Detection of Mysticete Calls: a Sparse Representation-Based Approach

François-Xavier Socheleau, Flore Samaran

IMT Atlantique
Dépt. Signal & Communications
Technopôle de Brest-Iroise - CS 83818
29238 Brest Cedex 3
Téléphone : +33 (0)2 29 00 13 04
Télécopie : +33 (0)2 29 00 10 12
URL : **www.imt-atlantique.fr**

Collection des rapports de recherche d'IMT Atlantique
RR-2017-04-SC

François-Xavier Socheleau
Flore Samaran

Detection of Mysticete Calls: a Sparse Representation-Based Approach

Date d'édition : 30 octobre 2017
Version : 1.0

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Contents

Figures

Tables

Abstract

This paper presents a methodology for automatically detecting mysticete calls. This methodology relies on sparse representations of these calls combined with a detection metric that explicitly takes into account the possible presence of interfering transient signals. Sparse representations can capture the possible variability observed for some vocalizations and can automatically be learned from the time series of the digitized acoustic signals, without requiring prior transforms such as spectrograms, wavelets or cepstrums. The proposed framework is general and applicable to any mysticete call lying in a linear subspace described by a dictionary-based representation. The potential of the detector is illustrated on North Atlantic blue whale D calls extracted from the DCLDE 2015 low frequency database as well as on "Madagascar" pygmy blue whale calls extracted from the OHASISBIO 2015 database. Receiver operating characteristic curves (ROC) are calculated and performance is compared with three other methods used for automatic call detection: the XBAT bank of matched spectrograms, a bank of matched filters derived from a generalized likelihood ratio approach and a kernel-based spectrogram detector. On the test data, the ROC curves show that the proposed detector outperforms these three methods.

1. Introduction

Long-term passive acoustic monitoring (PAM) has been used successfully in many oceans of the world to study the presence, distribution and migration patterns of baleen whales, or mysticetes [1, 2]. Mysticetes are ideal candidates for PAM since they are known to produce different types of distinctive sounds year-round or at a specific season, depending on the species [3, 4]. Their repertoire is composed of a wide variety of intense, often low-frequency sounds including tonal, frequency-modulated or pulsive sounds. PAM provides an alternative method to traditional visual surveys. It is less affected by weather and sighting conditions and acoustic recorders can collect data continuously throughout days, seasons or years. However, manual detection (aurally or by visual inspection of spectrograms) of the mysticete sounds in these large volume of recordings is a long and laborious task whose efficiency can potentially be affected by the experience and the degree of fatigue of the operator. For most long term acoustic recordings, manual detection is unrealistic. Therefore, the development of efficient and robust automatic detection methods is rising over the past decade [5, 6, 7, 8].

Mysticetes calls are usually not erratic and present some form of "structure" which is often revealed through a local concentration of energy on spectrograms. Such a structure can be formalized by expressing calls into linear expansions of elementary waveforms that belong to a *dictionary* of known functions. These functions are classically extracted from Fourier or chirp bases [9, 10, 11, 12] and represent the salient features of a specific vocalization. When automated detectors are applied on data, the prior knowledge of this dictionary is very useful to discriminate a particular signal of interest from other sounds. Kernel-based spectrogram detectors [9], (bank of) matched spectrograms [13], bank of matched filters [14], or subspace detectors [10] make (implicit) use of such a dictionary.

The way these dictionary-based detectors have been implemented has shown to be very efficient for stereotyped mysticete calls [9, 10]. However, they often fail to capture calls whose time-frequency pattern may differ depending on factors such as season, behavior of the animal, propagation or ambient noise conditions [15, 16, 17, 18]. In such scenarios, the difficulty for detectors is to take into account this variability while avoiding the detection of interfering transient sounds of no interest for the intended application. In [14], the detection of variable mysticete calls is addressed for North Atlantic right whale. It relies on a very general signal model whose parameters are not assumed to be known in advance by the detector. A polynomial-phase signal model is chosen and, for each observation, the unknown call characteristics (start frequency, frequency slope and curvature) are estimated with a maximum likelihood approach. Based on these estimates, a likelihood ratio test is then used. The main drawback of such an approach is that it requires a full statistical model of the acoustic data processed by the detector. Such a model may be very difficult to obtain when the data are heterogeneous and contain a wide variety of transient sounds. Moreover, if the data do not match the statistical model, interfering transient signals

can trigger the detector and generate many false alarms. Designing an ad-hoc post-processing algorithm to remove these false alarms is still possible but it usually requires tune the detector with numerous dataset-specific parameters [12]. Given the huge diversity of sound sources underwater, this kind of methods lacks of general applicability as the features of interferences may greatly vary from one dataset to another.

In this paper, we generalize the standard approach of dictionary-based detection by modeling mysticete calls with sparse representations. Such representations express a given signal as a linear combination of base elements in which many of the coefficients are zero [19]. In our context, the key idea behind sparse representations is to use a dictionary of large dimension spanning all possible variations of the call to detect, while modeling each occurrence of this call as a linear combination of very few elements of that dictionary so as to limit false alarms. Sparse representations have been successfully applied in various fields such as computer vision [20], image and speech denoising [21, 22], compressed sensing [23], and more. The elements of the dictionary can either be chosen from theoretical bases but can also be automatically learned from the time series of real data [19]. Both cases are considered in this work.

The proposed detector is an extension to the sparse framework of the decision statistic presented in [10]. This statistic was shown to offer optimal properties with respect to false alarm and detection probabilities and can be interpreted as an estimate of the signal-to-interference-plus-noise ratio (SINR). This SINR measures the match between the observed data and the assumed sparse representation of the call to detect. As opposed to fully parametric methods, it does not require to learn a priori the features of interfering signals. The potential of the proposed method is illustrated with data extracted from two databases annotated by human analysts and containing different types of blue whale (*Balaenoptera musculus*) calls: the DCLDE 2015 low frequency database [24] and the OHASISIBIO 2015 database [25, 26]. The first database contains North Atlantic blue whale D calls [27]. Repeated with no regular intervals, D calls have been recorded in the presence of blue whales in many locations and have been suggested as contact calls or feeding calls [28]. Unlike stereotyped calls, blue whale D calls have variable characteristics in duration, frequency content and frequency sweep (e.g. downsweeps [29], upsweeps [30], archsounds [31]) with no obvious geographic variation [30, 32]. The second database contains "Madagascar" call types produced by pygmy blue whales in the west and central part of the Indian Ocean [33]. This call type consists of a phrase with two long units repeated in patterned sequences every 90-100 s, over a period extending from a few minutes to hours [34]. These two databases exhibit complementary characteristics in terms of call variability, call complexity, signal-to-noise ratio and occurrence of interfering transient signals and are thus relevant to illustrate the general applicability of the proposed method.

The paper is organized as follows. In Sec. 2, the observation model as well as the dictionary-based sparse representation of mysticete calls are presented. Sec. 3 details the detection strategy, which is then assessed with real data in Sec. 4. Finally, conclusions are given in Sec. 5.

Notation: Throughout this paper, lowercase boldface letters denote vectors, e.g., $\boldsymbol{x}$, and uppercase boldface letters denote matrices, e.g., $\mathbf{A}$. The superscript ^{T} means transposition. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$. $\|\cdot\|_p$ designates the ℓ_p norm and $\|\cdot\|_F$ is the Frobenius norm. The symbol $\odot$ denotes the Hadamard (entrywise) product between matrices. The cardinality of a set $\mathcal{A}$ is denoted $|\mathcal{A}|$. Finally, the distribution of a Gaussian random vector with mean $\mathbf{m}$ and covariance matrix Σ is denoted $\mathcal{N}(\mathbf{m}, \Sigma)$.

2. Data model

2.1. Observation model

PAM systems process digitized time series representing underwater sounds received on hydrophones. The acquired data result from the mixture of signals of different nature. From a detector perspective, these signals can be classified into three categories:

- The mysticete sound of interest (sound to detect when present in the data).
- The transient noise or interference, which designates any transient signal of no interest for the intended application (e.g., ship noise, airguns, earthquakes, ice tremors, calls of other whales, etc.).
- The background noise that results from the mixture of numerous unidentifiable ambient sound sources. As opposed to what is called interference in this work, background noise does not include any transient signal.

Therefore, given an observation window of N samples, the observation vector $\boldsymbol{y} \in \mathbb{R}^N$ is here represented as

$$\boldsymbol{y} \overset{\Delta}{=} \mu \boldsymbol{s} + \epsilon \boldsymbol{\psi} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{s} \in \mathbb{R}^N$ designates the signal of interest, $\boldsymbol{\psi} \in \mathbb{R}^N$ is the interference and $\boldsymbol{w} \in \mathbb{R}^N$ is the background noise. μ and ϵ are random variables valued in $\{0, 1\}$ modeling the possible presence or absence of $\boldsymbol{s}$ and $\boldsymbol{\psi}$, respectively. Eq. (1) is here assumed to model the recorded time series after signal shaping. Classical shaping includes whitening or spectral equalization and band-pass filtering [6] (it sometimes includes baseband conversion as well, in this case, $\boldsymbol{y}$ is complex-valued).

As detailed in Sec. 2.2, the signal of interest is assumed to be "structured", which here means that it lies in a linear subspace that can be described by a dictionary-based representation. No parametric model is assumed for interferences since they can be very heterogeneous and are often random. However, to differentiate $\boldsymbol{\psi}$ from $\boldsymbol{s}$, some assumptions must be made. $\boldsymbol{\psi}$ is here defined as any transient signal whose energy lies mostly outside the subspace in which $\boldsymbol{s}$ resides (more details are given in Sec. 2.3). Finally, in agreement with several statistical analyzes conducted in the frequency range of mysticete sounds [14, 10], the background noise $\boldsymbol{w}$ is modeled as a Gaussian vector.

