Automatic Classification of Whistles Produced by Indo-Pacific Humpback Dolphins (*Sousa chinensis*)

Suranga Chandima Nanayakkara*, Mandar Chitre † S.H. Ong‡ and Elizabeth Taylor*

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

‡Department of Electrical and Computer Engineering, Division of Bioengineering
National University of Singapore, 9 Engineering Drive 1,
Singapore 117576
eleongsh@nus.edu.sg

***Abstract*—A fast, robust technique is needed to facilitate studies of vocalisations by dolphins and other marine mammals such as whales in which large quantities of acoustic data are commonly generated. It is sometimes necessary to be able to describe whistle contours quantitatively, rather than simply looking at descriptors such as start frequency, maximum frequency, number of inflection points, etc. This is important when whistles are to be compared using an automated classification system, and is an essential component of a real-time, automated classification system for use with a raw data stream. In this paper we describe a rapid and robust high order polynomial curve fitting technique which extracts features in preparation for automated classification. We applied this method to classify natural vocalizations of Indo-Pacific humpback dolphins (*Sousa chinensis*). We believe the method will be widely applicable to bioacoustic studies involving FM acoustic signals in both underwater and in-air environments.**

***Index Terms*—Feature extraction, Whistle classification**

I. Introduction

Underwater acoustic recordings containing dolphin vocalisations are often analysed in the time-frequency domain using spectrograms. Spectrogram feature extraction techniques are widely used in whistle classification studies because they provide a visual representation of the whistle's frequency variation over time, and feature vectors for whistles can be extracted from their spectrogram images [1]. In order to investigate the way in which information might be encoded by dolphins in their natural FM (Frequency Modulated) whistles, it is necessary to identify and systematically describe small variations among whistles that superficially sound similar. One way to do this is by clustering whistles into groups containing whistles with similar structures. Clustering approaches attempt to partition data into several sub-groups or clusters where clusters represent groups of data with a high degree of internal similarity that might provide insights into the structure of the data. The level of internal similarity deemed meaningful to either dolphins or researchers is critical because whistle spectrograms that look superficially similar or identical to the human observer might contain information within small frequency modulated variations that could be overlooked as unimportant variations on a theme. For example, this might be particularly important in studies of dolphin vocalisations described as 'signature whistles'.

As part of the on-going dolphin communication work at the Marine Mammal Research Laboratory, NUS (National University of Singapore), we have made a large number of recordings from Indo-Pacific humpback dolphins (*Sousa chinensis*) kept by Underwater World Singapore Pte. Ltd. at Dolphin Lagoon, Sentosa. The duration of these recordings was between 20 to 40 minutes and the number of whistles they contained varied from about 100 to 400 whistles per recording. In this paper, we have proposed a method for classifying these whistles into their natural groupings.

II. Related Work

Studies of the vocalisations produced by small cetaceans have mainly focused on the Bottlenose dolphin (*Tusiops truncatus*) and these include many attempts to cluster bottlenose dolphin whistles using qualitative methods (human observers) as well as quantitative methods [2], [3], [4], [5] and [6]. McCowan [2] proposed a quantitative method to categorize dolphin whistles using a technique based on contour similarity. She used 20 frequency measurements equally distributed over the whistle contour as the basic feature set and generated a set of correlation coefficients for each whistle and its similarity to every other whistle in a data set. This method is likely to be very sensitive to the presence of outliers and it's also time consuming because the points has to be manually determined. Janik [4] compared the above method with human observation in identifying whistle similarities, and found that human observers were still better at categorizing whistles than computers. In addition, Douaze et al. [7] have reported personal observations with bottlenose dolphins from different geographical locations. However, when a human observer is performing the classification, it is difficult to know the threshold for classification being used and difficult to replicate that threshold. Moreover, a human might miss a small parameter variation that could be important to the animal. Hence, we have used a high order polynomial curve fitting technique to analyze the dolphin whistles. As far as we aware, this is the first attempt to do such an analysis on natural vocalizations of Indo-Pacific Humpback dolphins.

