Acoustics'08
Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Unsupervized clustering of whistle-sounds of bottlenose dolphins (Tursiops truncatus)

Sebastian Huebner

Pestalozzistrasse 5, 14482 Potsdam, Germany
sebastian@sejona.de

An unsupervised symbolic clustering algorithm for whistles of bottlenose dolphins (*Tursiops truncatus*) was tuned on a small training data set. Afterwards, recordings containing a large number of natural whistles and a test data set containing 1520 instances of twelve different artificial whistle types were analyzed by the algorithm. Results are discussed with regard to the signature whistle hypothesis.

1 Introduction

The signature whistle hypothesis was first stated by Caldwell & Caldwell (1965) and is based upon two conceptions: (1) In the vocal repertoire of a group of bottlenose dolphins whistle types exist which are characterizable by a stable time-frequency contour-shape. (2) Each individual animal can be attributed exactly one such whistle type - it's *signature whistle*. The signature whistles of a group of animals are disjunct and about 90% of all whistles may be classified as signature whistles.

The hypothesis was supported by many later studies: Mimicry of signature whistles can be observed only in rare cases [13, 6, 12, 8]. In captivity calves coin their individual signature whistle within the first six months of live [1]. In nature the ontogeny of signature whistles takes as long as one year and whistles remain stable for up to 12 years [9]. Signature whistles of male calves are more similar to their mother's ones than those of female calves who remain members of the group [9, 10].

Newer findings suggest that whistles may be understood as contact barks ensuring cohesion of the group [6, 11, 5]. Also they may play an important role in the mother-calf relationship [12] and time-frequency characteristics of whistles appear to be correlated to stimuli-changes in discrimination tasks [4]. Furthermore, some authors claim that each individual animal has it's own whistle–repertoire [7].

These and other aspects of whistles have been discussed intensively in the literature, but the "contour-shape" conception still lacks proper specification. This, however, is fundamental for the discourse about the signature whistle hypothesis.

Goal of this study is to find and investigate the most frequent whistle types in the repertoire of a group of bottlenose dolphins that are subject to a well defined time-frequency contour-shape criterion. Furthermore, it is investigated if these types can be understood as signature whistles. In order to ensure a neutral treatment of the conception of contour-shape an unsupervised machine learning approach was chosen.

2 Material and Method

2.1 Material

The acoustic material used in this study is a large collection (87,9 GB) of hard disk recordings of the vocalizations of *Tursiops truncatus*. It was made in early summer 2002 during 28 days by the author at the Dolphin-Reef in Eilat (Israel). The corpus comprises 273,3 hours of recordings in 453 files. All files are mono, uncompressed 16 Bit PCM with a sample-rate of 96kHz. A broadband hydrophone (TC 4014, Reson) was used together with an external soundcard (DSP24, Hoontech). In summer 2002, about thirteen adult animals and three calves lived in the group. One of the adult animals was absent for most of the time. The hydrophone was located six meters away from the sandy shore. About 90% of the recordings were made from 10 a.m. until 10 p.m. In the following sections, the collection will be reffered to as the Eilat-Collection. It contains about 25.000 whistles.

2.2 Method

In order to infer from the Eilat-Collection those whistle types that are subject to a neutral time-frequency contour-shape criterion the following procedure was applied:

1. A set of twelve artificial "signature" whistles was generated with a synthesizer. Each of these whistles was between 570 ms and 980 ms long. Each whistle contained ten arbitrary changes in the time-frequency contour. Length, bandwidth and changes in the contour were determined by random procedures. Artificial whistle had no harmonics but in all other aspects resembled natural dolphin whistles (see figure 1).
2. A test data set containing 1520 representations of artificial whistles (152 of each) was generated by mixing whistle-free recordings from the Eilat group with instances of artificial whistles. These were inserted with different levels of quality (sometimes partially instantiated, overlapping and low-level) at random points in order to model true recordings of natural whistles.
3. An unsupervised cluster induction algorithm [3, 2] was tuned on a small set of training data containing both artificial and natural whistles. The algorithm induces clusters, each representing one whistle type with a stable time-frequency contour-shape. In a nutshell, the algorithm works as follows:
    (a) A set of annotations is generated by unsupervised classification of audio recordings with the help of a number of classifiers that respond to short sinusoidal wave patterns in different frequency bands. Classifiers are designed to form a raster that covers the full bandwidth of all whistles.
    (b) A set of well-formed chains of annotations neighboring in time-frequency space is created from the set of all annotations.
    (c) A set of symbolic representations of whistles is automatically extracted from the set of chains. Time of all representations is normalized. Each representation is now a time series of symbolic elements.