2.2. Dictionary-based sparse representation of mysticete calls

The vast majority of mysticete calls are not erratic and present some structures which are often visible on time-frequency representations such as spectrograms. Formally, these structures can be taken into account by assuming that the signal of interest can be decomposed into a linear expansions of $M < N$ waveforms, called *atoms*, that belong to a dictionary of functions, i.e.,

$$s(n) \approx \sum_{m=0}^{M-1} d_m(n) \times \theta_m, \tag{2}$$

where $0 \leq n \leq N-1$ is the time index and θ_m is a coefficient for the atom $d_m(n)$. Using matrices, (2) can also be expressed as

$$\boldsymbol{s} \approx \mathbf{D}\boldsymbol{\theta}, \tag{3}$$

where $\mathbf{D} \in \mathbb{R}^{N \times M}$ denotes the dictionary and $\boldsymbol{\theta} \in \mathbb{R}^M$ is the vector containing the weighting factors. In other words, $\boldsymbol{\theta}$ is the vector of coordinates of $\boldsymbol{s}$ in a linear subspace spanned by the columns of $\mathbf{D}$. When the sound of interest satisfies this model, the knowledge of $\mathbf{D}$, that characterizes the salient features of $\boldsymbol{s}$, is very useful to design an efficient detector.

Several existing detection methods implicitly rely on the dictionary-based framework of (3). For instance, the well known detectors based on spectrogram correlation or spectrogram kernels [9] assume that the energy of the signal to detect is well localized in the time-frequency plane. In that case, the underlying dictionary is a Fourier-like basis where each oscillating component is weighted by some

masking function whose values are chosen by analyzing the spectrogram of the signal of interest. Formally, $\mathbf{D}$ then satisfies the following relationship

$$\mathbf{D} = (\mathbf{C} \odot \mathbf{P})^T , \tag{4}$$

where the entries of $\mathbf{C} \in \mathbb{R}^{M \times N}$ contain M oscillating atoms, e.g., $[\mathbf{C}]_{mn} = \cos(\pi nm/M)$, with m the frequency index. $\mathbf{P} \in \mathbb{R}^{M \times N}$ denotes the time-frequency mask. In other approaches such as [10, 11], the dictionaries are built from oscillating atoms whose instantaneous frequencies are determined by known analytic expressions.

Model (3) has shown to be very efficient when the sound to detect is highly stereotyped and when its time-frequency pattern is rather simple [10]. In that case, the number M of atoms is rather small. However, such assumptions may be too restrictive. Within the same vocalization class, some mysticete sounds may exhibit a time-frequency variability depending on individuals [35], body condition [36], social behaviors [37], increase in ocean background noise [15], propagation conditions [18], etc. and may also exhibit non trivial frequency patterns. Therefore, the desired dictionary should incorporate enough variability to model all possible calls of the same type (i.e., M large), while limiting the range of variation for a single call so as not to design an interference-sensitive detector. Both goals can be achieved simultaneously by finding a *compact* representation of the signal of interest in terms of linear combination of atoms in a dictionary that can be of *large dimension*. Considering these remarks, model (3) is extended to

$$\boldsymbol{s} \approx \mathbf{D}\boldsymbol{\theta}, \text{ with } \|\boldsymbol{\theta}\|_0 \leq K << M, \tag{5}$$

where $\|\boldsymbol{\theta}\|_0$ denotes the ℓ_0 (pseudo-)norm that returns the number of non-zero coefficients in $\boldsymbol{\theta}$. When $\boldsymbol{s}$ can be represented by a small number of non-zero coefficients in the basis $\mathbf{D}$, model (5) is referred to as *sparse representation* in the signal processing literature [19]. The inequality $\|\boldsymbol{\theta}\|_0 \leq K$ is called the sparsity constraint.

D calls of blue whales [27] are typical examples that fit well with (5). Each single call is a simple frequency-modulated (FM) sweep that could well be approximated by a linear combination of a few atoms. However, such calls exhibit variability in initial frequency, FM rate, duration, and bandwidth. Therefore, the ℓ_0 norm of $\boldsymbol{\theta}$ is small for each single call but the active atoms, corresponding to non-zero entries of $\boldsymbol{\theta}$, can be different from one call to another.

There exist two main approaches to design relevant dictionaries for detection problems. The first approach relies on the choice of theoretical "preconstructed" atoms, usually chosen to match some time-frequency patterns. For PAM systems, this approach is rather standard: the atoms are chosen after analyzing a set of training data and are usually extracted from Fourier, chirp or wavelet bases [9, 10, 11, 14, 38].

The second approach, less common in the PAM context, relies on recent advances in signal processing and constructs empirically-learned dictionaries, in which the generating atoms are automatically designed from the data [19]. Given a set of $L \geq M$ training signals $\{\boldsymbol{s}_i\}_{i=1}^{L}$, such an approach seeks the dictionary $\mathbf{D}$ that leads to the "best possible" representation for each signal in this set with the sparsity constraint of (5). The training signals are stored in a training matrix $\mathbf{S} \in \mathbb{R}^{N \times L}$ and the dictionary $\mathbf{D}$ is found by solving the following minimization problem

$$\min_{\mathbf{D},\Theta} \|\mathbf{S} - \mathbf{D}\Theta\|_F^2 \text{ subject to } \|\boldsymbol{\theta}_i\|_0 \leq K, \forall\, 1 \leq i \leq L, \tag{6}$$

where $\Theta = \{\boldsymbol{\theta}_i\}_{i=1}^{L}$. This problem describes each given signal $\boldsymbol{s}_i$ as the best sparse representation $\boldsymbol{\theta}_i$ over the unknown dictionary $\mathbf{D}$, and aims to jointly find the proper representations and the dictionary [19]. Numerical solutions to problem (6) can be obtained with the method of optimized directions (MOD) [39], K-SVD [40] or other algorithms such as [41].

Choosing whether the dictionary should be designed from a mathematical model of the data or from the data itself is context dependent. For simple chirp-like calls, choosing the first approach may be a good option: the parameters of the model can easily be learned from the training data using standard approaches

such as maximum likelihood estimation and the risk of overfitting the training data may be reduced by choosing simple analytical models. However, for more complex calls, such mathematical models may be over-simplistic and/or not easy to build. In that case, designing the dictionary directly from the data may be a good option. The benefits of both approaches are discussed in Sec. 4.

2.3. Interference

Using model (5) for $\boldsymbol{s}$ is also very useful to bound our lack of knowledge on interfering transient signals. Such signals are often very heterogeneous (abiotic, biological or anthropogenic sound sources) and their features can rarely be represented by a single parametric model. Intuitively, interference $\boldsymbol{\psi}$ is defined as a transient signal having not much in common with the signal of interest $\boldsymbol{s}$. More formally, it can be defined as a signal whose energy lies mostly outside the subspace in which $\boldsymbol{s}$ resides. Based on model (5), this can be expressed as

$$\max_{\mathcal{U}:|\mathcal{U}|=K} \|\mathbf{P}_{\mathbf{D}(\mathcal{U})}\boldsymbol{\psi}\|^2 < \|\boldsymbol{\psi}\|^2/2, \tag{7}$$

where $\mathcal{U}$ is any K-combination of the set $\mathcal{M} \triangleq \{0, 1, \cdots, M-1\}$, $\mathbf{D}(\mathcal{U}) \in \mathbb{R}^{N\times K}$ is the submatrix of $\mathbf{D}$ with columns indexed in $\mathcal{U}$ and $\mathbf{P}_{\mathbf{D}(\mathcal{U})} = \mathbf{D}(\mathcal{U})\left(\mathbf{D}(\mathcal{U})^T\mathbf{D}(\mathcal{U})\right)^{-1}\mathbf{D}(\mathcal{U})^T$ is the projection matrix onto the subspace spanned by the columns of $\mathbf{D}(\mathcal{U})$. Note that $\|\mathbf{P}_{\mathbf{D}(\mathcal{U})}\boldsymbol{\psi}\|^2$ may be non null since $\boldsymbol{\psi}$ and $\boldsymbol{s}$ may not be orthogonal because of partial time-frequency similarities. As opposed to other approaches such as [14, 42, 12], no statistical model is assumed for $\boldsymbol{\psi}$, no particular interference subspace is considered, and it is not assumed that interfering transient sounds share common features that the detector can learn thanks to a training dataset.

3. Detection method

Based on the observation model (1) and for each observation window of size N, our detection problem is to decide whether μ equals 0 or 1. The decision must be made even if the observation contains interfering signals ($\epsilon = 1$). This problem can be cast in the standard binary hypothesis testing framework:

$$\begin{cases} \textbf{Observation : } \boldsymbol{y} = \mu\boldsymbol{s} + \epsilon\boldsymbol{\psi} + \boldsymbol{w}, \text{ with } \boldsymbol{w} \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}_N\right), \\ \textbf{Null hypothesis } \mathrm{H}_0 : \mu = 0, \\ \textbf{Alternative hypothesis } \mathrm{H}_1 : \mu = 1. \end{cases} \tag{8}$$

To be efficient, the detector can take advantage of its knowledge of $\mathbf{D}$. Note that it cannot know $\boldsymbol{\theta}$ *a priori* because (i) the index of the non-zero entries of $\boldsymbol{\theta}$ may change from one call to another and (ii) $\boldsymbol{\theta}$ contains random amplitudes that can strongly be affected by propagation conditions which are not perfectly known to the detector.