III. Method

A. De-noising and Tracing

Tracing dolphin whistles is often difficult due to the presence of environmental noise which degrades the signal to noise ratio. In the present analysis, Snapping shrimp in the dolphin habitat produced fairly loud, broadband noise, a common characteristic of warm shallow water acoustics. An automated technique developed by Malawaarachchi et al. [1] was used to de-noise the recordings and trace the whistles. However, because of the presence of broadband Snapping shrimp noise and occasional low whistle intensity, it was not possible to create perfect tracings of all whistle contours recorded. However, at this stage, it was assumed that we had a perfect match between the actual spectrogram of a whistle and the traced points. The underlying idea was to avoid any error caused by mismatch between an actual and a traced whistle (in the feature extraction phase) and observe the error arising from the clustering algorithm alone. This would enable the selection of appropriate features and fine-tune the clustering algorithm. A perfect match was obtained by first, manually adding points over a small temporal scale to fill obvious gaps in the sound recording in order to complete the whistle, and then by removing any obvious outliers.

B. Feature Extraction

The shape of whistle contours in the time-frequency plane is widely regarded as an important characteristic of a dolphin whistle [4]. This contour is made up of a fundamental frequency, as well as harmonics. Harmonics are a frequency component of the fundamental frequency, and are integer multiples or fractions of the frequency of the carrier wave. In this study, we concentrated solely on the fundamental frequency of a contour. A feature vector was constructed using a shape descriptor, which was a polynomial fit of the traced whistle contour (i.e. fundamental frequency contour). However, our preliminary testing showed that simple polynomial fitting is unlikely to give good results since such polynomial fitting is very sensitive to shifts in time and frequency. Therefore, the following procedure was used to obtain a more robust polynomial fit which is invariant to time and frequency shifts.

First, a traced whistle contour (time, frequency points) is shifted and scaled so that the middle point of the whistle corresponds to time = 0 and duration of the whistle is [-1 1]. Then a standard polynomial fitting is performed on the scaled time, frequency points. A threshold was set to provide a compromise between the accuracy of the polynomial fit and the order of the polynomial, to avoid over-fitting or under-fitting the whistle contour. We allow the order of the polynomial fit to vary between 1 and 14 depending on the complexity of the whistle contour. The algorithm shown in Figure 1 is used to get a reasonable polynomial fit while maintaining the feature vector size constant.

$n = 1,\ \epsilon = \infty$
calculate $p_n(t)$
$\epsilon = \frac{1}{k}\sum_{i=0}^{i=k}(p_n(t) - y)^2$
while ($\epsilon < \epsilon_t$ or $n \leq n_{max}$)
  $n = n + 1$
  calculate $p_n(t)$
  $\epsilon = \frac{1}{k}\sum_{i=0}^{i=k}(p_n(t) - y)^2$
end while
make $p_n(t) = a_0 + a_1 t + \ldots a_{14} t^{14}$ by zero padding

Fig. 1. Pseudocode for polynomial fit, where ϵ is the mean square error (MSE) between the polynomial fit and actual whistle, n is the polynomial order, $p_n(t)$ is n^{th} order polynomial fit, k is the number of data points (time, frequency) in the traced whistle, y is the actual data point (frequency value) of the whistle corresponding to t and ϵ_t is 5% of the error introduced by a 15^{th} order polynomial fit.

It was observed that the above polynomial fit is very sensitive to shifts in time. (i.e., two whistles, that are similar in shape but slightly shifted in time could result in very different polynomial fits). In order to overcome this problem, we shifted the whistle contours along the time axis by δ so that the maximum frequency of each whistle is aligned at the time origin. The formula to obtain the polynomial fit coefficients of the shifted whistle contour from the original polynomial fit coefficients is

$$b_i = \sum_{j=i}^{N} \binom{j}{i} \delta^{j-i} a_j \tag{1}$$

where b_i is the i^{th} coefficient of shifted polynomial, N is the order of the original polynomial fit and a_j is the j^{th} coefficient of the un-shifted polynomial fit. Finally, the constant term of the polynomial fit (y-intercept) was removed in order to make it invariant to frequency shifts. The proposed polynomial fit is shown in Figure 2, for 3 selected whistle contours. Instead of just taking the coefficients of the polynomial fit, we augmented the features by taking the value of the polynomial fit at $N(= 100)$ points. This is very similar to the features used by McCowan [2]. However, we essentially remove the noise by fitting a polynomial and taking the values of it at N points whereas McCowan has simply taken the frequency measurements from the original whistle contour. In addition,