(d) From all symbolic representations those are chosen that have a minimum length of 500 ms. The obtained set contains only well formed representations of whistles long enough to be signature whistles.

(e) The set of well formed representations is clustered by making use of a formal symbolic-similarity criterion. Symbolic-similarity is defined as the maximum number of common elements in two symbolic representations. If symbolic-similarity is above threshold two representations are merged into a new representation.

(f) Each merged representation is used for further comparisons and merges within an iterative procedure. The algorithm stops when no more merges are possible.

(g) Each resulting symbolic representation specifies exactly one cluster (i.e. one whistle-type). For each cluster an overall compression rate is computed from local compression rates of symbolic elements.

The clustering algorithm was applied to the Eilat-Collection and to the test data set. The similarity threshold was set to 80%. From resulting clusters those with a compression rate above 4 were selected. Bivariate plots of center frequency vs. bandwidth and of center frequency vs. gradient (in the time-frequency space) were computed from the set of all induced clusters and from the set of clusters with high compression rate.

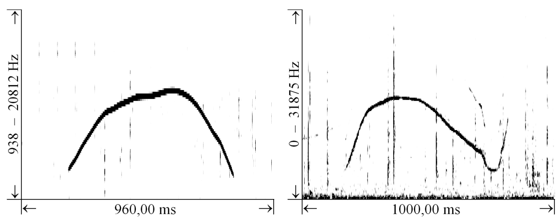

Figure 1: Artificial whistle (l) and natural whistle (r)

Rationale: The test data set with the twelve artificial whistles represents an ideal model of an acoustic environment in which the signature whistle hypothesis holds. All artificial whistles have a distinct fine grained time-frequency contour-shape which is blurred by natural noise, sonar clicks, mutual overlaps, partial instantiation and low signal-to-noise ratio. An algorithm that is able to correctly induce clusters for all artificial signature whistle types from such data should be able to do so for natural signature whistles as well. If the clustering algorithm performs correctly on the test data set, then the output obtained from clustering the entire Eilat-Collection can be interpreted only in three reasonable ways:

1. All presumed signature whistles are among the most frequent induced whistle types. The conception of contour-shape used to generate artificial whistles is correct. The signature whistle hypothesis holds.
2. Not all presumed signature whistles are among the most frequent induced whistle types. The conception of contour-shape used to generate artificial whistles is nevertheless correct. The signature whistle hypothesis does not hold.
3. Not all presumed signature whistles are among the most frequent induced whistle types *because* the conception of contour-shape used to generate artificial whistles is not correct.

All three possible outcomes of the experiment are relevant for the signature whistle hypothesis itself. In the first two cases, the signature whistle hypothesis either holds or fails. In the third case, it has to be concluded that the conception of contour-shape used to generate artificial whistles is not correct - but then it must be asked how an appropriate contour-shape-conception can be specified.

3 Results

3.1 Eilat-Collection

Classifiers for short sinusoidal wave patterns created 984.033 annotations in total. From these 1.072 well formed symbolic representations of long whistles were extracted. This equals 4.3% of the estimated 25.000 whistles in the Eilat-Collection. Symbolic representations contained an average of 39 elements. 42.149 annotations (4.3%) were used to create symbolic representations of long whistles.

Six automatically extracted well-formed symbolic representations are shown in figure 2. These representations can be interpreted as proper representations of signature whistles of different animals of the group.

Bivariate plots of center frequency vs. bandwidth and of center frequency vs. gradient of all found symbolic whistle-representations and of all generated symbolic cluster identifiers are shown in figure 3. The clustering algorithm generated 52 clusters with a compression rate higher than 4. The twelve most dense clusters could be interpreted as variations of only two basic whistle forms (see figure 3). These basic forms appear to be predominant in the vocal repertoire of the group.

3.2 Test data

Classifiers for short sinusoidal wave patterns created 97.458 annotations in total. From these 818 symbolic representations of whistles were extracted. This equals 54% of the 1520 instances of artificial whistles in the test data set. Symbolic representations contained an average of 47.9 symbolic elements (annotations). 39.160 annotations (about 40.2%) were used to create symbolic whistle-representations.