Based on the theory of subspace detector [43, 44], we recently proposed in [10] a detector that shows *optimal* properties with respect to false alarm and detection probabilities in the specific case where $K = M$, i.e., without sparsity constraints. This detector can be expressed, in our context, as the following threshold test [10, Eq. (12)]

$$\mathcal{T}_\eta(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \dfrac{\|\mathbf{D}\hat{\boldsymbol{\theta}}\|_2^2}{\|\boldsymbol{y} - \mathbf{D}\hat{\boldsymbol{\theta}}\|_2^2} > \eta, \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where η is the detection threshold and $\hat{\boldsymbol{\theta}}$ is the least-square estimate of $\boldsymbol{\theta}$ which satisfies $\hat{\boldsymbol{\theta}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{y}$. The decision to accept the alternative hypothesis is made when $\mathcal{T}_\eta(\boldsymbol{y}) = 1$. The detection metric can be interpreted as an estimate of the signal-to-interference-plus-noise ratio (SINR): $\mathbf{D}\hat{\boldsymbol{\theta}}$ is an estimate of the signal of interest $\boldsymbol{s}$ and $\boldsymbol{y} - \mathbf{D}\hat{\boldsymbol{\theta}}$ an estimate of the interference plus noise $\epsilon\boldsymbol{\psi} + \boldsymbol{w}$. When $\mu = 0$, the SINR equals and becomes significant for $\mu = 1$. Test (9) was shown in [10] to be asymptotically uniformly most powerful among invariant tests [45, Ch. 6 & 13]. This property states that when N is much larger than K

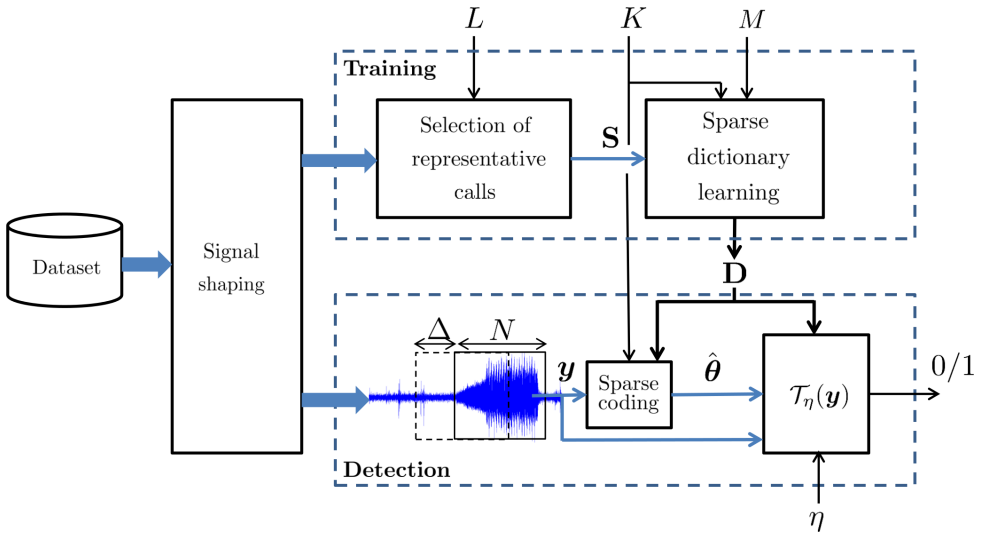

FIGURE 1– Functional block diagram of SRD.

and for a given false alarm probability, $\mathcal{T}_\eta(\boldsymbol{y})$ is the test that yields the highest probability of detection (among invariant tests) for the problem formalized in (5), (7) and (8). Note that in practice, N is often much larger than K (see Sec. 4 and [10] for instance).

When $K << M$, i.e., with a sparsity constraint, we suggest to keep the same threshold test (9) but with a different estimate of $\boldsymbol{\theta}$. This is needed because the least-square estimate of $\boldsymbol{\theta}$ is a vector that is very likely to be filled since no sparsity constraint is enforced. Finding an estimate of $\boldsymbol{\theta}$ with $K << M$ is known as sparse coding in the signal processing literature and is performed with pursuit algorithms [46, 47, 19]. For an observation vector $\boldsymbol{y}$ and a known dictionary $\mathbf{D}$, these algorithms obtain $\hat{\boldsymbol{\theta}}$ as the (approximate) solution of

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^M} \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\theta}\|_2^2 \text{ subject to } \|\boldsymbol{\theta}\|_0 \leq K. \tag{10}$$

The combination of Eq. (9) and (10) is next referred to as the *sparse representation-based detector* (SRD).

In practice, long time-series must be analyzed without knowing the time-of-arrival of each individual call. The standard detection approach in this situation is to repeat test $\mathcal{T}_\eta$ on a sliding window of size N with an overlap of $N - \Delta$ samples between consecutive windows. The overall detection strategy is illustrated in Fig. 1.

4. Illustration with real data

The performance of SRD is illustrated with data extracted from two databases: the DCLDE 2015 low frequency database [24] that contains annotated blue whale D calls and the OHASISBIO 2015 database [25, 26] that contains annotated "Madagascar" pygmy blue whale calls. These two datasets have been chosen because they exhibit complementary characteristics in terms of call variability, call complexity, signal-to-noise ratio and occurrence of transient interfering signals. For the experiments, we do not apply a cross validation procedure as usually done in supervised machine learning approaches (e.g., 60% of the available data for training the detector, another 20% for cross-validation, and the rest for test). We believe that the benefit of an automated detector is the ability, within a data set, to detect much of the data based on few labelled examples from that particular data (saving human analyst's effort). Hence, 10% (or less) training data are randomly selected for the validation. Details are given hereafter.

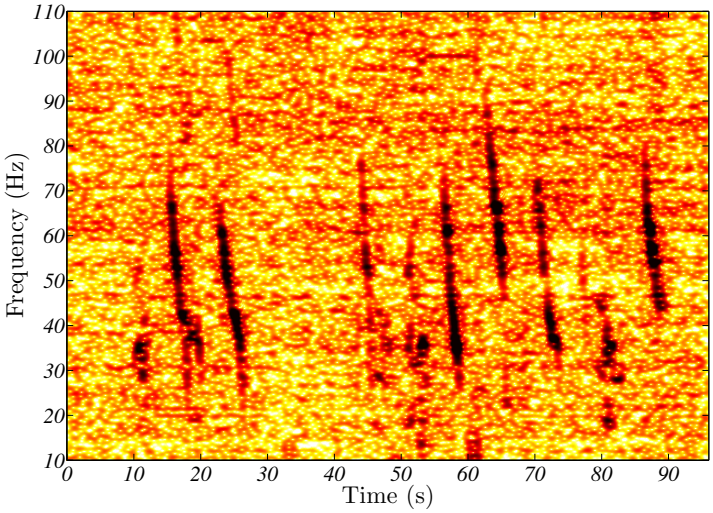

FIGURE 2– Spectrogram of seven consecutive D calls extracted from the DCLDE 2015 dataset.

4.1. DCLDE 2015 dataset

4.1.1. Dataset description

The DCLDE 2015 low frequency database contains annotated blue whale D calls (see Fig. 2) obtained with high-frequency acoustic recording packages deployed in the Southern California Bight. The analysis is conducted on data recorded at the CINMS B site (latitude: 34-17.0 N, longitude: 120-01.7 W) in summer 2012. This dataset contains 906 calls over more than 9 days (223.7 h) and the ambient soundscape is very rich and includes many different kind of interferences (a few examples are shown in Fig. 3). Note also that interferences occur much more often than D calls so that $\mathbb{P}[\epsilon = 1] >> \mathbb{P}[\mu = 1]$.

The raw data are sampled at 2 kHz. Prior to detection, the data are band-pass filtered between 10 and 120 Hz and down-sampled at $f_s = 250$ Hz. They are also whitened using a FIR filter whose time-varying impulse response is derived from the knowledge of the noise power spectral density estimated every 300 s as described in [10, App. A]. Fig. 4 shows the signal-to-noise ratio (SNR) distribution of the D calls. For each noisy observation $\boldsymbol{y}$ of a call, the SNR is estimated in the 10 to 120 Hz bandwidth as

$$\widehat{\text{SNR}} = \frac{\boldsymbol{y}^T\boldsymbol{y}}{N\widehat{\sigma^2}} - 1, \tag{11}$$

where $\widehat{\sigma^2}$ is given by the robust estimator detailed in [10, App. A].

4.1.2. Performance

In this section, the performance of SRD is analyzed for different dictionary sizes M as well as different sparsity constraints K. It is also compared with the spectrogram correlation-based detector of the eXtensible BioAcoustic Tool package (XBAT) developed by the Cornell University Laboratory of Ornithology [13] as well as with a bank of matched filters derived from the generalized likelihood ratio (GLRT) approach presented in [14]. A total of 815 calls were used to test the performance of the detectors while the other $L = 91$ calls (i.e., 10% of the dataset) were used for training the detectors. The training calls were randomly selected among those with a SNR greater than 5 dB.

