction

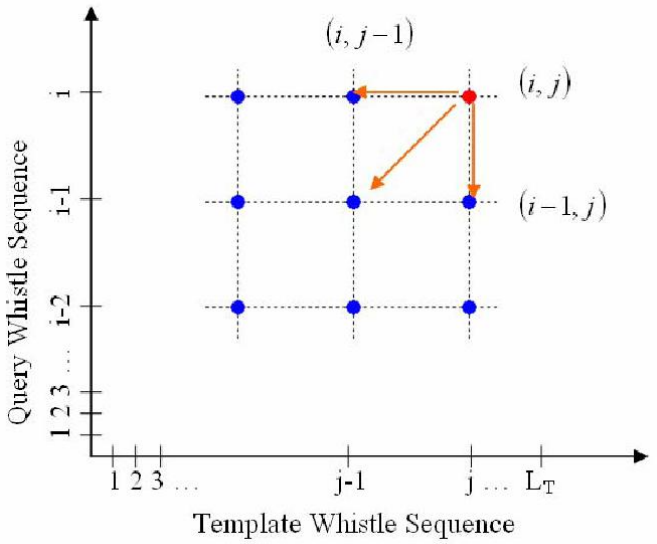
DOLPHIN whistle classification is a key step in study of dolphin recognition and behavior. Underwater acoustic recordings containing dolphin vocalizations are used to obtain whistle spectrums. On the short time Fourier transformed (STFT) spectrum image, a frequency-time representation known as 'whistle contour' is usually used to classify and recognize dolphins. In an experiment planned by the Marine Mammal Research Laboratory at the National University of Singapore, indo-pacific humpback dolphins (*Sousa chinesis*) will be trained to pair whistles with objects or actions. These dolphins will response and mimic the template whistles sent by trainer. Te level of similarity in whistle structure of the dolphin response to the template whistle has to be measured automatically.

Whistle recordings are contaminated with many kinds of background noises, such as snapping shrimp in the habitat, mechanical noises from boat, etc. Only fundamental frequencies are considered as the principle information as the information in the harmonics is redundant. Malawaarachchi et al. [1] has developed an automatic technique to remove unwanted noises, suppress harmonics, segment and trace whistle contours with good results. However the performance cannot be guaranteed with large scale analysis of whistle clips, where we do not have enough detailed information on the background noise and the whistle intensity, and therefore cannot tune the parameters accurately. Previous work [2] in whistle classification assumes whistle traces of high quality. The well-known MaCowan's N-points algorithm [3] samples 20 evenly distributed points along whistle contour as feature vector. In automatic tracing and classification, outliers and breaks may appear and make this approach error-prone.

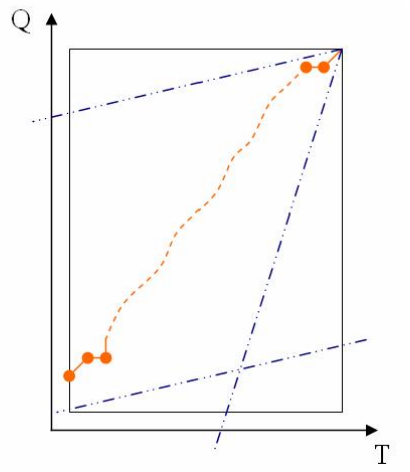
Whistle matching by visual inspection typically focuses on the general structure. Whistles may slightly vary locally like the variation of speech speed, but that does not affect the overall structure. Sampling points with an even distribution may not be the best matching criterion. Given the concern of imperfections in whistle extraction, dynamic time warping (DTW) is more suitable for matching whistles in non-linear time domain, looking for a match with minimum accumulated difference. This template matching algorithm presented in this paper uses whistle traces directly from the automatic traces generated by Malawaarachchi algorithm [1] with a certain degree of tolerance of tracing errors. Empirical tracing mistakes and non-standardizing by human tracing and curve fitting errors can also be avoided.

The template whistles are synthesized artificially in the dolphin research. An adaptive endpoint constraint is proposed to modify the basic DTW algorithm. A comparison experiment between modified and basic DTW was carried out to prove a much better matching for automatic whistle analysis tool.

978-1-4244-2126-8/08/$25.00 ©2008 IEEE



(a) : Difference is accumulated from the minimum of previous 3 by *0°-45°-90° warping*.



(b) Global warping path constraints – 'anchored beginning and free ending'.

Fig. 1. Cost Matrix Calculation.