Bivariate plots of center frequency vs. bandwidth and of center frequency vs. gradient of all found symbolic whistle-representations and of all generated symbolic cluster identifiers are shown in figure 5. The clustering algorithm generated 35 clusters with a compression rate higher than 4. The twelve most dense clusters represented exactly all artificial whistle types in the test data set (see figure 6).

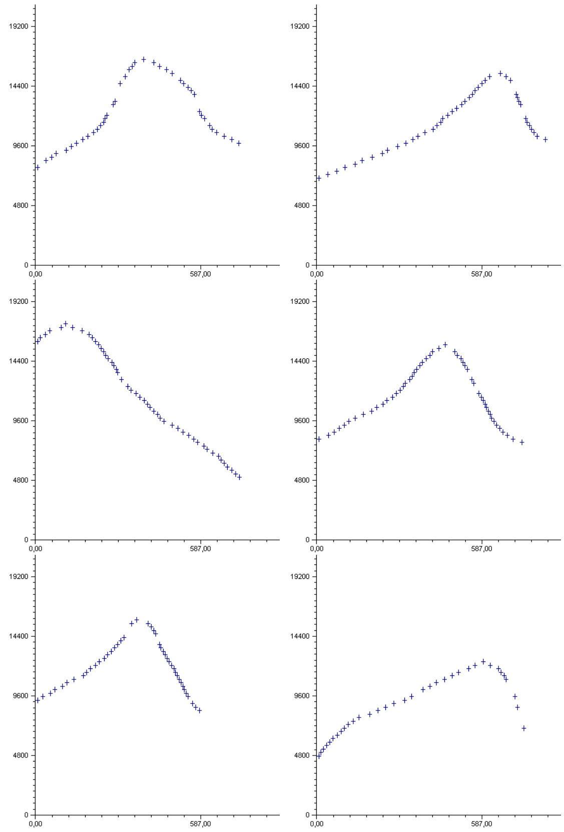

Figure 2: Six automatically extracted symbolic representations of natural whistles. Crosses represent symbolic elements.

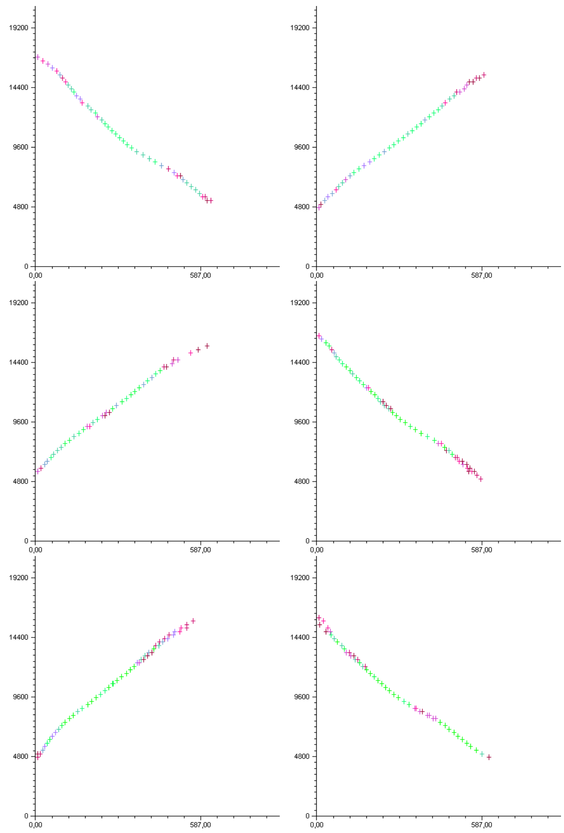

Figure 3: The six most frequent induced natural whistle types longer than 500ms. Green colors indicate high local compression rate. Red colors indicate low local compression rate.

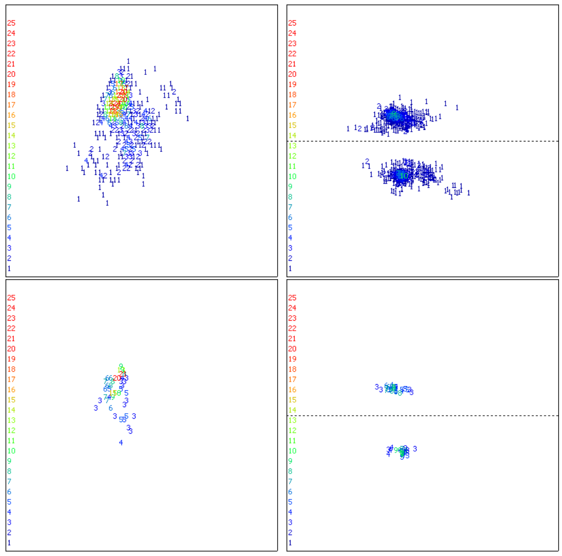

Figure 4: Bivariate plots of center frequency vs. bandwidth and of center frequency vs. gradient. Top: All 1072 long whistle representations. Bottom: The 52 cluster identifiers with a compression rate ≥ 4.