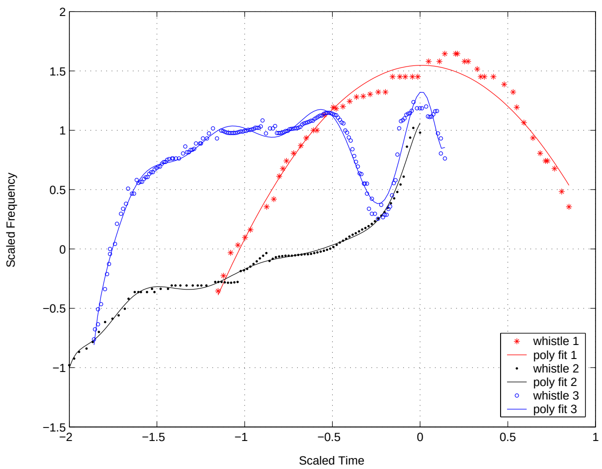


Fig. 2. Polynomial fit of three whistle contours after scaling and shifting the maximum point to time = 0

we generated another feature vector by taking the slope (the first derivative) of the polynomial fit at $N(= 100)$ points. Principal component analysis (PCA) was used to remove redundant information from the feature vectors and performance of the clustering was compared for the three different feature sets.

C. Clustering

The principle underlying clustering is to make sure that samples in the same cluster are in some specified ways more similar to each other than to samples in different clusters. One way to make this into a well-defined problem is to define a criterion function that measures the clustering quality of any partition of the data. In this paper, we have used the simplest and most widely used criterion function for clustering: the Sum-of-Squared Error (SSE) criterion [8]. The SSE criterion is defined by the total squared errors in representing the given set of data by the set of mean $\{\boldsymbol{m_1}, \ldots, \boldsymbol{m_k}\}$. Denoting SSE by J_e, it could be formulated as

$$J_e = \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in H_i} ||\boldsymbol{x} - \boldsymbol{m}_i||^2 \qquad (2)$$

where $\boldsymbol{x}$ represents a vector drawn from partition H_i and $\boldsymbol{m_i}$ is the mean vector of the partition H_i given by $\boldsymbol{m_i} = \frac{1}{n_i} \sum_{\boldsymbol{x} \in H_i} \boldsymbol{x}$, n_i is the number of samples in partition H_i. An optimal partition will minimize J_e, and the clustering obtained will be the best in SSE sense.

The form of J_e given in equation (2) gives an indication of the clustering quality in terms of compactness. However, its value depends on the total number of data feature vectors as well as the dimensionality of the feature vectors. It is desirable to normalize the J_e so that direct comparison is possible for data sets with different number of feature vectors and different dimensions. Therefore, J_e is normalized with respect to the total number of feature vectors and the data dimensions as follows:

$$\hat{J}_e = \frac{1}{d.\sum_i n_i} \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in H_i} ||\boldsymbol{x} - \boldsymbol{m}_i||^2 \qquad (3)$$

where d is the dimension of the feature vectors and $\sum_i n_i$ will give the total number of feature vectors in the data set.

To determine the appropriate number of clusters objectively, we use the percentage reduction, δ, in the cost function

$$\delta = \frac{\hat{J}_{e\ 1\ cluster} - \hat{J}_{e\ k\ clusters}}{\hat{J}_{e\ 1\ cluster}} \times 100 \qquad (4)$$

We allow 10% tolerance and set the value of δ as 90% to obtain the number of natural clusters.

IV. Results and Discussion

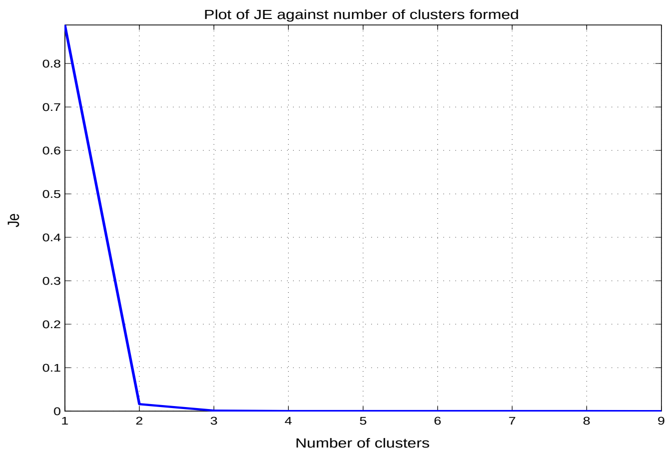
Nine sample whistles (recorded from Indo-Pacific humpback dolphins) were selected as the test cases to run the clustering algorithm in order to determine the best feature extraction method. These nine whistle samples include simple monotonically increasing contours as well as complex whistle contours with 1 or more turning points. We ran the proposed polynomial fitting technique on 9 test cases and obtained 3 different feature matrices as follows:

- Coefficients of the polynomial (without y-intercept)
- Value of the polynomial fit at 100 equally distributed points
- Slope of the polynomial fit at 100 equally distributed points

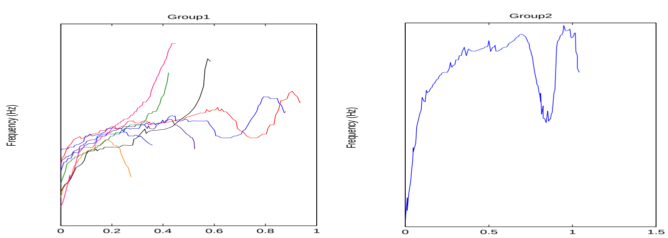
Results of the clustering algorithm for each of the above cases are shown in Figures 3, 4 and 5.

From the results shown in Figure 3, it is clear that the coefficient of the polynomial fit is not a good feature in clustering. This might be due to the fact that the polynomial coefficients are not unique. However, using the value of the polynomial fit at 100 points (evenly distributed along the time axis) leads to much better clustering as shown in Figure 4. This feature extraction method is similar to the method proposed by McCowan [2], in which she used 20 frequency measurements (evenly distributed along the time axis) as the feature vector and reported this is sufficient even to describe complex whistle contours with several turning points. We tried the proposed method using 20 frequency measurements as recommended, but the procedure produced the same results as when we used 100 points along the polynomial fit. By fitting a polynomial to the time-frequency contour and taking the values of polynomial fit at N points, we essentially remove the noise present in the traced whistle. In that sense, our proposed feature extraction method is more robust to noise compared to the feature extraction proposed in McCowan (1995) [2] and is also appropriate for use in a rapid automated classification system.

Results of clustering using the third feature extraction method show that it is quite sensitive to small variations in the whistle contour. In fact, group 2 of Figure 4(b) (when

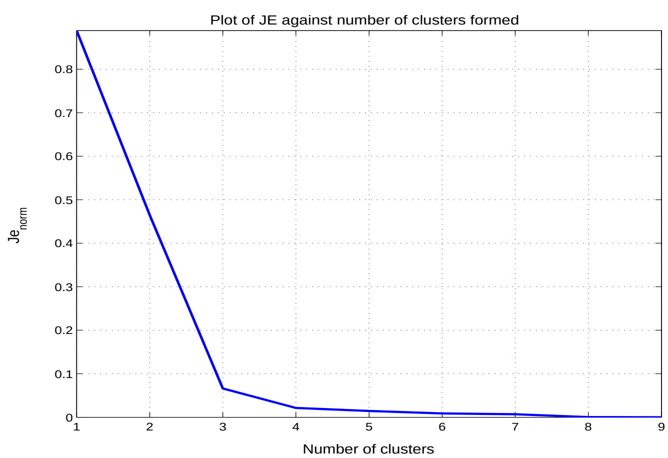


(a)

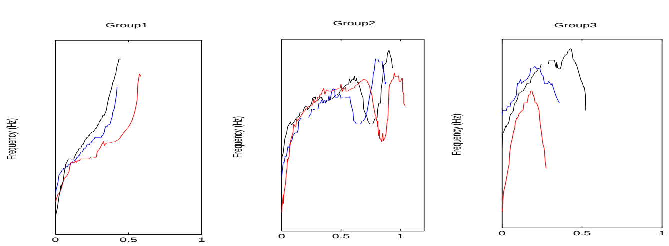


(b)

Fig. 3. Results of the clustering using coefficient of the polynomial fit as the features. (a) Error vs number of clusters. (b) Resulting clusters with $\delta = 90\%$.

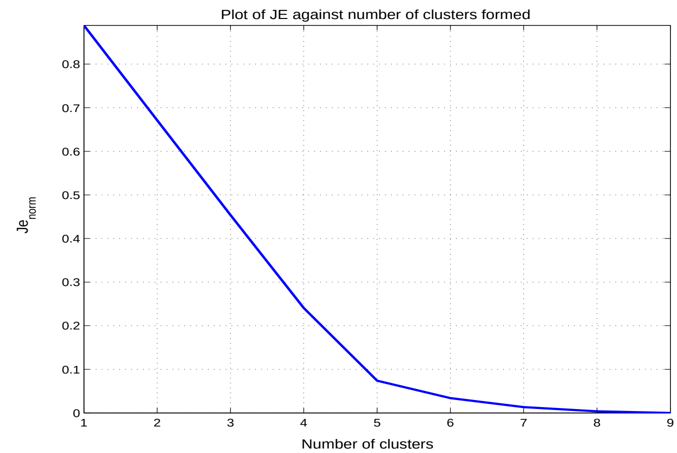


(a)

