Estimating the number of sperm whale (*Physeter macrocephalus*) individuals based on grouping of corresponding clicks

Carlos de Obaldía Pastor, Gediminas Simkus, Udo Zölzer

Abstract

An automated Passive Acoustic Monitoring (PAM) method for the detection and differentiation of sperm whale individuals is proposed. Various methods benefit from the correlation of multi-channel recordings to identify active whales. However, the proposed approach employs audio recordings from a single hydrophone and uses a correspondence analysis to differentiate between active individuals. Segments of a click are obtained by thresholding the RMS envelope of the Teager-Kaiser energy operator's output. Possible click reflections are detected by an iterative comparison of extracted click-like segments with each other. Corresponding clicks are grouped by comparing extracted click features and the cross-correlation of sequential time-segments of the click segments with one another. Signal characteristics which describe time and spectral signatures of clicks are employed to derive rules for an effectual correspondence between the click segments. Corresponding clicks are then grouped to determine the number of active sperm whales, hence identifying plausible signatures for individuals. Results with real-world recorded signals confirm that such correspondences can be used for further applications. For instance, intra-species classification in monaural recordings.

Introduction

Anthropogenic noise contamination of underwater environments affect marine life and is a growing concern. Marine mammals are extremely affected, since sound is used underwater as a primary sense for communication and foraging. Changes in the acoustical environment of cetaceans can thus bring behavioral and health problems in populations and individuals.

Toothed whales use pulse-like sounds, or clicks, to perform echolocation tasks. Sperm whales (*Physeter macrocephalus*) produce such clicks using their acoustic organ, the spermaceti organ, which is located on their head and is filled with a substance with a lower density than water. Each of these clicks can be characterized by a multi-pulse structure [1, 2]. The multi-pulse structure is a result of the reflection of a generated pulse between the frontal and distal air sacs in the acoustic organ of a sperm whale. In [3] the authors are able to estimate the size of a sperm whale based on their click structure.

In order to reduce the emission of noise in underwater environments, Passive Acoustic Monitoring (PAM) techniques are used to detect the presence of marine mammals. If animals are present, noise sources should be mitigated. In this work, a PAM system for grouping

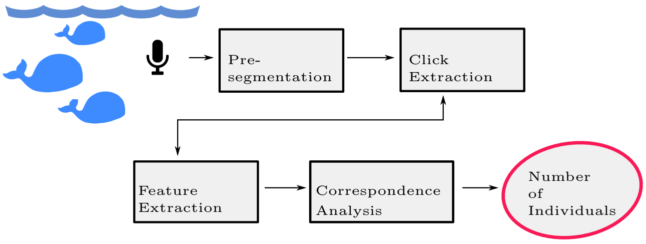


Figure 1: Proposed PAM system. The signal is pre-segmented in transient groups. Reflections are identified on this segments based on a comparison of features. For click correspondence, features for each click are used to collect the segments in groups of corresponding clicks for each individual in the waveform.

corresponding clicks of different active sperm whale individuals is presented. Estimation of the number of whales in the waveform is based on the number of individuals identified with the algorithm.

Proposed PAM System

Passive Acoustic Monitoring (PAM) in aquatic environments refers to the use of hydrophones to detect, monitor and localize vocalizing marine mammals. Methods for wildlife monitoring have been traditionally driven by visual observation but are also developed in combination with acoustic surveys [4].

In Fig. 1 the proposed monitoring system for grouping clicks in individuals is presented. A signal recorded by a single hydrophone is segmented into blocks of variable length which would contain a direct-path click. Similarities in the feature set of these segments are then used to arrange them into groups which would correspond to different individuals. The physical characteristics of whales are sufficient to differentiate them using rule-based algorithms. A pair-wise comparison of clicks was preferred to supervised learning solutions due to the unavailability of ground-truth labeled data, and to unsupervised learning algorithms [5] which assume an invariant feature set.

Click Segmentation

To detect the echolocation clicks, the signal is pre-segmented into blocks of variable length. These segments are analyzed to remove possible reflections caused by the underwater channel.