Two types of dictionary are tested for SRD. The first one is based on a parametric model known as the third degree polynomial-phase model (PPM) which represents each atom as [49]

$$d_m(n) = \cos\left(2\pi \sum_{k=1}^{3} \frac{f_{k-1,m}}{k} \left(\frac{n}{f_s}\right)^k + \alpha_{0,m}\right), \; 0 \le m \le M-1 < L, \; 0 \le n \le N-1, \tag{12}$$

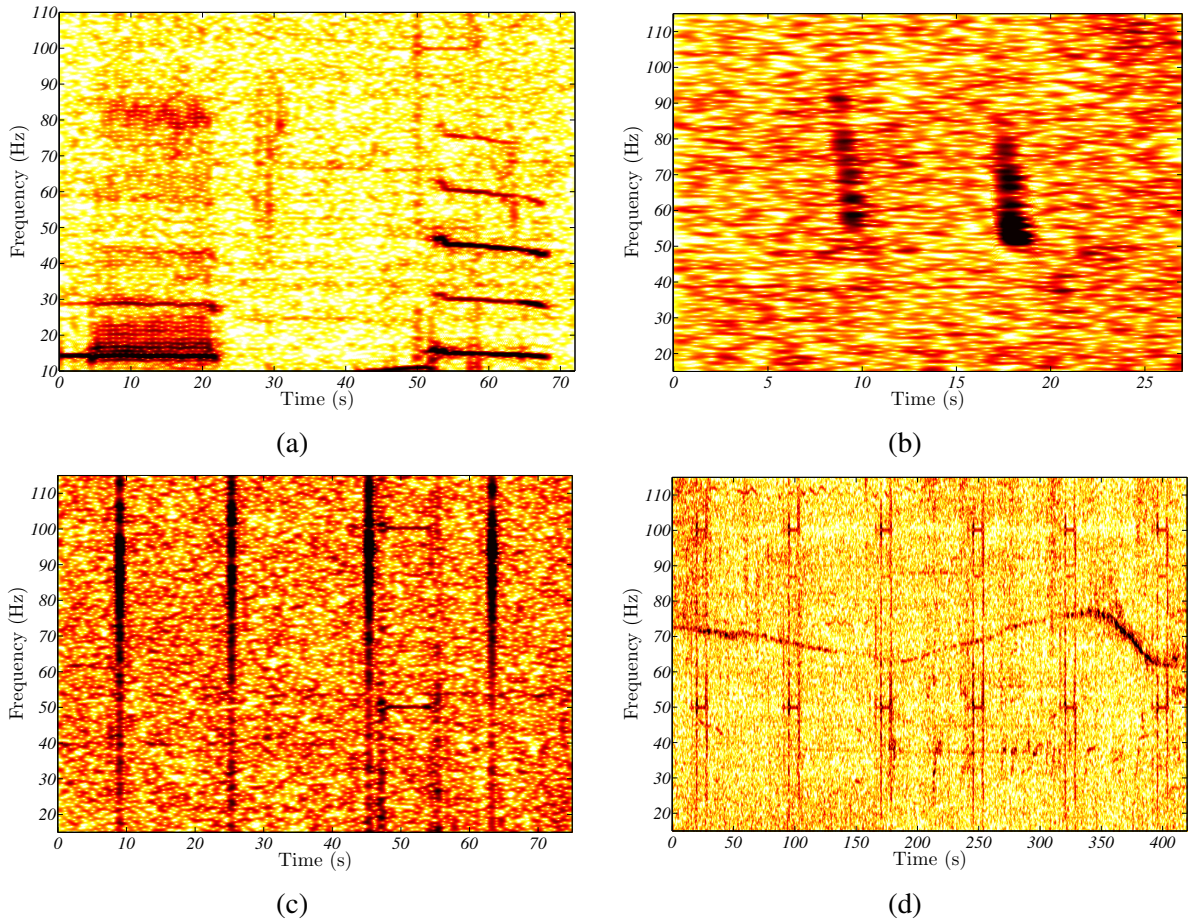

FIGURE 3– Spectrogram examples of non-D call transient signals. (a) Blue whale pulsed and tonal calls [28]. (b) Fin whale "40 Hz" calls [48]. (c) Unidentified recurrent transient sounds. (d) Unidentified periodic transient sounds plus a frequency modulated sound.

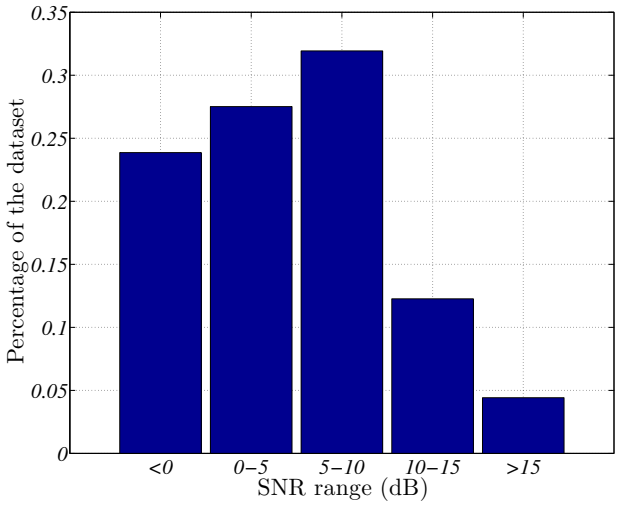

FIGURE 4– Signal-to-noise ratio distribution of the D calls (expressed in percentage of the total number of annotated D calls).

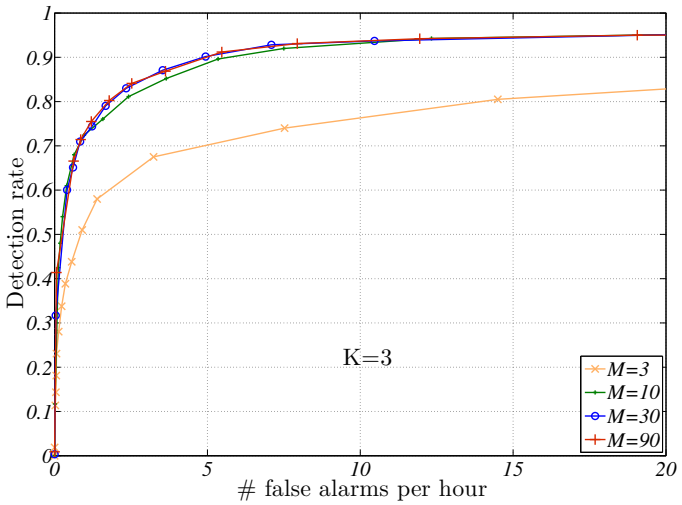

FIGURE 5– Receiver operating characteristic curve of SRD for several dictionary sizes M.

where $\alpha_{0,m}$ is the initial phase of the m-th atom, $f_{0,m}$ its start frequency, and $f_{1,m}$ and $f_{2,m}$ its frequency slope and curvature, respectively. This model is well adapted to simple chirps and has already been successfully applied to the detection of North Atlantic right whale contact calls [14]. The coefficients $f_{k-1,m}$ are found by minimizing the mean square error between (12) and each training signals. To limit the search space, it is assumed that $f_{0,m} \in [20; 110]$ Hz, $f_{1,m} \in [-50; 0]$ Hz/s, $f_{2,m} \in [0; 10]$ Hz/s^2. Note that to be invariant to the initial random phase, the in-phase and quadrature components of the signals are used. SRD combined with model (12) is next referred to as SRD-PPM. The second type of dictionary is directly build from the data itself where (6) is solved using the K-SVD algorithm [40]. SRD combined with K-SVD is next referred to as SRD-KSVD.

For the detection, N is set to 1250 (max. call duration=5 s) and $\Delta = 15$. In Eq. (10), $\hat{\boldsymbol{\theta}}$ is obtained by applying the orthogonal matching pursuit algorithm (OMP) [47] (the Matlab code for K-SVD and OMP is available at http://www.cs.technion.ac.il/~ronrubin/software.html). The performance is assessed by assuming that the annotations provided by the experienced human operators (EHO) represent the ground truth. Most results are displayed as receiver operating characteristic (ROC) curves, which show the detection rate of SRD as a function of the average number of false alarms per hour of processed signal. The number of false alarms is controlled by the choice of the threshold η. The performance analysis is first conducted with SRD-PPM and then a comparison with SRD-KSVD as well as with other detection methods is presented.

Fig. 5 shows the impact of the dictionary size M on the performance of SRD-PPM. The M atoms are built with the M best triplets $(f_{0,m}, f_{1,m}, f_{2,m})$ that minimize the mean square error between (12) and the $L = 91$ training signals. The choice of the size M must be made in relation to the level of call variability observed in the dataset. For truly stereotyped calls, SRD can work well with M small. However, if M is underestimated, it can degrade the detector performance. This is emphasized by Fig. 5, where the best performance for detecting D calls in this dataset is obtained for a dictionary size of at least 30 atoms. Choosing more than 30 atoms only adds some redundancy in the dictionary and therefore does not bring any performance improvement. However, it does increase the processing time as discussed at the end of this subsection.

The sparsity constraint K is directly related to the "complexity" of each single call to detect. Signals combining variability and high complexity (such as erratic signals) must be constructed from a large number of atoms while signals of low complexity should be composed of a few atoms. As shown in Fig. 6, the simple FM structure of D calls (see Fig. 2) can be well detected by setting $K = 3$ in (10). If K is

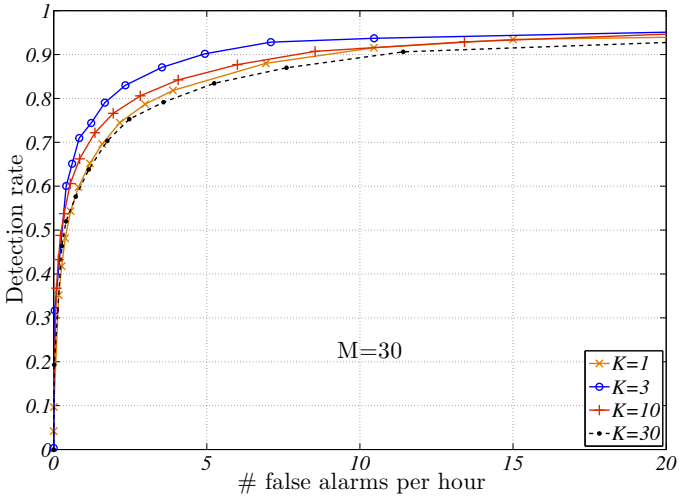

FIGURE 6– Receiver operating characteristic curve of SRD for several sparsity constraints K.

