scientific reports

Check for updates

OPEN

Automatic detection and annotation of eastern Caribbean sperm whale codas

Guy Gubnitsky[1,7,8], Yaly Mevorach[2,7,8], Shane Gero[3,4,7,8], David F. Gruber[5,7,8] & Roee Diamant[1,6,7,8]✉

A key technology for sperm whale (*Physeter macrocephalus*) monitoring is the identification of sperm whale communication signals, known as *codas*. In this paper we present the first automatic coda detector and annotator. The main innovation in our detector is graph-based clustering, which utilizes the expected similarity between the clicks that make up the coda. Results show detection and accurate annotation at low signal-to-noise ratios, separation between codas and echolocation clicks, and discrimination between codas from simultaneously emitting whales. Using this automatic annotator, insights into the characterization of sperm whale communication are presented. The results include new types of coda signals, analysis of the distribution of coda types among different whales and for different years, and evidence for synchronization between communicating whales in terms of coda type and coda transmission time. These results indicate a high degree of complexity in the communication system of this cetacean species. Source code and data for our system is publicly available

Keywords Animal language, Sperm whale vocalization, Detection of coda, Coda annotation, Echolocation clicks, Graph-based clustering, Generalized Gaussian mixture model (GGMM)

In recent years, considerable effort has been made to record and capture the bioacoustics of sperm whale (*Physeter macrocephalus*). For example, Project CETI (Cetacean Translation Initiative), founded in 2020, is an interdisciplinary research initiative that uses advanced machine learning and cutting-edge robotics to better understand sperm whale communication[1]. The backbone of this effort is custom-built passive bioacoustic arrays covering a 20×20 kilometer area where these whale families reside (collecting over 30 TB/month), in conjunction with robotic acoustic and video tags on the whales, underwater gliders, towed hydrophone arrays and aerial drones to augment the rich contextual communication data. This initiative has led to a theoretical model using unsupervised machine translation to understand cetacean communication[2]. In addition, the method in[3] modeled codas as (variable-length) Markov chains, revealing new patterns of inter-clan sperm whale social learning. The work in[4] examines vowel-like spectral properties of codas and that in[5] elucidated a sperm whale phonetic alphabet. The method in[6] trains a transformer on click timings (inter-click intervals), which is able to predict sperm whale codas in an exchange based on long-term dependencies, as well as future diving behavior.

More broadly, in the face of biodiversity loss comparable to previous mass extinction events[7], armed with rapidly miniaturizing hardware to record wildlife and wild spaces, scientists are challenged not by collecting increasing large, and continuous amounts of data; but by devoting the time and development of the software tools to curate and interpret it (see[8]). Automated species detection and annotation of vocalizations is a critical step to characterizing and monitoring our changing ecosystem, and this is even more true in the world's oceans. Further, it allows for rapid assessment of previously data depauperate areas. In this case, given the importance of culture in conservation[9,10], and that the cultural population structure of sperm whale clans is defined by acoustic repertoire[11,12], an automatic detection and annotation of sperm whales codas will allow for the definition of clans present in previously unrecorded waters using passive acoustics, and thus provide not only a tool for understanding this species' communication system, but also for their applied conservation and management.

[1]Department of Marine Technology, University of Haifa, 3498838 Haifa, Israel. [2]Morris Kahn Marine Research Station, Department of Marine Biology, University of Haifa, 3498838 Haifa, Israel. [3]Department of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada. [4]The Dominica Sperm Whale Project, Roseau, Dominica. [5]Baruch College, City University of New York, New York, NY 10010, USA. [6]Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, Zagreb, Croatia. [7]Project CETI, New York, NY 10003, USA. [8]Project CETI, Roseau, Dominica. ✉email: roee.d@univ.haifa.ac.il

Current approaches

Sperm whales are among the most acoustically active toothed whales, making them attractive species for passive acoustic monitoring (PAM). Many methods have been proposed to detect and classify sperm whale clicks[13–18]. However, existing techniques focus on the detection and classification of echolocation clicks, which whales use to navigate the darkness of the deep sea and hunt squid prey. However, when individuals communicate with each other, sperm whales use an acoustically distinct type of click[19,20], which are produced in short sequences with stereotyped rhythm and tempo called 'codas'[21–23]. Codas last less than two seconds and have been characterized by the varying number of constituent clicks and the intervals between them. Codas are typically produced by multiple sperm whales in bouts that can last from 10 seconds to over half an hour.

Coda recognition is usually done by automatic click detection and subsequent manual annotation to distinguish coda clicks from echolocation clicks, to separate vocalizers and to 'link' clicks to the same coda. Existing manual annotation systems include custom Matlab scripts[24,25], the Rainbow Click software developed by the International Fund for Animal Welfare[22,26] and more recently CodaSorter[12,27]. However, while all of these programs provided a user interface for manual annotation of codas by users based on feature extractions from the clicks (e.g. interpulse interval, angle of arrival, and spectra content of the clicks), none of these programs provided automatic detection of codas, classification of echolocation and coda clicks, or the ability to automatically separate sources and automatically annotate individual clicks as belonging to the same coda. While Beslin et al.[18] performed classification between echolocation and coda clicks, this solution omits codas as "bad" echolocation. As a result, there are no automated approaches for detecting coda clicks yet. This may be due in part to the difficulty of separating sources between vocalizers, as codas from multiple whales often overlap during exchanges[22,28].

In this work an automated detector for codas is offered. The detector is intended for both presence detection and characterization of codas, while separating overlapping sources. The detector is designed for a wide dynamic range to enable recordings from both an acoustic tag attached to a sperm whale (near field) and an acoustic recorder deployed from a boat or mooring (far field). The complexity of the detector is low, so that real-time operation is possible. Offline analysis of data from acoustic tags that were not previously processed revealed new insights into the synchronization of coda characterization between pairs of communicating whales as well as new types of coda signals.

Preliminaries: characteristics of the Coda signal

The codas of sperm whales consist of patterns of broadband, impulsive signals called *clicks*, which are characterized by a multi-pulsed structure[20]. The first impulse is generated by forcing air through the whale's phonic lips. The generated sound is reflected by the air sacs at the anterior and posterior ends of the spermaceti organ. The result is a series of pulses of decreasing amplitude that follow each other at equal time intervals; a product of the transit time between the nasal air sacs in both directions. The two most important features for distinguishing between coda and echolocation clicks are the decay rate between pulses within the two click types, which is probably the result of different signaling pathways in the whale's nose[29], and their rhythmic structure. The echolocation clicks are produced in long series with a periodic pattern with characteristic inter-click intervals (ICIs) which are a function of the whale's search range[19]. In contrast, codas are sequences of clicks with stereotyped rhythm and tempo[21], which are thought to serve communication during exchanges between multiple individuals[22], and may serve social recognition and identity[11,26,30].

One challenge in recognizing codas is managing distortions caused by multipath interference. These lead to high variability in coda features, making it difficult to distinguish between a coda, a sequence of echolocation clicks or other interfering transients such as snapping shrimp or shipping cavitation noises. These challenges are highlighted by the example in Fig. 1. The figure shows recordings of two codas from a Dtag[31] attached to a sperm whale (A Dtag is an acoustic tag that is attached to the skin of the whale with a suction cup to provide a time series of acoustic recordings as well as depth and temperature information). The bottom panel shows a coda originating from the tagged whale (the focal whale) and the top panel shows a coda received from a distant whale. The codas are represented by super-resolution spectrograms[32]. We observe that the codas received from the focal whale and the distant whale have a different tempo-spectral structure. We note that the short time delay of 1 ms between the strongest pulses of the clicks in the top panel may be due to a surface reflection since the whale was close to the sea surface.

While the structure of clicks associated with a coda may vary between different codas due to e.g., the emission beam pattern, an acoustic shadow and surface reflection effects[33–35], we argue that these variations are negligible within the clicks of each coda since these clicks roughly share the same emission and propagation channel. This is demonstrated in Fig. 1. Such similarity is also evident in the clicks' inter-pulse intervals (IPI), which are strongly correlated with the size of the whale's spermaceti organ[19,36]. These observations motivate us to build our coda detector on a similarity metric to identify sequences of statistically related clicks. The result is a clustering scheme that, given a set of detected clicks (e.g., the method in[37]), separates groups of clicks with similar structure based on a likelihood metric. To improve our ability to distinguish between coda and usual clicks, we set limits within our clustering solution on the resonance frequency and on the number of interpulses within a click.

Results

Description of the testbed

Fewer than 600 individuals have been identified in the sperm whale community in the Eastern Caribbean[38,39]. These are divided into at least two, but perhaps three cultural clans[27]. Matrilineal social units of sperm whales associate with other units of whales which share a similar dialect of codas, creating a culture-based population structure known as *clans*. In the Eastern Caribbean, there are more members of the EC1 clan (> 200 whales) than members of the EC2 clan (< 200 whales)[38]. Known social units of female and immature sperm whales[40]

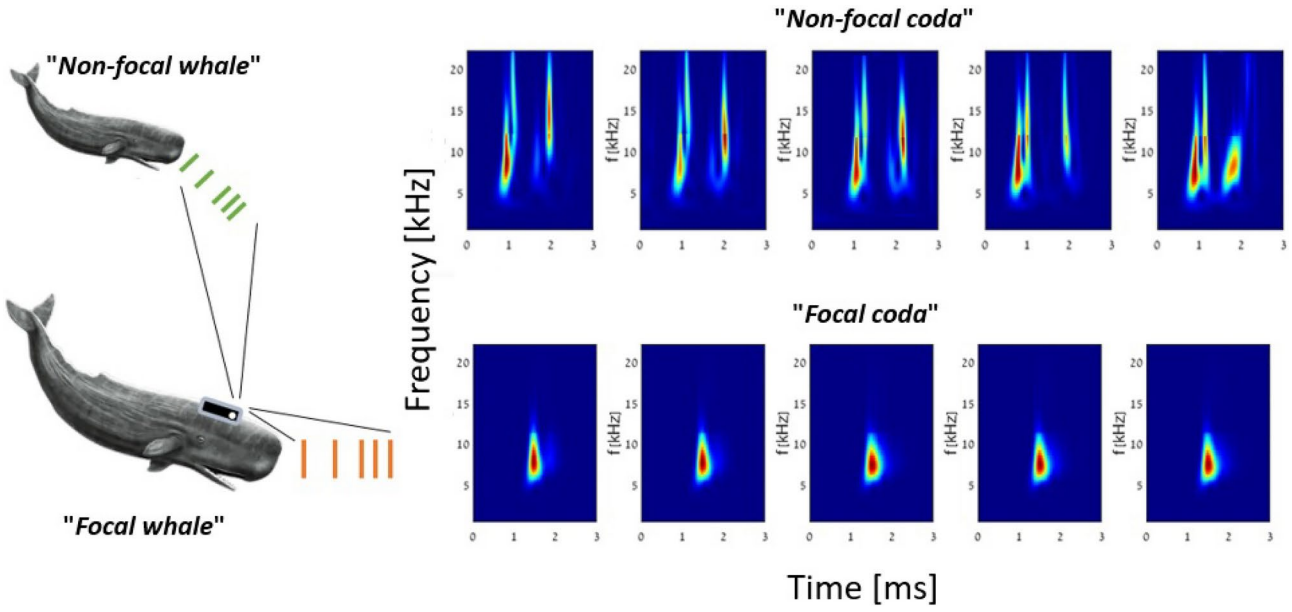


Fig. 1. Illustration of a scenario with a focal and a non-focal whale emitting 5-click codas. An example of a super-resolution spectrogram of individual clicks from a real tag recording of the codas of the focal and non-focal whale are shown in the top right and bottom right panels, respectively. A clear distortion of the structure of the clicks can be seen in the coda received from a distance. However, the similarity between the clicks is retained.

were located and tracked in an area covering approximately 2000 km^2 along the entire west coast of the island of Dominica (N15.30 W61.40).