II. RELATED WORK

Previously used for speech recognition, DTW is known for its tolerance to time shifting and partial variation. This works well for the same sentences spoken with different speeds. Some classification experiments have adopted basic DTW for killer whales [4] [5]. There are mainly two calculation steps – difference matrix, and minimum distance on cost matrix.

A. *Difference Matrix* $D[i, j]$

Whistle traces – the frequency pixel series from each time bin, are used as 1-D feature vector. The difference matrix D records the Euclidean distance between elements i and j from the query and template whistles respectively.

$$D[i, j] = \left| Q(i) - T(j) \right|^2 \qquad (1)$$

B. *Cost Matrix* $M[i, j]$ *and Minimum Distance*

A weighted sum of distances on the warped matching M is constructed as below, where k is the matched pairs at $D[i, j]$.

$$E(M) = \sum_{k=1}^{K} \omega_k D_k \qquad (2)$$

The cost matrix $M[i, j]$ records the matching difference up to i^{th} and j^{th} elements pair. During dynamic programming [5], a running tab is kept on the pair difference while adding up to a minimum accumulated cost measure. The simplest and most straightforward algorithm named by Ellis method [7] uses all unit weights and obtains the path from the minimum of three previously determined element differences.

$$M[i, j] = \min \begin{pmatrix} M[i-1.j-1] \\ M[i-1, j] \\ M[i, j-1] \end{pmatrix} + D[i, j] \qquad (3)$$

We call it *0°-45°-90° warping* shown in Figure 1(a). The cost path starts from the pair of last elements from the two sequences, i.e. the right top of different matrix D. Hence the allowable region of time warping path is constrained in Figure 1(b) with anchored beginning but free ending.

Warping can be altered to suit application purpose. For example, Sakoe and Chiba [6] altered the method and use a more complex and weighted sum as the cost function. This constrains the warping area accordingly.

III. MODIFIED DTW AND METHOD

A. *Whistle Extraction*

Large set of whistle clips may have different background and whistle intensity; therefore classification cannot directly implement on original signals. A short-time Fourier transform (STFT) interprets data into spectrogram. The 2D intensity image is automatically processed by Malawaarachichi et al. [1], removing commonly known noises such as mechanical noise and snapping shrimp. Whistle harmonics are suppressed and a transient suppress filter is used to segment whistle contour from the background, resulting in pixel sequences in terms of frequency variation over time.

However, due to occasionally low SNR or unknown conditions and image quantization, whistle traces may have few outliers beyond both whistle bandwidth and time duration (Figure 2a). Visual breaks are often seen at steep slope of whistle contour due to limited quantization of spectrum image (Figure 2b). The green crosses are traced points on the spectrograms. We can see a significant break in middle of whistle traces from #17. Our classification application will simply picks up these sequences as feature vector for template

978-1-4244-2126-8/08/$25.00 ©2008 IEEE

matching.

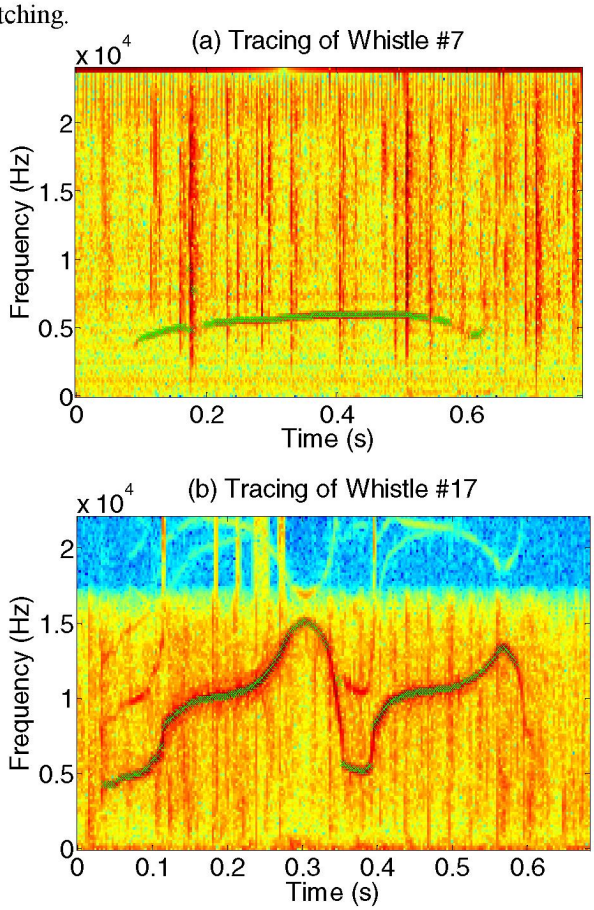


Fig. 2. Automatic tracing (*x*) with breaks and outliers: (a) Outliers beyond bandwidth near 0.2s; (b) breaks due to low SNR and limited image quantization.