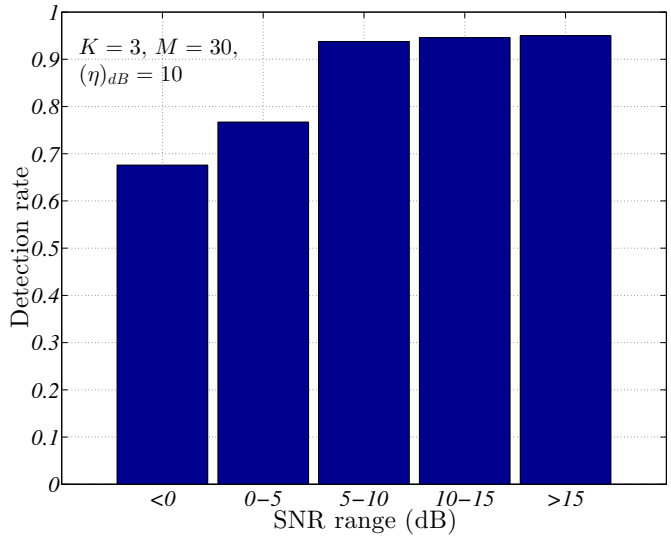

FIGURE 7– Detection rate of SRD as a function of the signal-to-noise ratio. The false alarm rate is set to 2.3 false alarms per hour of processed signal.

overvalued, SRD will tend to detect D calls as well as other more complex signals, which will generate more false alarms. If K is too small, the assumed signal subspace may be to small to capture the D call variability and SRD may miss some calls.

Fig. 7 details the performance of SRD-PPM as a function of the SNR. The detection threshold is set to -10 dB, which corresponds to 2.3 false alarms per hour for the setting $K = 3$ and $M = 30$. As expected, the detection rate increases with the SNR and becomes greater than 90% for calls with a SNR greater than 5 dB. Examples of false alarms and missed detections obtained with this setting are displayed in Fig. 8.

The performance of SRD-PPM and SRD-KSVD are compared in Fig. 9. Building the dictionary from the mathematical model (12) or from the data itself using K-SVD does not significantly impact the performance for the DCLDE 2015 dataset. However, SRD-KSVD performs slightly worse than SRD-PPM because of possible overfitting. Slight overfitting can occur when the number L of training signals is not sufficiently large (in comparison to the possible call variability) so that K-SVD does not learn enough to

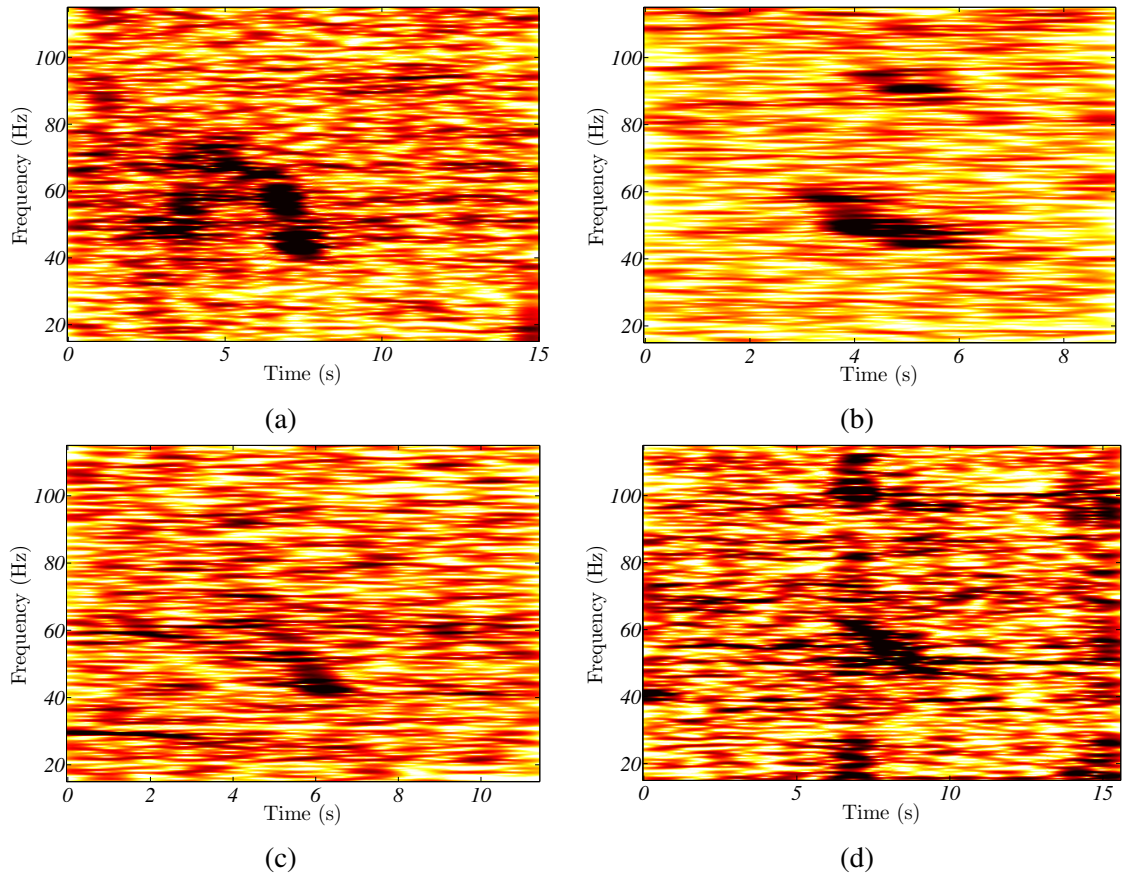

FIGURE 8– (a)-(b): spectrograms of signals detected by SRD but not selected by the experienced human operator as D calls. (c)-(d): spectrograms selected by the experienced human operator as D calls but rejected by SRD.

generalize from trend but tends to memorize too much the training data. Overfitting can also impact the sensitivity of the detector to the dictionary size and the sparsity constraint. The impact of these parameters on the ROC curve of SRD-KSVD (not shown here) is very similar to what is observed in Fig. 5 and Fig. 6, except that SRD-KSVD is slightly more sensitive to change in K and M. Therefore, when the call structure is quite simple, using a parametric model for the dictionary may be more robust than an empirical approach. However, as discussed Sec. 4.2, KSVD can be useful when the call structure is more complex.

SRD is also compared in Fig. 9 with two other detectors: the XBAT software [13], commonly used to analyze blue whale calls [50, 51, 52], and a bank of matched filters similar to the GLRT expressed in [14, Eq. (18)]. As opposed to SRD which computes its detection statistics on time series, XBAT performs spectrogram correlations between the data and some templates pre-selected in the dataset. For the tests, the 30 templates with the highest SNR in the learning dataset were selected. The spectrogram was computed using a Hamming window of 512 samples with 384 overlapping samples. The bank of matched filters is derived from a parametric model of the probability density function of the observation $\boldsymbol{y}$ under both hypothesis H_0 and H_1 as defined in (8). Given that D calls are simple FM sweeps like contact calls of North Atlantic right whale, the same model as in [14] is used for comparison. $\boldsymbol{y}$ is assumed to be a locally stationary Gaussian process whose mean is null under H_0 and is a third degree polynomial-phase signal under H_1. For each observation, the unknown parameters (f_0, f_1, f_2) are replaced by their maximum likelihood estimates in a likelihood ratio test [53, Ch. 6]. The search space for this estimation is the set of triplets (f_0, f_1, f_2) obtained by applying a maximum likelihood estimation on the $L = 91$ training signals.

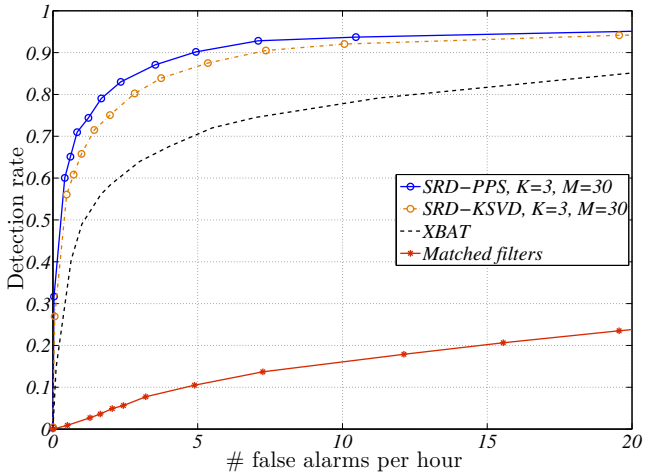

FIGURE 9– Performance comparison between SRD-PPS, SRD-KSVD, XBAT and a bank of matched filters.