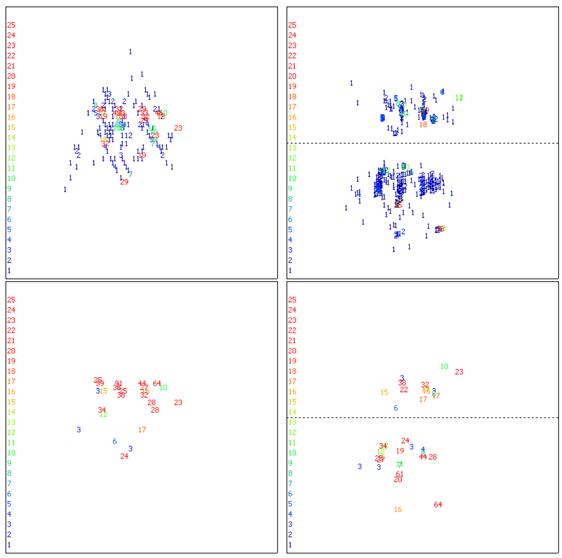

Figure 5: Bivariate plots of center frequency vs. bandwidth and of center frequency vs. gradient. Top: All 818 long whistle representations. Bottom: The 35 cluster identifiers with a compression rate ≥ 4.

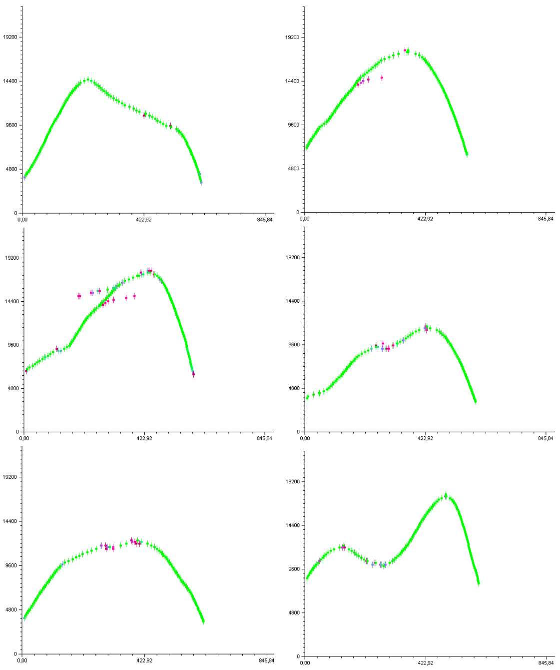

Figure 6: Symbolic identifiers of the six most frequent induced artificial whistle types. Green colors indicate high local compression rate. Red colors indicate low local compression rate.

4 Discussion

The Eilat-Collection contains about 25.000 recorded natural whistles. For most of these no symbolic representations were extracted by the algorithm. Main reasons were low signal-to-noise ratio, shortness, frequency jumps, interruptions, trills and other irregularities in the shape of the whistles. Extracted, however, was a sufficient number of full-length whistles with high signal-to-noise ratio. All presumed signature whistles of the group were present in the set of extracted symbolic representations.

Clustering almost perfectly reproduced all twelve artificial whistle types. The algorithm also built a comparable number of clusters for natural whistle types. Clusters formed from natural whistles, however, provide low support for the signature whistle hypothesis. Symbolic identifiers of the most frequent clusters look similar at first glance but are to different to be merged by the algorithm. The twelve most frequent induced cluster identifiers can be interpreted as variations of only two basic whistle forms predominant in the vocalizations of the group. These two basic forms are also clearly visible as two accumulations in the bivariate plot of center frequency vs. gradient (see figure 4). In addition, a variety of different but more rare types in automatically generated clusters could be observed.

It can be concluded that natural whistle types which fit the strict contour-shape conception used to generate artificial whistles do indeed exist. Most of these, however, cannot be interpreted as signature whistles. With regard to the three possible interpretations of the outcome of this experiment we can state that:

1. The first interpretation is not correct as only two of twelve presumed signature whistles are among the induced whistle types.
2. The second interpretation could be correct but then the signature whistle hypothesis is not correct.
3. The third interpretation could also be correct but then it must be asked what actually a "contour-shape" is. It has to be concluded that the signature whistle hypothesis lacks correct specification of one of it's two fundamental assumptions.