Click segmentation is performed in two steps. First, in the pre-segmentation, the transients which conform the click are detected and plausible boundaries for each click are set in the signal. Such delimited segments would contain clicks and click reflections. The second step is to select just the click segments by enforcing a set of rules with the extracted features, so that reflections due to underwater channel conditions could be discarded.

Pre-segmentation

A pre-segmentation method similar to the proposed one in [6] based on the Teager-Kaiser energy operator (TKEO) [7] is employed to detect click transients. The TKEO is applied on the discrete time signal $x(n)$ as:

$$\Psi(x(n)) = x^2(n) - x(n+1)x(n-1). \quad (1)$$

After applying the energy operator, all the peaks of $\Psi(x(n))$ which lie above a certain threshold Γ_{TK} are found. The threshold Γ_{TK} is defined as:

$$\Gamma_{TK}(x(n)) = \mu(x(n)) + h \cdot \sigma(x(n)), \quad (2)$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and the standard deviation of the signal. The constant h in Eq. 2 is a weight factor for setting the peak threshold lower or higher. This threshold is exploited in [6] to find significant peaks, group them, and segment the signal in clicks. However, in this way a segment could be formed by joining consecutive groups of clicks if the clicks are too near to each other. Also, if the peaks of a single click are too far away from each other, the click will be divided.

To overcome this drawback an approach based on signal smoothening by calculating the RMS envelope is employed. The RMS envelope is applied as in [8] to $\Psi(x(n))$ with a window length of 20 ms and a hop size of 4 samples. Peaks are then found on the RMS envelope of $\Psi(x(n))$. Preliminary boundaries for the segments of clicks are set at the first minima before and after each peak. The RMS envelope (red) in Fig. 2 is calculated from the TKEO signal (blue) of a particular click (black).

A Hamming window is then applied to a signal segment within the time positions given by these preliminary boundaries. The TKEO is again applied to the windowed segment and peaks are found inside such boundaries. Following experimental results, the beginning of a segment of a click starts at 0.5 ms before the first TKEO peak and ends at 5 ms after the last TKEO peak. The offset is necessary to prolong the segment as the click ending decays slowly in time and in amplitude. An example signal with its corresponding segment and click positions is shown in Fig. 3. The difference of the time positions of the segments is defined as the inter-segment interval.

Click Extraction

In ocean environments the speed of sound is a function of temperature, pressure and salinity. Therefore, in underwater acoustic channels, click and reflection power varies with time due to its multi-path characteristics. A reflection may thus achieve a slightly larger power level than the direct-path click signal. Although reflections behind the direct-path signal are plausible in underwater channel conditions, in the presented work only forward reflections are considered, i.e. the reflected segments are assumed to be after the detected segment.

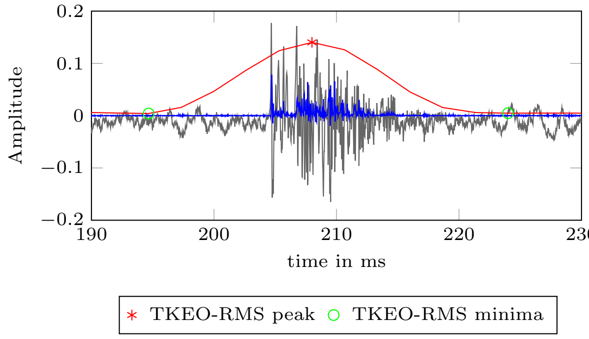


Figure 2: A RMS peak surrounded by two minima. The RMS envelope (red) is calculated on the TKEO signal (blue).

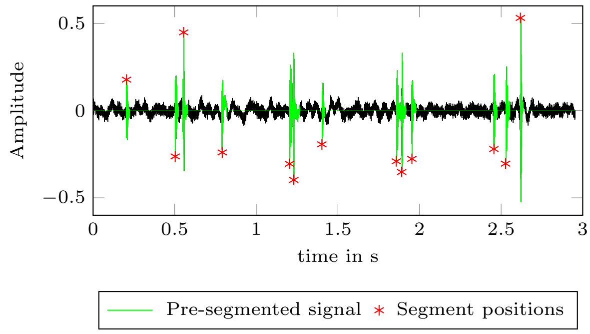


Figure 3: Detected click positions after pre-segementation.