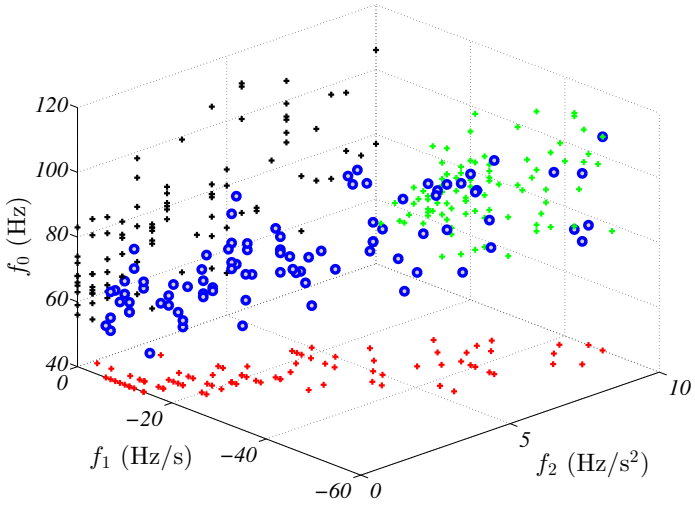

FIGURE 10– The blue circles represent the scatter plot of the triplets (f_0, f_1, f_2) obtained by maximum likelihood estimation on the training dataset. The green, black and red crosses represent the 2D projections of these triplets on the (f_0, f_1), (f_0, f_2), (f_1, f_2)-planes, respectively.

These triplets are shown in Fig. 10. As in [14], this parametric approach implicitly assumes that $\boldsymbol{\psi} + \boldsymbol{w}$ is a zero-mean Gaussian process but with a short period of stationarity to deal with possible impulsive $\boldsymbol{\psi}$. In the experiment, this period is set to 15 s so that the covariance matrix of the process is estimated every 15 s using the method described in [14, Sec. III].

As shown in Fig. 9, SRD largely outperforms both XBAT and the bank of matched filters. For less than 10 false alarms per hour, SRD detects between 15% and 20% more calls than XBAT and, for an average detection rate of 80%, XBAT produces six times more false alarms than SRD. Tests (not shown here) have also been conducted with a larger number of XBAT templates without showing any performance improvement. The bank of matched filters performs very poorly and cannot be used as such in rich soundscapes similar to the DCLDE 2015 dataset. Both methods mostly fail because they are very sensitive and generate more false alarms than SRD. These results are not imputable to a wrong signal modeling.

K=3				
M	3	10	30	90
SDRTR	301	264	179	38
M=30				
K	1	3	10	30
SDRTR	191	179	132	109

TABLE 1– Signal-duration-to-run-time ratio (SDRTR) as a function of the number of atoms M and the sparsity constraint K.

Indeed, the D call signal model underlying the XBAT detector can be seen as a specific case of (5) with $K = 1$ and real-data atoms, and the signal model for the bank of matched filters is also the one of (5), with $K = 1$, combined with the parametric model (12) for the atoms. The way the interfering transient signals are modeled or implicitly considered by the detectors are responsible for these bad results. For instance, to derive the GLRT approach of [14], a statistical model of the interference plus noise ($\boldsymbol{\psi} + \boldsymbol{w}$) must be chosen. A locally stationary Gaussian model is used for the test. Given the complexity of the underwater acoustic environment, a single parametric model can hardly encompasses all the diversity of interfering signals. It seems difficult to find a relevant model when the nature of $\boldsymbol{\psi}$ is very heterogeneous (Fig. 3) and when $\mathbb{P}[\epsilon = 1] >> \mathbb{P}[\mu = 1]$. This limitation was already identified in [12] and addressed by applying a multi-stage method: a GLRT followed by a classification algorithm. While this method can offer satisfying results, it suffers from a major drawback: the features of the interfering transient noise must first be identified or learned by the detector (e.g. "The models are based on the spectral properties of typical kinds of impulsive noise observed in the data." [12, pp. 360]). Such an approach lacks of general applicability because the features of $\boldsymbol{\psi}$ may greatly vary from one dataset to another. One strength of SRD is that it does not try to model the interference itself but defines it with respect to the model of $\boldsymbol{s}$ (see Eq. (7)).

Table 1 shows the signal-duration-to-run-time ratio (SDRTR) of SRD as a function of M and K. This ratio is computed as the dataset duration (223.7 h) divided by the total processing time. SRD is implemented in Matlab (without parallel computing) and runs on a workstation with an Intel Core i7 CPU M 620 @ 2.67GHz x4 with 5 Gio of RAM. Most of the computation time is spent solving (10) using OMP, which makes the SDRTR decrease with M and K. For $K = 3$ and $M = 30$, SRD can process 24 hours of signal in less than 12 minutes, which meets the requirements needed to ensure a wide scale deployment.

4.2. OHASISBIO 2015 dataset

4.2.1. Dataset description

The OHASISBIO network of hydrophones was initially deployed in December 2009 at fives sites in the Southern Indian Ocean to monitor low-frequency sounds, produced by seismic and volcanic events, and by large baleen whales [25, 26]. 50 hours of signals were extracted from the data recorded nearby La Reunion Island hydrophone in the Madagascar Basin (latitude: 26-05.0 S, longitude: 058-08 E) in May 2015. These data contain recurrent pygmy blue whale calls of Madagascar type. As shown in Fig. 11, these calls are composed of two units, named unit A and B is this paper. Unlike the DCLDE 2015 dataset, the 50 hours of signals extracted from the OHASISBIO dataset does not contain many interfering transient signals. There are only a few Z-calls of Antarctic blue whales [52] and some earthquakes. Therefore, to challenge the detector, only the detection of unit B will be considered so that the first unit will act as an interfering signal from a signal processing perspective. Out of the 50 hours of signals, 1040 B units were manually annotated by an EHO. The structure of B units is more complex than D calls. It starts with a pulse centered around 30 Hz, followed by a FM downsweep with harmonics. The pattern of B units slightly changes from one call to another. Its duration can fluctuate from 15 to 25 s and the relative amplitude of harmonics may also change but the frequency bandwidth remains the same.

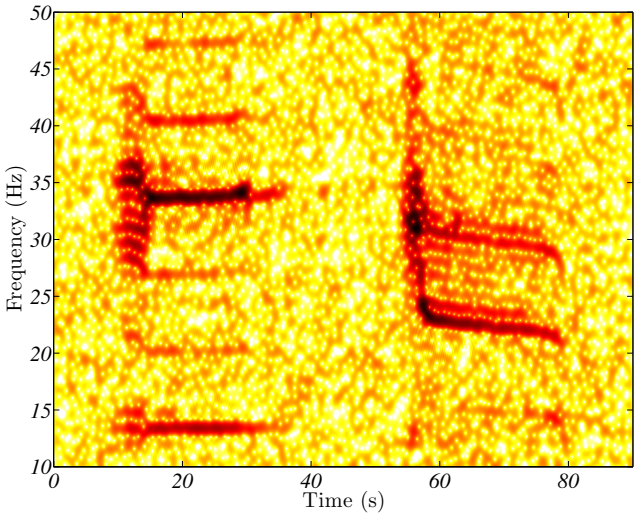

FIGURE 11– Spectrogram of "Madagascar" pygmy blue whale calls composed of units A (left) and B (right).

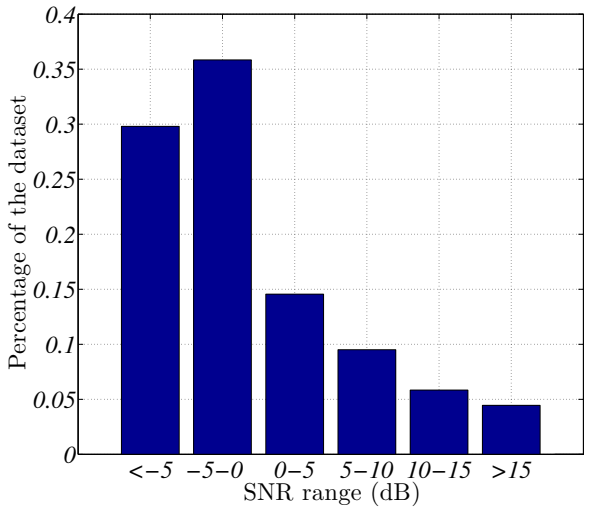

FIGURE 12– Signal-to-noise ratio distribution of B units (expressed in percentage of the total number of annotated calls).

The data are sampled at $f_s = 250$ Hz and band-pass filtered between 20 and 40 Hz before detection. They are also whitened using the approach described in Sec. 4.1. Fig. 12 shows the signal-to-noise ratio distribution of the B units. For each noisy observation $\boldsymbol{y}$ of a call, the SNR is estimated in the 20 to 40 Hz bandwidth. Note that it contains a large amount of low SNR calls.

4.2.2. Performance

In this section, the performance of SRD is compared with Mellinger and Clark's spectrogram-based detector [9]. This detector computes the correlation between a user-defined kernel with the spectrogram of the data. It has proven efficient to detect calls with relatively stable time-frequency patterns and is thus a good candidate for pygmy blue whale calls. For this method, the spectrogram is computed using a 512-samples sliding time Hamming window with 384 overlapping samples. 4 hours of signal containing

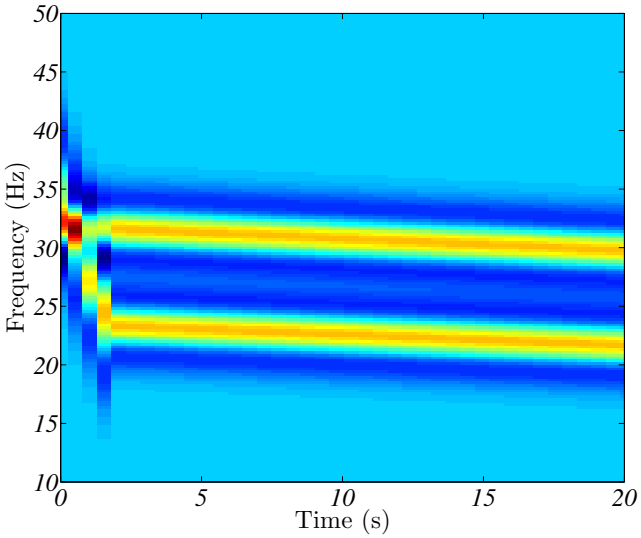

FIGURE 13– Time-frequency kernel used for the detection of "Madagascar" pygmy blue whale calls (unit B) with Mellinger and Clark's method [9].