Lowering the threshold of the similarity criterion underlying cluster formation does not yield better results. The maximum number of dense clusters is achieved when the symbolic similarity threshold is set to 80% (see figure 7). Lowering the threshold yields low-quality clusters with symbolic representations that model whistles less appropriate.

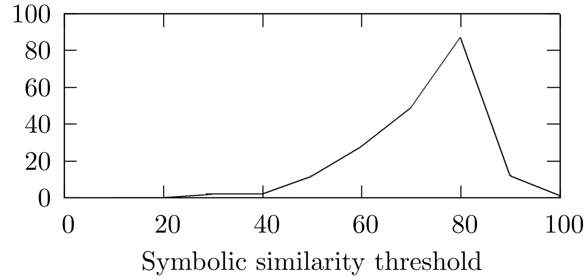

Figure 7: Number of clusters of natural long whistles with compression rate ≥ 2 in dependency of symbolic similarity threshold.

Changing the minimum length parameter in step (d) of the algorithm does also not yield results which are better for the signature whistle hypothesis. Manifold symbolic representations of short subsequences in whistles appear. These can be clustered with high overall compression rates. Interpretation of such clusters with statistical methods is up to future studies.

5 Acknowledgements

The author would like to acknowledge the support provided by Doug Saddy (Reading), Dietmar Todt (Berlin), Frank Veit (Berlin) and the crew at Dolphin Reef in Eilat / Israel.

References

[1] M. C. Caldwell and D. K. Caldwell. The whistle of the atlantic bottlenosed dolphin (tursiops truncatus): Ontogeny. In H. Winn and B. Olla, editors, *Behavior of Marine Animals*, pages 369–410. Plenum Press, New York, 1979.

[2] S. V. Huebner. *Wissensbasierte Modellierung von Audio-Signal-Klassifikatoren. Zur Bioakustik von Tursiops truncatus.* Dissertation, University of Potsdam, 2007.

[3] S. V. Huebner. *Wissensbasierte Modellierung von Klassifikatoren für Zeit-Frequenz-Muster in PCM-Daten.* Logos, Berlin, 2007.

[4] V. Janik, G. Dehnhardt, and D. Todt. Signature whistle variations in a bottlenosed dolphin, tursiops truncatus. *Behavioural Ecology and Sociobiology*, 35:243–248, 1994.

[5] V. Janik, L. S. Sayigh, and R. S. Wells. Signature whistle shape conveys identity information to bottlenose dolphins. *P Natl Acad Sci USA*, 103:8293–8297, 2006.

[6] V. Janik and P. Slater. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal Behaviour*, 56:829–838, 1998.

[7] B. McCowan and D. Reiss. Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (delphinidae, tursiops truncatus): A re-evaluation of the signature whistle hypothesis. *Ethology*, 100:194–209, 1995.

[8] D. Reiss and B. McCowan. Spontaneous vocal mimicry and production by bottlenose dolphins (tursiops truncatus): Evidence for vocal learning. *The Journal of Comparative Psychology*, 107(3):301–312, 1993.

[9] L. Sayigh, P. Tyack, and R. W. M. Scott. Signature whistles of free-ranging bottlenose dolphins, tursiops truncatus: mother-offspring comparisons. *Behavioural Ecology and Sociobiology*, 26(4):247–260, 1990.

[10] L. S. Sayigh, P. L. Tyack, R. S. Wells, M. D. Scott, and A. B. Irvine. Sex difference in signature whistle production of free-ranging bottlenose dolphins, tursiops truncatus. *Behavioral Ecology and Sociobiology*, 36:171–177, 1995.

[11] L. S. Sayigh, P. L. Tyack, R. S. Wells, M. D. Scott, and A. B. Irvine. Individual recognition in wild bottlenose dolphins: A field test using playback experiments. *Animal Behaviour*, 57:41–50, 1998.

[12] R. A. Smolker, J. Mann, and B. B. Smuts. Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants. *Behavioral Ecology and Sociobiology*, 33:393–402, 1993.

[13] P. L. Tyack. Whistle repertoires of two bottlenosed dolphins, tursiops truncatus: Mimicry of signature whistles? *Behav. Ecol. Sociobiol.*, 18:251257, 1986.