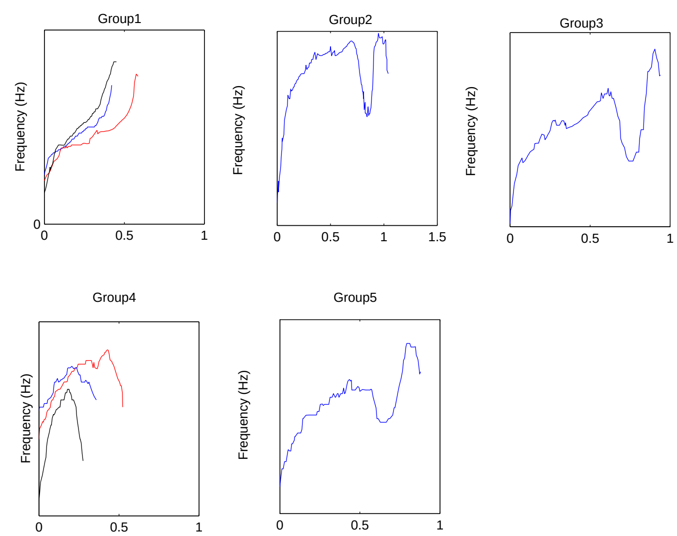


(b)

Fig. 4. Results of the clustering using value of the polynomial fit at 100 evenly spaced points as the features. (a) Error vs number of clusters. (b) Resulting clusters with $\delta = 90\%$.



(a)



(b)

Fig. 5. Results of the clustering using slope of the polynomial fit at 100 evenly spaced points as the features. (a) Error vs number of clusters. (b) Resulting clusters with $\delta = 90\%$.

using only the value of the polynomial fit) are separated in to three groups in Figure 5(b) (when using the slope of the polynomial fit). The reason for this is that although whistles in group 2 of Figure 4(b) look similar in general, they possess small variations in terms of the gradients at different points.

Comparing the clustering performance using three different feature sets, it is clear that the value of a polynomial fit at $N(= 100)$ evenly distributed points is an appropriate feature to use when the objective is to cluster whistles according to their general shape.

V. Conclusion and Future Work

This paper has introduced an algorithm, based on high order curve fitting techniques to extract features from dolphin whistle contours. Three different feature extraction methods are discussed in detail. In the next phase of our work, the proposed method will be used to classify a large number of Indo-Pacific Humpback dolphin whistles into their natural groupings. In addition, this classification technique will be combined with an automatic whistle tracing proposed by Mallaawarachchi et al. [1] to automatically classify the dolphin whistles in underwater recordings. The technique presented in this paper is likely to prove useful to researchers in this area and could be equally well be applied to other animal vocalizations, such as bird songs.

VI. Acknowledgement

We are grateful to Underwater World Singapore Pte. Ltd. for providing the opportunity to work with their dolphins, and thank the other Marine Mammal Research Laboratory team members working on dolphin communication and cognition for their help in recording the dolphin whistles used in this study.

References

[1] A. Mallawaarachchi, S. H. Ong, M. A. Chitre, and E. Taylor, "A method of tracing dolphin whistles," *OCEANS'06 Asia Pacific IEEE Conferencer*, 2006.

[2] B. McCowan, "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (delphinidae, tursiops truncatus," *Ethology*, vol. 100, pp. 177–19, 1995.

[3] B. McCowan and D. Reiss, "The fallacy of signature whistles in bottlenose dolphins: a comparative perspective of signature information in animal vocalizations," *Animal Behaviour*, 2001.

[4] V. M. Janik, "Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods," *Animal Behaviour*, 1999.

[5] V. B. Deecke and V. M. Janik, "Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, p. 645653, 2006.

[6] J. R. Buck and P. L. Tyack, "A quantitative measure of similarity for tursiops truncatus signature whistles," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2497–2506, 1993.

[7] E. Douaze, L. Ong, M. dos Santos, K. Schultz, and E. A. Taylor, "Description and comparison of geographically isolated bottlenose dolphin (tursiops truncatus) whistles," in *Proceedings of the 25th European Association for Aquatic Mammals Annual Symposium*, Duisburg Germany, March 1997.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-interscience Publication, 2000.