B. *Tracing Accuracy*

Before template matching, we define a criterion called 'Tracing Accuracy' for selecting proper traces generated by the automatic tracing. In our research, the 'outliers' describing tracing errors are defined as, either

1) 'Data values that have a low likelihood of being consistent with the rest' or,

2) 'Data points that are far from the main body in time domain'.

The first type of outliers usually has a high standard deviation from the mean. As shown in Figure 3, most outliers circled are far from the whistle contour in frequency. Besides this some outlier frequencies might be consistent with the main body, occurring before and after whistles. This makes outlier detection difficult to decide in presence of breaks within whistle contour. One measure of tracing accuracy records the percentage of outliers and breaks, compared with common agreed manual traces. Normalized root mean squared error (RMSE) evaluates the tracing error compared with '*spline*' interpolated manual traces. The tracing error between the auto-traces and reference is measured at the time instances at the tracing points that the former has. Scaling factor '***bw***' is whistle bandwidth.

$$error = \sqrt{\frac{\sum_{t_1}^{t_{end}} |f_R - f_A|^2}{n}} / bw \qquad (4)$$

The bounding $[t_1 \quad t_{end}]$ is defined by reference time duration and ***n*** is the number of total sampling instances.

Figure 3 shows the comparison between auto-traces and reference, where outliers are circled in blue. 'Missed' and 'extra' are counted and normalized by total number of tracing points. The tracing accuracies of 18 whistles for experiment are shown in Table I. Most of them have tracing accuracy below 0.1.

This measurement describes tracing performance in general. Detailed situation such as the difference between template whistles, affection of local noises affect the matching as well.

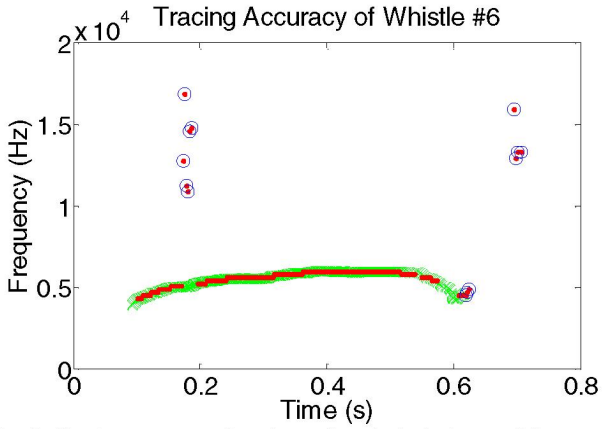


Fig. 3. Tracing accuracy plot: the outliers in both time and frequency are circled in blue (• – auto-traces, *x* – manual traces as reference).

C. *Modified DTW*

Basic DTW is modified in terms of its cost matrix and minimum path calculation. The cost matrix is constructed by a modified warping in 5 directions - 0°-27°-45°-63°-90° warping shown in Equation 5 and Figure 4. This combines and modifies methods of Ellis [7] and Sakoe [6], allowing one-to-many mapping for slightly different length matching and local variations. At the same time, single frequency outliers can be ignored by the 27°-63° direction without over-warping.

$$M[i,j] = \min \begin{pmatrix} M[i-1,j-1] \\ M[i-2,j-1] \\ M[i-1,j-2] \\ M[i,j-1] \\ M[i-1,j] \end{pmatrix} + D[i,j] \qquad (5)$$

We denote mapping between point ***p*** on template of length ***N*** and ***q*** on query whistle of length ***M*** in Equation 6.

978-1-4244-2126-8/08/$25.00 ©2008 IEEE

$$q = \omega(p) \qquad (6)$$

For both starting and ending of the matching path, we use an UE2-1 (unconstrained endpoints, 2-to-1 range of slope) method [6] and further define ourselves a dynamic parameter δ for adaptive endpoint matching.

$$\delta = M/12 \qquad (7)$$

The adaptive boundary condition is illustrated in Equation 8 in our experiment.

$$\begin{cases} 1 \le \omega(1) \le \delta + 1 \\ N - \delta \le \omega(M) \le N \\ \min[\omega(p) = 1] \quad 1 \le p \le 1 + 2\delta \\ \max[\omega(p) = N] \quad M - 2\delta \le p \le M \end{cases} \qquad (8)$$

Rather than starting from the pair of ending points in basic DTW, the minimum pair difference is selected in the range colored at right bottom in Figure 4. The back-trace start for minimum cost path is constrained to a range to avoid partial matching.