60 calls were randomly chosen to built the training dataset. A total of 46 hours containing 980 calls were therefore used to test the performance of both detectors.

For Mellinger and Clark's method, the learning consisted in designing several time-frequency kernels that match the time-frequency pattern of training signals. The time-frequency kernel resulting in the best ROC curve on the 4 hours of training data was chosen for the comparison with SRD. This kernel is shown in Fig. 13. Note that a posteriori verification showed that it was also the one performing the best on the test dataset. For SRD, the dictionary was learned using K-SVD. As opposed to D calls, the complexity of B units make it difficult to find a mathematical expression for the dictionary in (5). Although it may be possible to find one, this expression would probably require to tune many ad-hoc parameters, which is not desirable. Therefore, designing the dictionary directly from the data is a good option in this context. The dictionary parameters K and M were chosen as those maximizing the ROC curve on the training dataset, leading to $K = 2$ and $M = 10$. Note that since the B units are quite stable, the performance of SRD-KSVD is not very sensitive to change in K and M.

SRD-KSVD is compared with Mellinger and Clark's method in Fig. 14. On the OHASISBIO 2015 dataset, SRD performs the best. As shown in Fig. 15, for a false alarm rate set to the same value (1.6 false alarms per hour) for both detectors, it is more difficult for Mellinger and Clark's method to detect low SNR signals than it is for SRD-KSVD. Since these kind of signals represent a large proportion of the dataset, the global detection rate is significantly better for SRD-KSVD, especially when the false alarm rate is smaller than 10 false alarms per hour (note that these numerical results may slightly be affected by errors that EHO could possibly make in annotating low SNR signals). The signal representations used by the two detectors may explain the performance difference. On the one hand, the short-time Fourier transform used in [9] is, by definition, adapted to linear combinations of sine and cosine but is not guaranteed to be the best representation for signals with more complex structures. One the other hand, the sparse representation (5) is built from the data itself so it is expected to represent the real data well, and therefore to be more sensitive to low SNR signals. However, this performance improvement comes at the expense of computational complexity. Comparison of detector run times shows that, on average, Mellinger and Clark's method is 2.3 times faster than SRD-KSVD on the workstation whose configuration is detailed at the end of Sec. 4.1.

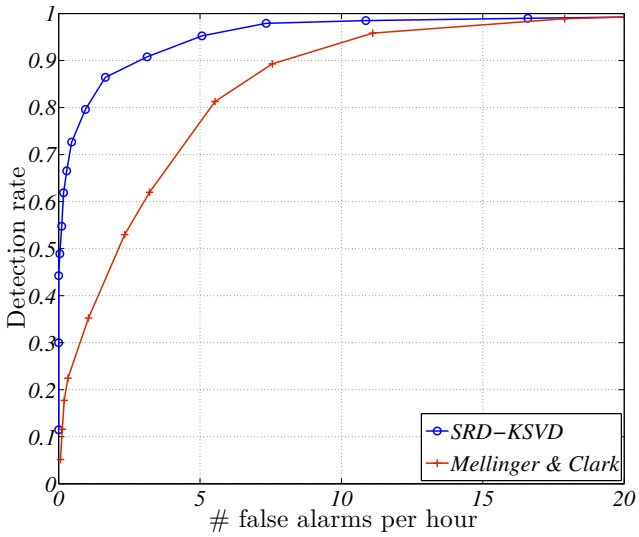

FIGURE 14– Performance comparison between SRD-KSVD (K=2, M=10) and Mellinger and Clark's method with the time-frequency kernel shown in Fig. 13.

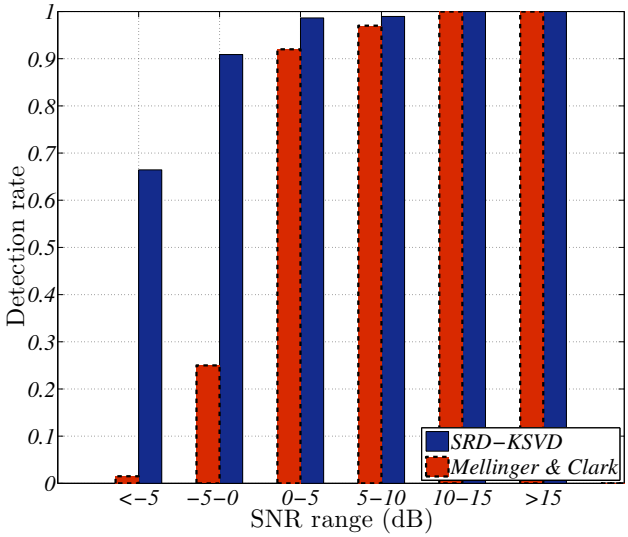

FIGURE 15– Detection rate as a function of the signal-to-noise ratio for SRD-KSVD (K=2, M=10) and Mellinger and Clark's method with the time-frequency kernel shown in Fig. 13. For both methods the false alarm rate is set to 1.6 false alarms per hour of processed signal.

5. Conclusion

Sparse representations offer new prospects for the detection of mysticete calls. They generalize standard representations based on time-frequency dictionaries, they can handle the variability of a given type of call, and, when no parametric model is available, they can automatically be learned from the digitized time-domain data without requiring prior transforms such as spectrograms, wavelets, cepstrums, etc. In addition, sparse representations are easy to design since they rely on two parameters only: the dictionary size and the sparsity constraint. These parameters reflect the degree of variability and complexity of the call to detect.

When combined with the SINR decision statistic, sparse representations can lead to good detection performance. Application of this approach to D calls of North Atlantic blue whales has shown that a 15% to 20% detection gain can be achieved compared to a bank of matched spectrograms and also that the proposed method is much more robust to interfering transient signals than a fully parametric bank of matched filters. The proposed framework is very general and is applicable to any mysticete call that lies in a linear subspace that can be described by a dictionary-based representation. The method has also been applied to "Madagascar" pygmy blue whale calls and has shown to outperform Mellinger and Clark's kernel-based spectrogram detector on low SNR calls.

A natural extension of this work would be to apply sparse representations to classify mysticete sounds [54]. Our intuition is that it may require to tune less (hyper-)parameters than time-frequency contour-based [7] and/or neural network-based classifiers [55]. In addition, applying a metric similar to the SINR statistic (9) at the output of a classifier may be relevant to reject any interfering signals, without trying to explicitly (and exhaustively) learn their features during a training phase. Further research needs to be conducted to validate these hypotheses.

Acknowledgments

The authors would like to thank Ana Širović and Simone Baumann-Pickering of the Scripps Institution of Oceanography for providing the DCLDE 2015 database as well as Jean-Yves Royer of the University of Brest, CNRS Laboratoire Domaines Océaniques for providing the OHASISBIO 2015 database.

References

[1] D. K. Mellinger, K. M. Stafford, S. E. Moore, R. P. Dziak, and H. Matsumoto, "An overview of fixed passive acoustic observation methods for cetaceans," *Oceanography*, vol. 20, pp. 36–45, Dec. 2007.

[2] A. Širović, J. A. Hildebrand, S. M. Wiggins, and D. Thiele, "Blue and fin whale acoustic presence around Antarctica during 2003 and 2004," *Marine Mammal Science*, vol. 25, no. 1, pp. 125–136, 2009.

[3] R. S. Payne and S. McVay, "Songs of humpback whales," *Science*, vol. 173, no. 3997, pp. 585–597, 1971.

[4] A. Širović, J. A. Hildebrand, S. M. Wiggins, M. A. McDonald, S. E. Moore, and D. Thiele, "Seasonality of blue and fin whale calls and the influence of sea ice in the western Antarctic peninsula," *Deep Sea Research Part II : Topical Studies in Oceanography*, vol. 51, no. 17–19, pp. 2327–2344, 2004,

[5] Gillespie, D., Mellinger, D.K., Gordon, J., McLaren, D., Redmond, P., McHugh, R., Trinder, P.W., Deng, X.Y. and Thode, A. "PAMGUARD : semiautomated, open source software for real- time acoustic detection and localisation of Cetaceans," *Proc. of the Institute of Acoustics*, pp. 1–9, 2008

[6] W. M. X. Zimmer, *Passive Acoustic Monitoring of Cetaceans*, Cambridge University Press, pp. 1–368, Cambridge, 2011.

[7] M. F. Baumgartner and S. E. Mussoline, "A generalized baleen whale call detection and classification system," *J. Acoust. Soc. Am.*, vol. 139, pp. 2889–2902, 2011.

[8] M. A. Roch, A. Širović, and S. Baumann-Pickering, "Detection, classification, and localization of cetaceans by groups at the scripps institution of oceanography and San Diego state university (2003-2013)," *Detection, Classification, Localization of Marine Mammals using passive acoustics*, Dirac NGO, pp. 27–52, Paris, 2013.

[9] D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am.*, vol. 107, pp. 3518–3529, 2000.

[10] F.-X. Socheleau, E. Leroy, A. Carvallo Pecci, F. Samaran, J. Bonnel, and J.-Y. Royer, "Automated detection of Antarctic blue whale calls," *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 3105–3117, 2015.

[11] M. Bahoura and Y. Simard, "Chirplet transform applied to simulated and real blue whale (balaenoptera musculus) calls," in *Proceedings of the 3rd International Conference on Image and Signal Processing*, 2008, ICISP '08, pp. 296–303.