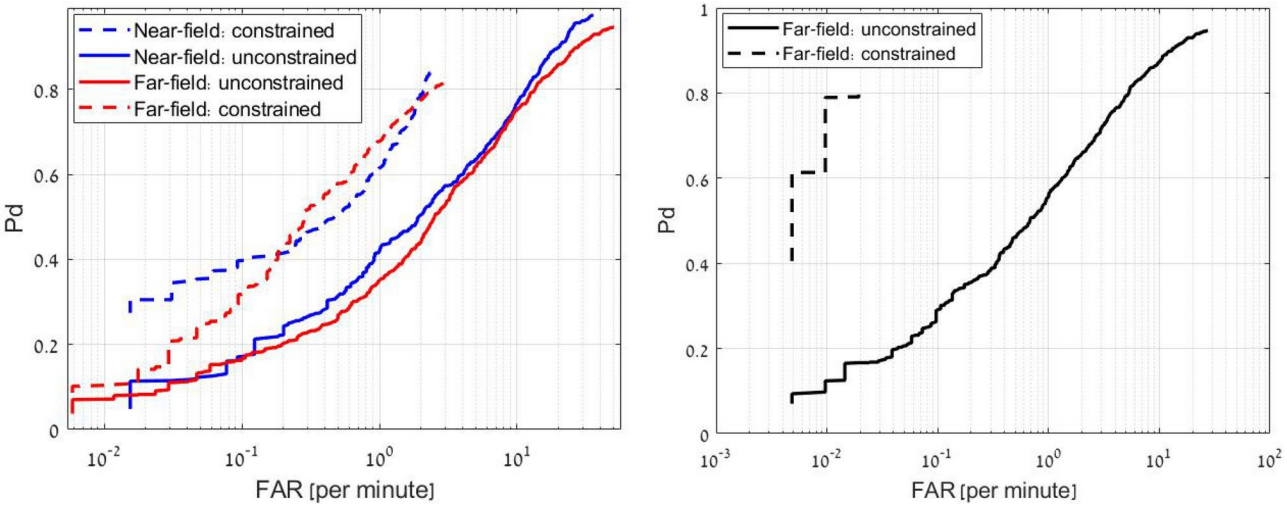
Our dataset includes both near-field data from 42 tags deployed on 25 different individuals in 11 different social units, with photo-identification to match the tagged whale, and far-field data by hydrophone deployed from a boat. Preliminary manual annotations for near-far data between 2014–2016 yielded a set of 3948 codas, while a set of 4930 codas was obtained for far-field data between 2005–2012. Additional annotations were supported by our automatic detector to analyze 34.5 h of near-field data collected in 2018, yielding a further set of 843 codas. Detector-assisted annotations also enabled the processing of 36.17 hours of far-field data, yielding 727 codas. For the assessment of false alarms, noise segments were used, which were divided into two categories: (1) four hours of raw acoustic signals that were manually verified to include ambient noise, ship noise, whistles and clicks from other marine mammals, and (2) three hours of echolocation clicks from several sperm whales, which were used to investigate the detector's ability to distinguish between codas and echolocation clicks.

Detection and annotation results

The detection performance is analyzed in Fig. 2 in terms of the Receiver Operating Characteristic (ROC) to investigate the trade-off between the detection rate (Pd) and the false alarm rate (FAR). The former is defined by the rate of true-positive cases verified by the manual annotations, and the latter by the rate of false-positive cases per minute. Pd and FAR were calculated based on manual annotations. As explained in "Methods", the detector applies a maximum likelihood approach, which is based on a distribution analysis of legacy database. To avoid overfitting, these distributions were evaluated from our near-field database represents recordings from a DTag from 2014-2016. Our far-field database represents recordings from an acoustic recorder deployed from a vessel from 2005–2012. The performance is shown for the most recent data collection, i.e. for the near-field (Dtags) from 2018 and the far-field (recorder deployed from vessel) from 2022–2023. We explore our detector's performance in two configurations: (1) when the detection solution is constrained by the maximum resonant frequency of the clicks and by the number of identified interpulses within the click. Here, following[19], we require the resonant frequency to be below 12 kHz and the average number of interpulses to be above 2.7, and (2) when the solution is unconstrained.

The performance is compared for noise with echolocation clicks (Fig. 2a) and for noise without echolocation clicks (Fig. 2b). The detection results are shown for both far-field and near-field data[1], and for constrained and unconstrained detection (see details in "Clustering" and "Interpulse analysis"). The results show that the FAR reduces by an order of magnitude when using constraints, at the cost of a slight decrease in the detection rate. For the echolocation noise (Fig. 2a), no significant difference in performance is observed between the near-field and far-field datasets. From this, we conclude that our detector can handle a low signal-to-noise ratio (SNR) well.

[1]The results for noise without echolocation clicks are for far-field only, as the data from tags almost always contained echolocation clicks.



(a) Noise including echolocation clicks. **(b)** Noise do not include echolocation clicks.

Fig. 2. ROC curve, evaluated for near- and far-field data. Performance is explored for two detector modes: (1) constrained, where we limit the clicks' resonant frequency and the number of interpulses, and (2) unconstrained, where we do not impose any restriction. The solid lines refer to the unconstrained detection mode, while the lines for the constrained mode are dashed. False alarm rate evaluated for noise data including (**a**) sperm whale echolocation clicks, (**b**) ambient noise without echolocation clicks (only far-field data).

	General noise				Echolocation clicks			
Constraint applied	$\mathscr{P}$	f_r	$\mathscr{P}$ & f_r	**None**	$\mathscr{P}$	f_r	$\mathscr{P}$ & f_r	**None**
Target: Pd=0.18 & FAR=0.12	0.0097	0.024	0	0.034	0.035	0.082	0.023	0.12
Target: Pd=0.4 & FAR=0.33	0.0097	0.275	0	0.33	0.199	0.987	0.152	1.5

Table 1. False alarm rate (FAR) per minute. Results are given for far-field data with detection thresholds set according to detection rate of {0.18,0.4} and a false alarm rate of {0.12,0.33}. The results are analyzed for two noise categories: without echolocation clicks (general noise) and with echolocation clicks (echolocation clicks). Performance is evaluated for three options for a constrained configuration of the detector: (1) constraining the number of identified interpulses within clicks, $\mathscr{P}$, (2) constraining the resonant frequency, f_r , (3) constraining both $\mathscr{P}$ and f_r, and (4) an unconstrained configuration of the detector..

Comparing the results for the two noise categories, we find that the FAR is twice as high for detection without constraints when the noise contains echolocation clicks, which is the harder case. Since the multipulse and resonant frequency constraints also allow the separation of noise transients, this difference is magnified when considering the performance differences between the constraint detectors.

In Table. 1, we examine the effects of each constraint for both cases of noise categories for the far-field dataset. Two detection thresholds are selected by a trade-off of (1) Pd = 0.18 and FAR = 0.12 per min and (2) Pd = 0.4 and FAR = 0.33 per min. To investigate the effects of the clustering constraints, we show the results without constraints (None), with the multipulse constraint only ($\mathscr{P}$), with the resonant frequency constraint only (f_r) and with both constraints ($\mathscr{P}$ & f_r). In both cases of noise categories, we find that the multipulse constraint is most dominant in distinguishing coda clicks from echolocation clicks and noise transients. The resonant frequency constraint also contributes, but less. We explain this result by the effect of the channel impulse response, which distorts the spectrum of the far-field signal and allows less separation by spectra analysis. For the more common case of "general noise" in far-field data, we observe a very low FAR for constraint detection.

Next, we examine the accuracy in the automatic annotation of all clicks within the coda. This attribute is important to define the type of coda, which in turn is determined by the number of clicks and the rhythm and tempo of the ICI pattern[23,28]. Especially in the presence of echolocation clicks or in the frequent case of overlapping codas, a true coda click can easily be disqualified as an echolocation. Fig. 3 shows the cumulative distribution function (CDF) of the ratio between the number of identified clicks and the actual clicks in the coda. We compare the performance for the far-field and near-field datasets. Identification of all clicks in the coda is achieved in about 70% of cases, while in another 10-20% of cases 80% of clicks are identified. The results show no significant difference between the near-field and far-field datasets. Although the accuracy level in detecting the coda clicks is promising, we note that misidentification of a single click may lead to misclassification of the coda type. We leave the improvement of this result to future work.

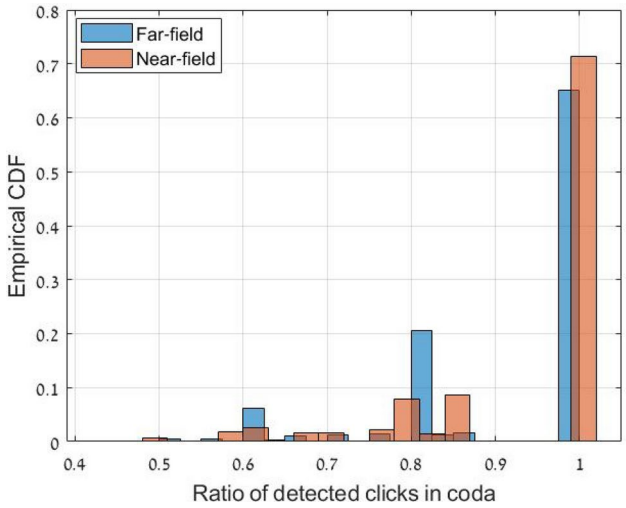


Fig. 3. Cumulative distribution function (CDF) of the ratio of the number of clicks detected in the coda in the near- and far-field datasets using constrained detection mode.

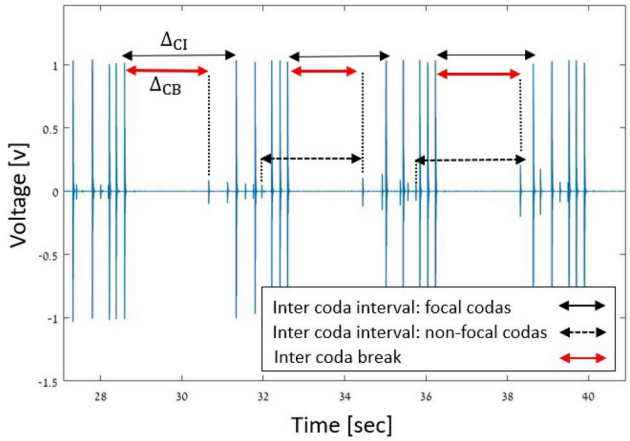


Fig. 4. An example of 13 seconds from a near-field recording showing a dyadic coda exchange between a focal and a non-focal whale. The Inter Coda Interval of the focal and non-focal whale are indicated by black solid and dashed arrows, respectively. The Inter Coda Break is indicated by red solid arrows.

Communication characterization: findings from the coda annotator

The ability of the developed automatic detector and annotator to provide large datasets of coda was used to investigate the characteristics of coda signals towards the understanding of sperm whale communication. To eliminate bias in our analysis, we filter our coda database to use only codas from the near field, which allows determination of the identity of the focal whale. Further, since no classifier is currently proposed to distinguish between codas of individual whales, the following analysis considers dyadic social interactions by codas involving only two whales. To that end, we only consider coda pairs (see definition below) between a focal and a non-focal whale. These cases are identified by considering pairs of codas, detected within a buffer of 7 sec and where there is a substantial difference in both codas' amplitude. Other codas that contained a single coda with no detected response or codas of more than two whales were discarded. An example of a dyadic interaction of 3 coda pairs is shown in Fig. 4 by the red arrow marks. This process resulted in 635 coda pairs used for the below analysis.