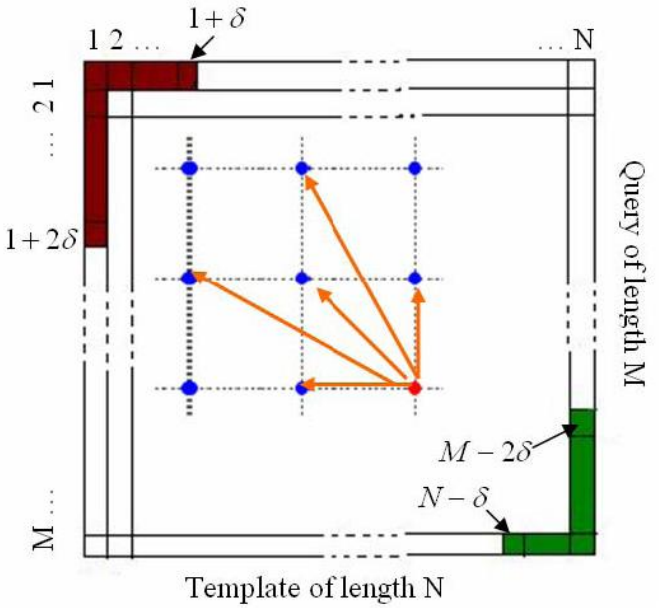


Fig. 4. UE2-1 DTW with 0°-27°-45°-63°-90° warping for calculation of cost matrix: the elements in green are initial pair selection area and red are ending area. Template sequence is of length N while query sequence is M.

D. *Whistle Template Matching*

To be frequency invariant for the frequency-modulated (FM) dolphin signals, whistle traces are shifted by their median frequency. The use of median over the mean is driven by the consideration of robustness to outliers. Changes to bandwidth due to outliers make frequency scaling impossible. The shifting range and step is adjustable for suitable capacity of computation complexity.

UE2-1 does not restrain the beginning or ending and hence the matching length is dynamic. The accumulated difference is scaled by N/N_s, where N_s is the matched length and N is the length of template whistle.

IV. EXPERIMENTS AND COMPARISON

Three methods are implemented for comparison and our modified DTW shows advantages in dynamic matching. It gives a smaller difference when pairing query whistles against the correct template. It also shows a larger differentiation ability in matching decisions.

Let the difference of query whistle from the correct template be d_D and from other template be d_o. The differentiation ability is defined in Equation 8 and is always positive.

$$diff = \frac{d_o - d_D}{d_D} \qquad (8)$$

Larger differentiation ability indicates easier decision on selecting matching template.

The three matching methods are:

1) *McCowan's N-point Feature Vector:* We use *N* as 20 and sample them along whistle contour.
2) *Basic DTW*
3) *Our Modified DTW*

In presence of breaks or time domain outliers, 20-point feature vector cannot evenly distributed on the actual whistle contour. This descriptor therefore cannot give nice matching. Therefore I only compare the performance of basic DTW and our modified one.

18 query whistles (Figure 5) are to be matched to 5 templates (Figure 6). Before matching, Euclidean distance is used for element difference in different matrix. Basic DTW mismatches one query (Whistle #1) and has larger pair difference values for the correct matching (Table II). Our modified DTW gives better matching in both pair difference and differentiation ability.

Figure 7 shows one example of the tolerance to outliers in both frequency and time from query whistle 1. This whistle was matched incorrectly by the basic DTW. By ignoring the single outlier inside the traced whistle, the modified DTW improves the matching performance. Figure 8 shows the matched path on the cost matrix as a somehow diagonal curve and accumulated difference as image intensity. We can see that modified DTW gives a better matching.