[12] I. R. Urazghildiiev, C. W. Clark, T. P. Krein, and S. E. Parks, "Detection and recognition of north Atlantic right whale contact calls in the presence of ambient noise," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 358–368, July 2009.

[13] XBAT, "eXtensible BioAcoustic Tool," www.birds.cornell.edu/brp/ (date last viewed 14/6/30), Cornell Laboratory of Ornithology, NY, U.S.A.

[14] I. R. Urazghildiiev and C. W. Clark, "Acoustic detection of north Atlantic right whale contact calls using the generalized likelihood ratio test," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 1956–1963, 2006.

[15] S. E. Parks, C. W. Clark, and P. L. Tyack, "Short- and long-term changes in right whale calling behavior : The potential effects of noise on acoustic communication," *J. Acoust. Soc. Am.*, vol. 122, no. 6, pp. 3725–3731, 2007.

[16] P. O Thompson, "Marine biological sound west of San Clemente Island : diurnal distributions and effects on ambient noise level during July 1963.," Tech. Rep., US Navy Electronics Laboratory Report, San Diego, CA, pp. 1–42, 1963.

[17] A. N. Gavrilov and R. D. McCauley, "Steady inter and intra-annual decrease in the vocalization frequency of Antarctic blue whales," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4476–4480, 2012.

[18] J. Bonnel and A. Thode, "Using warping processing to range bowhead whale sounds from a single receiver," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 3526–3526, 2013.

[19] M. Elad, *Sparse and Redundant Representations : From Theory to Applications in Signal and Image Processing*, Springer Publishing Company, Incorporated, 1st edition, pp. 227–246, New York, New York, 2010.

[20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[21] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[22] M. G. Jafari and M. D. Plumbley, "Fast dictionary learning for sparse representations of speech signals," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1025–1031, 2011.

[23] Y. C. Eldar and G. Kutyniok, *Compressed sensing : theory and applications*, Cambridge University Press, pp. 1–558, Cambridge, 2012.

[24] "Dclde 2015," http://www.cetus.ucsd.edu/dclde/datasetDocumentation.html, Accessed : 2016-07-01.

[25] E. Tsang-Hin-Sun, J.-Y. Royer, and J. Perrot, "Seismicity and active accretion processes at the ultraslow-spreading southwest and intermediate-spreading southeast indian ridges from hydroacoustic data," *Geophysical Journal International*, vol. 206, no. 2, pp. 1232–1245, 2016.

[26] E. Leroy, F. Samaran, J. Bonnel, and J.-Y. Royer, "Seasonal and diel vocalization patterns of Antarctic blue whale (balaenoptera musculus intermedia) in the southern Indian Ocean : A multi-year and multi-site study," *PloS one*, vol. 11, no. 11, pp. 1–20, 2016.

[27] P. O. Thompson, L. T. Findley, O. Vidal, and W. C. Cummings, "Underwater sounds of blue whales, balaenoptera musculus, in the gulf of california, mexico," *Marine Mammal Science*, vol. 12, no. 2, pp. 288–293, 1996.

[28] E. M. Oleson, S. M. Wiggins, and J. A. Hildebrand, "Temporal separation of blue whale call types on a southern california feeding ground," *Animal Behaviour*, vol. 74, no. 4, pp. 881 – 894, 2007.

[29] Stafford K. M., Nieukirk S. L., and Fox C. G., "Geographic and seasonal variation of blue whale calls in the north Pacific," *Journal of Cetacean Research and Management*, vol. 3, no. 1, pp. 65–76, 2001.

[30] Rankin S., Ljungblad D., Clark C., and Kato H., "Vocalisations of antarctic blue whales, balaenoptera musculus intermedia, recorded during the 2001/2002 and 2002/2003 iwc/sower circumpolar cruises, area v, Antarctica," *Journal of Cetacean Research and Management*, vol. 7, pp. 13–20, 2005.

[31] D. K. Mellinger and C. W. Clark, "Blue whale (balaenoptera musculus) sounds from the north Atlantic," *J. Acoust. Soc. Am.*, vol. 114, no. 2, pp. 1108–1119, 2003.

[32] C. L. Berchok, D. L. Bradley, and T. B. Gabrielson, "St. Lawrence blue whale vocalizations revisited : Characterization of calls detected from 1998 to 2001," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 2340–2354, 2006.

[33] D. Ljungblad, C. W. Clark, and H. Shimada, "A comparison of sounds attributed to pygmy blue whales Balaenoptera musculus brevicauda recorded south of the Madagascar Plateau and those attributed to true blue whales Balaenoptera musculus recorded off Antarctica," *Rep. Int. Whal. Comm.*, vol. 49, no. 6, pp. 439–442, 1998.

[34] F. Samaran, C. Guinet, O. Adam, J.-M. Motsch, and Y. Cansi, "Source level estimation of two blue whale subspecies in southwestern Indian Ocean," *J. Acoust. Soc. Am.*, vol. 127, no. 6, pp. 3800–3808, 2010.

[35] J. A. McCordic and Susan E. Parks, "Individually distinctive parameters in the upcall of the north Atlantic right whale (eubalaena glacialis)," *J. Acoust. Soc. Am.*, vol. 137, no. 4, pp. 2196–2196, 2015.

[36] B. S. Miller, R. Leaper, S. Calderan, and J. Gedamke, "Red shift, blue shift : Investigating doppler shifts, blubber thickness, and migration as explanations of seasonal variation in the tonality of antarctic blue whale song," *PloS one*, vol. 9, no. 9, pp. 1–11, 2014.

[37] K. M. Stafford and S. E. Moore, "Atypical calling by a blue whale in the gulf of Alaska (l)," *J. Acoust. Soc. Am.*, vol. 117, no. 5, pp. 2724–2727, 2005.

[38] F. Lelandais and H. Glotin, "Mallat's matching pursuit of sperm whale clicks in real-time using Daubechies 15 wavelets," in *New Trends for Environmental Monitoring Using Passive Systems, 2008*, Oct 2008, pp. 1–5.

[39] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. On 1999 IEEE International Conference - Volume 05*, Washington, DC, USA, 1999, pp. 2443–2446.

[40] M. Aharon, M. Elad, and A. Bruckstein, "K-svd : An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[41] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

[42] L. L. Scharf and B. Friedlander, "Matched Subspace Detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146 – 2157, 1994.

[43] L. L. Scharf, *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*, Addison-Wesley, pp. 1 – 524, Reading, Massachusetts, 1991.

[44] F.-X. Socheleau and D. Pastor, "Testing the energy of random signals in a known subspace : An optimal invariant approach," *IEEE Signal Process. Lett.*, vol. 21, no. 10, pp. 1182–1186, Oct. 2014.

[45] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses, 3rd edition*, Springer, pp. 1 – 784, New York, New York, 2005.

[46] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[47] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit : recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, Nov 1993, pp. 40–44 vol.1.

[48] W. A. Watkins, "Activities and underwater sounds of fin whales (balaenoptera physalus)," *Sci. Rep. Whales Research Inst. Tokyo*, vol. 33, pp. 83–118, 1981.

[49] Shimon Peleg and Boaz Porat, "Estimation and classification of polynomial-phase signals," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 422–430, 1991.

[50] O. Boisseau, D. Gillespie, R. Leaper, and A. Moscrop, "Blue (balaenoptera musculus) and fin (b. physalus) whale vocalisations measured from northern latitudes of the Atlantic Ocean," *J. CETACEAN RES. MANAGE*, vol. 10, no. 1, pp. 23–30, 2008.

[51] F. W. Shabangu and K. Findlay, "Overview of the IWC IDCR/SOWER cruise acoustic survey data," Tech. Rep., SC/65b/Forinfo 18, International Whaling Commission, pp. 1–6, 2014.

[52] F. Samaran, K. M. Stafford, T. A. Branch, J. Gedamke, J.-Y. Royer, R. P. Dziak, and C. Guinet, "Seasonal and geographic variation of southern blue whale subspecies in the Indian Ocean," *PLoS ONE*, vol. 8, no. 8, pp. 1– 10, 2013.

[53] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II, Detection Theory, 14th printing*, Prenctice Hall, pp. 1–560, 2009.

[54] Shu Kong and Donghui Wang, "A brief summary of dictionary learning based approach for classification (revised)," *arXiv preprint arXiv :1205.6544*, pp. 1–8, 2012.

[55] X. C. Halkias, S. Paris, and H. Glotin, "Classification of mysticete sounds using machine learning techniques," *J. Acoust. Soc. Am.*, vol. 134, no. 5, pp. 3496–3505, 2013.

OUR WORLDWIDE PARTNERS UNIVERSITIES - DOUBLE DEGREE AGREEMENTS

3 CAMPUS, 1 SITE

IMT Atlantique Bretagne–Pays de la Loire – **http://www.imt-atlantique.fr/**

Campus de Brest
Technopôle Brest-Iroise
CS 83818
29238 Brest Cedex 3
France
T +33 (0)2 29 00 11 11
F +33 (0)2 29 00 10 00

Campus de Nantes
4, rue Alfred Kastler
CS 20722
44307 Nantes Cedex 3
France
T +33 (0)2 51 85 81 00
F +33 (0)2 99 12 70 08

Campus de Rennes
2, rue de la Châtaigneraie
CS 17607
35576 Cesson Sévigné Cedex
France
T +33 (0)2 99 12 70 00
F +33 (0)2 51 85 81 99

Site de Toulouse
10, avenue Édouard Belin
BP 44004
31028 Toulouse Cedex 04
France
T +33 (0)5 61 33 83 65

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Imprimé à IMT Atlantique
Dépôt légal : Octobre 2017
ISSN : 2556-5060