Our main findings from the operation of the automatic coda annotator are evidence for synchronization between the focal and non-focal whales. This was established by examining the time delay between successive codas and between the clicks of the two codas. This type of correspondence arguably corresponds to the synchronization required for a stable telecommunication session. In particular, we note that the structure of the coda can be considered as a communication modulation signal. The time interval between codas can then be referred to as the communication baud rate, while variations between the click delays (ICI) of the codas can reflect the information encoding to obtain the channel's capacity outage. Another finding is the discovery of two new coda structures. This was revealed by comparing the extracted coda features with those of codas available in the literature. To describe these two findings, the following definitions are used, and examples of each of these structures are given in Fig. 4.

- Coda type: Categorical coda representation obtained by clustering codas based on their ICI vectors[26].
- Dyadic coda exchange: Period in which codas are made by a pair of whales.
- Inter-coda interval: A time interval, Δ_{CI}, between a consecutive pair of calls of a particular whale. As marked in Fig. 4, the inter-coda interval for a given whale is measured by the time delay between the last click of a coda and the first click of a subsequent coda of the same whale.
- Coda pair: Two consecutive codas originating from different whales in a dyadic exchange within a buffer of 7 sec, such that the first click of one coda is received before the last click of the second coda.
- Inter-coda break: The time interval, Δ_{CB}, between a coda pair. As marked in Fig. 4, the Inter-coda Break is defined by the time delay between the last click of one coda and the start of the first click of the interlocutor's next coda.

Modulation signal: analysis of coda type distribution

A coda type is characterized by the number of clicks it consists of (between 3 and 10) and by the sequence of time delays between the individual clicks. Examples of this are a sequence of 5 clicks in which the first three clicks are evenly spaced and at a different distance from each other than the last two clicks (type 1+1+3), or a sequence of 8 clicks with a gradually increasing or decreasing distance (type 8i or 8D, respectively). To date, 25 coda types have been identified[26]. These were determined by a clustering result applied to annotated ICI vectors of codas[26,41].

The analysis in Fig. 5 shows the distribution of codas for all considered coda pairs found during the periods of dyadic coda exchange. The results are presented as a matrix for the distribution of a 'signal and response' (S&R) session between the focal and non-focal whale. Without loss of generality, the coda type of the focal whale in the matrix's columns is considered here as 'signal' (S) and the coda type of the non-focal whale (hereafter referred to as the *Interlocutor*) in the matrix's columns is considered the 'response' (R). The bin's color corresponds to the probability (on a logarithmic scale) that a particular focal's coda type is observed before a particular interlocutor's coda type. The results show that the '1+1+3' type is the most typical coda type for both the signaler and the responder. Furthermore, the variation in the coda types is not large and the highest concentration is between types '1+1+3' and itself, between types '1+1+3' and '4R2' and between type '5R1' and itself.

The Baud-rate: time delay dependencies in whale codas

The changes in the structure of the identified codas arguably resembles that of encoding of information via a modulation signal. Here we use the analogy of a baud rate to correspond to the inter-coda Break, Δ_{CB}, as defined in 2.3. The probability density function (PDF) of Δ_{CB} is shown in Fig. 6 for three focal whales named 'Atwood' (whale #5586), 'Fork' (whale #5151) and 'Pinchy' (whale #5560). Figure 6a shows the PDF of $\Delta_{\mathrm{CB}} - \Delta_{\mathrm{CI}}$ separately for each whale. An observable shift in the expectation value of the three PDFs reflects differences in the duration of S&R sessions between whales. The longer tail towards the negative side of the distribution shows that sometimes whales skip a coda overlap and leave a gap of approximately similar to the exchanges Inter-coda Interval. This is further explored in Fig. 6b, where we show the PDF of Δ_{CB} for the three most frequent S&R coda types, combined for the three focal whales. The seemingly similar PDFs in the latter figure show that the observed bias in Fig. 6a is not affected by the coda type.

Taken together, these results suggest that the structure of the exchange in terms of the Inter-coda Interval can be maintained despite the variation in ICI within codas, or coda types being exchanged.

Capacity outage: variation in information gain encoded in coda types

The analogy with the characteristics of telecommunication performance is further explored by the concept of capacity outage, which is represented here by the change in the structure of the codas, which arguably corresponds to symbol modulation and thus to information gain. The change in the coda's structure is analyzed by variations

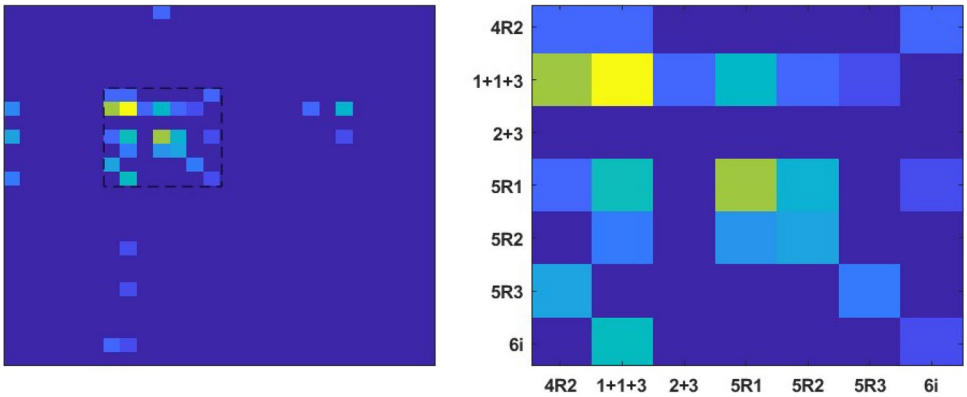
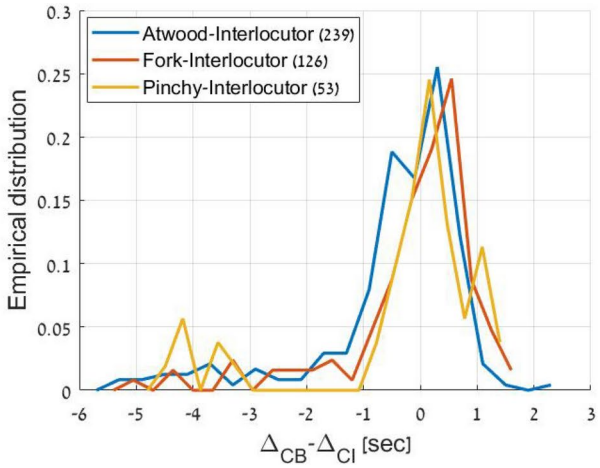
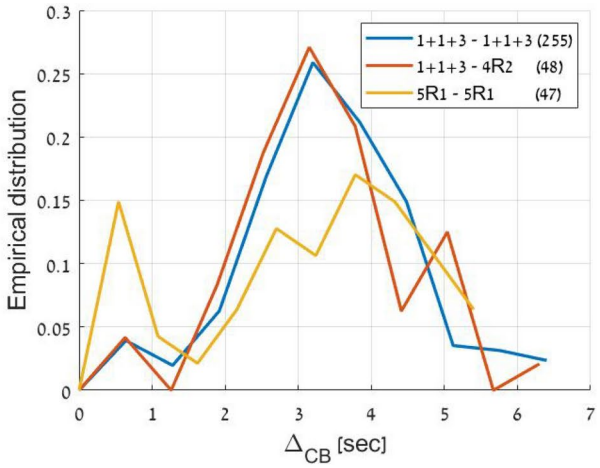


Fig. 5. Analysis of the distribution of coda types for all dyadic coda exchanges found. The left panel shows the distribution matrix of all pairs of coda types, with each bin representing a particular coda type from the focal whale (signal) and the subsequent coda type of the non-focal whale (response). The right panel is zoomed in on the most active coda types. The bins' color represents the probability of occurrence on a logarithmic scale.



(a) Probability density of $\Delta_{CB} - \Delta_{CI}$ for three focal whales, 'Atwood','Fork' and 'Pinchy'.



(b) Probability density of Δ_{CB} as a function of coda type.

Fig. 6. Distribution functions for the delay between the end of the 'S'-coda and the beginning of the 'R'-coda, Δ_{CB}.

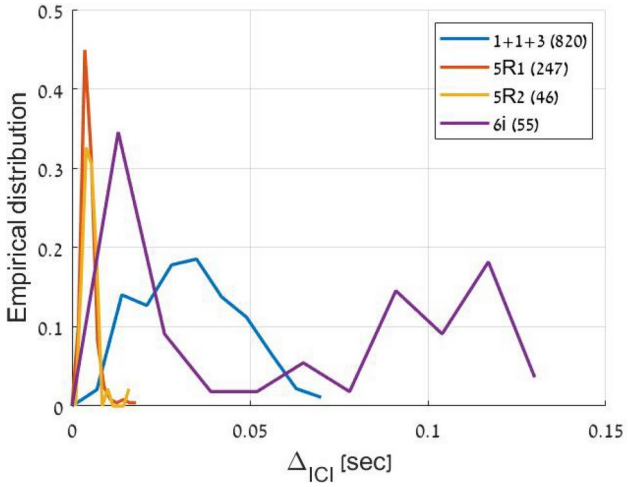


Fig. 7. PDF of Δ_{ICI} as a function of coda type. The number of detected Codas per type is given in the legend in brackets.

in the coda ICI relative to the typical ICI pattern for the specific coda type. This variation is measured by the metric

$$\Delta_{ICI} = ||\frac{1}{L-1}\left(\sum_{l=2}^{L} \tau_l(t) - \bar{\tau}_l(t)\right)||_2, \tag{1}$$

where $\tau_l(t)$ and $\bar{\tau}_l(t)$ are the ICI delays between the lth and $l-1$th coda's click of the measured and typical coda type t, L is the number of clicks comprising the coda, T is the number of coda types, and $||\cdot||_2$ denotes the $l2$-norm operator.

The PDF of Δ_{ICI} for four common S&R coda types is examined in Fig. 7. The values of Δ_{ICI} for the coda types 5R1 and 5R2 show small fluctuations which, according to our communication analogy, reflect a low degree of modulation leading to a low level of information transfer. In contrast, the PDF of Δ_{ICI} for coda type 1+1+3 shows higher ICI variation with greater diversity, while the highest diversity is shown in Δ_{ICI} for coda type 6i. This can be explained by the clicks structure of these coda types, where a greater ICI variation is observed the more complex the coda type is. That is, assuming that some information is encoded in the ICI variation (e.g. social identity information,[30]), the information capacity depends on the coda type. In other words, the more complex the modulation symbol (coda type) is, the more information (ICI variation) it can transport, to use the analogy with telecommunication once again.

The above observation is next investigated for various focal whales. The PDFs of Δ_{ICI} for the focal whales 'Atwood', 'Fork' and 'Pinchy' and their corresponding interlocutor are shown in Fig. 8a. The similar PDFs for the ICI of the focal and interlocutor whales reflect synchronization between the 'S'-coda and the 'R'-coda. Our results here, and in Fig. 5, suggest that whales often match coda type (and therefore Δ_{ICI}) during dyadic coda exchanges (as suggested by[22,28]. In fact, Sharma et al.[28], suggests that whales in dyadic coda exchanges, like those analysed here, are able to precisely match variation in both rhythm (relative timing of clicks in a coda) and tempo (total duration of a coda[28]), both which drive variation in ICI, and further can do so by matching tempo even when not matching the same coda type (rhythm). Here, we suggest that this synchronization of ICI variation may also reflect the dependence on the identity of the emitting whale. As is supported by the differences between the PDFs for 'Atwood' (upper panel), 'Fork' (middle panel) and 'Pinchy' (lower panel) and their interlocutor are matched to the identity of the whale.