978-1-4244-2126-8/08/$25.00 ©2008 IEEE

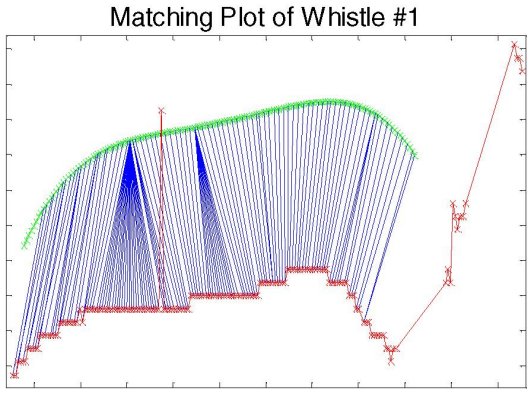
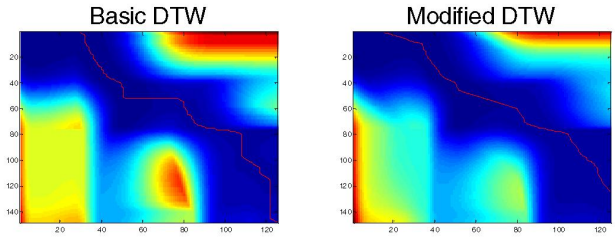
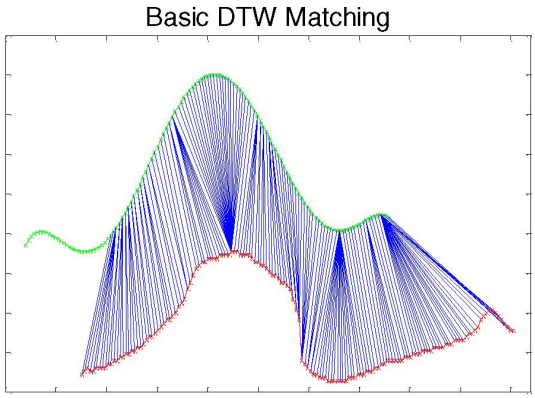


Fig. 7. Modified DTW matching: x (top curve) indicates template while x (bottom curve) are query traces. Point-to-point match is connected by blue line.





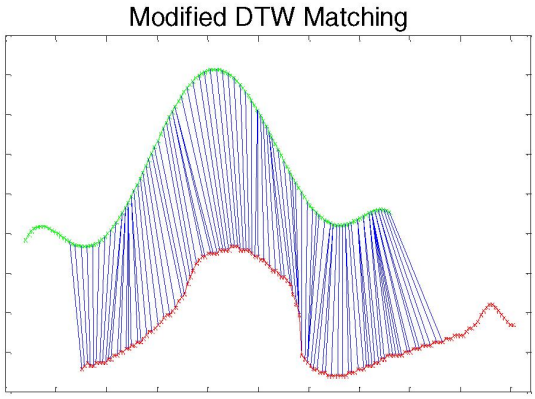


Fig. 8. Dynamic unconstrained endpoints for matching path: elements in green are selection range for minimum difference starting pair, while elements in red are ending range for selection. Difference is accumulated with 0°-27°-45°-63°-90° warping.

TABLE II
COMPARISON BETWEEN BASIC AND MODIFIED DTW

Query Whistle #	Basic DTW		Modified DTW	
1	3	0.0409	1	0.0002
2	1	0.0047	1	0.0002
3	2	1.4972	2	0.0017
4	3	0.0032	3	0.0004
5	4	0.0010	4	0.0004
6	1	0.0319	1	0.0067
7	4	0.0505	4	0.0038
8	2	0.1163	2	0.0483
9	4	0.0129	4	0.0002
10	5	0.0049	5	0.0021
11	5	0.0865	5	0.0463
12	5	0.1141	5	0.0522
13	5	0.0386	5	0.0110
14	3	0.0007	3	0.0003
15	1	0.2389	1	0.0003
16	5	0.0642	5	0.0375
17	5	0.2856	5	0.0663
18	1	0.0604	1	0.0004

Each method has the matching number in first column and pair difference (10^9) in second column.

An overall larger differentiation ability of our modified DTW for discriminating template matching compared with basic DTW is shown in Figure 9.

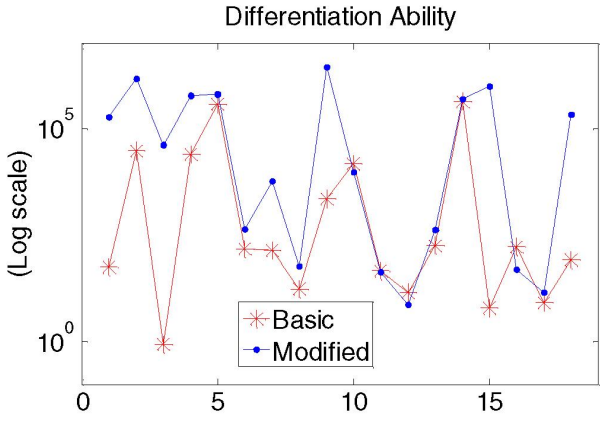


Fig. 9. Differentiation ability in log scale: the red curve is basic DTW and blue is our modified DTW for 18 query whistles.