In applications with a single hydrophone it is usually assumed that the reflected clicks propagates through multiple paths, thus making the reverberated click to be stretched in time[9]. This however depends on the whale position with regard to the hydrophone and on the quality of the signal. For consecutive clicks the second click may be falsely detected as a reverberated click of the first whale.

An iterative approach to detect reflections is introduced. Segments are compared in a sequential way. Let κ_m be each of the detected segments in the pre-segmentation step, and m the segment position, where $m = [1, \ldots, M]$ and M is the number of detected segments in $x(n)$. Reflections are searched to the right of κ_m. The segment is checked against a set of rules with the next segment κ_{m+1}. If κ_{m+1} is not an apparent reflection then m is increased. If on the other hand, κ_{m+1} is a click reflection candidate, the segment is flagged and skipped in the next search.

Reflected segments are identified based on a set of rules. Two groups of reflections are defined. A high power reflection is considered to have a short interval between two segments (< 35 ms), high negative magnitude of cross-correlated segments, and matching inter-pulse distances. Low power reflections on the other hand may have a large interval between segments (> 35 ms), the negative absolute magnitude of cross-correlated signal is higher than the positive magnitude, and usually a click followed by an echo has a significantly higher amplitude. After discarding reflections, the resulting set of κ_m segments should contain only segments with single direct-path clicks.

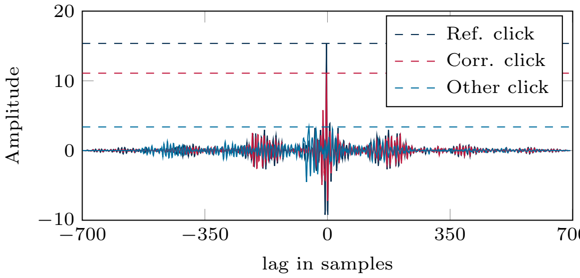


Figure 4: Correlation example. The black signal is the auto-correlation of a selected whale click. The red signal is the cross-correlation of two corresponding clicks. The blue signal is the cross-correlation of clicks from different whales. Dotted lines represent the peak correlation level for each signal.

Feature Extraction

Since the characteristics of the generated click depend on the trajectories inside the acoustic organ of the sperm whale, it can be implied that individual characteristics are to be found by analyzing the nature of the clicks [10]. In this sense, feature sets for each click are analyzed to find individual-specific characteristics. For example, the reflection of the pulses inside the acoustic organ characterize the impulse response of the sperm whale's head [11]. Assuming thus that the acoustic response in the spermaceti is characteristic of each individual, the click duration and the inter-pulse distance (IPD) are then considered as features.

Different acoustic characteristics of each individual are considered by evaluating the inter-segment interval, click duration and IPD, click correlation level, low frequency spectrum, and Linear Frequency Cepstral Coefficients (LFCC).

By calculating the auto-correlation of each segment and by cross-correlating the segments it is also possible to find a correspondence between the spectral energy content of two or more clicks. Fig. 4 shows the peak correlation level of a corresponding and a not corresponding segment of a click.

A cesptral smoothed spectrum is calculated for each click segment as in [8]. Liftering is done with a cut-off quefrency at $\frac{1}{8}$ times the cepstral length. The spectrum was truncated between 0.2 and 0.4 of the Nyquist frequency to demise irrelevant features. The spectral content of clicks appears to be clearly differentiable in this range. Fig. 5 shows the spectral envelope for two corresponding clicks sampled at 44.1 kHz.

The cepstrum is a good technique to determine time-invariant filter characteristics. This can also be helpful since the nature of a click is steady over a short period of time, and reflections due to channel-related phenomena are not characteristic of the click itself. A linear filter bank is then used to extract 12 LFCCs, from which the last 11 coefficients are used as features.