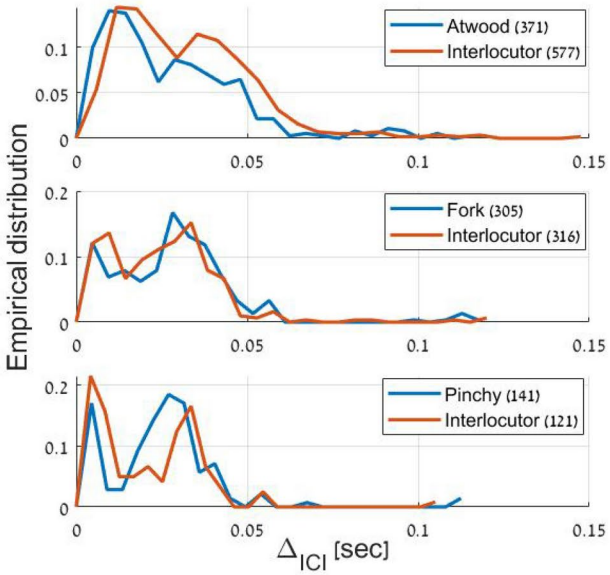
The above interpretation is supported by the results in Fig. 8b, where the PDF of the Inter-coda Interval, Δ_{CI}, is compared for the focal and the interlocutor, again for the whales 'Atwood', 'Fork' and 'Pinchy'. Again, a synchronization of Δ_{CI} is observed between the signaller and the responder, a synchronization that changes depending on the identity of the different focal whales. This suggests that the whales are able to vary their Inter-coda Interval depending on who they are interacting with.

Identification of new coda types

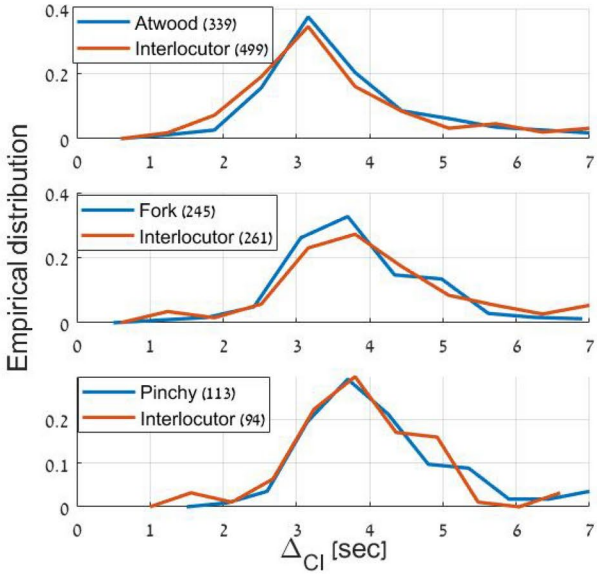
Coda types are a categorical representation of continuous ICI multivariate space. Previously, discrete coda types have been defined based on various clustering methods (see[11,12,25,26,42]). During these processes, some of the codas, whose ICI series are outliers or located in sparse areas between dense clusters, to be labelled as 'contamination' or 'noise' (rather than being 'forced' into categories) and thus only define more conservative, dense clusters in which all codas in a discrete 'type' are highly similar to each other. An example of this is shown in Fig. 9, where we visualize the three most dominant principal components, PC1, PC2 and PC3, for 427 codas with 6 clicks obtained by performing a principal component analysis (PCA) on our data. This PCA analysis allows the identification of clusters in the dataset; corresponding to different coda types. In the top right panel of Fig. 9 we show the analysis for the preexisting annotated dataset. We observe two prominent clusters labeled '6i' and '6R', while other scattered points are labeled as 'noise' simply because they did not meet thresholds of density or cluster shape based on the quantitative methods applied to define categories. By using our automatic detector for the near-field recordings from 2018, which were not previously annotated due to the large amount of data and the time required by manual methods, 843 codas were rapidly annotated. This then allows us to observe two more distinct clusters in the feature space, apart from the previous 6R and 6i coda types. This can be seen in the bottom right panel of Fig. 9. We define two more descending class, coda types as the new types '1+5' and '1+1+4' (rhythm plots are shown in the upper left and lower left panels of Fig. 9). We note that the codas within each of the new clusters have highly similar rhythm structure with some within type tempo variation; but that the two types have distinct rhythms and tempos.

Discussion

Our approach, arguably the first automatic detector and annotator for sperm whale codas, offers robust operation with few parameters to set, allowing analysis of data obtained by both acoustic tags and boat-mounted



(a) PDF of Δ_{ICI} for the focal and non focal whales.



(b) PDF of Δ_{CI} for the focal and non focal whales.

Fig. 8. Differences between variations of Codas generated by the focal and non focal whales. The number of coda sequences analyzed per whale is given in brackets in the legend..

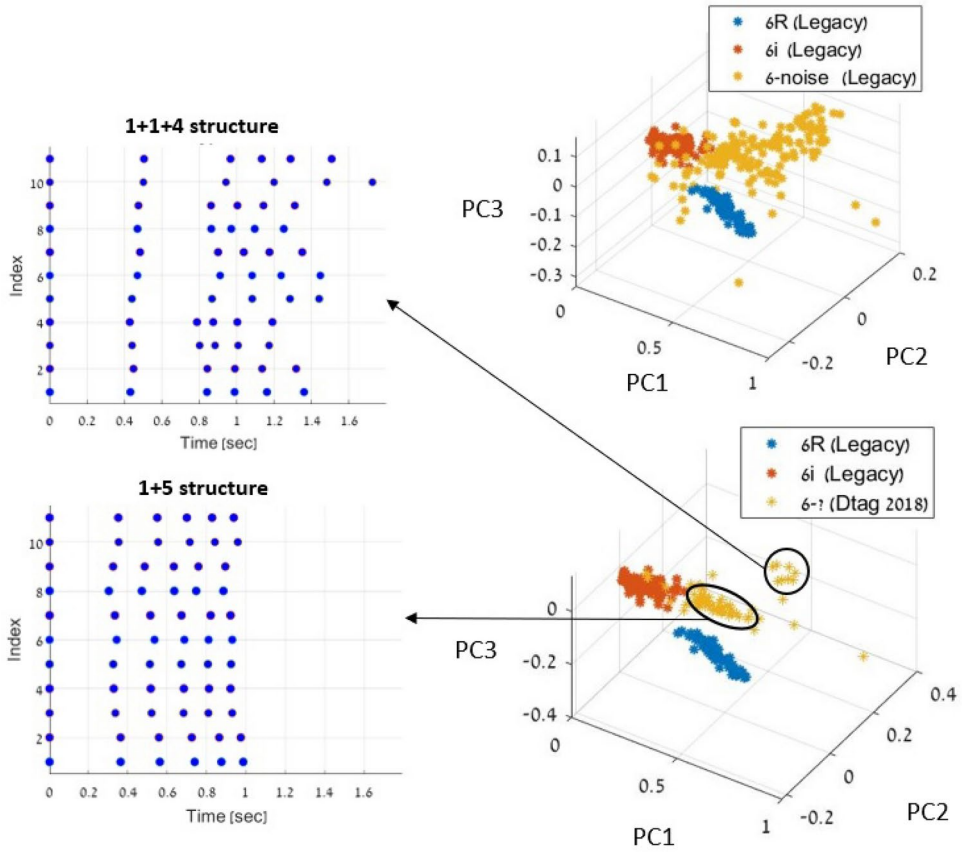


Fig. 9. Right panels: Distribution of the three dominant PCA components (PC1, PC2 and PC3) of 6 click codas out of 427 from our legacy database. The upper right panel shows the distribution of the annotated database. The bottom right panel shows the distribution of the unannotated dataset obtained with our automatic annotator and verified manually. Left panels: rhythm plots of ICIs from a few examples of the two new coda types.

hydrophones located far from the whale. Our detector is based on the assumption that the clicks within a coda roughly share the same emission and propagation channel. This assumption arises mainly from the short duration (<2 sec) of the coda. We argue that this time interval is short enough to assume same heading direction of the whale, especially since codas are produced during a socialization phase[43] which usually occurs at close range. Further, unlike other echolocation clicks, which are characterized by a sharp beam pattern[33,44–46], coda clicks are less directional[34], and are therefore less affected by changes in the orientation of the whale. It is worth noting that the IPI is the most reliable feature in terms of its stability across all clicks of a coda, as this feature is a product of the whale's anatomy and cannot be controlled by the whale, like other features such as the amplitude and frequency of clicks[4]. On the other hand, it is difficult to distinguish coda clicks from whales of similar head size based on IPI alone. The ROC performance of the detector shows that it is robust even at low SNR values and in high dynamic range, that it can handle the noise transients commonly found in the Dominica Island, and that it is able to distinguish between echolocation clicks and coda. As shown in Section 4 below, our annotator can also separate between overlapping codas. When the analyzed buffer is small, there are limitations when echolocation clicks from different whales overlap, as they may resemble a coda structure. Future developments will therefore focus on the inclusion of blind source separation to mitigate cases of interfering clicks, as well as a temporal likelihood prediction model to allow generalization to a larger number of coda types.

The detector provides automatic analysis of large datasets of raw acoustic data, thereby significantly reducing the processing time. The results of the automatic annotation were demonstrated to explore the characteristics of the codas. In particular, the distribution of coda types, the delays between successive codas and the ICI variation for different coda types and between different focal and non-focal whale pairs. An analogy was made for the resemblance of coda exchange to a telecommunications session, where the synchronization between the focal whale and its interlocutor in terms of delays between successive codas was compared with the baud rate of the communication, and the synchronization in terms of ICI variation was attributed to capacity outage. The large amount of codas collected with the new detector allowed a statistical analysis of coda features to comment on the diversity of codas. This led to the discovery of two new coda types.

Significantly, our results suggest that some of the within coda type variation, the structure of coda exchanges in terms of the inter-coda-intervals and inter-coda-breaks, vary based on the identity of interlocutors. While[26] suggested that within coda type ICI variation of the 5R1 coda could be used to potentially recognize individuals

within a small number of social units, these new findings suggest identity cues at the level of the structure of the coda exchange. This will require deeper investigation in relation to if these patterns are upheld depending on which whale vocalizes first (vocal asymmetries), such that whale B may match whale A's inter-coda-interval or inter-coda breaks when A vocalizes first, but A will match B if B vocalizes first. But also, how these patterns may perhaps reflect vocal cues of social dominance (see[47,48] or correlate with social or kin based relationships. Interestingly, Fork and Pinchy belong to the same social unit, while Atwood lives in a different one; and their distributions of Δ_{CI} are more similar to one another than to Atwood's distribution; so these patterns may reflect unit-level, rather than individual level patterns.

Looking towards the future plans for the detector, we intend a real-time implementation within the framework of Project CETI. This will include online detection from an offshore moorings to support the tagging work and create a live map of the whales' positions. The detector will also be implemented in real-time on board a sea glider and in surface vessels searching the area for sperm whales.