V. CONCLUSIONS AND FUTURE WORK

This paper introduced an improved DTW. The matching performance is improved in both pair difference and differentiation ability. This allows automated classification following automated tracing and hence gives a practical real-time dolphin whistle extraction and classification tool. In the next phase of our research, this method will be used for large scale analysis of dolphin whistle processing.

978-1-4244-2126-8/08/$25.00 ©2008 IEEE

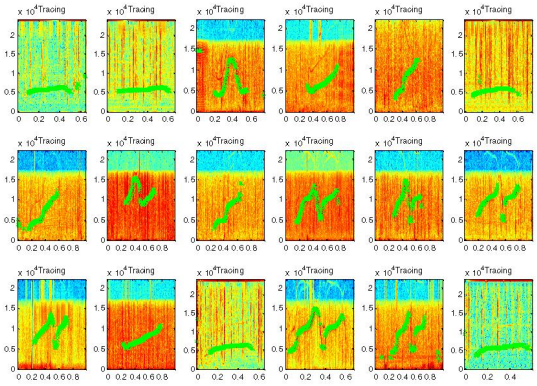


Fig. 5. 18 Query Whistle Traces

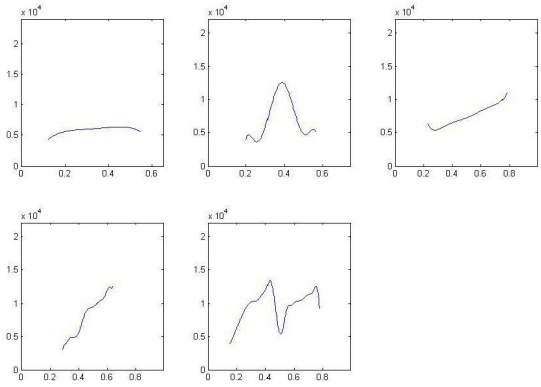


Fig. 6. 5 Template Whistle Traces

Whistle #	1	2	3	4	5	6	7	8	9
Error	0.16034	0.90598	0.01455	0.00614	0.00826	0.76409	0.00610	0.00940	0.01406
Missed	0	0	0	0.0204	0	0.0546	0.0158	0.0273	0
Extra	0.0646	0.0068	0.0667	0.0117	0.0607	0.0355	0.0290	0.0030	0.0103
Whistle #	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Error	0.03124	0.05493	0.00935	0.07712	0.01093	0,01921	0.03161	0.03798	0.02185
Missed	0	0.0097	0.0411	0	0.0145	0.0130	0.0170	0.0353	0.0409
Extra	0.0369	0.0195	0.0027	0.0233	0.0073	0.0487	0.049	0.1314	0.0446

Table I. Tracing Accuracy of 18 Query Whistles

REFERENCES

[1] Asitha Mallawaaarachchi, S.H. Ong, Mandar Chitre and Elizabeth Taylor, A Method for Tracing Dolphin Whistles, Oceans 06 Asia Pacific IEEE, 2006

[2] Nanayakkara Suranga Chandima, Mandar Chitre, S.H. Ong, Elizabeth Taylor. Automatic Classification of Whistles Produced by Indo-Pacific Humpback Dolphins (*Sousa chinensis*), Oceans 2007 Europe, pp1-5.

[3] McCowan, B. 2995: A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncates*). Ethology 100, 177-193.

[4] Judith C. Brown, Patrick J.O. Miller, Automatic Classification of Killer Whale Vocalizations Using Dynamic Time Warping, Acoustic Society of America, pp. 1201-1207

[5] Judith C. Brown, Andrea Hodgins-Davis and Patrick J.O. Miller, Classification of Vocalization of Killer Whales Using Dynamic Time Warping

[6] Hiroaki Sakoe and Seibi Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE* Trans. *Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, No. 1, Feb. 1978.

[7] D. Ellis (2003). Dynamic Time Warping (DTW) in Matlab. Available: http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/.

[8] Lawrence R. Rabiner, Aaron E. Rosenberg, and Stephen E. Levnson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition," *IEEE* Trans. *Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, No. 6, Dec. 1978.

978-1-4244-2126-8/08/$25.00 ©2008 IEEE