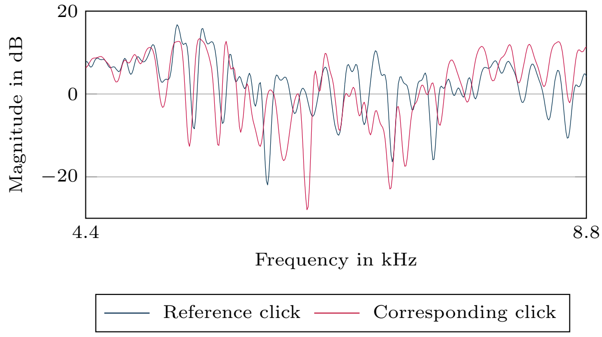
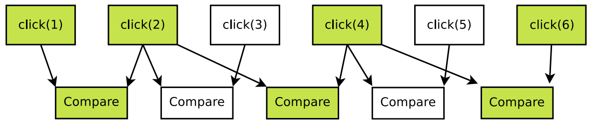
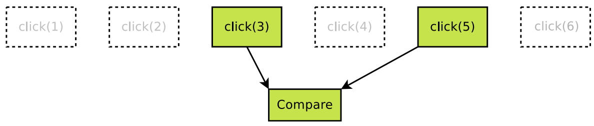


Figure 5: Spectrum of a reference and a corresponding click.



(a) Iteration 1



(b) Iteration 2

Figure 6: Iterative grouping example for corresponding click segments. Filled boxes show a positive correspondence. For each iteration one click group is found. Each group belongs to different whales.

Click Grouping

Decision rules are used to derive similarity between the click segments κ_m. Each feature set is exploited to develop a rule-based click grouping method.

Click grouping is done in an iterative fashion as depicted in Fig. 6. A segment κ_m is compared with the next segment κ_{m+1}. If the segments correspond, κ_{m+1} is added to the group and the comparison continues with the next click. If segments do not match, the segment is left aside for a next iteration. This process is repeated until no more clicks are left, or the ratio of remaining clicks to total clicks is very low. Depending on the amount and distance of matching features, clicks can have a strong, good, or low correspondence. A rule set is derived using inter-segment interval values, cross-correlation levels, IPD, and spectral and cepstral features.

The physical characteristics portrayed by the click features of the previous section do not take into account the time-variant characteristics of the underwater acoustic channel, nor the movement of the whale. To overcome feature variability, a time window $T_w = 10$ s is used to find correspondences. Groups of clicks in different time windows can then be merged using the same set of rules, or a derivation of them.

Results and Evaluation

During evaluation 24 signals of sperm whale recordings where used. The provided signals contain different num-

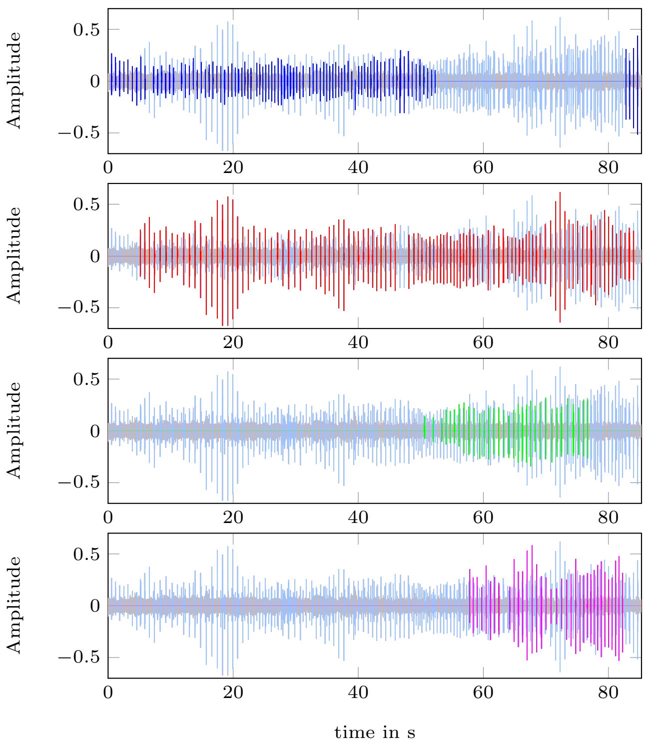


Figure 7: Original signal (gray), detected clicks (light blue), and grouped clicks for 4 detected individuals.