Methods

Method of data collection

Data collection was conducted in two ways: (1) acoustic tags attached to sperm whales (near-field), and (2) acoustic recorders deployed from a boat (far-field). In the first method, sound and movement tags (Dtag generation 3[31]) worn by the animals were deployed between 2014 and 2018 as a part of The Dominica Sperm Whale Project. These tags have a two-channel audio sampling at 120 kHz with a resolution of 16 bit, providing a flat (±2 dB) frequency response between 0.4 kHz and 45 kHz. Pressure and acceleration were also sampled at a rate of 500 Hz and a resolution of 16 bit, decimated to 25 Hz for offline analysis. Tagging was carried out from an 11 m rigid-hulled inflatable boat (RHIB). A 9 m hand-held carbon fiber pole was used to attach individual tags to a whale with four suction cups.

Data collection in the far field was carried out with a C75 Cetacean Research, Golden, CO 80401, USA, hydrophone deployed from the RHIB at a depth of 15 m. The hydrophone has a flat (±3 dB) frequency response between 10 Hz and 170 kHz. A MicPre-6 || data acquisition device from Sound Devices corporations, Wisconsin 53959, USA, was used to sample the signal at 96 kHz and 16-bit resolution.

All whales were identified from photographs of the trailing edge of their fluke ([49]). The identifications were used to ensure that only recordings from one of the two sympatric clans (EC1—the Eastern Caribbean Clan) were included in the analysis to control for any differences in repertoire between vocal clans[41].

Permitting and animal care approval

Data from The Dominica Sperm Whale Project were collected under scientific research permits from the Fisheries Division of the Government of Dominica. The field protocols for approaching, photographing, non-invasively tagging, and recording sperm whales were approved by either the University Committee on Laboratory Animals of Dalhousie University, Canada; the Animal Welfare and Ethics Committee of the University of St Andrews, Scotland; or Aarhus University, Denmark; and sometimes several or all of these across years. Data collected as part of Project CETI were collected under a scientific research permit from the Fisheries Division of the Government of Dominica. The field protocols for recording sperm whales for Project CETI were approved by The Institutional Animal Care & Use Committee at Harvard University. We confirm that all methods and experiments were performed in accordance with relevant guidelines and regulations. Specifically, under the ARRIVE guidelines (PLoS Bio 8(6), e1000412, 2010) and regulations relevant to wild whales.

Coda detection

A block diagram of the coda detector and annotator is shown in Fig. 10. The input to the detector is a buffer of 7 sec (configurable). First, a transient detector identifies a few regions- of interest (ROIs). For each ROI, three features are extracted: (1) the shape of the transient, (2) the intensity of the transient, and (3) a measure of the multipulse structure, if it exists. Clustering is then performed based on these features to group the ROIs based on their likelihood similarity. The clustering is performed iteratively until a stopping criteria is met. The details of the individual components are described below. A MATLAB implementation code for our scheme is available in[50]. Table 2 outlines the main notations of the method.

Click identification

Transient ROIs are determined by selecting identified peaks based on the expected attributes of valid coda clicks. The process starts with a bandpass filter for the desired frequency range of 2–24 kHz[15]. To enhance transients, following prior works for sperm whale click detection[15,51], the Teager-Kaiser energy operator (TKEO)[52],

$$z_n = x_n^2 - x_{n-1} \cdot x_{n+1} \quad , n = 1, \ldots, N. \tag{2}$$

is executed over the sampled signal $x_n, /n = 0, \ldots, N-1$. Peaks in z_n from (2) are identified as local energy maxima that fulfill a minimum distance between the peaks and a minimum threshold for the SNR. The former is set according to the maximum IPI of known codas (a value of 8 ms is used in the results). We limit the number of identified transients to an expected maximum of 20 peaks. An example of the transient identification process is shown in Fig. 10. Note that the features are extracted from the bandpassed signal x_n (e.g. the left panel in Fig. 10) and not from the enhanced signal z_n (e.g. the middle and right panels in Fig. 10).

Clustering transients into candidate codas

The clustering of transient ROIs into groups of possible codas is based on the observation that codas follow known ICI patterns[26] and on the assumption that the clicks forming a coda share statistical similarities. There

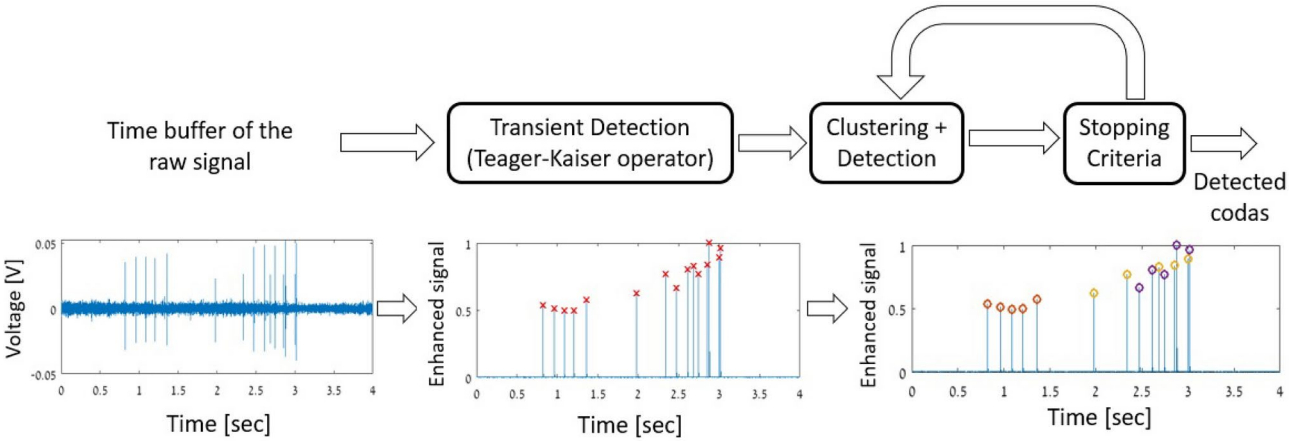


Fig. 10. A block diagram for the Coda detector.

Variable	Description
x_n	A sample of a measured signal at time index n
c_k	A binary association hypothesis of all clicks belonging to coda k
C	A set of all association hypothesis $\{\mathrm{c}_k\}_{k=1}^{K}$
K	The number of codas
$\mathscr{J}_{(\mathrm{c}_k)}$	A utility function that defines the overall likelihood that the clicks in c_k represent a valid coda
ρ_d	A detection threshold. Declares the detection when the utility function of a group of clicks exceeds this value.
$\mathscr{L}^{s}_{(\mathrm{c}_k)}$	Structural similarity likelihood. Defines how similar the clicks in c_k are to each other
$\mathscr{L}^{t}_{(\mathrm{c}_k)}$	Temporal likelihood. Defines how similar the rhythm of the clicks in c_k is to the rhythm of a known valid coda.
$F(\mathrm{c}_k)$	A penalty on the number of clicks in a coda that favors larger codas to prevent the clustering
	solution from converging on a subset of clicks rather than on the entire sequence.
$\mathscr{P}(\mathrm{c}_k)$	A measure of the median value of the number of identified inter-pulses of the clicks in group c_k
$\mathscr{P}_{\min}$	A minimum limit for the allowed value of $\mathscr{P}(\mathrm{c}_k)$
$f_r(\mathrm{c}_k)$	The average value of the measured resonant frequency of the clicks in in group c_k
$f_r^{\max}$	A maximum limit for the allowed value of $f_r(\mathrm{c}_k)$

Table 2. List of main variables.

are two explanations for the latter assumption: (1) each of the coda clicks is generated by the emitting whale in the same way, and (2) due to the short duration of the coda (0.1–2 s)[26,53,54], the channel's impulse response from the whale to the receiver is expected to be roughly time invariant, and thus the clicks of the same coda share the same channel distortion. An example of this is shown in Fig. 1, where different spectral shapes are observed for the far-field and near-field codas.

Our clustering solution is based on a likelihood definition for a valid coda's click sequence. This likelihood reflects (I) how similar the clicks in the sequence are and (II) how similar the rhythm of the clicks is to a known valid rhythm. The first metric is motivated by our expectation that the click elements of a valid coda would have a similar structure, while the second metric is motivated by the expected arrangement of the clicks based on a legacy database. Our clustering solution is formalized to search for groups of clicks that maximizes this likelihood while satisfying four constraints: (1) the likelihood value of each group must be high enough; (2) the clustering solution is unique, i.e., each click is assigned to a single group; (3) the number of inter-pulses within the click should be high enough; (4) the resonant frequency of the click should be within an expected range. This is formalized as the following optimization problem:

$$\hat{K}, \hat{\mathrm{C}} = \mathrm{argmax}_{K,\mathrm{c}_k,\ k=1,\ldots,K} \frac{1}{K} + \sum_{k=1}^{K} \mathscr{J}\left(K, \mathrm{c}_k\right)_{k=1,\ldots,K} \tag{3a}$$

$$\text{s.t. } \mathscr{J}_{(\mathrm{c}_k)} > \rho_d \tag{3b}$$

$$c_l \perp c_k,\ \forall l \neq k, \tag{3c}$$

$$\mathscr{P}(\mathrm{c}_k) > \mathscr{P}_{\min}\,, \tag{3d}$$

$$f_r(\mathrm{c}_k) < f_r^{\max}\,. \tag{3e}$$

where $\mathrm{c}_k,\ k = 1 \ldots K$ are M-dimensional binary assignment vectors such that

$$\mathrm{c}_k = [\omega_{k,1}, \ldots, \omega_{k,M}]\,, \quad \omega_{k,l} \in \{0,1\},\ l = 1, \ldots, M\,. \tag{4}$$

The utility function $\mathscr{J}_{(\mathrm{c}_k)}$ is defined by

$$\mathscr{J}_{(\mathrm{c}_k)} = \mathscr{L}^{\mathrm{s}}(\mathrm{c}_k) + \alpha_1 \mathscr{L}^{\mathrm{t}}(\mathrm{c}_k) - \alpha_2 F(\mathrm{c}_k)\,, \tag{5}$$

where $\mathscr{L}^{\mathrm{s}}(\mathrm{c}_k)$ and $\mathscr{L}^{\mathrm{t}}(\mathrm{c}_k)$ are the structural and temporal similarity likelihoods for a clustering vector c_k, respectively, $F(\mathrm{c}_k)$ is a penalty factor for low rank clusters such that

$$F(\mathrm{c}_k) = \exp\left(\mathrm{c}_k \mathrm{c}_k^{\mathrm{T}}\right)^{-1}, \tag{6}$$

and α_1 and α_2 are normalized weighting factors. Note that α_2 is a limiting factor intended to restrict the selection of clusters that define 'sub-codas'. For the case shown in the upper left panel in Fig. 9 for example, a large α_2 will favor the choice of the entire structure 1+1+4 over the sub-structure of 4 clicks with identical ICI, although both represent valid coda structures. On the other hand, a small α_2 allows the identification of new, unknown or rare coda types.

Four constraints are used. In (3b) a utility function above a detection threshold ρ_d is searched for. In (3c) we restrict the solution to a unique clustering. In (3d), we expect a valid coda click to contain at least $\mathscr{P}_{\min}$ multipulses and limit the number of multipulses in a click, averaged over c_k, to $\mathscr{P}(\mathrm{c}_k)$. While both coda and echolocation clicks consist of repetitive pulses of decreasing intensity, codas are characterized by a much lower decay rate[19]. This is possible thanks to the whale's ability to use the right nasal passage as an adaptive acoustic valve to direct the signal towards the junk in the case of echolocation clicks, or back in the spermaceti in the case of coda signals[55]. The low decay rate of the intermediate pulses is reflected in a larger number of pulse repetitions, resulting in a more pronounced periodic structure of the overall signal. A method for estimating the number of multipulses is presented in "Interpulse analysis". Our fourth condition in (3e) considers that the resonant frequency of the coda click, averaged over c_k, $f_r(\mathrm{c}_k)$, should be below a threshold $f_r^{\max}$, and following[19], we use $f_r^{\max} = 12$ kHz. The resonant frequency is calculated using a super-resolution spectrogram in[32].