ber of active individuals, as well as a different amount of clicks and are recorded in different environments. The number of individuals in the signals where previously estimated by an experienced listener. The signals are of different duration, from several locations, recorded at different depths and sampling rates. Only one channel on each signal is used for evaluation.

Results using the presented click grouping method correctly estimated the reference number of individuals in the evaluation signals. The plurality of individuals in the waveform is also found by analyzing the variability of the inter-segment interval of all waveforms. An unsteady inter-segment interval will indicate that more than one active individual is clicking.

Fig. 7 shows a click grouping result for an example signal. Signals are arranged in separate plots by groups of corresponding clicks. The reference information of the original signal indicates that three individuals are active in the given signal. The total number of whales which are detected is four. However, only three whales are active at once during the time slot between 58 s and 75 s.

Conclusion

The proposed Passive Acoustic Monitoring (PAM) technique for abundance estimation of sperm whales (*Physeter macrocephalus*) achieve good estimates of the number of active sperm whales in recorded signals. Abundance estimation is achieved by grouping the corresponding clicks according to the similarity of their feature sets. Rule-based decisions are used to perform grouping of the click segments. Such a rule set is derived by relating the physical characteristics of the clicks with the estimated number of individuals labeled in the signals. Characteristics found between correspondent clicks in the evaluated recordings suggest that a signature is present on clicks for different sperm whale individuals. The method could also be adapted for abundance estimation of other species.

References

[1] B. Møhl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund, "The monopulsed nature of sperm whale clicks," *The Journal of the Acoustical Society of America*, vol. 114, no. 2, p. 1143, 2003.

[2] R. Antunes, L. Rendell, and J. Gordon, "Measuring inter-pulse intervals in sperm whale clicks: consistency of automatic estimation methods.," *The Journal of the Acoustical Society of America*, vol. 127, pp. 3239–47, May 2010.

[3] V. Teloni, W. M. X. Zimmer, M. Wahlberg, and P. T. Madsen, "Consistent acoustic size estimation of sperm whales using clicks," *Journal of Cetacean Research and Management*, vol. 9, pp. 127–136, 2004.

[4] J. Barlow and B. Taylor, "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," *Marine Mammal Science*, vol. 21, no. July, pp. 429–445, 2005.

[5] X. Halkias and D. Ellis, "Estimating the Number of Marine Mammals Using Recordings of Clicks from One Microphone," *Proc. of the 2006 IEEE Int. Conf. on Acoustics Speech and Signal Processing Proceedings*, vol. 5.

[6] V. Kandia and Y. Stylianou, "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," *Applied Acoustics*, vol. 67, pp. 1144–1163, Nov. 2006.

[7] E. Kvedalen, "Signal processing using the Teager energy operator and other nonlinear operators," *Master, University of Oslo Department of Informatics*, no. May, 2003.

[8] U. Zölzer, *DAFX: Digital Audio Effects: Second Edition*. John Wiley & Sons, 2011.

[9] E. b. P. Strumillo, F. Bénard, H. Glotin, and P. Giraudet, "Highly defined whale group tracking by passive acoustic stochastic matched filter," in *Advances in Sound Localization*, InTech, 2011.

[10] L. Rendell and H. Whitehead, "Do sperm whales share coda vocalizations? Insights into coda usage from acoustic size measurement," *Animal Behaviour*, vol. 67, 2004.

[11] B. Møhl, "Sound transmission in the nose of the sperm whale Physeter catodon. A post mortem study," *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, vol. 187, June 2001.