Due to the requirement for unique clustering, the optimal solution of () is an NP-hard problem with similarities to the graph coloring problem[56]. In the case of many ROI candidates, a suboptimal solution is offered based on the following iterative procedure:

$$\hat{c}_k = \mathrm{argmax}_{c_k} \mathscr{J}_{(\mathrm{c}_k)} \tag{7a}$$

$$\text{s.t.}\quad \mathscr{J}_{(\mathrm{c}_k)} > \rho_d \tag{7b}$$

$$\mathscr{P}(\mathrm{c}_k) > \mathscr{P}_{\min}\,, \tag{7c}$$

$$f_r(\mathrm{c}_k) < f_r^{\max}\,. \tag{7d}$$

In each iteration, cluster candidates c_k that do not meet the orthogonality condition,

$$c_l \cdot \hat{c}_k^T = 0,\ \forall l \neq k. \tag{8}$$

are omitted. The iterations stop when no more valid clusters are found.

Structural likelihood analysis

To represent the similarity between a transient pair, each detected transient is treated as a node in an undirected graph. Let $s_{i,j}$ denote the similarity between a node pair *i* and *j*, where $s_{i,i} = 0$. These similarities combine an affinity matrix

$$S_{M\times M} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix}.$$

For a candidate cluster, c_k, the structural likelihood is defined as the average similarity between the cluster's nodes

$$\mathscr{L}^{\mathrm{s}}(\mathrm{c}_k) = \frac{\mathrm{c}_k S \mathrm{c}_k^{\mathrm{T}}}{0.5(\mathrm{c}_k \mathrm{c}_k^{\mathrm{T}} - 1)\mathrm{c}_k \mathrm{c}_k^{\mathrm{T}}}. \tag{9}$$

Three similarity measures are used: 1) waveform-based similarity, $s_{i,j}^{\mathrm{shape}}$, 2) IPI-based similarity, $s_{i,j}^{\mathrm{IPI}}$ and 3) intensity-based similarity, $s_{i,j}^{\mathrm{I}}$. The first one is calculated by the cross-correlation between the pair of clicks $y_i(t)$ and $y_j(t)$. Formally,

$$s_{i,j}^{\text{shape}} = \frac{\int y_i(t) \cdot y_j(t) dt}{\sqrt{\int y_i(t).^2 dt} \cdot \sqrt{\int y_j(t).^2 dt}}. \tag{10}$$

The second similarity metric requires the evaluation of the click's IPI e.g., using the method in[51], and is defined by the normalized difference

$$s_{i,j}^{\text{IPI}} = 1 - \frac{|\text{IPI}_i - \text{IPI}_j|}{\max(\text{IPI}_i, \text{IPI}_j)}. \tag{11}$$

Since the IPI is a function of the spermaceti size[57], the measure $s_{i,j}^{\text{IPI}}$ allows the separation between overlapping clicks of whales of different sizes. Similarly, the intensity-based similarity is defined by the difference

$$s_{i,j}^{\text{I}} = 1 - \frac{|\text{I}_i - \text{I}_j|}{\max(\text{I}_i, \text{I}_j)}, \tag{12}$$

where I_i is the RMS amplitude of the ith click in the cluster. Since underwater acoustic propagation is strongly dependent on distance, $s_{i,j}^{\text{I}}$ allows the discrimination of clicks emitted by whales at different distances from the receiver. The combined similarity measure $s_{i,j}$ is defined by the normalized weighted sum

$$s_{i,j} = \rho^{\text{corr}} s_{i,j}^{\text{corr}} + \rho^{\text{IPI}} s_{i,j}^{\text{IPI}} + \rho^{\text{I}} s_{i,j}^{\text{I}}, \tag{13}$$

where $\rho^{\text{corr}} + \rho^{\text{IPI}} + \rho^{\text{I}} = 1$.

Figure 11 shows a distribution function from our database for codas of 5 clicks and its projection into the temporal likelihood by the above procedure. Each point in the projection map represents the probability that a potential cluster of ICIs represented by the principal component represents a valid coda. We observe five non-overlapping clusters, each representing a different type of coda: 2+3, 5R1, 5R2, 5R3 and 1+1+3. By (20), the first four codas are represented by $\hat{K}_g = 1$, while for the 1+1+3 coda type we estimated $\hat{K}_g = 3$.

Temporal likelihood analysis

Let H be an $L_W \times W$ matrix containing the ICI patterns of all codas from a legacy database for codas containing $W + 1$ clicks. The number of rows, L_W, is the number of codas within the database, and each row in H is a vector of W ICI measurements for the coda, i.e. the time difference between each pair of successive clicks within the coda. We use H to evaluate the temporal likelihood of a possible cluster c_k to represent a coda signal. Rhythmic features of H are extracted by principle component analysis (PCA) to form a $Q \times W$ $(Q < L_W)$ representative matrix

$$\text{H}_{\text{R}} = \overline{\text{H}} \cdot \overline{\text{U}}\,, \tag{14}$$

where $\overline{\text{H}}$ is an unbiased version of H in which the sampled mean of each column of H is shifted to zero, and $\overline{\text{U}} \in \mathbb{R}^{W \times Q}$ is a matrix representation of the Q eigenvectors corresponding to the largest Q eigenvalues of H. Matrix $\overline{\text{U}}$ is used also to extract the same Q features from the ICI pattern, $\text{ICI}_{(c_k)}$, of the examined potential c_k cluster, denoted by

$$\text{O}_{(c_k)} = \text{ICI}_{(c_k)} \cdot \overline{\text{U}}\,. \tag{15}$$

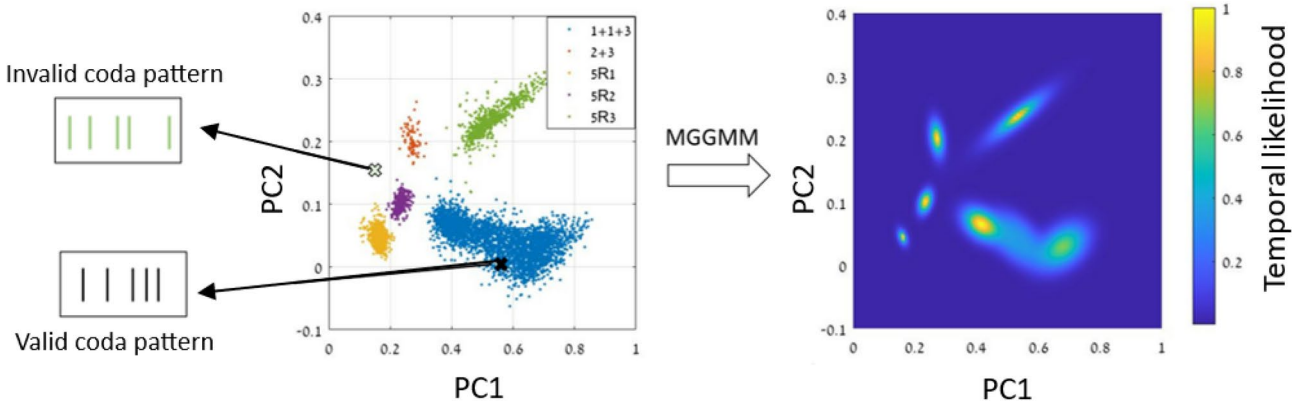


Figure 11. Left: distribution function from our database for the two significant eigenvectors, PC1 and PC2, for codas with 5 clicks. Right: projection map representing the temporal likelihood that an observed cluster of transients is associated with a valid coda.

To evaluate the probability of $\mathrm{O}_{(\mathrm{c}_k)}$ to be a valid coda representation, we calculate the temporal likelihood

$$\mathscr{L}^{\mathrm{t}}_{(\mathrm{c}_k)} = \sum_{g=1}^{G_W} P_r(\mathrm{O}_{(\mathrm{c}_k)} \implies g|\Theta_W), \tag{16}$$

where the $\mathrm{O}_{(\mathrm{c}_k)} \implies g$ term implies that $\mathrm{O}_{(\mathrm{c}_k)}$ is a feature representation of coda type g, G_W is the number of coda types available in H, and Θ_W is the set of parameters for the distribution of the coda features in H_R.

To calculate (16), we evaluate Θ_W from H_R. We estimate the distribution of each coda type g separately. Since the coda type is associated with the number of clicks, in the following we drop the sub-index W. Denote $h_g \in \mathbb{R}^{1\times Q}$ as a subset of H_R corresponding to coda type g, and let Θ_g be its distribution parameters. We model h_g as the mixture model

$$p(h_g|\Theta_g) = \sum_{k=1}^{K_g} \phi_k \cdot f_k(h_g|\theta_{g,k}) \tag{17}$$

of K_g clusters with prior

$$\sum_{k=1}^{K_g} \phi_k = 1$$

and the generalized Gaussian distribution[2][59],

$$f_k(h_g|\theta_{g,k}) = \frac{\Gamma(\frac{Q}{2})}{\pi^{\frac{Q}{2}}\Gamma(\frac{Q}{2\beta_{g,k}})2^{\frac{Q}{2\beta_{g,k}}}} \frac{\beta_{g,k}}{m^{\frac{Q}{2}}|\Sigma_{g,k}|^{\frac{1}{2}}} \cdot \exp\{-\frac{1}{2m^{\beta_{g,k}}}(h_g-\mu_{g,k})^T\Sigma^{-1}_{g,k}(h_g-\mu_{g,k})\}. \tag{18}$$

where $m \in \mathbb{R}^{1\times Q}$ is a vector of scale parameters, $\theta_{g,k} = \{\mu_{g,k}, \Sigma_{g,k}, \beta_{g,k}\}$ is the distribution parameter set with the expectation vector $\mu_{g,k} \in \mathbb{R}^{1\times Q}$, the shape vector $\beta_{g,k} \in \mathbb{R}^{1\times Q}$ and the symmetric covariance matrix $\Sigma_{g,k} \in \mathbb{R}^{Q\times Q}$.

For L_g being the number of codas of type g in H and h^l_g being the corresponding PCA-induced feature vector, we find the distribution parameters $\hat{\Theta}_g = \{\phi_k, m, \theta_{g,k}\}$ by the maximum likelihood

$$\hat{\Theta}_g = \arg\max_{\Theta_g} \sum_{l=1}^{L_g} \log\left(p(h^l_g|\Theta_g)\right), \tag{19}$$

To solve (19) we use a combination of the expectation maximization (EM) approach with the Riemannian averaged fixed-point (RA-FA) learning method according to the steps presented in[60] and[61], respectively.

Note that the above method requires the estimation of the number of clusters, K_g, in (17). To estimate K_g, we consider the Bayesian information criterion (BIC)[62], which is effective when the dimensionality is small and the sample size L_g is larger than the distribution parameters. Applied to the above case

$$\hat{K}_g = \underset{K_g}{\mathrm{argmin}}(\mathrm{BIC}), \tag{20}$$

where

$$\mathrm{BIC} = \ln(L_g)\left(\sum_{k=1}^{K_g}(|\theta_{g,k}|) - 1\right) - 2\cdot\ln\left(\sum_{l=1}^{L_g}\log\left(p(h^l_g|\hat{\Theta}_g)\right)\right). \tag{21}$$

Finally, plugging (17) and (19) into (16), the temporal likelihood is calculated by

$$\mathscr{L}^{\mathrm{t}}_{(\mathrm{c}_k)} = \sum_{g=1}^{G_W} p(\mathrm{O}_{(\mathrm{c}_k)}|\hat{\Theta}_g). \tag{22}$$

Interpulse analysis

Next, we discuss the process of calculating the average number of multipulses $\mathscr{P}(\mathrm{c}_k)$ for (3d) to separate codas from echolocation clicks. To identify interpulses, we use phase-slope analysis. Based on the transient nature of pulses, an interpulse is determined at a positive zero crossing of the phase slope function (PSF)[17]

[2]The General Gaussian distribution is chosen by its flexibility to represent multiple distribution types through its shape parameter β[58]

$$\mathrm{PSF}_n = -\frac{1}{J}\sum_{\omega_1}^{\omega_J} \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{X_R^2(\omega) + Y_R^2(\omega)}, \tag{23}$$

where $X(\omega) = X_R(\omega) + jX_I(\omega)$, and $Y(\omega) = Y_R(\omega) + jY_I(\omega)$, are the Fourier transforms of z_n from (2) and nz_n, respectively, and R and I stand for the real and imaginary parts, respectively.

The number of identified interpulses $\mathscr{P}(c_k)$, is calculated by the number of positive zero crossings in the PSF across the ROI of the cluster c_k. Figure 12 shows an example of the calculated PSF of code clicks with high and low SNR as well as of echolocation click. The examples illustrate the robustness of the method to weak signals, with more interpulses observed for coda clicks even at low SNR. We note that more pulses are expected for an echolocation click in the case of on-axis, i.e., when the whale's head is pointed at the receiver. However, this is a rare scenario.

Examples of detection and annotation

Examples of the operation of the detector and annotator are shown in Fig. 13. Figure 13a shows the ability to detect both focal and non-focal whales in a wide dynamic range, where the intensity of the coda signal of the focal whale (red markers) is 200 times stronger than that of the non-focal whale (yellow markers). An example of the annotator's ability to separate overlapping codas can be seen in Fig. 13b, where three overlapping codas with similar SNR can be seen. Figure 13a shows an example of the separation of a coda signal from an echolocation signal with similar acoustic intensity.

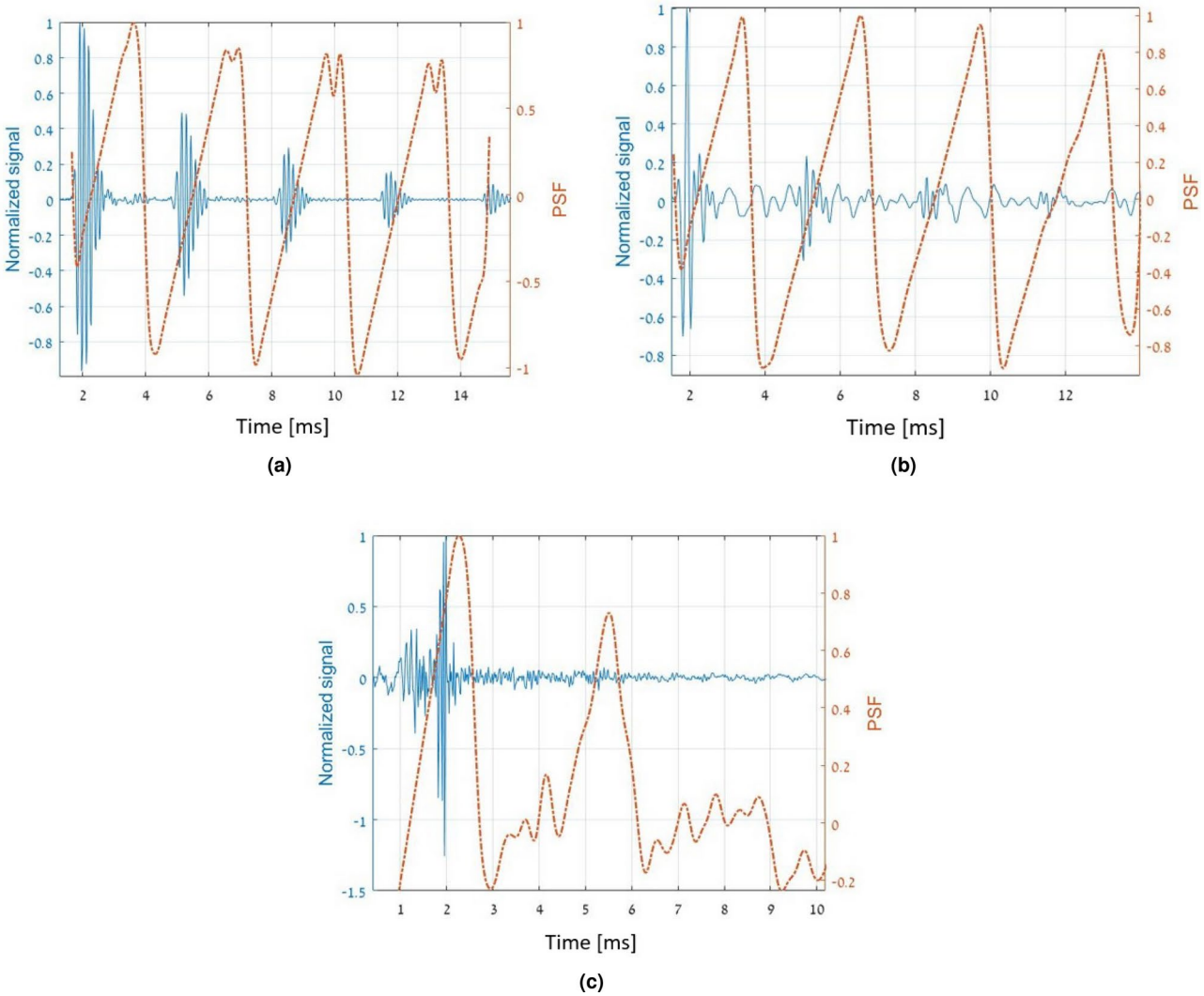
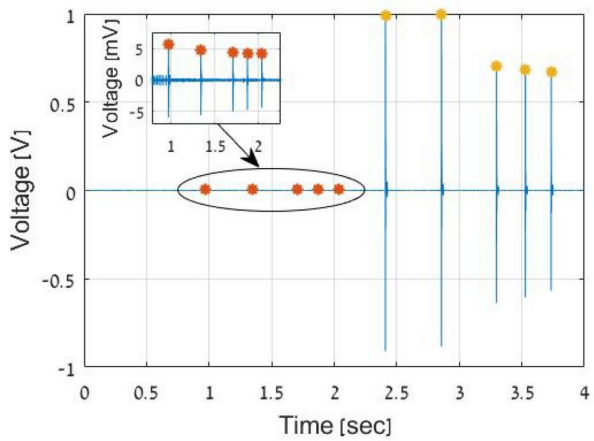
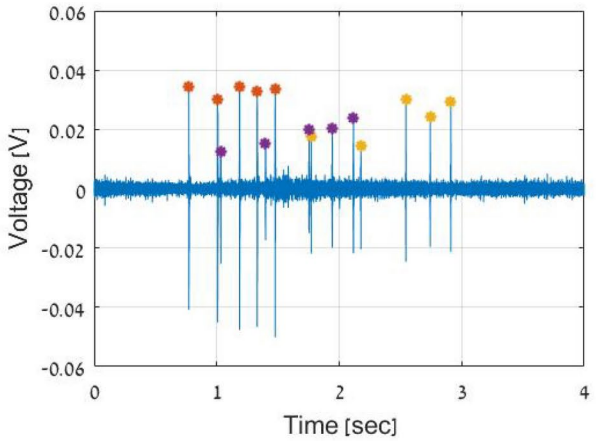


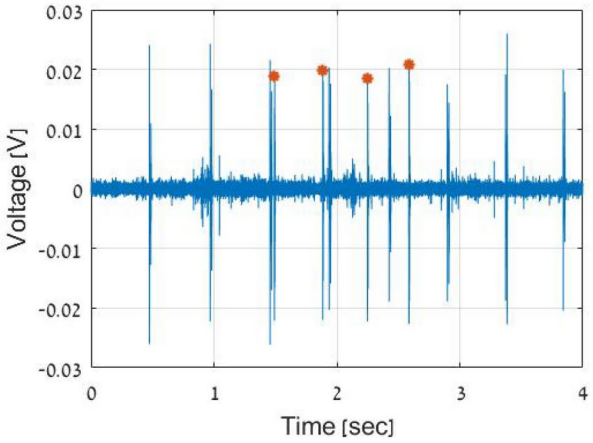
Fig. 12. (**a**) Tag measurement of a coda click, (b) Remote measurement of a coda click, (**a**) remote measurement of echolocation click. The corresponding estimated PSF from (23) is marked in dashed red lines. Each positive zero crossing of the PSF indicate the presence of a pulse.



(a) ***Wide dynamic range***: example of near-field annotation for two codas, one with high SNR from the focal whale (red markers) and one with low SNR from a distant non-focal whale (yellow markers). A zoom in of the low SNR coda is shown in the top left box.



(b) ***Cocktail-party***: example of far-field annotation of three overlapping codas, each marked with a different color.



(c) ***High interference***: example of far-field annotation of a coda signal in the presence of echolocation clicks of similar intensity. Red circles mark the detected coda clicks.

Fig. 13. Examples for the operation of the coda detector and annotator in three challenging scenarios.

Data availability

Code availability of the implementation as well as recording examples used in the paper are available in the link https://github.com/GuyGuva/Coda-detector[50].

Code availability

Readers can access the implementation code with no restrictions. The source code of our implementation is publicly available on GitHub (https://github.com/Project-CETI/Coda-detector/tree/v1).

Received: 18 March 2024; Accepted: 1 April 2025
Published online: 14 April 2025

References

1. Andreas, J. et al. Towards understanding the communication in sperm whales. *iScience* **25**, 104393. https://doi.org/10.1016/j.isci.2022.104393 (2022).
2. Goldwasser, S., Gruber, D., Kalai, A. T. & Paradise, O. A theory of unsupervised translation motivated by understanding animal communication. In *Thirty-Seventh Conference on Neural Information Processing Systems* (2023).
3. Leitao, A. et al. Evidence of social learning across symbolic cultural barriers in sperm whales. *eLife* **13** (2024).
4. Beguš, G., Sprouse, R. L., Leban, A. & Gero, S. Vowels and diphthongs in sperm whales. *OSF* (2023).
5. Sharma, P. et al. Contextual and combinatorial structure in sperm whale vocalisations. *Nat. Commun.* **15**, 3617 (2024).
6. Sharma, P., Gero, S., Rus, D., Torralba, A. & Andreas, J. A machine learning model of sperm whale communication predicts vocal exchanges and behaviour. *bioRxiv* (2024).

7. Jablonski David, E. S. M. Mass extinctions and their rebounds: A macroevolutionary framework. *Paleobiology* 1–14. https://doi.org/10.1017/pab.2024.13 (2025).
8. Müller, J. et al. Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests. *Nat. Commun.* **14**, 6191 (2023).
9. Brakes, P. et al. Animal cultures matter for conservation. *Science* **363**, 1032–1034 (2019).
10. Brakes, P. et al. A deepening understanding of animal culture suggests lessons for conservation. *Proc. R. Soc. B* **288**, 20202718 (2021).
11. Rendell, L. E. & Whitehead, H. Vocal clans in sperm whales (Physeter macrocephalus). *Proc. R. Soc. B* **270**, 225–231. https://doi.org/10.1098/rspb.2002.2239 (2003).
12. Hersh, T. A., Gero, S., Rendell, L. & Whitehead, H. Using identity calls to detect structure in acoustic datasets. *Methods Ecol. Evolut.* **12**, 1668–1678 (2021).
13. Sánchez-García, A., Bueno-Crespo, A. & Sancho-Gómez, J. An efficient statistics-based method for the automated detection of sperm whale clicks. *Appl. Acoust.* **71**, 451–459 (2010).
14. Roch, M. A., Soldevilla, M. S., Hoenigman, R., Wiggins, S. M. & Hildebrand, J. A. Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes. *Can. Acoust.* **36**, 41–47 (2008).
15. Lohrasbipeydeh, H., Dakin, D. T., Gulliver, T. A., Amindavar, H. & Zielinski, A. Adaptive energy-based acoustic sperm whale echolocation click detection. *IEEE J. Ocean. Eng.* **40**, 957–968 (2014).
16. Kandia, V. & Stylianou, Y. Detection of sperm whale clicks based on the Teager-Kaiser energy operator. *Appl. Acoust.* **67**, 1144–1163 (2006).
17. Kandia, V. & Stylianou, Y. A phase based detector of whale clicks. In *2008 New Trends for Environmental Monitoring Using Passive Systems*. 1–6 (IEEE, 2008).
18. Beslin, W. A., Whitehead, H. & Gero, S. Automatic acoustic estimation of sperm whale size distributions achieved through machine recognition of on-axis clicks. *J. Acoust. Soc. Am.* **144**, 3485–3495 (2018).
19. Madsen, P. et al. Sperm whale sound production studied with ultrasound time/depth-recording tags. *J. Exp. Biol.* **205**, 1899–1906 (2002).
20. Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A. & Lund, A. The monopulsed nature of sperm whale clicks. *J. Acoust. Soc. Am.* **114**, 1143–1154 (2003).
21. Watkins, W. A. & Schevill, W. E. Sperm whale codas. *J. Acoust. Soc. Am.* **62**, 1485–1490 (1977).
22. Schulz, T. M., Whitehead, H., Gero, S. & Rendell, L. Overlapping and matching of codas in vocal interactions between sperm whales: Insights into communication function. *Anim. Behav.* **76**, 1977–1988 (2008).
23. Weilgart, L. & Whitehead, H. Coda communication by sperm whales (Physeter macrocephalus) off the Galapagos Islands. *Can. J. Zool.* **71**, 744–752 (1993).
24. Rendell, L. & Whitehead, H. Do sperm whales share coda vocalizations? Insights into coda usage from acoustic size measurement. *Anim. Behav.* **67**, 865–874 (2004).
25. Rendell, L. & Whitehead, H. Comparing repertoires of sperm whale codas: A multiple methods approach. *Bioacoustics* **14**, 61–81 (2003).
26. Gero, S., Whitehead, H. & Rendell, L. Individual, unit and vocal clan level identity cues in sperm whale codas. *R. Soc. Open Sci.* **3**, 150372 (2016).
27. Vachon, F., Hersh, T. A., Rendell, L., Gero, S. & Whitehead, H. Ocean nomads or island specialists? Culturally driven habitat partitioning contrasts in scale between geographically isolated sperm whale populations. *R. Soc. Open Sci.* **9**, 211737 (2022).
28. Sharma, P. et al. Contextual and combinatorial structure in sperm whale vocalisations. *bioRxiv* 2023–12 (2023).
29. Huggenberger, S., André, M. & Oelschläger, H. H. A. The nose of the sperm whale: Overviews of functional design, structural homologies and evolution. *J. Mar. Biol. Assoc. UK* **96**, 783–806. https://doi.org/10.1017/S0025315414001118 (2016).
30. Hersh, T. A. et al. Evidence from sperm whale clans of symbolic marking in non-human cultures. *Proc. Natl. Acad. Sci.* **119**, e2201692119. https://doi.org/10.1073/pnas.2201692119 (2022).
31. Johnson, M. P. & Tyack, P. L. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Ocean. Eng.* **28**, 3–12 (2003).
32. Moca, V. V., Bârzan, H., Nagy-Dăbâcan, A. & Mureșan, R. C. Time-frequency super-resolution with superlets. *Nat. Commun.* **12**, 337 (2021).
33. Schulz, T. M., Whitehead, H. & Rendell, L. Off-axis effects on the multi-pulse structure of sperm whale coda clicks. *J. Acoust. Soc. Am.* **125**, 1768–1773 (2009).
34. Jacobs, E. R. et al. The active space of sperm whale codas: Inter-click information for intra-unit communication. *J. Exp. Biol.* jeb-246442 (2024).
35. Johnson, M. P. & Tyack, P. L. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Ocean. Eng.* **28**, 3–12 (2003).
36. Møhl, B., Larsen, E. & Amundin, M. Sperm whale size determination: Outlines of an acoustic approach. *Mamm. Seas Fish. Ser. No. 5 Mamm. Seas* **3**, 327–331 (1981).
37. Gubnitky, G. & Diamant, R. *Detecting the presence of sperm whales' echolocation clicks in noisy environments* (IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024).
38. Vachon, F., Rendell, L., Gero, S. & Whitehead, H. Abundance estimate of eastern caribbean sperm whales using large scale regional surveys. *Mar. Mamm. Sci.* https://doi.org/10.1111/mms.13116.
39. Gero, S., Gordon, J., Carlson, C., Evans, P. & Whitehead, H. Population estimate and inter-island movement of sperm whales, Physeter macrocephalus, in the eastern Caribbean sea. *J. Cetacean Res. Manag.* **9**, 143–150 (2007).
40. Gero, S. et al. Behavior and social structure of the sperm whales of Dominica, West Indies. *Mar. Mamm. Sci.* **30**, 905–922 (2014).
41. Gero, S., Bøttcher, A., Whitehead, H. & Madsen, P. T. Socially segregated, sympatric sperm whale clans in the Atlantic Ocean. *R. Soc. Open Sci.* **3**, 160061 (2016).
42. Schulz, T. M., Whitehead, H., Gero, S. & Rendell, L. Individual vocal production in a sperm whale (Physeter macrocephalus) social unit. *Mar. Mamm. Sci.* **27**, 149–166 (2011).
43. Whitehead, H. Consensus movements by groups of sperm whales. *Mar. Mamm. Sci.* **32**, 1402–1415 (2016).
44. Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A. & Lund, A. The monopulsed nature of sperm whale clicks. *J. Acoust. Soc. Am.* **114**, 1143–1154 (2003).
45. Au, W. W., Branstetter, B., Moore, P. W. & Finneran, J. J. The biosonar field around an Atlantic bottlenose dolphin (Tursiops truncatus). *J. Acoust. Soc. Am.* **131**, 569–576 (2012).
46. Au, W. W., Branstetter, B., Moore, P. W. & Finneran, J. J. Dolphin biosonar signals measured at extreme off-axis angles: Insights to sound propagation in the head. *J. Acoust. Soc. Am.* **132**, 1199–1206 (2012).
47. Cheng, J. T., Tracy, J. L., Ho, S. & Henrich, J. Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *J. Exp. Psychol. Gen.* **145**, 536 (2016).
48. Kavanagh, E. et al. Dominance style is a key predictor of vocal use and evolution across nonhuman primates. *R. Soc. Open Sci.* **8**, 210873 (2021).
49. Arnbom, T. Individual photographic identification : A key to the social organization of sperm whales (1987).
50. Project-CETI. Link to our Matlab implementation code. 10.5281/zenodo.14902261 (2024). https://github.com/Project-CETI/Coda-detector/tree/v1

51. Gubnitsky, G. & Diamant, R. Inter-pulse estimation for sperm whale click detection. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5 (IEEE, 2023).
52. Kaiser, J. F. On a simple algorithm to calculate the 'energy' of a signal. In *International Conference on Acoustics, Speech, and Signal Processing*. 381–384 (IEEE, 1990).
53. Pavan, G. et al. Time patterns of sperm whale codas recorded in the Mediterranean sea 1985–1996. *J. Acoust. Soc. Am.* **107**, 3487–3495 (2000).
54. Amano, M., Kourogi, A., Aoki, K., Yoshioka, M. & Mori, K. Differences in sperm whale codas between two waters off Japan: Possible geographic separation of vocal clans. *J. Mammal.* **95**, 169–175 (2014).
55. Huggenberger, S., André, M. & Oelschläger, H. H. An acoustic valve within the nose of sperm whales (Physeter macrocephalus). *Mamm. Rev.* **44**, 81–87 (2014).
56. Garey, M. R. Computers and intractability: A guide to the theory of np-completeness, Freeman. *Fundamental* (1997).
57. Gordon, J. C. Evaluation of a method for determining the length of sperm whales (Physeter catodon) from their vocalizations. *J. Zool.* **224**, 301–314 (1991).
58. Diamant, R., Tan, H.-P. & Lampe, L. Los and nlos classification for underwater acoustic localization. *IEEE Trans. Mobile Comput.* **13**, 311–323 (2012).
59. Kotz, S. Multivariate distributions at a cross road. In *A Modern Course on Statistical Distributions in Scientific Work: Volume 1-Models and Structures Proceedings of the NATO Advanced Study Institute held at the University of Calgagry, Calgary, Alberta, Canada* July 29–August 10, 1974. 247–270 (Springer, 1975).
60. Boukouvalas, Z., Said, S., Bombrun, L., Berthoumieu, Y. & Adalı, T. A new Riemannian averaged fixed-point algorithm for mggd parameter estimation. *IEEE Signal Process. Lett.* **22**, 2314–2318 (2015).
61. Najar, F., Bourouis, S., Zaguia, A., Bouguila, N. & Belghith, S. Unsupervised human action categorization using a Riemannian averaged fixed-point learning of multivariate ggmm. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15*. 408–415 (Springer, 2018).
62. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* 461–464 (1978).

Acknowledgements

This study was funded by Project CETI via grants from Dalio Philanthropies and Ocean X; Sea Grape Foundation; Virgin Unite, and Rosamund Zander/Hansjorg Wyss through The Audacious Project: a collaborative funding initiative housed at TED. We thank the Chief Fisheries Officers and the Dominica Fisheries Division officers for research permits and their collaboration in data collection; all the crews of R/V Balaena and The DSWP team for data collection, curation, and annotation of audio recordings across 15 years; as well as Dive Dominica, Al Dive, and W.E.T. Dominica for logistical support while in Dominica. We are grateful to Kristian Beedholm for CodaSorter, as well as Mark Johnson and Peter L. Tyack for their in-kind contribution of Dtags and associated code during DSWP field research in 2014-2018.

Author contributions

G.G conceived the detector and annotator design and processed the data. Y.M. assisted in CETI data collection and manual annotation. S.G conducted and funded DSWP fieldwork, collected, annotated, and curated DSWP dataset, assisted in CETI data collection, edited the manuscript, and assisted in deriving conclusions. D.F.G provided funding and management and assisted in deriving conclusions. R.D conceived the detector and annotator design, assisted in CETI data collection and was in charge of manuscript write up.

Funding

Project CETI via grants from Dalio Philanthropies and Ocean X; Sea Grape Foundation; Virgin Unite, and Rosamund Zander/Hansjorg Wyss through The Audacious Project

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to R.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025